# Effects of Heterogeneity on Measuring Efficiency Scores : The Case of China?s Banking Sector

**Chapter** · July 2015

DOI: 10.1057/9781137454638.0007



Yizhe Dong
The University of Edinburgh



Douglas Cumming
Florida Atlantic University



Wenxuan Hou
The University of Edinburgh

# Effects of Heterogeneity on Measuring Efficiency Scores: the Case of China's Banking Sector

Yizhe Dong*
The School of Management and Business, Aberystwyth University

Douglas Cumming,
Schulich School of Business, York University

Alessandra Guariglia,
Department of Economics, University of Birmingham

Wenxuan Hou,
University of Edinburgh Business School

Edward Lee
Manchester Business School, University of Manchester

---

Abstract

This study aims to investigate the significance of adequately accounting for heterogeneity in banking efficiency. It compares different specifications of stochastic cost frontier models which attempt to account for bank heterogeneity in various ways. This study compares the estimated parameters from five different specifications, conducts a specification test and discusses the effect of accounting for the heterogeneity on efficiency scores and rankings. The findings indicate that heterogeneity across banks could influence production directly, through the cost frontier, or indirectly through its effect on efficiency. This study concludes that accounting for the heterogeneity in the sample of Chinese banks is an important issue which, if not taken into account, may lead to biased estimates of banking efficiency.

*Keywords:* Cost efficiency; Stochastic frontier analysis; Heterogeneity; Chinese banking

*JEL classification:C52 D24 G21*

---

* The School of Management and Business, Aberystwyth University, SY23 3AL, UK;
Email addresses: yid1@aber.ac.uk

# 1 Introduction

Banks play a central role in the financial system and also in the real economy, as the 2008 financial crisis has vividly illustrated. The measurement of banking performance has gained importance for both policy-makers and practitioners in relation to their decision making. Hence, empirical assessment of the efficiency of banking institutions has received considerable attention in the banking literature (see Berger and Humphrey 1997; Berger 2007; Fethi and Pasiouras 2010; for a comprehensive review of banking efficiency studies). However, efficiency measures can vary substantially across different samples and empirical specifications, which may limit the use of efficiency measures by decision makers. Some previous studies (e.g. Bauer et al. 1998; McKillop et al 2005; Bos et al. 2009) have compared the efficiency of financial institutions which is obtained using different approaches or specifications. They reported that efficiency scores vary considerably across different models. Therefore, choosing an appropriate frontier model is very important for measuring banking efficiency.

Recent studies have observed that estimates of banking efficiency could be biased if heterogeneity across banks is ignored in estimation models. Systematic differences across banks could affect either the bank's costs directly, or the bank's ability to operate efficiently, or both. For example, Mester (1996) argued that the quality and degree of risk of a bank's outputs should be taken into account when modelling banking production. If such factors are not controlled for, some banks might be mislabelled as inefficient simply because they are operating in a more risk-averse manner than others, while others might be mislabelled as efficient because they are producing lower quality outputs than others (Mester 1997). Berger and Mester (1997) and Hughes and Mester (2008) also suggested that the inclusion of financial capital is very important for estimating banking efficiency and helping to account for the difference in a bank's risk preferences. Moreover, some heterogeneous environmental variables which are not under the control of management may significantly affect a bank's efficiency. Including these environmental variables could reliably control for differences in: banking type; bank size; regulatory, market and demographic conditions, etc. Various environmental variables have been tested in many banking efficiency studies, for example those by Berger and Mester (1997), Cavallo and Rossi (2002), Casu and Molyneux (2003) and more recently Fries and Taci (2005), Bonin et al. (2005), Sensarma (2006), Kumbhakar and Wang (2007) and Tecles and Tabak (2010).

Although some studies seek to control for systematic differences across banks in efficiency frontier models, there are only a few studies that have investigated the significance of accounting for heterogeneity in banking efficiency analysis. Mester (1997) studied the U.S. banking sector and compared cost efficiency derived from a single cost function model (common benchmark) and the separate cost functions model. He tested and rejected the single cost function model and suggested that the separate cost functions model is a more appropriate one to use. More importantly, he concluded that bank cost efficiency scores can be biased if sample heterogeneity is ignored in estimation models. Bos et al. (2008) focused on the effects of accounting for heterogeneity on the efficiency scores of German banks. They compared different specifications of stochastic cost and profit frontier models which attempt to account for systematic differences among banks in various ways. They found that bank heterogeneity significantly influenced the efficiency results and the (general specification) model which accounts for heterogeneity by including exogenous variables in both the frontier and inefficiency term could be considered as the most preferred specification. The empirical studies documented which have looked at the effects of heterogeneity on efficiency scores to date are primarily based on the U.S. and E.U. markets, with much less insight into, and discussion of, the banking sector in emerging or transitional economies. A study by Poghosyan and Kumbhabar (2010) may be one of the few exceptions. They investigated the cost efficiency of banks in twenty former socialist emerging economies by employing a latent class stochastic efficiency frontier model. The model explicitly accounts for differences in technological regimes caused by the heterogeneity of the economic environments in which the banks operate. They concluded that, if the heterogeneity of technological regimes is not accounted for, it can create downward-bias in efficiency score estimates.

Over the last thirty years, China has been one of the world's fastest growing emerging economies and has become the world's second largest economy in terms of nominal GDP in 2010. China's banking sector has played a very important role in its economic growth. In order to create a sound and effective banking system, the Chinese authorities have implemented a series of reforms designed to address the institutional, political and organizational problems faced by the banking industry. The reforms have included inter alia: establishing a two-tier banking system; separating so-called policy lending from commercial lending; removing the credit ceiling on deposits and loans; reducing the systemic risk of the banking sector; gradually privatising state-owned banks; encouraging state-owned banks to seek a listing on the stock exchange; and relaxing the restrictions on foreign bank entry into the local market[1]. It is important, therefore, to employ an appropriate frontier model to measure the efficiency levels of Chinese banks over the reform period. This efficiency analysis will assist Chinese government policy-makers and bank managerial staff in the decision making process.

A few studies have investigated the efficiency of Chinese banks using either a non-parametric or a parametric frontier approach (Chen et al. 2005; Fu and Heffernan 2007; Kumbhakar and Wang 2007; Ariff and Can 2008; Berger et al.2009; Jiang et al. 2009 and Luo et al. 2011). However, to the best of my knowledge, no studies to

[1] For more general and detailed information about the background to the Chinese banking sector, see Fu and Heffernan 2007; Cousin 2007; and Berger et al. 2009.

date have assessed and compared the (cost) efficiency of Chinese banks resulting from different model specifications. Therefore, this study aims to fill this gap in the banking literature by comparing the results of a number of well-established stochastic cost frontier specifications. Such a comparison provides exhaustive empirical evidence on the effects of heterogeneity on efficiency scores as well as useful and persuasive information for efficiency analysis.

This study employs an unbalanced data set that consists of 41 Chinese banks over the period from 1994 to 2007 and estimates the cost frontier using five different specifications which attempt to account for bank heterogeneity in different ways. Following that, it examines the effects of sample heterogeneity on estimated efficiency scores and endeavours to identify the most appropriate frontier model specification for our sample of Chinese banks.

The remainder of the paper is organized as follows: Section 2 explains the methodology to be employed; Section 3 defines the variables and presents the data; Section 4 discusses the empirical results; and Section 5 presents the conclusions.

# 2 Methodology

## 2.1 Stochastic cost frontier approach

It is well known that the cost (or profit) function approach for determining the optimal combination of factors of production is the dual of the production function approach[2]. A general version of the minimum cost function (also known as the cost frontier) can be written as:

$$TC_i \geq TC^* = f(Q_{it}, W_{it}; \beta), \qquad i = 1,\dots, \mathrm{I}, \tag{1}$$

where $TC_i$ is the observed total cost of the individual bank, $i$; $Q_{it}$ is a vector of the outputs of the bank, $i$; $W_i$ is an input price vector of the bank, $i$, $f(Q_{it}, W_{it}; \beta)$ is the cost frontier common to all banks representing the minimum cost of producing outputs $Q_i$ when the banks face input prices $W_i$, and $\beta$ is a vector of the technology parameters to be estimated. Because the above cost frontier is deterministic, such a formulation ignores measurement errors and other sources of statistical noise, and so all deviations from the frontier are attributed to

[2] The production function summarizes the technology of a bank; that is, the relationship between outputs and inputs under which it operates. Comparing this with the production function, the cost function adds the economic dimension of determining the technically efficient combinations of factors of production which minimize the total cost of particular output levels. This latter aspect of cost minimization is referred to as allocative efficiency.

inefficiency. To overcome this drawback, Aigner, Lovell and Schmidt (1977) and Meeuse and van den Broeck (1977) simultaneously proposed the stochastic frontier model (SFA). Their model adds a symmetric error term to the deterministic frontier, which accounts for statistical noise. Taking into consideration the characteristics of our data set and the purposes of this study, we decided to apply this classical frontier model to the panel of Chinese banking data on which our empirical analysis is based. We start by outlining a baseline cost frontier model (named M1) for measuring the efficiency of Chinese banks. Then we discuss a number of alternative stochastic cost frontier models that have been developed to account for sample heterogeneity. The basic stochastic cost frontier model for a panel data set can be written as:

$$lnTC_{it} = f\,(Q_{it}, W_{it}; \beta) + v_{it} + u_{it} \quad i = 1,\ldots,\mathrm{I},\; t = 1,\ldots\mathrm{T} \tag{2}$$

where $lnTC_{it}$ is the logarithm of the total cost of bank *i* at time *t*; $v_{it}$ is a two-sided normal disturbance term with zero mean and variance $\sigma_v^2$ representing the effects of noise, and $u_{it}$ is a non-negative random disturbance term capturing the effects of cost inefficiency and is assumed as a half-normal distribution, $N^+(0, \sigma_u^2)$. Additionally, $v_{it}$ and $u_{it}$ are independently distributed from each other.

Cost efficiency (CE) is measured relative to the estimated frontier and formally defined as the ratio of the best practice minimum cost to the cost actually incurred, or :

$$CE_{it} = \frac{f\,(Q_{it}, W_{it}; \beta)\exp(v_{it})}{f\,(Q_{it}, W_{it}; \beta)\exp(v_{it} + u_{it})} = \exp(-u_{it}) \tag{3}$$

For the half-normal case, Battese and Coelli (1988) proposed an appropriate point estimator for cost inefficiency which involves the conditional expectation of $\exp(-u_{it})$ given the entire error term $\varepsilon_{it} = v_{it} + u_{it}$ [3]. This is expressed as:

$$CE_{it} = \mathrm{E}\left[\exp(-u_{it})\middle|\varepsilon_{it}\right] = \left[\frac{1-\Phi(\sigma_* - \varepsilon_{it}\gamma / \sigma_*)}{1-\Phi(-\varepsilon_{it}\gamma / \sigma_*)}\right].\exp\left\{-\varepsilon_{it}\gamma + \frac{1}{2}\sigma_*\right\} \tag{4}$$

where $\Phi(.)$ is the standard normal cumulative distribution function and $\sigma = \sqrt{\sigma_v^2 + \sigma_u^2}$ , $\sigma_* = \sigma_v^2\sigma_u^2 / \sigma^2$ and $\gamma = \sigma_u^2 / \sigma^2$ . The value of $\gamma$ must lie between zero and one. A value of one indicates that the deviation from the

[3] An alternative point estimator for efficiency is given by Jondrow et al.(1982) (JLMS). Battese and Coelli (1988) and Kumbhakar and Lovell (2000) point out that Battese and Coelli (1988)'s estimator is to be preferred, particularly when $u_i$ is not close to zero. This is because the JLMS estimator includes only the first term in the power-series expansion of exp (-u).

frontier is due to cost inefficiency, while a value of zero indicates that the deviation is explained purely by noise[4].

The baseline model assumes that efficiency follows a half-normal distribution with a mean of zero. This specification implies that the likelihood of inefficient behaviour monotonically decreases for increasing levels of inefficiency and that most units are likely to be concentrated close to the cost frontier, causing artificially high efficiency levels. However, there is no theoretical reason to support the ex ante monotonicity assumption invoked by Aigner, Lovell and Schmidt (1977). In light of this, Stevenson (1980) argues that inefficiency is not likely to be distributed with such a monotonically declining density function. He also argues that the half-normal distribution assumption used by Aigner, Lovell and Schmidt (1977) is unnecessarily restrictive. Therefore, the second model (M2) used in this study relaxes the half-normal assumption and assumes $u$ to be normally distributed with a non-zero (constant) mean truncated at zero from above (Stevenson 1980). This truncated normal distribution requires one more parameter $\mu$ (its mean) to be estimated and the point estimate of cost efficiency for each bank is given by the following formula:

$$CE_{it} = \mathrm{E}\left[\exp(-u_{it})\middle|\varepsilon_{it}\right] = \left[\frac{1-\Phi(\sigma_{*}-(-\sigma_{u}^{2}\varepsilon_{it}+\mu\sigma_{v}^{2})/\sigma\sigma_{*})}{1-\Phi(-\sigma_{u}^{2}\varepsilon_{it}+\mu\sigma_{v}^{2}/\sigma\sigma_{*})}\right].\exp\left\{-\frac{-\sigma_{u}^{2}\varepsilon_{it}+\mu\sigma_{v}^{2}}{\sigma}+\frac{1}{2}\sigma_{*}\right\} \qquad (5)$$

We prefer the truncated specification to the half-normal model, because the former provides a somewhat more flexible representation of the pattern of efficiency in the data. Nonetheless, both models face an important limitation which fails to account for heterogeneity across banks, an issue which will now be developed further[5].

The two models described above assume that banks operate in perfectly competitive input-output markets. Thus, banks' input prices are taken as exogenous. However, this assumption may not be valid when banks are heterogeneous. Some of the factors which contribute to banks' heterogeneity (e.g. level of equity) could also make their input prices partially endogenous and thus influence both their technical and allocative efficiencies. Additionally, under the conventional frontier model, different banks are assumed to produce equivalent quality in terms of outputs. However, there are likely to be differences across banks in the quality of their outputs. Because the traditional output variables do not fully capture heterogeneity in bank outputs, differences in production quality may be incorrectly measured as differences in cost inefficiency (Berger and Mester 1997). Some banks might be incorrectly categorized as inefficient merely because they produce higher quality outputs than other banks. Thus, failure to recognize the heterogeneity across banks may bias estimates of cost efficiency. In order to

[4] The efficiency measure from equation (4) takes values over the interval $[1,\infty)$ and a value equal to one means that it is fully efficient. Given this, the cost efficiency score can be calculated as $1/CE_{it}$.

[5] The models absorb all unmeasured heterogeneity through the inefficiency term ($u_{it}$).

overcome these problems, we can incorporate these differences into our efficiency models. Formally, it is appropriate to account for the heterogeneity by including control variables, $Z_{it}$, along with the outputs and input prices in a stochastic cost frontier model (M3), which can be written as follows:

$$lnTC_{it} = f\,(Q_{it}, W_{it}, Z_{it}; \beta) + v_{it} + u_{it} \qquad i = 1,\ldots,\mathrm{I},\ \ t = 1,\ldots\mathrm{T} \qquad (6)$$

Apart from including control variables in the deterministic kernel of the stochastic cost frontier, the model given by equation (6) is structurally indistinguishable from the conventional stochastic cost frontier model given by equation (2). In this model, it is assumed that $Z_{it}$ directly influences the cost of production and thus it may be more precise in its estimates of the parameters and cost efficiency. Until now, all the models that have been presented assume that all banks within an industry use the same production technology to convert inputs into outputs and that all banks face similar environmental conditions; that is, the shape of the cost frontier is the same across all banks.

We know, however, that some heterogeneous environmental variables (or exogenous variables), which are neither inputs to the production process nor outputs of it, may influence the performance measures obtained. For example, variations in market structure, regulation and type of ownership may cause variations in banking performance. The omission of such heterogeneity may also lead to biased estimates of the parameters describing the cost frontier, and consequently, cost inefficiency is likely to be stated inaccurately.

According to Kumhakar and Lovell (2000), generally there are two main ways in which environmental variables can be incorporated into efficiency measurement models. In the simplest case, if the environmental variables which are not under the control of management directly influence the structure of the production process itself, it is appropriate to incorporate these variables into the cost function as regressors (e.g., Good *et al*., 1993). In this case the stochastic cost frontier model (M4) is given as:

$$lnTC_{it} = f\,(Q_{it}, W_{it}, Z_{it}, E_{it}; \beta) + v_{it} + u_{it} \qquad i = 1,\ldots,\mathrm{I},\ \ t = 1,\ldots\mathrm{T} \qquad (7)$$

where $E_{it}$ is a vector of exogenous variables in the deterministic kernel of the stochastic production frontier accounting for systematic differences across banks due to ownership structure, size and market structure etc. By including the additional variables, the cost frontier incurs a parallel shift. This is different from the influence of incorporating control variables, which changes the shape of the frontier. In other words, each bank faces a different cost frontier, but we still assume that the shape of the frontier is the same for all banks. One limitation of this model is that the additional variables do not explicitly explain the variations in the efficiency levels of banks.

Another way to account for heterogeneity and to achieve an explicit explanation of efficiency is to estimate the parameters of the stochastic frontier and inefficiency models simultaneously. This approach is called a one-stage approach, and was developed by Battese and Coelli (1995). It assumes that the environmental variables influence the degree of cost inefficiency (the distribution of inefficiency) and hence that cost inefficiencies are expressed as a function of these factors and are integrated into the stochastic frontier model (Battese and Coelli 1995). By comparing this with the incorporation of exogenous variables into the cost function, this method allows the raw efficiency scores to be adjusted in order to reflect the nature of the operational environments in which banks conduct business (Kumhakar and Lovell 2000). This approach also avoids the omitted variables and independence problems which plague the two-stage estimation procedure[6,7]. The general Battese and Coelli (1995) model is specified in the same way as equation (6) with one exception, which is that the inefficiency term $u_{it}$ is expressed as an explicit function of a vector of exogenous variables, $E_{it}$, and a random error term. The model (M5) can be expressed as follows:

$$lnTC_{it} = f\,(Q_{it}, W_{it}, Z_{it}, E_{it}; \beta) + v_{it} + u_{it}, \quad \text{where } u_{it} = \delta E_{it} + w_{it} \qquad i = 1,\ldots,\mathrm{I},\ t = 1,\ldots\mathrm{T} \qquad (8)$$

where the random error term $w_{it}$ captures the effect of the 'unobserved' factors and is defined by a truncated normal distribution with a zero mean and constant variance; $E_{it}$ captures the observed factors which explain differences in cost efficiency across banks and $\delta$ is a vector of the parameters to be estimated. Since the inefficiency term $u_{it}$ is non-negative, the truncation point is $-\delta E_{it}$. In the above model, the truncated inefficiency term $u_{it}$ is independently but not identically distributed and takes the form: $u_{it} \sim (\delta E_{it}, \sigma_u)$. The cost efficiency of the *i*th bank becomes:

$$CE_{it} = \exp(-u_{it}) = \exp(-\delta E_{it} - w_{it}) \qquad (9)$$

When considering the different models summarized above, it is not possible to find a convincing theoretical argument which suggests that one particular specification for assessing efficiency is better than another. Hence, the choice of frontier models is 'frequently a judgement call' (Kumhakar and Lovell, 2000, p.266). Thus, it is the case that the specification of inefficiency in frontier modelling is usually ad hoc and is based on tractability rather than on any optimal theoretical criteria for assessing efficiency (Kumhakar et al., 1997). Consequently, in

[6] The two-stage approach seeks to explain the variation in estimated inefficiencies (Kalirajan 1981; Pitt and Lee 1981). In the first stage, a cost frontier and banks' efficiency levels are estimated, ignoring the exogenous variables. In the second stage, the estimated efficiency scores are then regressed against the exogenous variables.

[7] See Kumhakar and Lovell (2000) and Wang and Schmidt (2002), for a more detailed discussion of serious econometric problems which were encountered in the two-stage approach.

this study, we employ the different stochastic cost frontier models briefly summarized in Table 1. The models use different assumptions for the distribution of (in)efficiency terms, and different ways of incorporating control and environmental variables which account for sample heterogeneity. Importantly, however, we also compare the results obtained from each model in order to assess the reliability and robustness of our results and, in particular, to determine the most appropriate model with which to measure Chinese banking efficiency.

**Table 1 SFA Model Specifications**

| Models | Specification | Inefficiency *u* | Heterogeneity |
|---|---|---|---|
| Baseline (M1) | $f(Q_{it}, W_{it}; \beta)$ | $u_{it} \sim N^{+}(0, \sigma_u^2)$ | None. |
| Truncated (M2) | $f(Q_{it}, W_{it}; \beta)$ | $u_{it} \sim N^{+}(0, \sigma_u^2)$ | None. |
| Controlled (M3) | $f(Q_{it}, W_{it}, Z_{it}; \beta)$ | $u_{it} \sim N^{+}(0, \sigma_u^2)$ | Bank specific observed factors in cost function. |
| Kernel (M4) | $f(Q_{it}, W_{it}, Z_{it}, E_{it}; \beta)$ | $u_{it} \sim N^{+}(0, \sigma_u^2)$ | Bank specific observed factors in cost function. |
| Error effects (M5) | $f(Q_{it}, W_{it}, Z_{it}; \beta)$ | $u_{it} \sim N^{+}(\mu + \delta E_{it}, \sigma_u^2)$ | Bank specific observed factors in cost function. Heterogeneity in the mean of inefficiency distribution. |

### 2.2 Econometric specifications

When parametric methods are used to estimate efficiency, we should first consider the choice of a functional form for the cost function. In this study, we use the translog (transcendental logarithmic) form, which is the most commonly used functional form in the banking efficiency literature, as our specification of the cost frontier. For example, the cost frontier for the model (M3) is specified as:

$$\ln\frac{TC_{it}}{W_{it,1}} = \beta_0 + \sum_{l=1}^{L}\beta_l \ln(Q_{it,l}) + \sum_{m=1}^{M}\chi_m \ln(\frac{W_{it,m}}{W_{it,1}}) + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M}\eta_{mn}\ln(\frac{W_{it,m}}{W_{it,1}})\ln(\frac{W_{it,n}}{W_{it,1}}) + \frac{1}{2}\sum_{l=1}^{L}\sum_{j=1}^{L}\varphi_{lj}\ln(Q_{it,l})\ln(Q_{it,j})$$

$$+\sum_{l=1}^{3}\sum_{m=1}^{2}\iota_{lm}\ln(Q_{it,l})\ln(\frac{W_{it,m}}{W_{it,1}}) + \sum_{k=1}^{K}\rho_k \ln Z_{it,k} + \frac{1}{2}\sum_{k=1}^{K}\sum_{s=1}^{K}\xi_{ks}\ln(Z_{it,k})\ln(Z_{it,s})$$

$$+\sum_{k=1}^{K}\sum_{m=1}^{M}\theta_{km}\ln(Z_{it,k})\ln(\frac{W_{it,m}}{W_{it,1}}) + \sum_{k=1}^{K}\sum_{l=1}^{L}\psi_{kl}\ln(Z_{it,k})\ln(Q_{it,l}) + u_{it} + v_{it} \qquad i = 1,\ldots,I,\ t = 1,\ldots T$$

where ln*TC* is the natural logarithm of total (operating and financial) costs; $Q_l$ and $W_m$ are the output quantities and input prices, respectively; *Z* is a vector of the control variables; and $\beta, \chi, \varphi, \iota, \eta, \rho, \xi, \theta,$ and $\psi$ are the parameters to be estimated. The duality theorem requires that the cost function must be linearly homogeneous in input prices while continuity requires that the second order parameters must be symmetric. Thus, the total costs and input price terms are normalised by the first input price $W_1$, in order to impose a linear homogeneity restriction on the model. In addition, the standard symmetry restrictions $\varphi_{lj} = \varphi_{jl}$, $\eta_{mn} = \eta_{nm}$ and $\xi_{ks} = \xi_{sk}$ apply to the above cost function.

# 3 Data and Variables

## 3.1 Data

Our sample consists of an unbalanced panel which covers 41 Chinese banks over the period from 1994 to 2007, and contains a total of 397 observations. The sample comprises the big four state-owned banks, three policy banks, twelve national and regional joint-stock banks, sixteen city commercial banks and six foreign banks. At the end of 2007, these 41 banks owned almost 80% of the total assets of Chinese banking institutions. The bank-level data were mainly extracted from the Almanac of China's Finance and Banking issued by the China Finance Society (1995-2008) and BankScope – Fitch's International Bank Database. Missing information, additional data and double checks were made from other data sources, such as individual banks' statutory annual financial reports, the China Banking Regulatory Commission's internal database, the KPMG (China) internal database, the China Economic Information Network (www.cei.gov.cn), the China Statistical Yearbook and the China Labour Statistical Yearbook. All monetary data have been deflated to a common year 1994 using the Chinese GDP deflator. The stochastic frontier models described in the previous section require three sets of variables: (1) input prices and outputs; (2) control variables; and (3) environmental variables. The following section provides a detailed definition and description of the three sets of variables used in our modelling procedures.

## 3.2 Variable definitions

### *3.2.1 Outputs, input prices and total costs*

Following most other empirical banking efficiency studies, we adopt the intermediation approach in order to define the outputs and inputs (price) of banking services[8]. The outputs are specified as total loans ($Q_1$), other earning assets ($Q_2$) and non-interest income ($Q_3$)[9]. The inputs consist of labour ($X_1$), total physical capital ($X_2$) and deposits plus other borrowed funds ($X_3$). The methods used for measuring cost efficiency also require the

[8] The intermediation approach, suggested by Sealey and Lindley (1977), treats a bank as an intermediary, which collects funds from savers and transforms those funds into earning assets. This approach particularly emphasizes the overall costs of banks, and is appropriate for addressing questions related to cost minimization by the affected banks (Ferrier and Lovell, 1990).

[9] The non-interest income which acts as a proxy for non-traditional activities; that is, off-balance sheet items. Although off-balance sheet items are technically non-earning assets, they increase the bank's income and are an important component of banking business. Therefore, it should be included when modelling a bank's cost characteristics; otherwise, total output would be understated (Jagtiani and Khanthavit, 1996, Rogers, 1998 and Clark and Siems, 2002).

total cost and the market prices of inputs for all banks. Total cost (TC) includes both interest and operating expenses. The price of labour ($W_1$) is measured by the ratio of personnel expenses to the number of employees[10]. The price of physical capital ($W_2$), which is also called the user cost of capital, is defined as the ratio of other operating expenses to the book value of fixed assets (net of depreciation)[11]. The price of deposits plus other borrowed funds ($W_3$) is calculated by the ratio of total interest expenses on borrowed funds to total borrowed funds.

### *3.2.2 Control Variables*

In addition to the above input and output variables, this study also incorporates three control variables which are used to attempt to address the omitted variables problem and to account for the heterogeneity of our sample of banks. The first control variable is the level of equity ($Z_1$) which is included as a quasi-fixed input in the banking cost function. At this point it needs to be emphasized that the level of equity is an important aspect of efficiency measurement. Berger and Mester (1997) argue that a bank's insolvency risk depends on the level of its equity capital since it provides a cushion against portfolio losses and financial distress. Insolvency risk (non-performing loans) influences the bank's costs through the risk premium which the bank has to pay for its borrowings. This issue is particularly important in the Chinese banking sector where the insolvency risk of a bank could potentially be very high because of a large proportion of non-performing loans in its asset portfolio. However, equity capital is more than just a cushion against insolvency. The level of a bank's equity capital also provides an alternative to deposits and other borrowed funds as a source of loanable funds. Thus, the level of a bank's equity capital may have a direct impact on the bank's other borrowing costs. Incorporating the level of equity capital into the estimated cost function is also intended to control for a bank's different risk preferences[12]. Banks lever their equity capital with demandable debt to reflect their attitudes toward risk. If some banks are more risk averse than others, they may choose a higher level of equity capital than those that are less risk averse. Since a bank's equity capital is typically more expensive than deposits, this could lead to the conclusion that the risk averse bank produces its outputs in an allocatively inefficient manner; that is, by using the wrong input mix. However, an alternative explanation is that the relative levels of equity capital across banks are actually due to

[10]Some personnel expenses figures were not available for the early years of the sample period (about 20% of data missing in our sample). Hence, when the personnel expenses figures were not available from a bank's financial statement we assume that the growth rate in the unit price of labour matches the growth rate in the average wage rate for the Chinese financial sector. Detailed information about average wages and salaries for China's financial sector is published in the China Statistical Yearbook and the China Labour Statistical Yearbook.
[11] Other operating expenses are calculated as the operating expenses less expenses on employees (that is, wages, salaries and other benefits provided to employees).
[12] Hughes and Mester (1993) and Hughes et al. (1996) tested and rejected the assumption of risk neutrality for banks.

different risk preferences (Mester 1996). Therefore, the level of equity capital should be taken into account in the bank's production process.

Following Hughes and Mester (1993) and Mester (1996, 1997), another important control variable included in the cost function is that of non-performing loans ($Z_2$). This captures the quality of a bank's assets as well as the probability of bank failure and can influence a bank's costs in a number of ways. On the one hand, problem loans would be endogenous to the bank[13]. A large proportion of problem loans may be due to 'bad management'. Inefficient banks do not practice adequate loan underwriting and monitoring and hence will sustain higher losses due to non-performing loans. Problem loans may also be caused by short-run cost savings on the initial credit evaluation and loan monitoring ('skimping'). This would produce short term cost efficiencies that would be artificially inflated to higher levels than a bank which spends adequate resources to ensure that its loans are of good quality. In other words, some banks might be incorrectly labelled as inefficient merely because they produce higher quality outputs than other banks. On the other hand, problem loans are equally likely to be exogenous to the bank due to negative economic shocks ('bad luck'). That is, exogenous events can increase the amount of problem loans. As a consequence, the bank incurs extra administrative expenses and managerial efforts in order to alleviate the effects that these problem loans have on their operating activities. These extra operating costs lead to a reduction in cost efficiency. Controlling for non-performing loans in cost functions offers a way of removing, by statistical means, the costs of dealing with problem loans. Finally, the time trend variable ($Z_3$) is included in the stochastic cost function in order to control for the effects of technical progress over time. The time trend is a 'catch all' variable which captures the effect of technological factors, such as 'learning by doing' and organisational changes allowing for the more efficient use of existing inputs.

### *3.2.3 Environmental variables*

In our study, the five environmental variables are also incorporated into the model to account for heterogeneity across banks. These environmental variables ($E_i$) cannot usually be controlled by bank managers, or at least are partially exogenous. The first category of environmental variables included in our analysis is the ownership structure of banks. Specifically, the ownership structure variable is designed to capture differences that may arise between state-owned, domestic private and foreign banks (see DeYoung and Nolle 1996; Hasan and Marton 2003; and Berger et al. 2009 for details). The second environmental variable included in our efficiency models

[13] Berger and DeYoung (1997) tested the bad management, skimping and bad luck hypotheses and found mixed evidence for the exogeneity of non-performing loans. See Berger and DeYoung (1997) for further discussion.

is the size of banks. This is taken into account in order to control for potential scale biases in the estimating process (Berger and Mester 1997, Casu and Girardone 2006 and Kumbhakar and Wang 2007). Bank size may be an important determinant of net interest margins and spreads if there are economies of scale in the Chinese banking sector. In other words, one bank may be more efficient than another as a result of the economies of scale that arise from size rather than because of better management. A key objective of deregulation and liberalisation of banking operations is to improve resource allocation and banking performance (Berger and Humphrey, 1997). Therefore, we also include an environmental variable designed to capture the impact of World Trade Organisation (WTO) accession on Chinese banking efficiency[14]. The next environmental variable is included in order to capture the fact that the efficiency of listed banks may be improved because of the market discipline mechanism and better corporate governance imposed by listing the company on the stock exchange. Once a bank goes public, it becomes subject to legal, regulatory, and disclosure requirements which usually lead to better corporate governance practices and which impose additional external monitoring procedures on the management of the bank (Berger and Mester 1997;Uchida and Satake 2009). Therefore, it might be expected that banks with shares listed on the stock exchange would be more efficient, all else being equal. The characteristics of the market structure may also influence a bank's profitability and operational efficiency (see Demsetz 1973; Berger 1995; Berger and Mester 1997; Isik and Hassan 2003). Thus, this study also includes two environmental variables which characterise the competitive conditions and structure of the market in which banks operate. The Herfindahl-Hirschman index (HHI) measures the degree of market concentration and market share proxies for relative market power.

The definitions of all the variables used are presented in Table 2 and the statistics relating to those variables are summarized in Table 3. Table 3 shows the mean, standard deviation and other statistics of the variables across the 41 banks employed in our empirical analysis for the period from 1994 until 2007. All monetary variables have been deflated to a common year – 1994 - using the Chinese GDP deflator.

[14] After accession to the WTO, the Chinese accelerated the pace of deregulation and liberalization in the banking industry. For example, the geographical and client restrictions on foreign banks were gradually lifted and interest rates were further

**Table 2 Variable definitions**

| Variable name | Description |
| --- | --- |
| Total loans ($Q_1$) | Including total customer loans, trade bills, bills discounted, entrusted loans and impaired loans, but excluding loan loss reserves. |
| Total other earning assets ($Q_2$) | Including balances due from the central bank and other depository institutions, inter-bank loans, investment in securities and other investments, but excluding investment loss reserves. |
| Non-interest income ($Q_3$) | Including net fees and commissions, gains on foreign exchange transactions, gains on investment and other operating income. |
| Price of labour ($W_1$) | Total interest expenses on borrowed funds divided by total borrowed funds. |
| Price of physical capital ($W_2$) | Other operating expenses divided by fixed assets. |
| Price of funds ($W_3$) | Personnel expenses divided by the number of employees. |
| Total costs | Including total interest expenses, personnel expenses and other operating expenses. |
| The control variables | |
| Equity capital ($Z_3$) | The level of equity |
| Non-performing loans ($Z_2$) | The level of non-performing loans |
| Time trend ($Z_3$) | T=1 for 1994, T=2 for 1995,……, T=14 for 2007 |
| Environmental variables | |
| State-owned banks ($E_1$) | A dummy variable that takes the value of one if a bank's government agency controlled ownership is greater than 50% of total ownership and the value of zero otherwise. |
| Foreign banks ($E_2$) | A dummy variable that takes the value of one if a bank's foreign ownership is greater than 50% of total ownership and the value of zero otherwise. |
| Domestic private banks | A dummy variable that takes the value of one if a bank's private domestic ownership is greater than 50% of total ownership and the value of zero otherwise. |
| Bank size ($E_3$) | Natural logarithm of total assets. |
| WTO accession ($E_4$) | A dummy variable that takes the value of one for banks in the post WTO period (2002-2007) and zero for the pre-WTO period (1994-2001). |
| Listed banks ($E_5$) | A dummy variable that takes the value of one if a bank's shares are publicly traded on a stock exchange and the value of zero if they are not. |
| HHI ($E_6$) | HHI is defined as the sum of the squared asset market shares of all banks. |
| Market share ($E_7$) | MS is defined as the ratio of an individual bank's total assets to the total assets of all banks in a given year. |

Note: The domestic private bank dummy is dropped from the regression equation (used as a reference group) to avoid problems with multicollinearity.

**Table 3. Descriptive Statistics of the Variables**

| Variable name | Mean | St. Dev | Min | Max |
| --- | --- | --- | --- | --- |
| Total costs* | 24584 | 65577 | 13 | 633080 |
| Total loans * | 296686 | 530281 | 182 | 2572235 |
| Total other earning assets * | 176090 | 377239 | 77 | 2887446 |
| Non-interest income * | 2270 | 4863 | 1 | 43680 |
| Price of funds* | 0.0331 | 0.0311 | 0.0052 | 0.2421 |
| Price of physical capital* | 0.5612 | 0.5084 | 0.0852 | 6.9304 |
| Price of labour* | 0.0736 | 0.0441 | 0.0114 | 0.2611 |
| Equity* | 24592 | 55867 | 79 | 571795 |
| Non-performing loans* | 52723 | 128523 | 0 | 644503 |
| Time trend | 8.751 | 3.758 | 1 | 14 |
| State-owned banks | 0.5202 | 0.5002 | 0 | 1 |
| Foreign banks | 0.0682 | 0.2524 | 0 | 1 |
| Private banks (as a reference group) | 0.4116 | 0.4931 | 0 | 1 |
| Bank size | 5.7102 | 5.9818 | 2.4634 | 6.7377 |
| Listed banks | 0.1607 | 0.3677 | 0 | 1 |
| Post-WTO period | 0.5592 | 0.4971 | 0 | 1 |
| Herfindahl-Hirschman index | 0.1489 | 0.0307 | 0.1133 | 0.2405 |
| Asset market share | 0.0353 | 0.0664 | 0.00002 | 0.3460 |

Note: all financial values are inflation-adjusted to the base year 1994.
[a] Unit: RMB one million

liberalized.

# 4 Empirical results

### 4.1 Cost frontier estimates

All the stochastic frontier models are estimated using maximum likelihood techniques, based on the computer program FRONTIER 4.1 (Coelli, 1996). The maximum likelihood estimates of the parameters of the stochastic frontier cost functions are presented in Appendix. Before proceeding to analyse the parameter estimates of the various cost functions, it is worth noting that although the translog cost function is more flexible than other functional forms, multicollinearity may exist among the variables, thus leading to inconsistent parameter estimates[15]. However, multicollinearity may not be a serious problem when efficiency scores are used purely for forecasting purposes. The results shows that the parameter estimates of output quantities and input price terms are positive and highly significantly different from zero across all five model specifications. This suggests that the cost function is non-decreasing both for outputs ($Q$) and input prices ($W$) which are the theoretical requirements for a valid cost function[16]. Therefore, the domain of applicability for the estimated parameters is at least congruent with the data points. In addition, the empirical estimates of the translog cost functions summarized in Appendix are compatible with the intuition since the output and input price variables have the expected signs (both positive).

### 4.2 Key Estimation Results

Table 4 summarises some additional key information besides the parameter estimates for our five cost frontiers. In particular, the parameters determining the shape and location of the inefficiency distribution are shown in the first three columns of the table. The inefficiency location parameter, $\mu$, is significantly different from zero for environmental factors in error specification (M5). This may be explained by the fact that we introduced heterogeneity into the efficiency distribution for this model. However, estimates of $\mu$ in alternative models are not significantly different from zero at the 5% level. The $\gamma$ parameters corresponding to the estimated proportion of bank inefficiency in the composite total error term are significantly different from zero in both the baseline model (M1) and all the alternative models. This parameter produces high values (close to unity) in the models (M2, M3, M4 and M5) which account for heterogeneity, revealing that most of the variations in observed costs

[15] If the multicollinearity problem is mainly created by a strong positive correlation between the second order terms in the translog form of the cost function, maximum likelihood estimates are still unbiased and efficient. But, in such circumstances, multicollinearity problems cause the estimated standard error of the coefficients to be large, leading to small values for the t-ratios. This in turn biases results towards accepting the null hypothesis that the coefficients are equal to zero (see Gujarati 2003 for more detail).

[16] The Hessians of the cost function with respect to input prices for all models are negative semi-definite. This suggests that the concavity of the cost function in input prices is satisfied.

from the frontier are due to bank inefficiency. The difference between the $\gamma$ coefficients for Models 4 and 5 is likely to be explained by the way in which the environmental factors are included in these two models. In addition, the magnitude of the variance parameter $\sigma^2$ in Models 2, 3, 4 and 5 is larger when compared to the baseline specification (Model 1).

**Table 4 Key Estimation Results**

| Model specification | μ | γ | $\sigma^2$ | log likelihood | LR test of one-sided error |
|---|---|---|---|---|---|
| Baseline (M1) | 0 | 0.5115*** | 0.0557*** | 88.8902 | 2.4076 (3.841) |
| Truncated (M2) | -6.9109* | 0.9637*** | 0.7414* | 91.6750 | 7.977 (5.991) |
| Controlled (M3) | -7.4554* | 0.9821*** | 0.8376* | 164.3368 | 26.2449(5.991) |
| Kernel (M4) | -0.8215 | 0.9888*** | 0.1001 | 185.0729 | 36.2558 (5.991) |
| Error effects (M5) | 2.2074*** | 0.8639*** | 0.1237*** | 175.1729 | 47.9171(16.919) |

Notes:
1. $\sigma^2 = \sigma_u^2 + \sigma_v^2$; $\gamma = \sigma_u^2 / (\sigma_u^2 + \sigma_v^2)$;
2. $\chi^2$ critical values for 5% significance level are in parentheses;
3. ***, ** and * indicate 1%, 5% and 10% significance levels, respectively.

The logarithmic values of the likelihood function, a frequently used criterion for calculating more accurate statistical properties of an econometric model estimated though the maximum likelihood technique, are presented in the fourth column of Table 4. We find that the log-likelihood values for Models 3, 4 and 5 are higher than the baseline (M1) and truncated specifications (M2), suggesting that including a set of explanatory variables in the specification to account for heterogeneity improves the fit significantly. The last column in Table 4 reports the results of one-sided log-likelihood ratio (LR) tests of the standard response function (OLS) versus the full frontier model. The null hypothesis in this test is $\gamma = 0$ versus the alternative of $\gamma > 0$. If the null hypothesis is accepted, this could indicate that $\sigma_u^2$ and $\delta_i$ are zero and hence that there are no inefficiency effects present in the cost function, leaving a specification with parameters that can be appropriately estimated using the method of ordinary least squares (OLS) (Coeli, 1996). If, however, the null hypothesis is rejected, this could suggest that a standard mean response function is not an adequate representation of the data. In the case of the baseline model (M1), the null hypothesis is accepted at the 5% level of significance, suggesting that the stochastic frontier analysis (SFA) provides an inappropriate specification for the Chinese banking data. However, in the case of alternative models, the null hypothesis is rejected in favour of the stochastic frontier cost function. Therefore, the results of the baseline model should be treated with caution or perhaps even discarded, while the results of the alternative models (M2, M3, M4 and M5), which successfully account for heterogeneity, appear to provide a more faithful fit to the available data from the Chinese banks.

### 4.3 Model Specification Tests

In order to investigate whether one model specification provides a better fit to the sample data than the other, generalised likelihood ratio tests were conducted. These tests provide a convenient way to check whether a reduced (restricted) model provides the same fit as a general (unrestricted) model. Table 5 presents the steps taken, and the results of the log-likelihood tests which were carried out.

**Table 5 Specification Tests**

| Model Description | Restriction | Log likelihood | LR test of one-side error | Critical value $\alpha = 5\%$ | $H_0$: restricted model is better than unrestricted model |
|---|---|---|---|---|---|
| Step 1: Half-normal vs. Truncated normal | | | | | |
| -Truncated normal distribution (M2) | Unrestricted | 91.675 | | | |
| -Normal distribution (M1) | 1 | 88.890 | 5.57 | 3.841 (1) | Reject $H_0$ |
| Step 2: Model without control variables vs. Model with control variables | | | | | |
| Truncated normal with all control variables (M3) | Unrestricted | 164.337 | | | |
| -without all control variable parameters (M2) | 24 | 91.675 | 145.324 | 36.415 (24) | Reject $H_0$ |
| Step 3: Models without environmental variables vs. Models with environmental variables | | | | | |
| Including environmental variables in cost function (M4) | Unrestricted | 185.073 | | | |
| - not including environmental variables in cost function (M3) | 7 | 164.337 | 41.472 | 14.067 (7) | Reject $H_0$ |
| Including environmental variables in error effects (M5) | Unrestricted | 175.172 | | | |
| - not including environmental variables in error effects (M4) | 7 | 164.337 | 27.67 | 14.067 (7) | Reject $H_0$ |

In Table 5, the first step involves testing the half-normal model (M1) against the truncated normal model (M2). The generalized log likelihood ratio test statistic is 5.57, which is greater than the $\chi^2$ critical value at the 5% level with one degree of freedom. Therefore, this suggests that the truncated-normal distribution model is more compatible with the data than the half-normal distribution model. Based on the first step decision, the second step is to test the hypothesis which states that the truncated model without any control variables is more compatible with the data than the model that includes all three control variables (equity, non-performing loans, and time trend data being equal to zero). We reject the hypotheses according to the log-likelihood ratio statistics. Thus, the results suggest that the control variables have significant effects on total costs and should be included in the cost function frontier. The next step involves examining whether including the environmental variables in the model specification has significant explanatory power. The tests are done by comparing the model including the environmental variables in the deterministic kernel of the frontier (M4) and/or in the distribution of the inefficiency term (M5) with the truncated model including all the control variables (M3). The tests showed that both hypotheses can be rejected and thus we can conclude that environmental factors should not be ignored in the analysis of Chinese banking efficiency.

The question of whether the environmental variables should be treated as explanatory variables in cost function (M4) or as determinants of cost inefficiency (M5) is not directly answered by the generalised log likelihood ratio test. These two model specifications are not nested and hence no set of restrictions can be imposed which allow a test of one specification against the other[17]. Therefore, it is difficult to provide an unequivocal assessment as to whether the stochastic cost function specification M4 or the stochastic cost function specification M5 is more compatible with the Chinese banking data that is available to us.

On the basis of the above empirical results, however, it can be seen that sample heterogeneity significantly influences stochastic cost frontier estimation. Thus, any model of Chinese banking efficiency which is used for policy purposes should explicitly account for sample heterogeneity by introducing control variables and/or environmental variables as part of its argument. Failure to do so would result in mis-specification, leading to inappropriate parameter and efficiency estimates and, more importantly, potentially flawed policy decisions. However, choosing between a model that considers the environmental variables as a part of the deterministic kernel of the frontier and a model that considers the environmental variables as determinants of cost efficiency is a difficult issue. Here, we would tend to favour the model that treats environmental variables as explanatory variables of cost efficiency. The reasons for this are: first, the improved significance of the critical parameters $\mu$ $\gamma$ and $\sigma^2$ in Model 5 and the insignificance of the coefficients for most of the environmental variables in Model 4; second, it also appears that the estimated frontier represents the outer boundary of the cost possibility set, irrespective of environmental issues (Coelli *et al*., 1999).

### 4.4 Efficiency Level

In this section, we will compare the cost efficiency levels derived from the five different models. Table 6 provides a statistical summary of the estimated efficiency scores of all the banks for the various models. Thus, the mean, median and the lowest and highest levels of efficiency for the models are presented in the table.

[17] We have constructed an artificial nested model that includes environmental variables both in the cost function and also as factors explaining cost inefficiency. We then tested the null hypotheses associated with the M4 and M5 against the artificial model. Both null hypotheses were rejected. Thus it cannot be argued that one approach provides a better fit to the sample data than the other.

Regarding the overall mean values of the cost efficiency scores for the entire period, the range is relatively small spanning from 87.26 % to 91.14%, indicating that the average bank in the sample could reduce its costs by approximately 9% to 13%, in order to match its performance with the best possible bank practice. Model 5 (heterogeneity in the inefficiency term) yields the highest mean and medium efficiency estimates, while the baseline (half-normal) model (M1) generates the lowest efficiency estimates. This result indicates that neglecting heterogeneity across banks may create a downward bias in efficiency scores. Moreover, explicitly accounting for heterogeneity in terms of ownership, size, market structure, etc. in the distribution of the inefficiency component leads to a mean cost efficiency that is approximately 2 to 4 percentage points higher than in the other specifications. However, the mean efficiency scores in Models 2, 3 and 4 are similar, suggesting that accounting for the heterogeneity in the efficiency frontier did not greatly influence efficiency estimates for these models. The maximum efficiency scores are relatively high for Models 3, 4 and 5, suggesting that heterogeneity across banks is an important driver of cost differences.

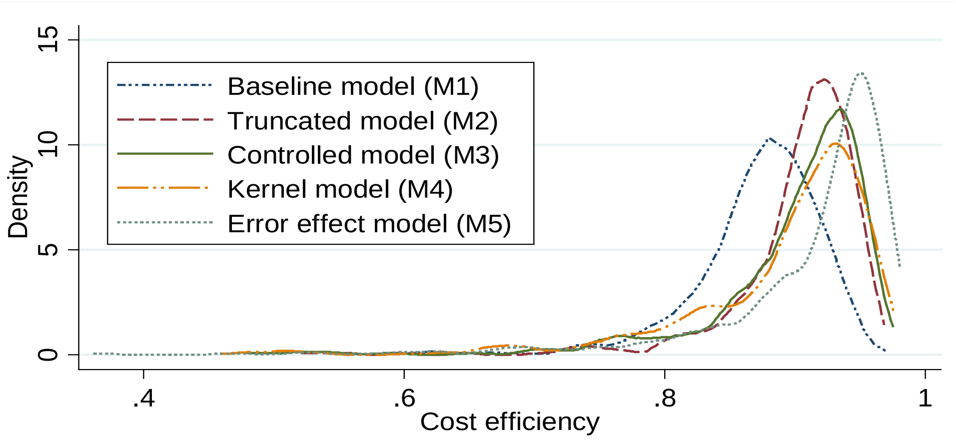
**Table 6 Summary Statistics of the Mean Efficiency Estimates**

| | Model 1 Baseline | Model 2 Truncated | Model 3 Controlled | Model 4 Kernel | Model 5 Error |
|---|---|---|---|---|---|
| Mean | 0.8726 | 0.9029 | 0.8983 | 0.8900 | 0.9114 |
| Median | 0.8795 | 0.9149 | 0.9166 | 0.9139 | 0.9379 |
| Standard deviation | 0.0480 | 0.0527 | 0.0668 | 0.0777 | 0.0789 |
| Maximum | 0.9581 | 0.9685 | 0.9752 | 0.9755 | 0.9808 |
| Minimum | 0.6179 | 0.5217 | 0.4802 | 0.4595 | 0.3625 |

Note: Efficiencies are calculated by using 14 years worth of data for 44 banks (397 observations) and figures in table are based on average efficiency for each bank over sample period.

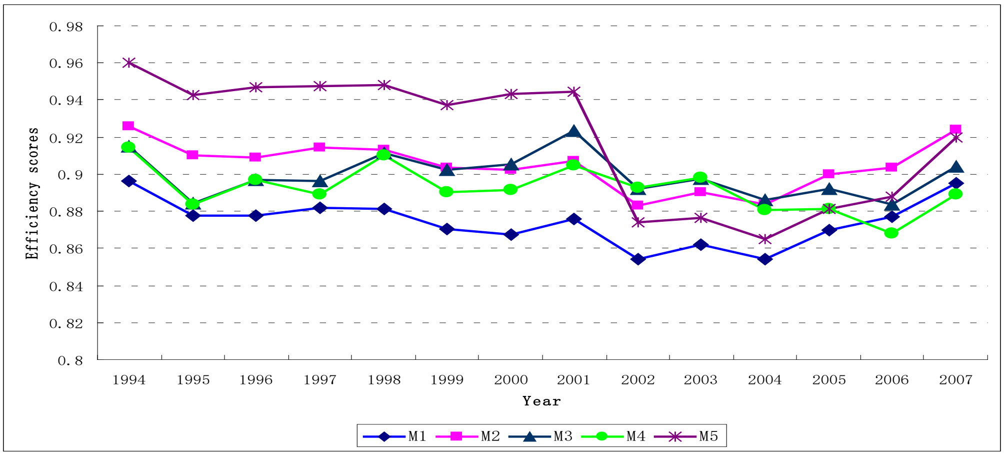
Figure 1 plots the estimated distribution of the cost efficiency scores for the five models. The different patterns of the distribution of estimated cost efficiency scores are due to the different ways of accounting for sample heterogeneity. Inefficiency may be significantly overestimated by the baseline model, as shown in Figure 1, in which sample heterogeneity is not taken into account. After accounting for the bank heterogeneity (Models 3, 4 and 5), the distribution of estimated cost efficiency scores shifts to the right. In particular, Model 5, which absorbs the bank heterogeneity both in the cost frontier and the inefficiency term, generates the highest level of estimated cost efficiency. These results suggest that controlling for heterogeneity is very important for estimating the level of efficiency. However, the efficiency estimates may be sensitive to the way in which we account for sample heterogeneity.

**Figure 1 Distributions of Estimated Cost Efficiency across Models**




The yearly mean cost efficiency of all the banks for the five different models is plotted in Figure 2. The trends or patterns of efficiency levels obtained from the five different specifications are broadly similar over time, especially for Models 3 and 4. In general, most banks showed relatively high efficiency in the early years (before 2002) but substantially less efficiency in the later years (after 2002). A significant decrease in efficiency levels appeared between 2001 and 2002 across all models. These emerging patterns may provide evidence that the 2001-2002 calendar year appeared to be associated with a structural change in the trend of cost efficiency and that this is associated with China's entry into the WTO which occurred at around this time.

**Figure 2 Average Efficiency Scores over Time**




### 4.5 Spearman's Correlation for Different SFA Models

Another potentially interesting comparison is whether the ranks obtained for the efficiency scores across the different specifications show any compatibility. The ranking of banks according to their cost efficiency scores can provide important information about the impact of structural change on banking efficiency. If different models rank banks completely differently, then it becomes difficult to draw any generalised conclusions. The Spearman rank correlation coefficients of the efficiency estimates are summarised in Table 7. These coefficients capture the similarities in the efficiency rankings across the various model specifications. In general, the rank correlations according to the efficiency scores among the first three models (M1, M2 and M3) are lower than the rank correlations between them and the last two models (that is M4 and M5). The near perfect correlation of efficiency rankings between the half-normal (M1) and truncated (M2) models suggests that these models identify the same banks as the best and worst performers. This shift in the inefficiency distribution seems to influence all banks in the sample to a very similar degree. The inclusion of control variables in the efficiency estimation leads to a decline in the rank correlation coefficient to around 0.8, indicating that this inclusion not only absorbs some heterogeneity but also affects competitive rankings for some banks. However, Models 3 and 4 show a very high correlation (0.94) in the estimated efficiency scores. This may suggest that introducing environmental factors into the kernel specification leads to only minor changes in the ranking order. We also find that Model 5 shows a relatively low correlation with other alternative models with rank order correlation coefficients ranging from 56% to 63%. These results suggest that Model 5, which includes environmental factors in the inefficiency term specification, ranks banking efficiency in a markedly different way when compared to the other four models. In sum, these results further improve our understanding of the effect of heterogeneity on efficiency estimates. It seems that accounting for heterogeneity is an important issue which, if not taken into account, may lead to biased estimates of bank efficiencies.

**Table 7 Spearman Rank Correlation between Efficiency Estimates**

| | Model 1 Baseline | Model 2 Truncated | Model 3 Controlled | Model 4 Kernel | Model 5 Error |
|---|---|---|---|---|---|
| Half-normal (M1) | 1.0000 | | | | |
| Truncated (M2) | 0.9994 | 1.0000 | | | |
| Controlled (M3) | 0.8148 | 0.8150 | 1.0000 | | |
| Kernel (M4) | 0.7623 | 0.7626 | 0.9433 | 1.0000 | |
| Error (M5) | 0.5778 | 0.5901 | 0.6360 | 0.5664 | 1.0000 |

Note: All correlations significant at 1% level

# 5 Conclusions

In this paper we estimated the cost efficiency of 41 Chinese banks over the period from 1994 to 2007 and examined the effects of sample heterogeneity on bank cost efficiency scores. The measures of cost efficiency were obtained from a number of well-established stochastic cost frontier models which attempt to account for heterogeneity across banks in different ways. This paper conducted a specification test to examine whether one model specification provides a better fit to the sample data than the others and discusses the effect of accounting for heterogeneity on parameter estimates for the cost frontier, the level of cost efficiency and efficiency ranking order.

For the sample of Chinese banks, it was found that the sample heterogeneity significantly influences some key stochastic cost frontier estimates and thus it can be concluded that, if the heterogeneity across banks is taken into account in models, then estimates of the stochastic cost frontier are likely to improve. In order to reach the best-specified stochastic cost frontier model, we followed a step-by-step specification testing procedure. The results suggest that the appropriate frontier model should incorporate both control variables and environmental variables in order to control for systematic differences across banks. We subsequently discussed the effects of applying different SFA specifications to banks' efficiency scores and ranking order. The results also indicate that it is important to control for heterogeneity across banks in frontier models. However, efficiency estimates are sensitive to the way in which we account for environmental variables.

This paper is important for more than just methodological matters. It also generated two findings which may be useful for policy-makers and bank mangers. First, it was found that the level of non-performing loans, a bank's ownership structure and its size all significantly affect both the bank's optimal costs as well as its ability to operate efficiently. Second, it was found that a significant decrease in efficiency levels appeared between 2001 and 2002 across all models. China's entry into the WTO occurred at around this time. This result suggests that the external environmental changes introduced in 2002 have had a significant negative impact on Chinese banking efficiency.

Finally, this study models heterogeneity in the stochastic frontier model framework by incorporating bank specific heterogeneity variables (observed heterogeneity) either in the cost function itself or as explanatory variables in a simultaneous regression model where cost inefficiency is the dependant variable. It is entirely possible, however, that the heterogeneity variables employed in our regression procedures are not complete and that our empirical analysis is therefore undermined by a problem caused by omitted variables. This, in turn, may also create potential biases in the estimates of our inefficiency scores. A potential way to address this problem is to use the 'true effects model' proposed by Greene (2005). Greene's model integrates an additional stochastic term into the traditional SFA model in order to distinguish all time invariant unobserved heterogeneities from the inefficiency term. Therefore, it would be very valuable to apply the 'true effects model' to our dataset in the future so as to shed some light upon the robustness of the results obtained in this study.

# Reference


Aigner D, Lovell CAK, Schmidt P (1977) Formulation and Estimation of Stochastic Frontier Production Function Models. J Econometrics 6: 21-37

Altunbas Y, Gardener EPM, Molyneux P, Moore B (2001) Efficiency in European Banking, Eur Econ Rev 45:1931-1955.

Battese GE, Coelli, TJ (1995) A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. Empir Econ 20: 325-332.

Battese GE, Coelli TJ (1988) Prediction of Firm-level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data. J Econometrics 38: 387-399.

Bauer PW, Berger AN, Ferrier GD, Humphrey DB (1998) Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Methods. J Econ Bus 50: 85-114.

Berger AN (1995) The Profit-Structure Relationship in Banking -Tests of Market-Power and Efficient-Structure Hypotheses. J Money Credit Bank 27: 404-431.

Berger AN (2007) International Comparisons of Banking Efficiency, Financial Markets, Inst, Instrum16:119-144

Berger AN, DeYoung R (1997) Problem Loans and Cost Efficiency in Commercial Banks. J Bank Financ 6: 849-870

Berger AN, Hasan I, Zhou M (2009) Bank Ownership and Efficiency in China: What will happen in the world's Largest Nation? J Bank Financ 33:113-130

Berger AN, Humphrey, DB (1997) Efficiency of Financial Institutions: International Survey and Directions for Future Research. Eur J of Operat Res 98: 175-212.

Berger AN, Mester LJ (1997) Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions? *J Bank Financ* 2: 895-947.

Bonin JP, Hasan I, Wachtel P (2005) Bank performance, efficiency and ownership in transition countries. J Bank Financ 29:31-53

Bos JWB, Koetter M, Kolari JW, Kool CJM (2009) Effects of Heterogeneity on Bank Efficiency Score. Eur J of Operat Res 195:251-261

Casu B, Molyneux P (2003) A Comparison Study of Efficiency in European Banking. Appl Econ 35: 1865-1876

Cavallo L, Rossi SPS (2002) Do Enviornmental variables affect the performance and technical efficiency of the European banking systems? A parametrical analysis using the stochastic frontier approach. Eur J of Financ 8:123-146

Casu B, Girardone C (2006) Bank Competition, Concentration and Efficiency In The Single European Market. Manchester School 74: 441-468.

Coelli TJ (1996) A Guide to FRONTIER version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation. CEPA Working Paper, University of New England

Coelli TJ, Perelman S, Romano E (1999) Accounting for Environmental Influences in. Stochastic Frontier Models: With Application to International Airline. J Prod Anal 11: 251-273.

Cousin V (2007) Banking in China. Palgrave Macmillan, UK

DeYoung R and Nolle DE (1996) Foreign-Owned Banks in the United States: Earning Market Share or Buying It?. J Money Credit Bank 28: 622-636

Fethi M, Pasiouras F (2010) Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. Eur J of Operat Res 204: 189-198.

Fries S, Taci A (2005) Cost Efficiency of Banks in Transition: Evidence from 289 Banks in 15 Post-communist Countries. J Bank Financ 29: 55-81.

Fu X, Heffernan S, (2007) Cost X-efficiency in China's Banking Sector. China Econ Rev 18: 35-53.

Greene WH (2005) Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model. J Econometrics 126: 269-303.

Hasan I, Marton K (2003) Development and Efficiency of the Banking Sector in a Transitional Economy: Hungarian Experience. *J Bank Financ* 27: 2249-2271.

Hughes JP, Lang W, Mester LJ, Moon C (1996) Efficient banking under interstate branching. J Money Credit Bank 28:1045-1071

Hughes JP, Mester LJ (1993) A Quality and Risk-adjusted Cost Function for Banks: Evidence on the too-big-to-fail Doctrine. J Prod Anal 4: 293-315

Hughes JP, Mester LJ (2008) *Efficiency in Banking: Theory, Practice, and Evidence*. In Berger AN, Molyneux P, Wilson JOS (eds.) The Oxford Handbook of Banking, Oxford

Isik I, Hassan MK (2003) Efficiency, Ownership and Market Structure, Corporate Control and Governance in the Turkish Banking Industry. J Bus Financ Account 30:1363-1421.

Jiang CX, Yao SJ, Zhang ZY (2009) The effects of governance changes on bank efficiency in China: a stochastic distance function approach. China Econ Rev 20:717-731

Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. *J Econometrics*19: 233-238.

Kalirajan K (1981) An econometric analysis of yield variability in paddy production. Canadian J Agric Econ 29:

283-294

Kumbhakar SC, Lovell CAK (2000) Stochastic Frontier Analysis, Cambridge University Press, Cambridge

Kumbhakar S, Wang, D (2007) Economic Reforms, Efficiency and Productivity in Chinese Banking J of Regulatory Econ 32:105-129.

McKillop DG, Glass JC, Ward A (2005) Cost Efficiency, Environmental Influences and UK Credit Unions, 1991 to 2001. Managerial Financ 31: 72-86

Meeusen W, van Den Broeck J (1977) Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. Int Econ Rev 18: 435-444.

Mester LJ (1993) Efficiency in the Savings and Loan Industry. J bank Financ 17: 267-286.

Mester LJ (1996) A Study of Bank Efficiency Taking into Account Risk-preferences. J Bank Financ 20:1025-1045

Mester LJ (1997) Measuring Efficiency at U.S. Banks: Accounting for Heterogeneity is Important. Eur J of Operat Res 98: 230-242.

Pitt MM, Lee L (1981) The measurement and sources of technical inefficiency in the Indonesia weaving industry. J Dev Econ 9: 43-64

Poghosyan T, Kumbhakar S (2010) Heterogeneity of technological regimes and banking efficiency in former socialist economies. J Prod Anal 33:19-31

Rezvanian R, Mehdian S (2002) An Examination of Cost Structure and Production Performance of Commercial Banks in Singapore. J Bank Financ 26: 79-98.

Sensarma R (2006) Are Foreign Banks Always the Best? Comparison of State-owned, Private and Foreign Banks in India Econ Modelling 23: 717-735.

Stevenson RE (1980) Likelihood Functions for Generalised Stochastic Frontier Estimation. J Econometrics 13: 57-66.

Tecles PL, Tabak BM (2010) Determinants of bank efficiency: The case of Brazil. Eur J of Operat Res 207:1587-1598

Uchida H, Satake M (2009) Market Discipline and Bank Efficiency. *J* International Financial Market Inst Money 19: 792-802.

## Appendix

**Maximum Likelihood Parameter Estimates for Stochastic Frontier Cost Functions**

| Variables | Parameter | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|---|
| Constant | $\beta_0$ | -.2289*** (0.0347) | -0.1944*** (0.0177) | -0.2515*** (0.0239) | -3.6257*** (0.9652) | -0.2173*** (0.0253) |
| $\ln Q_1$ | $\beta_1$ | 0.6431*** (0.0218) | 0.6419*** (0.0216) | 0.5484*** (0.0261) | 0.3429*** (0.0582) | 0.5526*** (0.0264) |
| $\ln Q_2$ | $\beta_2$ | 0.3101*** (0.0253) | 0.3117*** (0.0248) | 0.2693*** (0.0246) | 0.1600** (0.0371) | 0.2941*** (0.0250) |
| $\ln Q_3$ | $\beta_3$ | 0.0524*** (0.0148) | 0.0536*** (0.0148) | 0.0385*** (0.0140) | 0.0316*** (0.0136) | 0.0314** (0.0140) |
| $\ln(W_1/W_3)$ | $\chi_1$ | 0.7903*** (0.0156) | 0.7919*** (0.0155) | 0.7073*** (0.0207) | 0.7022*** (0.0214) | 0.7259*** ()0.0211 |
| $\ln(W_2/W_3)$ | $\chi_2$ | 0.1074*** (0.0206) | 0.1075*** (0.0203) | 0.1298*** (0.0190) | 0.1413*** (0.0185) | 0.1246*** (0.0195) |
| $0.5 \ln Q_1 \ln Q_1$ | $\varphi_{11}$ | 0.1016*** (0.0241) | 0.1037*** (0.0238) | 0.1037 (0.0639) | 0.1072* (0.0629) | 0.1013 (0.0648) |
| $\ln Q_1 \ln Q_2$ | $\varphi_{12}$ | -0.0615** (0.0276) | -0.0640** (0.0267) | -0.1043*** (0.0371) | -0.0789** (0.0365) | -0.1288*** (0.0368) |
| $\ln Q_1 \ln Q_3$ | $\varphi_{13}$ | -0.0152 (0.0093) | -0.0156* (0.0092) | 0.0067 (0.0204) | -0.0067 (0.0197) | 0.0148 (0.0207) |
| $0.5\ln Q_2 \ln Q_2$ | $\varphi_{22}$ | 0.0664* (0.0400) | 0.0689* (0.0385) | 0.1108*** (0.0381) | 0.0784** (0.0379) | 0.1067*** (0.0370) |
| $\ln Q_2 \ln Q_3$ | $\varphi_{23}$ | -0.0021 (0.0176) | -0.0022 (0.0174) | -0.0206 (0.0160) | -0.0145 (0.0155) | -0.0098 (0.0154) |
| $0.5\ln Q_3 \ln Q_3$ | $\varphi_{33}$ | 0.0141 (0.0123) | 0.0145 (0.0121) | 0.0112 (0.0120) | 0.0074 (0.0117) | 0.0070 (0.0112) |
| $0.5\ln(W_1/W_3)\ln(W_1/W_3)$ | $\eta_{11}$ | 0.0412 (0.0240) | 0.0403* (0.0237) | 0.0605* (0.0337) | 0.0410 (0.0327) | 0.0656* (0.0345) |
| $\ln(W_1/W_3) \ln(W_2/W_3)$ | $\eta_{12}$ | -0.0359* (0.0309) | -0.0363 (0.0305) | -0.0293 (0.0334) | 0.0057 (0.0324) | -0.0253 (0.0335) |
| $0.5\ln(W_2/W_3)\ln(W_2/W_3)$ | $\eta_{22}$ | -0.0077 (0.0377) | -0.0071 (0.0374) | 0.0184 (0.0338) | -0.0288 (0.0337) | 0.0101 (0.0343) |
| $\ln Q_1 \ln(W_1/W_3)$ | $\iota_{11}$ | 0.0223 (0.0208) | 0.0213 (0.0201) | 0.0584* (0.0347) | 0.0468 (0.0338) | 0.0480 (0.0344) |
| $\ln Q_1 \ln(W_2/W_3)$ | $\iota_{12}$ | -0.0034 (0.0216) | -0.0027 (0.0207) | -0.0046 (0.0304) | 0.0199 (0.0294) | 0.0097 (0.0313) |
| $\ln Q_2 \ln(W_1/W_3)$ | $\iota_{21}$ | 0.0118 (0.0282) | 0.0120 (0.0272) | 0.0431 (0.0345) | 0.0113 (0.0326) | 0.0238 (0.0323) |
| $\ln Q_2 \ln(W_2/W_3)$ | $\iota_{22}$ | -0.0304 (0.0338) | -0.0281 (0.0326) | -0.0684** (0.0329) | -0.0442 (0.0315) | -0.0349 (0.0323) |
| $\ln Q_3 \ln(W_1/W_3)$ | $\iota_{31}$ | -0.0140 (0.0135) | -0.0143 (0.0134) | -0.0282 (0.0176) | -0.0217 (0.0165) | -0.0223 (0.0167) |
| $\ln Q_3 \ln(W_2/W_3)$ | $\iota_{32}$ | 0.0095 (0.0165) | 0.0100 (0.0163) | 0.0221 (0.0151) | 0.0221 (0.0143) | 0.0155 (0.0151) |
| Control variables | | | | | | |
| $\ln Z_1$ | $\rho_1$ | - | - | 0.0343 (0.0213) | 0.0384* (0.0208) | 0.0361* (0.0211) |
| $\ln Z_2$ | $\rho_2$ | - | - | 0.0942*** (0.0120) | 0.0796*** (0.0117) | 0.0876*** (0.0121) |
| T | $\rho_3$ | - | - | 0.0032 (0.0043) | -0.0231*** (0.0117) | -0.0024 (0.0044) |
| $0.5\ln Z_1 \ln Z_1$ | $\xi_{11}$ | - | - | 0.0083 (0.0192) | 0.0148 (0.0189) | 0.0067 (0.0199) |
| $\ln Z_1 \ln Z_2$ | $\xi_{12}$ | - | - | 0.0002 (0.0115) | -0.0030 (0.0111) | -0.0070 (0.0118) |
| $\ln Z_1 T$ | $\xi_{13}$ | - | - | -0.0252*** (0.0076) | -0.0136* (0.0077) | -0.0206*** (0.0078) |
| $0.5 \text{Ln} Z_2 \ln Z_2$ | $\xi_{22}$ | - | - | 0.0141*** (0.0025) | 0.0109*** (0.0026) | 0.0098*** (0.0028) |
| $\ln Z_2 T$ | $\xi_{23}$ | - | - | 0.0068** (0.0031) | 0.0012 (0.0032) | 0.0008 (0.0035) |
| $0.5\ln Z_3 T$ | $\xi_{33}$ | - | - | 0.0002 (0.0024) | 0.0010 (0.0027) | -0.0004 (0.0024) |

Maximum Likelihood Parameter Estimates for Stochastic Frontier Cost Functions (continued)

| Variables | Parameter | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|---|
| Control variables | | | | | | |
| $\ln Z_1 \ln Q_1$ | $\psi_{11}$ | - | - | 0.0092 (0.0366) | -0.0199 (0.0375) | 0.0125 (0.0373) |
| $\ln Z_1 \ln Q_2$ | $\psi_{12}$ | - | - | 0.0309 (0.0324) | 0.0351 (0.0323) | 0.0352 (0.0327) |
| $\ln Z_1 \ln Q_3$ | $\psi_{13}$ | - | - | -0.0114 (0.0200) | 0.0002 (0.0194) | -0.0226 (0.0202) |
| $\mathrm{Ln} Z_2 \ln Q_1$ | $\psi_{21}$ | - | - | -0.0179 (0.0138) | -0.0096 (0.0134) | -0.0086 (0.0145) |
| $\ln Z_2 \ln Q_2$ | $\psi_{22}$ | - | - | -0.0020 (0.0115) | -0.0027 (0.0112) | 0.0066 (0.0120) |
| $\ln Z_2 \ln Q_3$ | $\psi_{23}$ | - | - | 0.0027 (0.0065) | 0.0068 (0.0061) | 0.0010 (0.0066) |
| $T \ln Q_1$ | $\psi_{31}$ | - | - | 0.0073 (0.0084) | 0.0088 (0.0081) | 0.0132 (0.0088) |
| $T\ln Q_2$ | $\psi_{32}$ | - | - | 0.0077 (0.0078) | 0.0006 (0.0074) | 0.0074 (0.0079) |
| $T\ln Q_3$ | $\psi_{33}$ | - | - | -0.0017 (0.0053) | 0.0000 (0.0049) | -0.0012 (0.0053) |
| $\ln Z_1 \ln(W_1/W_3)$ | $\theta_{11}$ | - | - | -0.0784*** (0.0293) | -0.0500* (0.0283) | -0.0616** (0.0298) |
| $\ln Z_1 \ln(W_2/W_3)$ | $\theta_{12}$ | - | - | 0.0390 (0.0301) | 0.0194 (0.0293) | 0.0266 (0.0305) |
| $\ln Z_2 \ln(W_1/W_3)$ | $\theta_{21}$ | - | - | 0.0106 (0.0104) | 0.0098 (0.0101) | 0.0099 (0.0110) |
| $\ln Z_2 \ln(W_2/W_3)$ | $\theta_{22}$ | - | - | 0.0014 (0.0092) | -0.0134 (0.0094) | -0.0066 (0.0097) |
| $T\ln(W_1/W_3)$ | $\theta_{31}$ | - | - | 0.0005 (0.0069) | 0.0020 (0.0067) | 0.0032 (0.0074) |
| $T\ln(W_2/W_3)$ | $\theta_{32}$ | - | - | 0.0030 (0.0063) | 0.0056 (0.0063) | 0.0023 (0.0066) |
| Environmental variables | | | | | | |
| state-owned banks | $\delta_1'$ | - | - | - | -0.0209 (0.0257) | - |
| foreign banks | $\delta_2'$ | - | - | - | -0.2591*** (0.0732) | - |
| size | $\delta_3'$ | - | - | - | 0.3156*** (0.0861) | - |
| listed banks | $\delta_4'$ | - | - | - | -0.0116 (0.0299) | - |
| deregulation | $\delta_5'$ | - | - | - | 0.1114*** (0.0340) | - |
| HHI | $\delta_6'$ | - | - | - | -1.8682 (1.2535) | - |
| market share | $\delta_7'$ | - | - | - | -3.6257 (0.9652) | - |
| intercept | $\delta_0$ | - | - | - | - | 2.2074*** (0.5199) |
| state-owned banks | $\delta_1$ | - | - | - | - | -0.2516** (0.1065) |
| foreign banks | $\delta_2$ | - | - | - | - | -0.1061*** (0.2899) |
| size | $\delta_3$ | - | - | - | - | -0.3563*** (0.1031) |
| listed banks | $\delta_4$ | - | - | - | - | -0.0066 (0.0857) |
| deregulation | $\delta_5$ | - | - | -- | - | 1.2117** (0.5284) |
| HHI | $\delta_6$ | - | - | - | - | -1.5270 (1.2195) |
| market share | $\delta_7$ | - | - | - | - | -7.2047*** (2.4588) |

Notes: 1. ***, ** and * indicate 1%, 5% and 10% significance levels, respectively.
2. Asymptotic standard errors in parentheses.