X-efficiency of Chinese Listed Commercial Banks: A Recursive Thick Frontier Approach

by

Rui Chen

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Arts, Honours
in the Department of Economics
University of Victoria
April 2016

Supervised by Dr. Paul Schure

for

Dr. Chris Auld, Honours co-advisor

Dr. Merwan Engineer, Honours co-advisor

Abstract

This paper empirically analyzes the X-inefficiency of Chinese 16 listed commercial banks in 2010-2014 by use of a Recursive Thick Frontier Approach (RTFA). This approach examines how banks perform compared to their cost frontier, which is the cost realized by the 'best practice' banks as a function of the output bundle. X-inefficiency measures the distance from the cost frontier. The empirical results indicate that, during 2010-2014, Chinese commercial banks have a good X-efficiency level. Specifically, joint-stock banks have the relatively low X-inefficiency. State-owned banks generally reduce their X-inefficiency and improve their efficiency. Conversely, city banks have the relatively high X-inefficiency (lowest X-efficiency) compared with other two types of commercial banks.

Special Thanks

Gratefully thanks Dr. Chris Auld and Dr. Merwan Engineer for everything they have taught me. I would like to show my sincere gratitude for my supervisor Dr. Paul Schure for all the help and patience on my thesis. I would like to say thanks to all my fellow honours students for their feedback and support.

Table of Contents

1. Introduction

This paper analyzes an aspect of the efficiency of Chinese commercial banks. In particular, I examine the X-inefficiency of these banks to reflect their X-efficiency. X-efficiency is a concept that looks at the costs of producing a given level of output. Following the recent literature (e.g., Berger and Mester (1997)), a firm is said to be X-inefficient if its costs of producing a given level of output are greater than the minimum costs. This outcome can arise in non-competitive markets where firms are somewhat insulated from competitive pressures to minimize costs. The banking industry in China is oligopolistic and hence has scope for considerable X-inefficiency. X-efficiency is distinct from production efficiencies related to scale and scope of the bank operation. Berger and Humphrey (1997) point out that the scale and scope inefficiency only account for no more than 5% of total costs on low efficiency or inefficiency of the banking sector, whereas reducing X-inefficiency helps banks save 20% costs. Hence, X-efficiency is the most important factor for the efficiency of commercial banks.

In this paper, instead of directly studying X-efficiency, I evaluate the X-inefficiency of Chinese commercial banks to reflect their X-efficiency levels in 2010-2014, and compare the X-inefficiency levels among three types of commercial banks. Besides, I use relatively new data and methodology for the X-inefficiency study. The particular method I employ is called the Recursive Thick Frontier Approach (RTFA). This involves collecting data on costs and outputs of 16 listed Chinese commercial banks in 2010-2014. Outputs, chosen based on a value-added approach, include key categories of banks: total loans, total deposits, fees and commission income, and investment securities. My RTFA analysis involves examining how banks perform compared to their cost frontier, which is the cost realized by the ‘best practice’ banks as a function of the output bundle. X-inefficiency measures the distance from the cost frontier. RTFA hinges on the observation that the best practice banks do not have efficiency levels that deviate significantly from the estimated frontier. My estimated cost frontier indicates that, as expected, when total deposits or fee and commission income increase, total costs will increase. Whereas, I find that total loans and investment securities are negatively related with total costs. After getting the estimated cost frontier function, I measure the X-inefficiency of different types of banks. The results show that, over the period of 2010-2014, Chinese commercial banks perform with a good efficiency level. Specifically, joint-stock banks have the relatively low X-inefficiency. State-owned banks generally reduce their X-inefficiency and improve their efficiency. By contrast, city banks have the relatively high X-inefficiency compared with other two types of commercial banks.

My contribution is relative to the specific paper Chen, Qiu, and Fu (2011), which uses RTFA to analyze the X-inefficiency of 19 Chinese commercial banks in 1996 -2005. They find that joint-stock banks are more X-efficient than state-owned banks. Also, city banks have relatively low X-inefficiency. Whereas, I will use more recent data in 2010-2014 to examine the current X-inefficiency levels of Chinese commercial banks as Chinese commercial banking sector has grown substantially and has also experienced increased competition from foreign banks as well as a turbulent macroeconomy. Also, I exclude input prices as variables in my cost function and just consider how outputs influence costs of banks. Finally, I will give more extensive analysis on the cost frontier and the X-inefficiency.

Harvey Leibenstein (1966) first introduced the concept of X-efficiency, but was vague defining it. Berger and Mester (1997) and others defined X-efficiency related to the cost frontier, i.e. the cost realized by the 'best practice' firms as a function of the output and input bundle. X-efficiency measures the extent to which the firm's realized cost is in line with the cost frontier. X-inefficiency measures the distance from the cost frontier. The estimation methods for X-efficiency (or X-inefficiency) can be categorized into two groups: full frontier approaches and thick frontier approaches. Full frontier approaches assume that all deviations from the cost frontier represent inefficiency, while thick frontier approaches assume that, besides inefficiency, cost levels might deviate from the frontier because of random errors. Some scholars employ full frontier approaches such as Data Envelopment Analysis (DEA). Other studies focus on thick frontier approaches such as Stochastic Frontier Analysis (SFA), Distribution Free Approach (DFA), and Thick Frontier Approach (TFA). For example, in the case of studying Chinese banks, Chen, Skully, and Brown (2005) employ DEA to examine the X-efficiency of 43 Chinese banks over the period 1993 to 2000. They find that large state-owned banks and city banks are more X-efficient than joint-stock banks. Fu and Heffernan (2007) study the X-efficiency of the Chinese banking sector from 1985 to 2002 based on SFA. They find the opposite result that state-owned banks have relatively low X-efficiency, while joint-stock banks perform better than other banks.

Wagenvoort and Schure (2006) created the Recursive Thick Frontier Approach (RTFA) for the X-inefficiency measurement, which is a thick frontier approach. The main regression model based on this method is the Cobb-Douglas cost frontier function, which indicates how bank outputs and inputs influence costs. Wagenvoort and Schure (2006) point out that RTFA is superior to SFA, especially when efficiency levels change over time. Also, RTFA is feasible even if we just have small number of time observations. Schure, Wagenvoort, and O'Brien (2004) apply this method to the EU banking sector.

My paper employs RTFA to study commercial banks in China in recent years, especially after 2009. I analyze the X-inefficiency of 16 listed commercial banks in China from 2010 to

2014. The 16 listed banks include 5 state-owned banks, 8 joint-stock banks, and 3 city banks[1]. By 2014, total assets of 16 listed banks take up more than 75% of all commercial banks. Thus, their efficiency could represent the situation of the banking sector. The main regression model I use is similar as Schure et al.'s (2004) Cobb-Douglas cost frontier function, but, according to Chinese banks' characteristics, I consider four outputs (total loans, total deposits, fee and commission income, and investment securities) in the function. I pick up output variables based on the value-added approach, which indicates that outputs are those producing services and costs (e.g., payment services of deposits, screening and monitoring services of loans). I employ type dummies instead of size dummies to reflect three main types of commercial banks in China, and also include time dummies to allow for shifts over time. More importantly, I exclude the input price terms in my function because I cannot use the market input prices. Mountain and Thomas (1999) prove that discarding input prices, which are measured in actual expenses for each bank in the function, is better when we cannot obtain or use market input prices. I will explore how explanatory variables in the cost function impact banks' costs. Also, I will estimate the X-inefficiency value for each commercial bank in 2010-2014. Under RTFA, the individual X-inefficiency value for each bank in each year does not mean anything because this individual value include both X-inefficiency and random errors. Only the averages over years and over types can represent X-inefficiency as I assume the mean of error terms is zero. So, I focus on comparing the X-inefficiency levels of three types of banks.

I use the RTFA algorithm of Schure et al.'s (2004) to obtain the final regression function, which is the cost frontier. As I scale outputs and costs with total assets (e.g., total costs/total

[1] For the background of 16 listed commercial banks, see Appendix A.

assets, total loans/total assets), the estimated cost function indicates that, as expected, when total deposits/total assets or fee and commission income/total income (income is a flow, so scaling over total income) increases, total costs/total assets will increase. Whereas, total loans/total assets and investment securities/total assets are negatively related with total costs/total assets. The negative relation between total loans and total costs is still a mystery at this stage. For investment securities, they do not incur high costs; thus, their negative effects might exceed positive effects. Besides, I find that state-owned banks have much (21%) lower cost frontier than joint-stock banks, and the cost frontier shifts up year by year. After getting the estimated cost frontiers, I measure the X-inefficiency levels of banks in 2010-2014. The results show that, over the period of 2010-2014, Chinese commercial banks perform with a good efficiency level. Specifically, joint-stock banks have the relatively low X-inefficiency. State-owned banks generally reduce their X-inefficiency and improve their efficiency, which is contrary to some studies showing state-owned banks have quite low X-efficiency (high X-inefficiency). By contrast, city banks have the relatively high X-inefficiency (lowest X-efficiency) compared with other two types of commercial banks.

The first part of my thesis is an introduction of X-efficiency and its measurement methods. Then, what is the recursive thick frontier approach (RTFA) and why it is better than other traditional methods are explained in the paper. Following that section is the empirical analysis of the cost frontiers and the X-inefficiency levels of Chinese 16 listed commercial banks between 2010 and 2014 based on RTFA. Finally, the conclusion of the research paper comes out.

2. The Concept of X-efficiency

2.1. Theories about X-efficiency

In 1966, Harvey Leibenstein first introduced X-efficiency, but his definition on it was vague. Since that time, most papers focused on the production frontier studies in the case of banks. That is, they studied how the output was realized by the 'best practice' banks as a function of the input bundle. X-efficiency measures the extent to which the bank's realized output is in line with the production frontier. If a bank's actual production point lies on the frontier it is efficient. If it lies below the frontier then it is inefficient, with the ratio of the actual to potential production defining the level of efficiency of the individual bank (Fiorentino, Karmann, & Koetter, 2006).

Since 1990s, Berger and Mester (1997) pointed out that the X-efficiency is similar as the cost efficiency. Allen and Rai (1996), Lozano-Vivas and Pasiouras (2010), and other people stated that the cost efficiency could represent the X-efficiency. In recent years, most studies turn to the cost frontier. This frontier measurement is similar as the production frontier, which shows how the cost is realized by the 'best practice' banks as a function of the output and input bundle. X-efficiency measures the extent to which the bank's realized cost is in line with the cost frontier. X-inefficiency measures the distance from the cost frontier.

Early literature show that the bank efficiency arises from the scale efficiency and scope efficiency; however, Berger and Humphrey (1997) find that the scale and scope inefficiency only account for no more than 5% of total costs, whereas reducing X-inefficiency help banks save 20% costs. Thus, X-efficiency is the most appropriate indicator for measuring the efficiency of commercial banks.

2.2. Methods for Measuring X-efficiency

The estimation of X-efficiency (or X-inefficiency) can be categorized according to the assumptions and techniques used to construct the efficient frontier. The main two groups are full frontier approaches and thick frontier approaches.

Full frontier approaches rely on linear programming to calculate piecewise linear segments of the efficient frontier. They assume that all deviations from the cost frontier represent inefficiency, such as the most popular method called Data Envelopment Analysis (DEA). This kind of method does not require explicit functional forms for both the frontier and deviations from it (X-inefficiency). So, full frontier approaches have some critical drawbacks. For example, they cannot cope with the random error term, which will lead to the final efficiency values deviate from the true efficiency.

Whereas, thick frontier approaches assume that, apart from inefficiency, cost levels might deviate from the frontier because of measurement errors or factors beyond the control of the firm’s management. Thus, observations might lie on both sides of the cost frontier. Common methods are Stochastic Frontier Analysis (SFA), Distribution Free Approach (DFA), and Thick Frontier Approach (TFA). Thick frontier approaches usually impose assumptions about the functional form of the frontier and distributional assumptions about the inefficiency. Hence, this kind of method has some advantages over the full frontier approach. For example, empirical models based on thick frontier approaches contain random disturbance terms, which deal with effects from random elements and give a better measurement of X-inefficiency. The most important thing we need to concern about this method is that if some mistakes occur when we set the frontier function, the X-efficiency (or X-inefficiency) measure will be misleading (Hjalmarsson, Kumbhakar, & Heshmati, 1996).

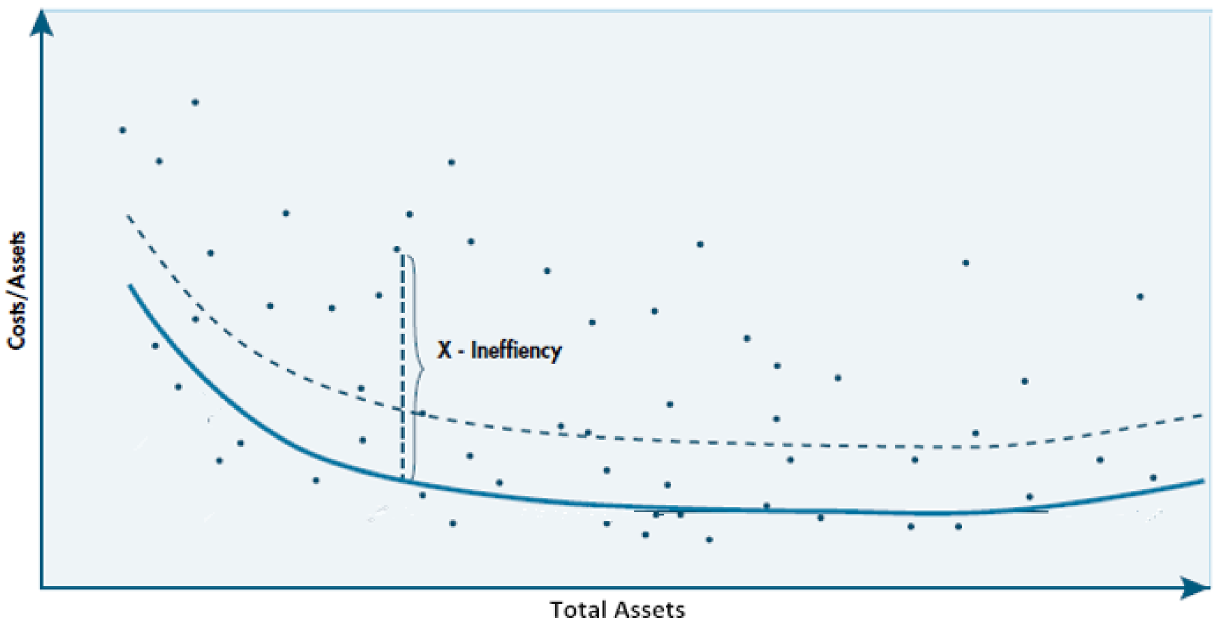
3. The Recursive Thick Frontier Approach (RTFA)

3.1. The Introduction of RTFA

My paper employs a Recursive Thick Frontier Approach (RTFA) for the X-efficiency measurement, which was created by Wagenvoort and Schure (1999a). Their formal article on this approach "A Recursive Thick Frontier Approach to Estimating Production Efficiency" was first published in 2006. RTFA is a thick frontier approach which hinges on observations that the best practice firms do not have efficiency levels that deviate significantly from the frontier. The core assumption is that there exists a group of firms that are efficient in the time period of the data. This seems a reasonable assumption even if the data set covers a short time span. This approach applies to both production frontiers and cost frontiers.

Specifically, according to Figure 1, the cost frontier is the solid blue line. The dotted line means that the frontier could move due to technical or managerial changes. During a short time period, it is unlikely that the management undergoes a significant change or new technologies are implemented. There are many dots in the graph, which represent different observations, such as different sizes of commercial banks (categorized by total assets) over different times, in the sample. If the dot represents individual bank, the distance from above the frontier consists of X-inefficiency and the random error term. In contrast, if I classify banks into different groups (e.g., different types, different sizes), the deviation from the cost frontier only refers to the X-inefficiency because I assume that the mean of stochastic error terms for each group of banks is zero. If the observation lies below the frontier, it is perfectly X-efficient. That is, for each group, any deviations from the cost frontier are due to X-inefficiency or really high X-efficiency.

Figure 1. X-inefficiency measure which can be derived from the cost frontier



Source: Wagenvoort & Schure (1999b)

3.2. Why RTFA is Better?

When compared with other thick frontier approaches, RTFA shows some appealing features. For example, the Thick Frontier Approach (TFA) looks at average costs, while RTFA only estimates the frontier parameters from efficient banks. This is an important benefit as the frontier function may not be a neutral transformation of the average function (Timmer, 1971). Additionally, compared to the most frequently used Stochastic Frontier Analysis (SFA), RTFA does not need specific distributional assumptions on the inefficiency term. SFA requires priori distributional assumptions regarding inefficiency, which are difficult to test. Then it would affect the estimated cost frontier and the inefficiency estimates. The only assumption that is made under RTFA is that there exists a group with a number of banks that are on the cost frontier in the given time period (Wagenvoort and Schure, 2006). Moreover, the SFA model tends to be mistakenly specified because this approach assumes that the inefficiency term is independent of the explanatory variables of the frontier function such as outputs, which is not suitable for the real banking data. Whereas, RTFA allows X-inefficiency to be dependent on the explanatory variables.

As the approach for panel data, RTFA displays merits over other standard panel data approaches, such as the fixed effects and the random effects models. The fixed effects model assumes that the inefficiency does not vary over time, which leads to the estimate of the inefficiency contains both inefficiency and all other time invariant firm-specific factors. For the random effects model, the inefficiency term is also assumed to be independent of regressors. Also, this method only works when we have many time observations. By contrast, RTFA works well even if we only have short time periods in the panel dataset.

Finally, Wagenvoort and Schure (2006) prove that RTFA finds the relevant parameters as well as the set of efficient banks in a robust way.

4. Empirical Analysis of Chinese Commercial Banks

4.1. Data Selection

In my paper, I choose 16 listed commercial banks in China including 5 state-owned banks, 8 joint-stock banks, and 3 city commercial banks[2]: Industrial and Commercial Bank of

[2] For the background of 16 listed commercial banks, see Appendix A.

China (ICBC), China Construction Bank (CCB), Bank of China (BOC), Agricultural Bank of China (ABC), Bank of Communications (BOCOM); China Merchants Bank (CMB), China CITIC Bank, China Minsheng Bank, Shanghai Pudong Development Bank (SPD Bank), Industrial Bank, China Everbright Bank, Hua Xia Bank, Ping An Bank; Bank of Beijing, Bank of Nanjing, and Bank of Ningbo. I construct my bank dataset from the annual report for each bank. I use the annual balance-sheet and income-statement data for this 16 listed banks for the period 2010–2014.

China has many commercial banks; however, most of them are unlisted, and these banks do not disclose complete data to the public. So, I can only get the 16 listed banks' full data. By 2014, the total assets of 16 listed Chinese commercial banks take up around 78.5% of all commercial banks, which indicates that their efficiency could represent the situation of the banking sector. I choose the time period of 2010-2014 because I can use the latest sample data. Also, after financial crisis, Chinese economy begins to revive, and new financial reform policies are implemented, which means that it is essential to study the efficiency of the banking sector. Moreover, under RTFA, it is suitable to employ this short time span of 5 years.

4.1.1. Bank Outputs

Berger and Humphrey (1992), based on prior literature, find that we have five main choices when considering how to define bank outputs and inputs. They are asset approaches, intermediation approaches, user cost approaches, production approaches, and value-added approaches.

Under the asset approach, bank liabilities are inputs because they offer raw materials of investable funds, while bank assets are considered to be outputs as they are uses of funds that generate the banks' revenue. Banks are regarded only as financial intermediaries between liability holders and those who receive bank funds. Sealey and Lindley (1977), based on the asset approach, created the intermediation approach. They state that banks purchase funds from depositors and turn these funds into loans, which is one of main bank outputs. Also, they first consider about interest payments, and define that inputs should include costs from labor, borrowing funds, and capital. The user cost approach determines whether a financial product is an input or an output on the basis of its net contribution to bank revenue (Hancock, 1985). If returns on an asset exceed the opportunity cost of funds or if costs of a liability are less than the opportunity cost, then the product is regarded as a financial output. Otherwise, it is a financial input. Benston, Hanweck, and Humphrey (1982) point out "output should be measured in terms of what banks do that causes operating expenses to be incurred." Thus, they define bank outputs as those activities that will cost lots of labor and physical capital; however, they also acknowledge that deposits have input characteristics. They regard deposits as both outputs and inputs. The value-added approach arises from the production approach, and becomes a popular method in recent years. Under value-added approaches, the banking sector is considered to be the service industry; therefore, all liability and asset categories have some output characteristics. The categories have substantial value added (Berger and Humphrey, 1992). Outputs usually consist of loans, deposits, investment securities, and fee and commission income. For example, bank loans will incur screening and monitoring activities, and bank deposits contain payment services.

In my paper, I use the value-added approach, and I choose four outputs in my cost model: total loans, total deposits, investment securities, as well as fee and commission income. Total loans comprise total loans plus total other lending (e.g., due from other banks and financial institutions). Total deposits consist of demand, savings, and time deposits. Investment securities include all securities listed on the investment account of bank annual reports (e.g., debt securities, equity securities, and others). Contrary to other three output variables available from the bank's balance sheet, fee and commission income is a flow variable that is taken from the bank's income statement.

4.1.2. Bank Inputs

The choice of my inputs conforms to standard practice in other literature. In general, three input prices will be considered into the cost function, including the labor price, fund price, and capital price. Many literature employ input prices based on the actual expenses incurred by each bank. However, Mountain and Thomas (1999) point out that inputs based on the market prices is superior to that based on the actual expenses.

I cannot use the market prices for inputs because I only study banks in one country, China. That is, each bank will face the same market prices, which will lead to regression not running due to the perfect collinearity issue. Hence, I choose to exclude inputs in my cost function. If I use actual expenses as input prices rather than get rid of input variables, it would incur misspecification for my cost function. Also, it would cause misleading regression estimating results, and my efficiency measure would be wrong (Mountain and Thomas, 1999). Including input variables based on the market prices in the cost function is more applicable to studies focusing on banks in different countries.

4.2. Model Specification

4.2.1. The Regression Model — Cobb–Douglas Cost Frontier Function

The cost frontier describes the relationship between costs and outputs of the relatively efficient commercial banks. I use RTFA and assume that deviations from the frontier can be caused by random errors as well as X-inefficiency. Random errors include measurement errors for outputs, or other unobserved factors. In my paper, I estimate a (transformed) Cobb–Douglas cost function, which has been augmented with dummies to allow for shifts over time and differences in the type of banks:

$$\frac{TC_{ti}}{TA_{ti}} = \gamma \left(\frac{y_{ti,1}}{TA_{ti}}\right)^{\beta_1} \left(\frac{y_{ti,2}}{TA_{ti}}\right)^{\beta_2} \left(\frac{y_{ti,3}}{TA_{ti}}\right)^{\beta_3} \left(\frac{y_{ti,4}}{TA_{ti}}\right)^{\beta_4} \sigma_1^{s_{1,ti}} \sigma_2^{s_{2,ti}} \delta_1^{t_1} \delta_2^{t_2} \delta_3^{t_3} \delta_4^{t_4} \varepsilon_{ti} \qquad (1)$$

$$i \in E$$

In Equation (1):

E: the set of best-practice banks in the dataset (efficient banks);

Period t: 2010-2014 annually;

Bank i: 16 listed commercial banks in China;

TC_{ti} and TA_{ti}: total costs and total assets, respectively, of bank i in period t;

$y_{ti,k}$: bank i's amount of output of type k in year t, k=1, 2, 3, 4 (four outputs: total loans, total deposits, fee and commission income, and investment securities);

$\frac{y_{ti,k}}{TA_{ti}}$: Total loans/TA; Total deposits/TA; Fee and commission income/Total income; Investment securities /TA;

(In general, my cost model should include input prices, such as the price of funds, the price of labor, and the price of buildings. As I have discussed early, choosing to exclude price terms is better in my paper.)

$S_{1,ti}$ and $S_{2,ti}$: two type dummies refer to three types: state-owned banks, joint-stock banks, and city banks (Choose 8 joint-stock banks as the base);

When bank i is one of 5 state-owned banks, $s_1 = 1$;

When bank i is one of 3 city commercial banks, $s_2 = 1$;

t_1, t_2, t_3, t_4: four time dummies, as the dataset comprises t = 5 years (2010 as the base year);

When period t is 2011, $t_1 = 1$, and other t=0;

ε_{ti} : a random symmetrically distributed disturbance term.

For the subset of inefficient banks in the dataset, that is, the set of banks that are not on the cost frontier represented by Equation (1), no specific structure is assumed. RTFA allows each individual inefficient bank to adopt any available technology and be inefficient to any possible degree (Wagenvoort et al, 2004).

My cost function is general. Different products can attribute differently to the costs of an efficient bank. I will estimate Equation (1) in logs. The group of parameters I need to estimate includes γ, β_1, β_2, β_3, β_4, σ_1, σ_2, δ_1, δ_2, δ_3, δ_4. Under the null hypothesis of statistical test, (σ_1, σ_2, δ_1, δ_2, δ_3, δ_4) $= 1$.

In Equation (1), I scale total costs and outputs over total assets rather than just using total costs and outputs in both sides of the equation, which is common in some other papers. Scaling is important for my analysis to allow items to be comparable between different banks. Specifically, instead of scaling over total assets, fee and commission income scale over total income because it is a flow variable.

In general, I need to adjust total costs and the output variables according to the inflation because these items are in nominal (RMB) terms, while a cost function requires them to be in real units. Thus, without the adjustment, the nominal terms would not be comparable between different years. In my model, I employ the ratio of total costs (outputs) and total assets. So, I do not need to adjust for the inflation.

Table 1.1-1.5 and Table 2 in the Appendix B section show the descriptive statistics regarding to main variables in this cost function. I define total costs of each bank as the sum of interest expense, operating expense, fee and commission expense, and impairment losses on assets from the bank annual report. Table 1.1-1.5 contain the descriptive statistics regarding each main variable for the period 2010–2014. From this five tables, I find all sample variables change as years. Total costs/total assets, fee and commission income/total income, and investment securities/total assets increase as year goes; whereas, total loans/total assets and total deposits/total assets decrease as year goes.

Before doing the regression, based on previous literature, I expect to see the following results: explained variable (total costs/total assets) positively depends on explanatory variables (total loans/total assets, total deposits/total assets, fee and commission income/total income, and investment securities/total assets). These items, such as loans, deposits, investment securities, and fee and commission, will cost banks' resources and services (e.g., payment services of deposits, screening and monitoring services of loans). Thus, as these items increase, costs should also increase.

4.2.2. The RTFA Algorithm

I use the estimation procedure similar as that of Wagenvoort et al. (2004). Below, let j indicate the number of the iteration and n_j be the number of commercial banks left in the sample of 16 banks during iteration j. Under RTFA, let I_{ij} be an indicator function that takes on the value of 1 if four or five out of the five OLS residuals for bank *i* in iteration j have the same sign (positive or negative). Each bank has five OLS residuals because my studying period contains five years (2010-2014). Banks having two or three out of the five residuals with the same sign are on the cost frontier. Define $Z_j = \sum_{i=1}^{n_j} I_{ij}$.

The RTFA algorithm:

Step 1: Set j=0 and $n_0 = N$, where *N* represents the number of banks in the dataset (N=16). Choose α as the speed of the data reduction. In my study, to get the steady results, I will get rid of banks one by one in each iteration. So, I set α = 1/16 ≈ 0.06, which means that n_j is reduced by 6% in each iteration.

Step 2: Compute the OLS estimates for (1− j*α)*100% of the data.

Step 3: Compute the Chow test statistic $\lambda_j = \frac{(Z_j - 0.375 * n_j)^2}{0.375 * (1 - 0.375) * n_j}$ and compare it with $\chi^2_{0.95}(1)$, the 95th percentile of the chi-squared distribution with one degree of freedom (I choose the 5% significance level in my test). If $\lambda_j < \chi^2_{0.95}(1)$, then stop the iterations and report the last OLS regression as the final result. Otherwise, *go to Step 4*.

Step 4: Compute the mean m_{ij} of the five regression residuals of each bank i, including banks which were omitted in the previous iteration. Set j=j+1. Select the n_j=(1−j*α)*N banks that enter the next iteration by getting rid of the j*α*N banks with the largest values of m_{ij}, i=1,…, N. *Go to Step 2.* (In each regression, the bank with the largest residuals has the lowest efficiency. Thus, after discarding this kind of banks, the remaining group of banks include those that have really high efficiency and are on the cost frontier).

I will explain the test statistic in the Step 3 of the algorithm. First, Z_j is large when the remaining sample still contain both efficient and inefficient banks (in my sample, relatively efficient banks tend to have four or five negative residuals, while inefficient banks tend to have four or five positive residuals). Whereas, when only best-practice banks are in the remaining group of banks, the five residuals of each bank tend to be scattered evenly around the cost frontier. That is, the probability of a single residual falls on either side of the regression line is 0.5. In my case, there is a probability of $(0.5)^5 + 4 * (0.5)^5 + (0.5)^5 + 4 * (0.5)^5$=0.375 that a bank has either four or five positive (negative) residuals. Also, these probabilities are independent; thus, the number of banks with four or five residuals of the same sign follows a binomial distribution B (n, p) with p = 0.375. According to the Law of large Numbers and the Central Limit Theory, when p is around 0.5, the binomial distribution approaches the normal regardless of the value of sample n; Or, when p largely deviates from 0.5, if n*p > 5 and n*(1-p) > 5, the binomial distribution also approaches the normal. Hence, in my sample, λ_j is asymptotically $\chi^2(1)$ distributed.

4.2.3. The X-inefficiency Measurement

In my paper, I study the X-inefficiency for each bank. I measure the degree of X-inefficiency of a bank by its distance from the cost frontier. Define $\widehat{TC_{ti}}/\ TA_{ti}$ to be the costs over the assets of bank i in year t if it were on the *cost frontier* (based on the final OLS regression in *Step 3*):

$$\frac{\widehat{TC_{ti}}}{TA_{ti}} = \hat{\gamma}\ \left(\frac{y_{ti,1}}{TA_{ti}}\right)^{\widehat{\beta}_1} \left(\frac{y_{ti,2}}{TA_{ti}}\right)^{\widehat{\beta}_2} \left(\frac{y_{ti,3}}{TA_{ti}}\right)^{\widehat{\beta}_3} \left(\frac{y_{ti,4}}{TA_{ti}}\right)^{\widehat{\beta}_4} \hat{\sigma}_1^{s_{1,ti}} \hat{\sigma}_2^{s_{2,ti}} \hat{\delta}_1^{t_1} \hat{\delta}_2^{t_2} \hat{\delta}_3^{t_3} \hat{\delta}_4^{t_4} \qquad (2)$$

The degree of X-inefficiency of bank i in year t ($X - ineff_{ti}$) represents the mean percentage cost reduction that the bank could have achieved without sacrificing any output:

$$X - ineff_{ti} = \frac{\left(\frac{TC_{ti}}{TA_{ti}} - \frac{\widehat{TC_{ti}}}{TA_{ti}}\right)}{\frac{TC_{ti}}{TA_{ti}}} \qquad (3)$$

4.3. Results

4.3.1. Regression Results

After the first iteration, I get $\lambda_0 = 13.07 > 3.84; that\ is, \lambda_0 > \chi^2_{0.95}(1)$. Thus, I need to go to Step 4. In this step, I will discard one bank with the largest mean residual. I get rid of Bank of Communications (BOCOM). Then, I have 15 banks left in the next iteration.

After the second iteration, I get $\lambda_1 = 8.22 > 3.84; that\ is, \lambda_1 > \chi^2_{0.95}(1)$. Thus, I need to go to Step 4. In this step, I get rid of Bank of Ningbo. Then, I have 14 banks left in the next iteration.

After the third iteration, I get $\lambda_2 = 2.31 < 3.84; that\ is, \lambda_2 < \chi^2_{0.95}(1)$. I stop the iteration and report the OLS regression based on remaining 14 banks:

$$\frac{\widehat{TC_{ti}}}{TA_{ti}} = 0.02 \left(\frac{y_{ti,1}}{TA_{ti}}\right)^{-0.68} \left(\frac{y_{ti,2}}{TA_{ti}}\right)^{0.47} \left(\frac{y_{ti,3}}{TA_{ti}}\right)^{0.05} \left(\frac{y_{ti,4}}{TA_{ti}}\right)^{-0.16} 0.79^{s_{1,ti}}\ 0.95^{s_{2,ti}}\ 1.27^{t_1}\ 1.38^{t_2}\ 1.42^{t_3}\ 1.61^{t_4} \quad (4)$$

Table 3. Parameter Estimations for the Cost Frontier					
Variable	**Coefficient**		**Std. Error**	**t-Statistic**	
Constant	0.019617	γ	0.206707	-19.019	*
Total Loans/TA	-0.679056	β_1	0.14014	-4.846	*
Total Deposits/TA	0.468675	β_2	0.214401	2.186	*
Fee and Commission Income/TI	0.045722	β_3	0.022171	2.062	*
Investment Securities/TA	-0.164458	β_4	0.057293	-2.870	*
Dummy S1 (State-owned Banks)	0.789619	σ_1	0.043066	-5.485	*
Dummy S2 (City Banks)	0.945101	σ_2	0.054059	-1.044	
Dummy 2011	1.271266	δ_1	0.040586	5.914	*
Dummy 2012	1.384062	δ_2	0.042053	7.729	*
Dummy 2013	1.424101	δ_3	0.045546	7.762	*
Dummy 2014	1.612018	δ_4	0.050668	9.424	*
Number of banks: 14	*Number of Observations: 14*5=70*				
R-squared	0.810746		F-statistic	25.27496	
Adjusted R-squared	0.778669		Prob(F-statistic)	0.000000	
Sum squared resid	0.595467				
TA= total assets, TI = total income					
***Significant at the 95% confidence level**					
Coefficients are estimated by the recursive thick frontier method					

According to Table 3, my model's significance test shows that the value of F-statistic is 25.27496, which is far larger than the critical value (p-value=0.00000). Thus, my model setting passes the significance test and it is suitable. With an adjusted R-squared of 78%, my model explains the variation in total costs over total assets of efficient banks very satisfactorily. From the parameter estimations, except dummy S_2, most of estimated coefficients are statistically significant at 5% significance level (95% confidence level).

According to the estimated coefficients of four output variables, as I expected, total deposits/total assets, and fee and commission income/total income have positive relations with total costs/total assets. As total deposits/total assets increases by 10%, total costs/total assets will increase by 4.7%. As fee and commission income/total income rises by 10%, total costs/total assets will increase by 0.5%. Total deposits contribute more to the total costs of banks. Conversely, total loans/total assets and investment securities/total assets have negative relations with total costs/total assets. For investment securities, they do not require too much effort to collect investment information and they do not incur high transaction costs; therefore, their negative effects might exceed positive effects. Investment securities are not important outputs compared with loans and deposits, so the negative sign is believable. Whereas, for total loans, they should have positive effects and their effects on total costs should be more than effects from total deposits because screening and monitoring activities of loans consume banks' resources. This is still a mystery at this stage.

For the two type dummies, S_1 and S_2, they reflect Chinese commercial banks' ownership structures and business scales. In my paper, based on prior literature on Chinese banks, I separate three types of banks and choose joint-stock banks as the base. From Table 3, the effect from dummy S_2 is not statistically significant. But when just focusing on its estimated coefficient 0.945, I find that city banks have 5.5% lower cost frontier than joint stock banks. By contrast, for dummy S_1, I can conclude that state-owned banks have 21% lower cost frontier than joint stock banks. This result is consistent with many Chinese studies on banks, which state that state-owned banks should have lower costs than other types of banks because they have large business areas and more systematic, specialized management systems. Time dummies represent shifts of cost frontiers over different years. For example, year 2011 has 27% higher cost frontier than year 2010. Year 2012 has about 11% higher cost frontier than year 2011. The cost frontier rises by 11%-19% year by year since 2011. These results might partly reflect the inflation effect, as the inflation increases by about 3% each year[3]. Moreover, these changes are reflected in total costs/total assets, which is a ratio. Such a high percentage rate might mean that the increasing speed of costs over years is far beyond that of assets. This is true when I refer to the original dataset and calculate the changes of 'total costs/total assets' over years.

4.3.2. The X-inefficiency Analysis

After obtaining the regression results, I can get the X-inefficiency results based on Equation (3), which are shown in Table 4.

The final OLS regression is based on remaining 14 banks. Thus, the most X-inefficient (have the largest value of X-inefficiency) is Bank of Communications (BOCOM) and then Bank of Ningbo. Some banks have negative values of X-inefficiency, which mean than these banks have quite low X-inefficiency and they are highly X-efficient. That is, the bank with the smallest negative X-inefficiency value has the highest X-efficiency, while the bank having the largest positive X-inefficiency value is least X-efficient. It is important to notice here that the individual X-inefficiency value for each bank in each year does not mean anything because this individual value include both X-inefficiency and random errors. Hence, only the averages over years and over types can represent X-inefficiency as I assume the mean of error terms is zero.

First, according to the average of the X-inefficiency value for each bank during the period of 2010-2014, I find that eight banks have quite low X-inefficiency (with negative values). Specifically, the least X-inefficient (the most X-efficient) three banks in this five years are China Merchants Bank (CMB), China CITIC Bank (CITIC), and Industrial and Commercial

[3] For the inflation in China, see http://www.tradingeconomics.com/china/consumer-price-index-cpi

Bank of China (ICBC). Two of them are joint-stock banks and one is state-owned bank. Next, I will focus on three different types of banks between 2010 and 2014. From Table 4, based on the weighted average of each type of banks, I find that joint-stock banks have negative X-inefficiency values from 2010 to 2013, and they also have a small positive value in 2014. Whereas, state-owned banks have a relatively big positive X-inefficiency value in 2010. Since then, overall, this X-inefficiency value falls and there exists a negative value in 2014. This means that the X-inefficiency of state-owned banks has decreased in recent years. City banks have the highest X-inefficiency. Except year 2010, city banks have large positive X-inefficiency values in 2011-2014.

In a whole, in 2010-2014, Chinese commercial banks have a good X-efficiency level. Specifically, joint-stock banks have the relatively low X-inefficiency. State-owned banks generally reduce their X-inefficiency and improve their efficiency, which is contrary to many studies (e.g., Fu and Heffernan (2007), Weilan (2004)) showing state-owned banks have quite low X-efficiency (high X-inefficiency). The distance of state-owned banks' X-inefficiency level from joint-stock banks' becomes smaller. They even become more efficient than joint-stock banks in 2014. Conversely, city banks have the relatively high X-inefficiency (lowest X-efficiency). These results are consistent with current situations in the Chinese banking sector. Joint-stock banks always have good efficiency levels in recent years. In general, state-owned banks are usually regarded as low efficient banks because they are regulated by the government; however, in recent year, five state-owned banks speed up internal reforms and own more freedom, which lead to a more cost-effective banking system. Whereas, city banks have low efficiency due to merger and reorganization[4].

Table 4. X-inefficiency of 3 Types Commercial Banks (2010-2014)

Unit: %

		2010	2011	2012	2013	2014	*Average*	
5 State-owned Banks	**ICBC**	**1.44**	**-8.14**	**-7.62**	**-7.96**	**-17.54**	*-7.96*	3rd
	CCB	**6.37**	**-3.54**	**2.84**	**-1.46**	**-6.36**	*-0.43*	7th
	BOC	**12.03**	**5.43**	**13.21**	**1.61**	**-2.52**	*5.95*	**11th**
	ABC	**7.34**	**3.50**	**1.22**	**-1.23**	**-3.81**	*1.41*	**10th**
	BOCOM	**44.88**	**37.76**	**39.37**	**36.55**	**35.55**	*38.82*	**16th**
	W. Average	*9.53*	*1.97*	*4.74*	*0.72*	*-4.60*		
8 Joint-stock Banks	**CMB**	**3.06**	**-6.49**	**-19.12**	**-13.56**	**-5.92**	*-8.40*	1st
	CITIC	**-13.06**	**-22.86**	**2.20**	**-10.14**	**2.13**	*-8.35*	2nd
	Minsheng	**12.12**	**20.49**	**-7.82**	**11.11**	**0.32**	*7.25*	**14th**
	SPD	**-16.97**	**2.26**	**3.37**	**6.80**	**8.74**	*0.84*	**9th**
	Industrial	**-12.37**	**-15.54**	**-7.36**	**0.81**	**2.39**	*-6.42*	4th
	Everbright	**1.75**	**-7.58**	**-6.52**	**3.83**	**-4.16**	*-2.54*	5th
	Hua Xia	**15.11**	**17.25**	**3.21**	**-5.37**	**2.18**	*6.48*	**12th**
	Ping An	**2.79**	**-3.11**	**4.11**	**13.08**	**17.40**	*6.85*	**13th**
	W. Average	*-2.75*	*-3.66*	*-4.71*	*-0.10*	*2.12*		
3 City Banks	**Beijing**	**-17.11**	**0.61**	**11.95**	**1.04**	**2.70**	*-0.16*	8th
	Nanjing	**-10.60**	**8.68**	**1.65**	**-2.31**	**0.29**	*-0.46*	6th
	Ningbo	**19.56**	**45.30**	**28.28**	**32.21**	**26.69**	*30.41*	**15th**
	W. Average	*-8.00*	*9.89*	*13.34*	*6.84*	*7.19*		

5. Conclusion

This paper uses a relatively new method called the Recursive Thick Frontier Approach (RTFA) to study the X-inefficiency levels of 16 listed commercial banks (5 state-owned banks, 8

[4] For the information on the Chinese banking sector, see China Banking Association, http://www.china-cba.net/

joint-stock banks, and 3 city banks) during the period of 2010-2014. The X-efficiency of the bank measures the extent to which the bank's realized cost is in line with the cost frontier, i.e. the cost realized by the 'best practice' banks as a function of the output bundle. X-inefficiency measures the distance from the cost frontier. Under RTFA, if I calculate the X-inefficiency for individual bank, the distance from the frontier consists of X-inefficiency and the random error term. In contrast, if I classify banks into different groups (e.g., different types), the deviation from the cost frontier only refers to the X-inefficiency because I assume that the mean of stochastic error terms for each group of banks is zero. Thus, what I focus on in this paper is comparing the X-inefficiency levels of different types of banks in 2010-2014.

I use the RTFA algorithm of Schure et al.'s (2004) to estimate the cost frontier. After two iterations of this algorithm, I obtain the final OLS regression function, which is the estimated cost frontier function. This cost frontier function indicates that total deposits as well as fee and commission income are positively related to total costs, while total loans and investment securities are negatively related to total costs. The negative relation between total loans and total costs is completely out of expectation, and is still a mystery at this stage. Moreover, I find that state-owned banks have much lower cost frontier than joint-stock banks. This result is reliable because state-owned banks should have lower costs than other types of banks due to their large business areas and more systematic, specialized management systems. Also, the cost frontier shifts up year by year. Based on the estimated cost frontiers, I measure the X-inefficiency levels of 16 listed banks in 2010-2014, and compare the X-inefficiency among three types of commercial banks. The results show that, over the period of 2010-2014, Chinese commercial banks perform with a good efficiency level. Specifically, joint-stock banks have the relatively low X-inefficiency. State-owned banks generally reduce their X-inefficiency and improve their efficiency, which is contrary to some studies showing state-owned banks have quite low X-efficiency (high X-inefficiency). However, my result for state-owned banks is consistent with current situations in the Chinese banking sector. In recent year, five state-owned banks speed up internal reforms and own more freedom, which lead to a more cost-effective banking system. Finally, city banks have the relatively high X-inefficiency (lowest X-efficiency) compared with other two types of commercial banks.

My paper still requires more extensions in the future. For example, I need to figure out why total loans and total costs are negatively related. Furthermore, in recent years, Chinese commercial banks are experiencing big reforms, and their qualities of assets have improved a lot. Hence, I can consider risk effects in my regression model, such as the non-performing loan ratio. Risk effects might influence the X-inefficiency levels.

Bibliography

Allen, L., & Rai, A. (1996). Operational efficiency in banking: An international comparison. *Journal of Banking & Finance*, 20(4), 655-672.

Benston, G. J., Hanweck, G. A., & Humphrey, D. B. (1982). Scale economies in banking: A restructuring and reassessment. *Journal of money, credit and banking*, 14(4), 435-456.

Berger, A. N., & Humphrey, D. B. (1992). Measurement and efficiency issues in commercial banking. In *Output measurement in the service sectors* (pp. 245-300). University of Chicago Press.

Berger, A. N., & Humphrey, D. B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European journal of operational research*, 98(2), 175-212.

Berger, A. N., & Mester, L. J. (1997). *Efficiency and productivity change in the US commercial banking industry: A comparison of the 1980s and 1990s* (No. 97-5).

Chen, D. F., Qiu, Z. X., & Fu, Y. (2011). X-Efficiency of China's Commercial Banks: A Recursive Thick Frontier Approach. *Research on Economics and Management*, 7, 013.

Chen, X. G., Skully, M., & Brown, K. (2005). Banking efficiency in China: Application of DEA to pre-and post-deregulation eras: 1993–2000. *China Economic Review*, 16(3), 229-245.

China Banking Regulatory Commission 2014 Annual Report (2015). Retrieved from: http://www.cbrc.gov.cn/chinese/files/2015/0F19960DD41D4206A246251A7225773E.pdf

China's city commercial banks: Opportunity knocks? (2007). Retrieved from KPMG: http://www.kpmg.com/CN/en/IssuesAndInsights/ArticlesPublications/Documents/China-banks-Opportunity-knocks-200703.pdf

Cho, Y. J. (1999). The banking system of the People's Republic of China. *Rising to the Challenge in Asia: A Study of Financial Markets*, 35.

Edwards, Jim. (2016, Feb 16). 'Some super bears are now expecting an imminent collapse of

China’s banking system and currency’. *Business Insider*. Retrieved from http://uk.businessinsider.com/china-banking-hits-290-of-gdp-banks-bailout-2016-2

Fiorentino, E., Karmann, A., & Koetter, M. (2006). The cost efficiency of German banks: a comparison of SFA and DEA. *Available at SSRN 947340.*

Fu, X. Q., & Heffernan, S. (2007). Cost X-efficiency in China's banking sector. *China Economic Review*, 18(1), 35-53.

Hancock, D. (1985). The financial firm: Production with monetary and nonmonetary goods. *Journal of Political Economy*, 93(5), 859-880.

Hjalmarsson, L., Kumbhakar, S. C., & Heshmati, A. (1996). DEA, DFA and SFA: a comparison. *Journal of Productivity Analysis*, 7(2-3), 303-327.

Kumaravadivel, Tusaani. (2013, July 9). An Overview of the Chinese Banking Industry. *The Queen's Business Review*. Retrieved from http://www.qbreview.org/an-overview-of-the-chinese-banking-industry/

Leibenstein, H. (1966). Allocative efficiency vs." X-efficiency". *The American Economic Review*, 56(3), 392-415.

Lozano-Vivas, A., & Pasiouras, F. (2010). The impact of non-traditional activities on the estimation of bank efficiency: international evidence. *Journal of Banking & Finance,* 34(7), 1436-1449.

Mainland China Banking Survey 2015. (2015). Retrieved from KPMG: https://www.kpmg.com/CN/en/IssuesAndInsights/ArticlesPublications/Documents/Mainland-China-Banking-Survey-201507-high.pdf

Mountain, D. C., & Thomas, H. (1999). Factor price misspecification in bank cost function estimation. *Journal of international financial markets, institutions and money*, 9(2), 163-182.

Sealey, C. W., & Lindley, J. T. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *The Journal of Finance,* 32(4), 1251-1266.

Schure, P., Wagenvoort, R., & O'Brien, D. (2004). The efficiency and the conduct of European banks: Developments after 1992. *Review of Financial Economics*, 13(4), 371-396.

Timmer, C. P. (1971). Using a probabilistic frontier production function to measure technical efficiency. *The Journal of Political Economy*, 776-794.

Turner, G., Tan, N., & Sadeghian, D. (2012). The Chinese banking system. *RBA Bulletin*, 53-64.

Wagenvoort, R., & Schure, P. (1999a). *The recursive thick frontier approach to estimating efficiency* (No. 1999/02). Economic and financial reports/European Investment Bank.

Wagenvoort, R., & Schure, P. (1999b). Who are Europe's efficient bankers?. *EIB Papers*, 4(1), 105-126.

Wagenvoort, R. J., & Schure, P. H. (2006). A Recursive Thick Frontier Approach to Estimating Production Efficiency*. *Oxford Bulletin of Economics and Statistics*, 68(2), 183-201.

Weilan, L. C. S. (2004). A Study on the Efficiency of China's Commercial Bank Based on SFA [J]. *Journal of Finance*, 6, 016.

Yao, S., Han, Z., & Feng, G. (2008). Ownership Reform, Foreign Competition and Efficiency of Chinese Commercial Banks: A Non-parametric Approach. *The world economy*, 31(10), 1310-1326.

Appendix A

The Chinese Banking Sector

The banking sector in China is large relative to the size of the Chinese economy and has expanded significantly over the past decade ("Mainland China Banking Survey," 2015). According to the China Banking Association, by 2015, there are 832 banks in China, including one central bank, three policy banks, one postal savings bank, many commercial banks, and rural/city credit cooperatives[5]. The banking system assets stand at 340% of GDP. Specifically, if the banking sector lost only 10%, that would be the equivalent of wiping out over 30% of China's annual GDP or more than $3.5 trillion (Edwards, 2016). Among these banks, the biggest contributors to Chinese GDP are commercial banks. They mainly have seven categories: 5 state-owned banks, 12 joint-stock banks, more than 100 city commercial banks, over 100 rural commercial banks, many rural cooperative banks, community banks, and small number of foreign banks.

Between 1978 and 1984, four largest state-owned commercial banks were established and the central bank of China, the People's Bank of China (PBOC), remained a vehicle for the implementation of monetary policy. The four state-owned banks are Industrial and Commercial Bank of China (ICBC), China Construction Bank (CCB), Bank of China (BOC), and Agricultural Bank of China (ABC). At that time, all the four are fully government-owned. The Chinese Government permitted to establish a number of other domestic banking institutions since the late 1980s, including the Bank of Communications (BOCOM), which was the first

[5] For the data of Chinese banks, see China Banking Association, available at: http://www.china-cba.net/

joint-stock bank in China (it is majority-owned by the Chinese but have private sector shareholders). In the following years, other 11 joint-equity banks were created and developed rapidly. In the 21st century, city commercial banks grew fast and became important members in the commercial banks. Currently, the primary players in the banking sector are the central bank, five state-owned commercial banks, joint-stock commercial banks, and city commercial banks.

Over past 15 years, the Chinese banking sector went through a number of reforms. After joining the World Trade Organization (WTO) in 2001, China agreed to open its banking sector to foreign competition. By 2006, foreign-owned banks were allowed to apply for a license to provide unrestricted services for all Chinese clients. This liberalization led to more public listings of commercial banks on the Chinese stock exchanges (Shanghai Stock Exchange and Shenzhen Stock Exchange) except Ping An Bank, Shanghai Pudong Development Bank (SPD Bank), and China Minsheng Bank, which were listed on the Chinese stock exchanges before 2001. Moreover, since 2003, the Chinese Government commenced on restructuring the four largest state-owned banks and facilitated them to convert into commercial banks. The state-owned commercial banks, although still heavily regulated by the government, became more commercially oriented. Four asset management companies (AMCs), which were established in 1999, disposed plenty of non-performing loans (NPLs) in these state-owned banks in the 2000s to help them successfully complete their initial public offerings (IPOs). By 2015, there are 16 listed banks in China, which consist of ‘big five’ state-owned banks, 8 joint-stock banks, and 3 city commercial banks. This 16 listed banks have many branches in China and account for around 61.37% of Chinese banking system assets in the end of 2014. Today, only three policy banks in China remain fully owned by the government. Many other banks have mixed ownership structures where the government may or may not have a controlling stake. Most of these reforms aimed to make the banking sector more commercially attractive (Kumaravadivel, 2013)[6].

As the banking sector in China becomes more liberalized in recent years, Chinese commercial banks' efficiency has become a high-profile issue because these banks, as the primary players in the sector, are facing unprecedented fierce competition, especially from the foreign financial institutions. As the rise of city commercial banks, state-owned banks, joint-stock banks, and them started competition in the financial market ("China's city commercial banks: Opportunity knocks?", 2007). Also, since China cancelled all the restrictions on foreign banks after 2006, these banks formed a huge threat to Chinese commercial banks because of their advantages of strong management experience as well as rich product and service types (Yao, Han, & Feng, 2008). Besides, due to financial crisis in 2008, although the national economy revived, its GDP growth rate went downward from 10.45% after 2010 to 6.9% in 2015[7]. The weak economy requires commercial banks to pay more attention to their efficiency. Furthermore, new implemented financial reform policies, such as the new round interest rate liberalization, push banks to concern about their efficiency. Confronted with the fierce competition and new macroeconomic environment, commercial banks should improve their efficiency. Therefore, measuring the bank efficiency benefits both the banking sector and the whole Chinese economy. The total assets of 16 listed Chinese commercial banks take up around 78.5% of all commercial banks, which indicates that their efficiency could represent the situation of the banking sector[8].

---

[6] For the history of the Chinese banking sector, see Cho (1999); Turner, Tan, & Sadeghian (2012); China Banking Regulatory Commission Annual Report (for 2010-2014), available at: http://www.cbrc.gov.cn/showannual.do

[7] For the history of China GDP growth, see The World Bank, available at: http://data.worldbank.org/

[8] China Banking Regulatory Commission 2014 Annual Report (2015)

Appendix B

Table 1.1. Total Costs/Total Assets Descriptive Statistics

	Mean	***Median***	***Max***	***Min***	***Stdev.***
2010	0.02862	0.02759	0.04001	0.02375	0.00408
2011	0.03625	0.03614	0.04841	0.02811	0.00640
2012	0.03862	0.03975	0.04355	0.03062	0.00383
2013	0.03984	0.03952	0.05403	0.03058	0.00659
2014	0.04413	0.04536	0.05976	0.03134	0.00728

Source: Annual Bank Report for each bank in 2010-2014

Table 1.2. Total Loans/Total Assets Descriptive Statistics

	Mean	***Median***	***Max***	***Min***	***Stdev.***
2010	0.55295	0.54939	0.69382	0.39752	0.08257
2011	0.57444	0.57379	0.71282	0.43701	0.07338
2012	0.54105	0.53028	0.69302	0.41522	0.07664
2013	0.52963	0.53468	0.64299	0.37886	0.07417
2014	0.51933	0.52941	0.63123	0.33987	0.07980

Source: Annual Bank Report for each bank in 2010-2014

Table 1.3. Total Deposits/Total Assets Descriptive Statistics

	Mean	*Median*	*Max*	*Min*	*Stdev.*
2010	0.74530	0.75422	0.85978	0.55390	0.08255
2011	0.71207	0.71178	0.82397	0.55849	0.07244
2012	0.68464	0.68717	0.82019	0.55696	0.08156
2013	0.68282	0.67819	0.81111	0.55233	0.07114
2014	0.66761	0.67040	0.78460	0.51466	0.07235

Source: Annual Bank Report for each bank in 2010-2014

Table 1.4. Fee Income/Total Income Descriptive Statistics

	Mean	*Median*	*Max*	*Min*	*Stdev.*
2010	0.05542	0.05015	0.09120	0.02573	0.02095
2011	0.05863	0.05459	0.09329	0.02958	0.02099
2012	0.05878	0.05801	0.08477	0.03431	0.01773
2013	0.07100	0.07584	0.09985	0.03823	0.01815
2014	0.07502	0.07584	0.11224	0.04361	0.02024

Source: Annual Bank Report for each bank in 2010-2014

Table 1.5. Investment Securities/Total Assets Descriptive Statistics

	Mean	*Median*	*Max*	*Min*	*Stdev.*
2010	0.17555	0.15892	0.27731	0.08382	0.06169
2011	0.16253	0.16702	0.25302	0.09161	0.05367
2012	0.18060	0.17944	0.26370	0.07543	0.05077
2013	0.21323	0.20784	0.34579	0.09446	0.05388
2014	0.24957	0.22201	0.47161	0.14898	0.07967

Source: Annual Bank Report for each bank in 2010-2014

Table 2. Descriptive Statistics of the Variables of the Cost Function

	Mean	*Median*	*Maximum*	*Minimum*	*Std. Dev.*
Total Costs/Total Assets	0.037492	0.036920	0.059757	0.023750	0.007791
Total Loans/Total Assets	0.543481	0.542928	0.712820	0.339867	0.080217
Total Deposits/Total Assets	0.698487	0.703967	0.859781	0.514656	0.081475
Fee Income/Total Income	0.063770	0.063669	0.112238	0.025733	0.021265
Investment Securities/Total Assets	0.196296	0.204546	0.471606	0.075431	0.068925
Observations	80 (5 years*16 banks)				

Source: Annual Bank Report for each bank in 2010-2014