ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Efficiency, Technical Change, and Returns to Scale in Large U.S. Banks: Panel Data Evidence from an Output Distance Function Satisfying Theoretical Regularity

Guohua Feng and Apostolos Serletis

June 2009

Working Paper 5/09

Efficiency, Technical Change, and Returns to Scale in Large U.S. Banks: Panel Data Evidence from an Output Distance Function Satisfying Theoretical Regularity

Guohua Feng
Department of Econometrics and Business Statistics
Monash University
Victoria 3800,
Australia

and

Apostolos Serletis
Department of Economics
University of Calgary
Calgary, Alberta, T2N 1N4
Canada

June 24, 2009

Abstract

This paper provides parametric estimates of technical change, efficiency change, economies of scale, and total factor productivity growth for large banks (those with assets in excess of $1 billion) in the United States, over the period from 2000 to 2005. This is done by estimating an output distance function subject to theoretical regularity within a Bayesian framework. We find that failure to incorporate theoretical regularity conditions results in mismeasured shadow revenue and/or cost shares, which in turn leads to perverse conclusions regarding productivity growth. Our results from the regularity-constrained model show that total factor productivity of the large U.S. banks grew at an average rate of 1.98% over the sample period. However, our estimates also show a clear downward trend in the growth rate of total factor productivity and our decomposition of the primal Divisia total factor productivity growth index into its three components — technical change, efficiency change, and economies of scale — indicates that technical change is the driving force behind this decline.

JEL classification: C11; D24; G21.

Keywords: Productivity decomposition; Translog output distance function.

1 Introduction

The great transformation of the U.S. banking industry, caused by fundamental regulatory changes together with technological and financial innovations, has stimulated a substantial body of productivity studies of this industry. Regarding regulation, major changes include, but are not limited to, the removal of geographic restrictions and the permission of combinations of banks, securities firms, and insurance companies. On the other hand, the industry has widely adopted various innovations in technology and applied finance. One of the most important consequences of these regulatory changes and technological and financial innovations has been financial consolidation, leading to larger and more complex banking organizations. In fact, according to Jones and Critchfield (2005), the asset share of large banks in the United States (those with more than $10 billion in assets) increased dramatically from 42 percent in 1984 to 73 percent in 2003. When the large banks with assets between 1 billion and 10 billion are also included, then the asset share of large banks in the U.S. would be as high as 86%. The increasing dominance of large banks in the U.S. commercial banking industry makes the productivity analysis of them particularly attractive.

The literature investigating productivity in the banking industry has been dominated by two methodologies: nonparametric Data Envelopment Analysis (DEA for short) and the parametric Stochastic Frontier Analysis (SFA for short). First put forward by Charnes *et al.* (1978), the DEA approach is a linear programming technique where the efficient frontier is formed as the piecewise linear combination that connects the set of best-practice observations in the dataset under analysis, yielding a convex production possibility set; see Berger and Humphrey (1997). Due to its non-parametric nature, however, the DEA approach does not provide as much insight into market structure and firm behavior as the parametric SFA approach does. For example, returns to scale has to be imposed a priori when the DEA approach is employed, thus rendering the identification of the contribution of scale economies impossible. See Ray and Desli (1997) and Atkinson *et al.* (2003).

The SFA approach, based on the ideas of Aigner *et al.* (1977) and Meeusen and van den Broeck (1977), involves the estimation of a specific parameterized efficiency frontier with a composite error term consisting of non-negative inefficiency and noise components. Two commonly used efficiency frontiers in this literature are cost and profit functions. See, for example, surveys in Berger and Humphrey (1997) and Berger *et al.* (1999). More recently, output distance functions are gaining increasingly popularity in the measurement and analysis of productivity in the banking industry. See, for example, Orea (2002) and Lovell (2003). Compared with the cost or profit function approach, the distance function approach has the major advantage of not requiring information on prices, and therefore can be used in situations where price information is missing, distorted or inaccurate.

Despite its increasing popularity, the output distance function used in previous studies suffers from the problem of inconsistency with theoretical regularity conditions. Microeconomic theory requires that the output distance function satisfies the regularity conditions of monotonicity and curvature. In particular, monotonicity requires that the output distance function be non-increasing in input and non-decreasing in outputs; and curvature requires that the output distance function be quasi-convex in inputs and convex in outputs. See Färe and Grosskopf (1994, p.38). However, none of the previous studies in the banking productivity literature have checked or imposed those theoretical regularity conditions when violated — see, for example, Orea (2002), Lovell (2003) and Chaffai *et al.* (2001).

While permitting a parameterized function to depart from the neoclassical function space is usually fit-improving, it can lead to misleading conclusions regarding productivity and efficiency. First, failure to incorporate theoretical regularity conditions into the estimation causes the hypothetical best practice frontier not to be fully efficient at those data points where theoretical regularity conditions are violated. This will result in mismeasured magnitudes of efficiency levels, which in turn leads to mismeasured productivity growth. See Feng and Serletis (2009). In addition, an inaccurately estimated hypothetical best practice frontier, on which technical change is measured, also implies mismeasured technical change. Second, in the case of the distance elasticity based productivity index used in this paper, the violation of the monotonicity condition can also result in mismeasured distance elasticity shares (shadow revenue/cost shares), which are used as weights to aggregate output and input growth. This will also result in mismeasured productivity growth. For example, a negative elasticity of the output distance function with respect to some input implies that an increase in the use of the corresponding input (with all other inputs and outputs held constant) will increase the (measured) productivity of that bank, which is economically implausible.

Motivated by the widespread practice of ignoring the theoretical regularity conditions, the purpose of this paper is to reinvestigate the efficiency, technical change, and returns to scale of large banks in the United States with more recent panel data over the sample period from 2000 to 2005, and by addressing the above theoretical regularity violation problem inherent in previous studies. In doing so, we first start with a parametric output-oriented (output distance function based) productivity index, by drawing on ideas suggested by Denny *et al.* (1981). We then decompose this productivity index into three components: the contribution of scale economies, technical change, and technical efficiency change.

To obtain the estimates of the three productivity components, we use a translog output distance function, estimated subject to full theoretical regularity. There are three approaches to incorporating theoretical regularity conditions into flexible functional forms: the Cholesky factorization approach, the nonlinear constrained optimization approach, and the Bayesian approach. For most flexible functional forms, the Cholesky factorization approach can only guarantee the curvature conditions in a region around the reference point (that is, a data point where curvature is imposed), and satisfaction of curvature at data points far away from the reference point can only be obtained by luck.[1] See Feng and Serletis (2008).

[1] The normalized quadratic flexible functional form, introduced by Diewert and Wales (1987), is an ex-

This is not satisfactory, especially when the sample size is large and violations of curvature are widespread. The nonlinear constrained optimization approach, originally proposed by Gallant and Golub (1984) and recently used by Feng and Serletis (2009) in the context of a stochastic cost frontier model, develops computational methods for imposing curvature and monotonicity restrictions at any arbitrary set of points. Though powerful, a problem with this approach is that it is difficult to obtain statistical inference when the sample size is large. This is because, in this case, the bootstrapping method needed to obtain statistical inference is unaffordably time consuming. See Feng and Serletis (2009). The Bayesian approach imposes theoretical regularity conditions either by using the accept-reject algorithm — see Terrel (1996) — or the random-walk Metropolis-Hastings algorithm — see, for example, O'Donnell and Coelli (2005). In this paper, we use the latter algorithm to impose monotonicity and curvature on the translog output distance function.

In this paper we take a different approach than that used in Feng and Serletis (2009). First, a Bayesian approach is used to impose the theoretical regularity conditions, while the nonlinear constrained optimization approach is used in Feng and Serletis (2009). An important advantage of the Bayesian approach is that it can provide exact statistical inference on the productivity components (i.e. firm efficiency, technical change, and returns to scale). In contrast, the constrained optimization approach used in Feng and Serletis (2009) provides only point estimates of the productivity components without statistical inference, which is apparently unsatisfactory. In fact, there are two methods for obtaining confidence intervals when the nonlinear inequality constrained optimization approach is used: by inversion of the chi-squared Wald or likelihood ratio statistics and by bootstrapping — see Schoenberg (1997) and Gallant and Golub (1984) for more details. However, the former method is very limited in that it works only when the unknown parameters are not on the boundary; while the latter method, as argued by Feng and Serletis (2009), is unaffordably time consuming when the sample size is large. To get around the statistical inference problem associated with the constrained optimization approach used in Feng and Serletis (1009), we use the Bayesian approach in this paper.

Another difference between this paper and Feng and Serletis (2009) is that here we estimate an output distance function, while Feng and Serletis (2009) estimate a cost function. A drawback with the cost function approach, when used in the banking efficiency literature, is that it may suffer from a severe measurement problem in calculating input prices — see Koetter (2006) and Mountain and Thomas (1999). In particular, in estimating cost functions, almost all existing studies in the banking efficiency literature assume that the input (factor) markets are competitive. This assumption requires a bank to be a price taker, purchasing inputs at given prices, determined exogenously in the respective input markets. In the absence of true input prices, most studies proxy them by relating individual banks' factor payments to employed production factors (i.e. dividing expenses by their respective stock of

ception where global curvature conditions can be imposed using the Cholesky factorization approach.

inputs). Consequently, the input prices used in these studies are bank specific and can vary greatly across banks. Clearly, these bank specific input prices contradict the assumption of competitive input markets and may be poor proxies for input prices — see also Berger and Mester (2003). As pointed out by Greene (1993), the poor measurement of true explanatory variables (i.e. input prices in this case) may distort efficiency estimations substantially. To overcome the potential problem occurring in the measurement of input prices, in this paper we use an output distance function which requires quantity information only.

The rest of the paper is organized as follows. In Section 2, we briefly discuss the parametric output-oriented productivity index and decompose it into three components: contribution of scale economies, technical change, and technical efficiency change. In Section 3 we present the translog output distance function and specify the homogeneity, monotonicity, and curvature constraints. In Section 4 we discuss Bayesian estimation procedures for imposing theoretical regularity on the parameters of the translog output distance function. Section 5 deals with data issues. In Section 6 we apply our methodology to a panel data of 292 large banks in the United States, discuss the effects of incorporating monotonicity and curvature, and also report our estimates of total factor productivity growth and its components. The last section summarizes and concludes the paper.

2 Theoretical Framework

2.1 Output Distance Functions

Before introducing the parametric output-oriented (output distance function based) productivity index, we first define the production technology and the output distance function. Assuming $\boldsymbol{x}\in R_{+}^{N}$ and $\boldsymbol{y}\in R_{+}^{M}$ represent the input and output vectors at time $t = 1, 2, ..., T$, the feasible production technology can be defined

$$P^{t}\left(\boldsymbol{x}^{t}\right) = \left\{\boldsymbol{y}^{t} : \boldsymbol{y} \text{ is producible from } \boldsymbol{x}\right\}.$$

The production technology satisfies a standard set of axioms including convexity, strong disposability, closedness and boundedness. See Färe and Primont (1995) for more details. An output distance function can then be defined as in Shephard (1970)

$$D_{o}^{t}\left(\boldsymbol{y}^{t}, \boldsymbol{x}^{t}\right) = \inf_{\theta}\left\{\theta > 0 : \frac{\boldsymbol{y}^{t}}{\theta} \in P^{t}\left(\boldsymbol{x}^{t}\right)\right\}. \tag{1}$$

It gives the minimum amount by which an output vector can be deflated and remains producible with a given input vector. It also coincides with the Farrel type output oriented measure of technical efficiency. Consistent with the assumptions satisfied by the production technology, the output distance functions is non-decreasing, convex and linearly homogeneous in outputs, and non-increasing and quasi-convex in inputs — see Färe and Grosskopf (1994, p38).

For notational simplicity, we follow the common practice in the literature and model the effect of time through an exogenous time variable, t. Thus, the output distance function defined in (1) can be rewritten as $D_o(\boldsymbol{x}, \boldsymbol{y}, t)$. As suggested by (1), $D_o(\boldsymbol{x},\boldsymbol{y}, t) \leq 1$. Deviations of the output distance function from one, due to technical inefficiency, can be accommodated as follows,

$$D_o(\boldsymbol{x}, \boldsymbol{y}, t)\psi(t) = 1, \tag{2}$$

where $\psi(t) \geq 1$.

2.2 The Parametric Output-Oriented Productivity Index and Its Decomposition

The parametric output-oriented productivity index used in this paper draws on ideas suggested by Denny *et al.* (1981) who developed measures of productivity growth from an estimated multi-output cost function. In Denny *et al.* (1981), the revenue shares, which are used as weights to aggregate output growth in the conventional Divisia productivity index, were replaced by cost elasticities to allow for nonmarginal cost pricing. A detailed discussion can be found in Fuss (1994).

Replacing the cost elasticities in Denny *et al.* (1981) by their corresponding elasticities of the output distance function with respect to outputs and the observed cost shares by their corresponding normalized elasticities of the output distance function with respect to inputs, as in Lovell (2003) and Orea (2002), we can obtain the parametric output-oriented (output distance function) based productivity index:

$$\left.\frac{d\ln TFP}{dt}\right|_{\text{Primal}} = \sum_{m=1}^{M} \tilde{\omega}_m \dot{y}_m - \sum_{n=1}^{N} \omega_n \dot{x}_n, \tag{3}$$

$$\tilde{\omega}_m = \frac{\partial \ln D_o\left(\boldsymbol{y}, \boldsymbol{x}, t\right)}{\partial \ln y_m}, \tag{4}$$

and

$$\omega_n = \frac{\partial \ln D_o\left(\boldsymbol{y}, \boldsymbol{x}, t\right)/\partial \ln x_n}{\sum_{k=1}^{N} \partial \ln D_o\left(\boldsymbol{y}, \boldsymbol{x}, t\right)/\partial \ln x_k} \tag{5}$$

A desirable characteristic of the parametric output-oriented productivity index in (3) is that it does not make restrictive assumptions about returns to scale and market structure, since elasticities are used as weights for both output and input growth.

Since $\tilde{\omega}_m$ and ω_n are actually the shadow revenue for output m and the shadow cost share for input n, respectively, the following requirements are needed

$$\sum_{m=1}^{M} \tilde{\omega}_m = 1 \text{ and } \sum_{m=1}^{M} \omega_n = 1, \tag{6}$$

where the former can be easily shown to be guaranteed by the linear homogeneity of the output distance function in outputs, and the latter is satisfied by definition.

To further simplify the notation in (5), we define

$$\varepsilon_n = \frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial \ln x_n}, \tag{7}$$

and

$$\varepsilon = -\sum_{n=1}^{N} \varepsilon_n,$$

so that ω_n in equation (7) can thus be rewritten as

$$\omega_n = -\frac{\varepsilon_n}{\varepsilon},$$

where ε has been shown by Färe and Grosskopf (1994, p. 103) to be the returns to scale (RTS) in terms of the output distance function.

Equations (3), (4), and (5) provide a basic framework for further decomposing the total factor productivity growth index using the output distance function. In particular, totally differentiating equation (2) with respect to time (after taking logs of both sides) and rearranging yields

$$\sum_{m=1}^{M} \frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial \ln y_m} \dot{y}_m = -\frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial t} - \frac{d \ln \psi(t)}{dt} - \sum_{n=1}^{N} \frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial \ln x_n} \dot{x}_n. \tag{8}$$

Substituting (8) into (3) yields

$$\left.\frac{d \ln TFP}{dt}\right|_{\text{Primal}} = TC + \Delta TE + SC, \tag{9}$$

where

$$TC = -\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t) / \partial t; \tag{10}$$

$$\Delta TE = -\partial \ln \psi(t) / \partial t; \tag{11}$$

$$SC = (\varepsilon - 1) \sum_{n=1}^{N} \left(-\frac{\varepsilon_n}{\varepsilon}\right) \dot{x}_n. \tag{12}$$

The first term in (9) is a primal measure of the rate of technical change. In terms of the output distance function, it captures the shift in the best practice distance frontier. In fact, it is a continuous time version of the technical change term in the Malmquist productivity index, which measures the shift in technology between the two periods evaluated at x_t and x_{t+1}. The second term is a primal measure of the change in technical efficiency. It represents the rate at which an observed firm is moving towards or away from the frontier. It is positive (negative) as technical efficiency increases (decreases) over time. It should be noted that what matters to productivity growth is not the level of technical efficiency, but its improvement over time. The third term captures the contribution of economies of scale. It is positive when increasing returns to scale prevails ($\varepsilon > 1$ in this case), negative when decreasing returns to scale prevails ($\varepsilon < 1$ in this case), and vanishes when constant returns to scale is present.

3 The Translog Output Distance Function

In order to implement the decomposition of total factor productivity growth, we need to parameterize and calculate the parameters of an output distance function. Here we choose to parameterize $D_o(\boldsymbol{y}, \boldsymbol{x}, t)$ as a translog function, which is the functional form often employed to model bank technology. See, for example, Orea (2002) and Chaffai *et al.* (2001). The translog output distance function, defined over M outputs and N inputs can be written as

$$\begin{aligned}\ln D_o(\boldsymbol{y}, \boldsymbol{x}, t) = a_0 &+ \sum_{m=1}^{M} a_m \ln y_m + \frac{1}{2}\sum_{m=1}^{M}\sum_{p=1}^{M} a_{mp} \ln y_m \ln y_p \\ &+ \sum_{n=1}^{N} b_n \ln x_n + \frac{1}{2}\sum_{n=1}^{N}\sum_{j=1}^{N} b_{nj} \ln x_n \ln x_j + \delta_t t + \frac{1}{2}\delta_{tt} t^2 \\ &+ \sum_{n=1}^{N}\sum_{m=1}^{M} g_{nm} \ln x_n \ln y_m + \sum_{m=1}^{M} \delta_{ym} t \ln y_m + \sum_{n=1}^{N} \delta_{xn} t \ln x_n, \end{aligned} \tag{13}$$

where t denotes a time trend. Symmetry requires $a_{mp} = a_{pm}$ and $b_{nj} = b_{jn}$. The restrictions required for homogeneity of degree one in outputs are

$$\sum_{m=1}^{M} a_m = 1;$$

$$\sum_{p=1}^{M} a_{mp} = 0 \quad \text{for all } m = 1, 2, \cdots, M;$$

$$\sum_{m=1}^{M} g_{nm} = 0 \quad \text{for all } n = 1, 2, \cdots, N;$$

$$\sum_{m=1}^{M} \delta_{ym} = 0.$$

One way of imposing these restrictions is to normalize the function by one of the outputs — see, for example, Lovell *et al.* (1994) and O'Donnell and Coelli (2005). This specific transformation through normalization has the advantage of converting equation (13), which is difficult to estimate directly, into an estimable regression model. We choose the Mth output for normalization, which leads to the following expression

$$\ln D_o\left(\frac{\boldsymbol{y}}{y_M}, \boldsymbol{x}, t\right) = \ln\left[\frac{1}{y_M} D_o\left(\boldsymbol{y}, \boldsymbol{x}, t\right)\right].$$

Using the homogeneity restriction, replacing $-\ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)$ with $u = \ln(\psi)$, and adding a random error, v, yields the stochastic output distance function

$$-\ln y_M = a_0 + \sum_{m=1}^{M-1} a_m \ln\left(\frac{y_m}{y_M}\right) + \frac{1}{2}\sum_{m=1}^{M-1}\sum_{p=1}^{M-1} a_{mp} \ln\left(\frac{y_m}{y_M}\right) \ln\left(\frac{y_p}{y_M}\right)$$

$$+ \sum_{n=1}^{N} b_p \ln x_p + \frac{1}{2}\sum_{n=1}^{N}\sum_{j=1}^{N} b_{nj} \ln x_n \ln x_j + \delta_t t + \frac{1}{2}\delta_{tt} t^2$$

$$+ \sum_{n=1}^{N}\sum_{m=1}^{M-1} g_{nm} \ln x_n \ln\left(\frac{y_m}{y_M}\right) + \sum_{m=1}^{M-1} \delta_{ym} t \ln\left(\frac{y_m}{y_M}\right) + \sum_{n=1}^{N} \delta_{xn} t \ln x_n + u + v, \quad (14)$$

where the v's are assumed to be independently and identically distributed (iid) as $N(0, \sigma^2)$, intended to capture statistical noise; $u = -\ln D$ is a nonnegative random variable, intended to capture technical inefficiency. We assume that u follows an exponential distribution with scale parameter λ, which we will discuss in more detail in Section 4. Further, we assume that v and u are independent of each other, an assumption we maintain throughout this paper.

Technical efficiency, technical change, and returns to scale can thus be shown, respectively, to be

$$TE = \exp(-u); \tag{15}$$

$$TC = -\frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial t} = -\left(\delta_t + \delta_{tt} t + \sum_{m=1}^{M} \delta_{ym} \ln y_m + \sum_{n=1}^{N} \delta_{xn} \ln x_n\right); \tag{16}$$

$$RTS = -\sum_{n=1}^{N} \frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial \ln x_n}. \tag{17}$$

Equation (15) can then be used to obtain efficiency change, $\Delta TE = -du/dt$, and (17) can be used to obtain the scale effect,

$$(\varepsilon - 1) \sum_{n=1}^{N} \left(-\frac{\varepsilon_n}{\varepsilon}\right) \dot{x}_n.$$

3.1 Monotonicity Constraints

As required by microeconomic theory, the output distance function (13) has to satisfy the theoretical regularity conditions of monotonicity and curvature. Monotonicity requires that $D_o(\boldsymbol{y}, \boldsymbol{x}, t)$ is non-increasing in $\boldsymbol{x}$ and non-decreasing in $\boldsymbol{y}$. That is,

$$\frac{\partial D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial x_n} \leq 0 \quad \text{and} \quad \frac{\partial D(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial y_m} \geq 0, \tag{18}$$

or, equivalently,

$$\frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial \ln x_n} \leq 0 \quad \text{and} \quad \frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial \ln y_m} \geq 0, \tag{19}$$

since $x_n / D_o(\boldsymbol{y}, \boldsymbol{x}, t) > 0$ and $y_m / D_o(\boldsymbol{y}, \boldsymbol{x}, t) > 0$. The monotonicity restrictions in (18) ensure that the outputs and inputs have nonnegative shadow prices. When expressed in elasticities as in (19), these monotonicity restrictions are critically important in ensuring that the shadow revenue and cost shares are nonnegative when decomposing the productivity growth index in (3).

The monotonicity conditions in (18) can be understood intuitively within the context of banking inputs and outputs. In particular, the condition $\partial D_o(\boldsymbol{y}, \boldsymbol{x}, t) / \partial x_n \leq 0$ in (18) means that the technical efficiency of a bank does not increase when usage of any input increases. In other words, for the same amount of outputs produced, a bank that requires more inputs cannot be more technically efficient. Similarly, $\partial D(\boldsymbol{y}, \boldsymbol{x}, t) / \partial y_m \geq 0$ in (18) means that the technical efficiency of a bank does not decrease when production of any output increases. Put it differently, for the same amount of inputs required, a bank that produces more outputs cannot be less technically efficient.

We now explicitly produce the monotonicity conditions for the output distance function

$$k_n = \frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial \ln x_n}$$

$$= b_n + \sum_{j=1}^{N} b_{nj} \ln x_j + \sum_{m=1}^{M} g_{nm} \ln y_m + \delta_{xn} t \leq 0, \text{ for } n = 1, \cdots, N; \quad (20)$$

$$r_m = \frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial \ln y_m}$$

$$= a_m + \sum_{p=1}^{M} a_{mp} \ln y_p + \sum_{n=1}^{N} g_{nm} \ln x_n + \delta_{ym} t \geq 0, \text{ for } m = 1, \cdots, M. \quad (21)$$

Noting that [see (3)]

$$\sum_{m=1}^{M} \frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial \ln y_m} = 1,$$

the monotonicity condition for the Mth output can be also rewritten as

$$1 - \sum_{m=1}^{M-1} \frac{\partial \ln D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial \ln y_m} \geq 0.$$

3.2 Curvature Constraints

Curvature requires that the output distance function $D_o(\boldsymbol{y}, \boldsymbol{x}, t)$ be quasi-convex in inputs and convex in outputs. See Färe and Grosskopf (1994, p. 38) for more details. For $D_o(\boldsymbol{y}, \boldsymbol{x}, t)$ to be quasi-convex in $\boldsymbol{x}$ it is sufficient that all the principal minors of the following bordered Hessian matrix

$$\boldsymbol{F} = \begin{bmatrix} 0 & f_1 & \cdots & f_N \\ f_1 & f_{21} & \cdots & f_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ f_N & f_{N1} & \cdots & f_{NN} \end{bmatrix},$$

are negative, where

$$f_n = \frac{\partial D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial x_n} = \frac{k_n D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{x_n},$$

and

$$f_{nj} = \frac{\partial^2 D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial x_n \partial x_j} = (b_{nj} + k_n k_j - \phi_{nj} k_n) \frac{D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{x_n x_j},$$

with $\phi_{nj} = 1$ if $n = j$ and 0 otherwise. Noting that factoring out $D_o(\boldsymbol{y}, \boldsymbol{x}, t)/x_n$ from the rows and $1/x_j$ from the columns of $\boldsymbol{F}$ does not change the signs of its principal minors, we can consider the following matrix

$$\widetilde{\boldsymbol{F}} = \begin{bmatrix} 0 & \tilde{f}_1 & \ldots & \tilde{f}_N \\ \tilde{f}_1 & \tilde{f}_{11} & \ldots & \tilde{f}_{1N} \\ \vdots & \vdots & \ldots & \vdots \\ \tilde{f}_N & \tilde{f}_{N1} & \ldots & \tilde{f}_{NN} \end{bmatrix}$$

where $\tilde{f}_n = k_n$, and $\tilde{f}_{nj} = b_{nj} + k_n k_j - \phi_{nj} k_n$. Thus, for $D_o(\boldsymbol{y}, \boldsymbol{x}, t)$ to be quasi-convex in $\boldsymbol{x}$ it is sufficient that all the principal minors of $\widetilde{\boldsymbol{F}}$ are negative.

Convexity in outputs will be ensured if and only if all the principal minors of the Hessian matrix,

$$\boldsymbol{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ h_{21} & h_{22} & \cdots & h_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ h_{1M} & h_{2M} & \cdots & h_{MM} \end{bmatrix},$$

are non-negative, where

$$h_{mp} \equiv \frac{\partial^2 D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{\partial y_m \partial y_p} = (a_{mp} - r_m r_p - \phi_{mp} r_m) \frac{D_o(\boldsymbol{y}, \boldsymbol{x}, t)}{y_m y_p},$$

for $m, p = 1, \cdots, M$ and $\phi_{mp} = 1$ if $m = p$ and 0 otherwise. Note that factoring out $D_o(\boldsymbol{y}, \boldsymbol{x}, t)/y_m$ from the rows and $1/y_p$ from the columns of $\boldsymbol{H}$ does not change the signs of its principal minors. Hence, we can simplify the problem by considering the following matrix

$$\widetilde{\boldsymbol{H}} = \begin{bmatrix} \tilde{h}_{11} & \tilde{h}_{12} & \cdots & \tilde{h}_{1M} \\ \tilde{h}_{21} & \tilde{h}_{22} & \cdots & \tilde{h}_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ \tilde{h}_{1M} & \tilde{h}_{2M} & \cdots & \tilde{h}_{MM} \end{bmatrix},$$

where

$$\tilde{h}_{mp} = a_{mp} - r_m r_p - \phi_{mp} r_m. \tag{22}$$

Thus, the distance function will be convex in outputs if and only if $\widetilde{\boldsymbol{H}}$ is positive-semidefinite.

Compared with the monotonicity conditions, which are easy to understand intuitively within the context of banking inputs and outputs, the curvature conditions are more technical in that they are mainly used in proving the duality theorem — see Färe and Grosskopf (1994) and Kuosmanen (2003). In particular, convexity in outputs of the output distance function imposes convexity on output correspondence, while quasi-convexity in inputs of the output distance function imposes convexity on input correspondence. At a more intuitive level, convexity in outputs implies that if two combinations of bank output levels can be produced with a given bank input vector $\boldsymbol{x}$, then any average of these output vectors can also be produced. This implicitly requires the bank output to be continuously divisible — see Greene and Heller (1981) and Kuosmanen (2003) for a more detailed discussion of convexity and its implications in economics. Similarly, quasi-convexity in inputs, which imposes convexity on input correspondence, implies that if a given output vector $\boldsymbol{y}$ can be produced with two combinations of bank input levels, then it can be produced with any average of these input vectors. This implicitly requires the bank inputs to be continuously divisible.

4 Bayesian Estimation

As noted above, we need to use an estimation method that is capable of imposing the above theoretical regularity conditions. In this paper, we choose the Bayesian method, whose merits have been discussed above. With the translog function for $D_o(\boldsymbol{y},\boldsymbol{x},t)$, the stochastic output distance function in (14) can be rewritten in a panel data framework as

$$q_{it} = \boldsymbol{z}_{it}'\boldsymbol{\beta} + u_{it} + v_{it}, \tag{23}$$

where $i = 1, \cdots, K$ indicates firms, $t = 1, \cdots, T$ indicates time, $q_{it} = -\ln y_{3,it}$, $\boldsymbol{z}_{it}$ is a vector comprising all the variables which appear on the right hand side of (14), and $\boldsymbol{\beta}$ refers to the corresponding vector of coefficients of the translog function (including the intercept).

The formulation of our empirical model as a random effects model (23) is convenient for Bayesian analysis. Although equation (14) can also be formulated as a fixed effects model, we prefer a random effects model. This is because, with a fixed effects model, we have to specify the same number of intercepts as that of observational units, which makes the implementation of Bayesian estimation methods cumbersome, since we have 292 banks in this study. The Bayesian procedures for estimating stochastic frontier models with and without constraints can be found in Koop and Steel (2001) and O'Donnell and Coelli's (2005). Given that there are relatively few Bayesian applications in the banking efficiency literature to date, we briefly detail the estimation procedure.

The use of Bayesian procedures requires choosing prior parameter values. Following Koop and Steel (2001) and O'Donnell and Coelli (2005), we adopt the following prior for $\boldsymbol{\beta}$

$$p\left(\boldsymbol{\beta}\right) \propto I\left(\boldsymbol{\beta} \in R_j\right), \tag{24}$$

where $I(\cdot)$ is an indicator function which takes the value 1 if the argument is true and 0 otherwise, and R_j is the set of permissible parameter values when no theoretical regularity constraints ($j = 0$) are imposed and when both the monotonicity and curvature constraints ($j = 1$) must be satisfied. This particular prior for $\boldsymbol{\beta}$ allows us to slice away the portion of posterior density that violates monotonicity and curvature of the output distance function. We also follow O'Donnell and Coelli (2005) and adopt the following prior for h

$$p(h) \propto h^{-1}, \quad \text{where } h = \frac{1}{\sigma^2} > 0. \tag{25}$$

As stated above, we choose an exponential distribution for u_{it}. This is mainly because van den Broek *et al.* (1994) argue that this distribution for inefficiency u_{it} is more robust to prior assumptions about parameters than other distributions. Since the exponential distribution is a special case of the gamma distribution, the prior for u_{it} is

$$p\left(u_{it} \left| \lambda^{-1}\right.\right) = f_{\text{Gamma}}\left(u_{it} \left| 1, \lambda^{-1}\right.\right), \tag{26}$$

where f_{Gamma} is a gamma density function. According to Fernandez *et al.* (1997), in order to obtain a proper posterior we need a proper prior for the remaining parameter, λ. Accordingly, we use the proper prior

$$p(\lambda^{-1}) = f_{\text{Gamma}}(\lambda^{-1} \left| 1, -\ln \tau^*\right.), \tag{27}$$

where τ^* is the prior median of the efficiency distribution.

With the priors (24)-(27), our joint prior probability density function is therefore

$$f\left(\boldsymbol{\beta}, h, \boldsymbol{u}, \lambda^{-1}\right) = p\left(\boldsymbol{\beta}\right) p(h) p\left(\boldsymbol{u} \left| \lambda^{-1}\right.\right) p(\lambda^{-1})$$

$$\propto h^{-1} I\left(\boldsymbol{\beta} \in R_j\right) f_{\text{Gamma}}\left(\lambda^{-1} \left| 1, -\ln \tau^*\right.\right) \prod_{i=1}^{K} \prod_{t=1}^{T} f_{\text{Gamma}}(u_{it} \left| 1, \lambda^{-1}\right.). \tag{28}$$

Finally, our best prior for the efficiency of large banks in the United States is the mean efficiency value of 0.899 reported by Tsionas (2006) who applied a Bayesian cost frontier (without constraints) to 128 large U.S. banks. In fact, after reviewing the results of 50 U.S. bank efficiency studies, Berger and Humphrey (1997) found that the annual average efficiency is 0.84 with a standard deviation of 0.07. So we are comfortable following Tsionas (2006), setting $\tau^* = 0.899$ in this study.

The likelihood function can be shown to be

$$L\left(\boldsymbol{q} \left| \boldsymbol{\beta}, h, \boldsymbol{u}, \lambda^{-1}\right.\right) = \prod_{i=1}^{K} \prod_{t=1}^{T} \left\{ \sqrt{\frac{h}{2\pi}} \exp\left[-\frac{h}{2} \left(q_{it} - \boldsymbol{z}_{it}' \boldsymbol{\beta} - u_{it}\right) \right]^2 \right\}$$

$$\propto h^{K \times T/2} \exp\left[-\frac{h}{2} \boldsymbol{v}' \boldsymbol{v} \right], \tag{29}$$

where $\boldsymbol{v} = (\boldsymbol{q} - \boldsymbol{z}'\boldsymbol{\beta} - \mathbf{I}_{KT}\boldsymbol{u})$, with $\mathbf{I}_{KT}$ being the $KT \times KT$ identity matrix.

Using Bayes's Theorem and combining the likelihood function in (29) and the joint prior distribution in (28), we obtain the posterior joint density function

$$f\left(\boldsymbol{\beta}, h, \boldsymbol{u}, \lambda^{-1} \left| \boldsymbol{q}\right.\right) \propto h^{(KT/2-1)} \exp\left[-\frac{h}{2}\boldsymbol{v}'\boldsymbol{v}\right] I\left(\boldsymbol{\beta} \in R_j\right) \times$$

$$\times\, f_{\text{Gamma}}\left(\lambda^{-1}|1, -\ln\tau^*\right) \prod_{i=1}^{K}\prod_{t=1}^{T} f_{\text{Gamma}}(u_{it}\left|1, \lambda^{-1}\right.). \tag{30}$$

Also, technical change (TC), elasticities (ε_n), returns to scale (RTS), and total factor productivity growth are all functions of $\boldsymbol{\beta}, h, \boldsymbol{u}$, and λ^{-1}. We are particularly interested in the posterior marginal densities of $\boldsymbol{\beta}$, $\boldsymbol{u}$, TE, ε_n, RTS, and TFP growth, and the means and standard deviations of these posterior densities.

Let $g(\boldsymbol{\beta}, h, \boldsymbol{u}, \lambda^{-1})$ represent these functions of interest. In theory, we could obtain the moments of $g(\boldsymbol{\beta}, h, \boldsymbol{u}, \lambda^{-1})$ from the posterior density through integration. Unfortunately, these integrals cannot be computed analytically. Therefore, we use the Gibbs sampling algorithm which draws from the joint posterior density by sampling from a series of conditional posteriors. Essentially, Gibbs sampling involves taking sequential random draws from full conditional posterior distributions. Under very mild assumptions [see, for example, Tierney (1994)], these draws then converge to draws from the joint posterior. Once draws from the joint distribution have been obtained, any posterior feature of interest can be calculated.

The full conditional posterior distributions for $\boldsymbol{\beta}$, h, $\boldsymbol{u}$, and λ^{-1}can be shown to be

$$p\left(\lambda^{-1} \left| \boldsymbol{q}, \boldsymbol{\beta}, h, \boldsymbol{u}\right.\right) \propto f_{\text{Gamma}}\left(\lambda^{-1} \left| KT + 1, \boldsymbol{u}'\boldsymbol{\iota}_{KT} - \ln\tau^*\right.\right); \tag{31}$$

$$p\left(h \left| \boldsymbol{q}, \boldsymbol{\beta}, \boldsymbol{u}, \lambda^{-1}\right.\right) \propto f_{\text{Gamma}}\left(h \left| \frac{KT}{2}, \frac{1}{2}\boldsymbol{v}'\boldsymbol{v}\right.\right); \tag{32}$$

$$p\left(\boldsymbol{\beta} \left| \boldsymbol{q}, h, \boldsymbol{u}, \lambda^{-1}\right.\right) \propto f_{\text{Normal}}\left[\boldsymbol{\beta} \left| \boldsymbol{b}, h^{-1}\left(\boldsymbol{z}'\boldsymbol{z}\right)^{-1}\right.\right] I\left(\boldsymbol{\beta} \in R_j\right); \tag{33}$$

$$p\left(\boldsymbol{u} \left| \boldsymbol{q}, \boldsymbol{\beta}, h, \lambda^{-1}\right.\right) = f_{\text{Normal}}\left(\boldsymbol{u} \left| \boldsymbol{q} - \boldsymbol{z}'\boldsymbol{\beta} - (h\lambda)^{-1}\boldsymbol{\iota}_{KT}, h^{-1}\mathbf{I}_{KT}\right.\right) \prod_{i=1}^{K}\prod_{t=1}^{T} I\left(\boldsymbol{u}_{it} \geq 0\right), \tag{34}$$

where $\boldsymbol{b} = \left(\boldsymbol{z}'\boldsymbol{z}\right)^{-1}\boldsymbol{z}'\left[\boldsymbol{q} - \mathbf{I}_{KT}\boldsymbol{u}\right]$, with $\boldsymbol{\iota}_{KT}$ being the KT vector of ones, and f_{Normal} is a normal density function.

The Gibbs sampler for Bayesian estimation without monotonicity and curvature constraints can be implemented by setting $I\left(\boldsymbol{\beta} \in R_0\right)$ in (33) equal to one and then drawing sequentially from the conditional posteriors in (31)–(34). Sampling from (31), (32), and (33) is straightforward. However, sampling from (34), a multivariate truncated normal distribution, is more complicated. Luckily, in our particular case, sampling from the multivariate truncated normal distribution (34) can be simplified as KT independent draws from the following univariate truncated normal distribution

$$p\left(\boldsymbol{u}_{it}\left|\boldsymbol{q}, \boldsymbol{\beta}, h, \lambda^{-1}\right.\right) = f_{\text{Normal}}\left(\boldsymbol{q}_{it} - \boldsymbol{z}_{it}'\boldsymbol{\beta} - (h\lambda)^{-1}, h^{-1}\right) I\left(\boldsymbol{u}_{it} \geq 0\right), \tag{35}$$

by noting that the covariance matrix is a scalar times an identity matrix, and the truncations are independent. Sampling from univariate truncated normal distributions can be easily implemented, using procedures discussed in Albert and Chib (1996).

The Gibbs sampler for Bayesian estimation with monotonicity and curvature constraints also involves taking sequential random draws from the above full conditional posterior distributions. Sampling from (31), (32), and (34) is the same as in the case without monotonicity and curvature constraints. However, sampling from the multivariate normal distribution (33) is even more involved than sampling from the multivariate normal distribution (34) in that the region to which $\boldsymbol{\beta}$ is truncated cannot be explicitly specified. There are two approaches in this literature which can be used to handle the sampling from the truncated multivariate normal distribution like (33): the accept-reject algorithm [see Terrell (1996)] and the Metropolis-Hastings (M-H) algorithm — see, for example, O'Donnell and Coelli (2005). The accept-reject algorithm has been criticized for its inefficiency in that it needs to generate an extremely large number of candidate draws before finding one that is acceptable — see O'Donnell and Coelli (2005). In this paper, we follow O'Donnell and Coelli (2005) and sample the truncated multivariate normal distribution (33) using the Metropolis-Hastings algorithm.

5 The Data

The data used in this study are obtained from the Reports of Income and Condition (Call Reports) over the six-year period ($T = 6$) from 2000 to 2005. We examine only continuously operating banks to avoid the impact of entry and exit and to focus on the performance of a core of healthy, surviving institutions during the sample period. In this paper, we selected the subsample of large banks, namely those with total assets in excess of one billion dollars (in 2000 dollars) in the last three year in the sample. This gives a total of 292 banks ($K = 292$) observed over 6 years.

To select the relevant variables, we follow the commonly-accepted intermediation approach proposed by Sealey and Lindley (1977), which treats deposits as inputs and loans as outputs. On the input side, three inputs are included. The quantity of labor, x_1; the quantity of purchased funds and deposits, x_2; and the quantity of physical capital, x_3, which includes premises and other fixed assets. On the output side, three outputs are specified.

These are securities, y_1, which includes all non-loan financial assets (i.e., all financial and physical assets minus the sum of consumer loans, non-consumer loans, securities, and equity); consumer loans, y_2; and non-consumer loans, y_3, which is composed of industrial, commercial, and real estate loans. All the quantities are constructed by following the data construction method in Berger and Mester (2003). These quantities are also deflated by the CPI to the base year 2000, except for the quantity of labor.

While non-traditional activities are clearly increasing in importance, the wide range of activities and imperfect data make the measurement of non-traditional activities problematic. See Stiroh (2000) for a discussion of the different approaches to the measurement of non-traditional activities. To avoid the uncertainties associated with the introduction of non-traditional activities, we choose not to include it as an output. But we do run an alternative model where non-traditional activities are considered as an extra output to check the robustness of the estimates of technical change.

6 Empirical Results

6.1 Regularity Tests

We start with the unconstrained parameter estimates and make 50,000 draws discarding the first 20,000 as a burn in. Table 1 presents the estimated parameters and also reports both standard deviations and 90% posterior density regions calculated as the 5th and 90th percentiles of the MCMC sample observations. We calculate 90% posterior density regions because it provides a better indication of likely values of the parameters when the marginal posterior distributions are asymmetric — see O'Donnell and Coelli (2005).

Regularity tests can be implemented by analyzing the estimated unconstrained marginal posterior pdfs of k_n and r_m and the principal minors of $\widetilde{\boldsymbol{F}}$ and $\widetilde{\boldsymbol{H}}$. We first evaluate the posterior means of k_n and r_m and the principal minors of $\widetilde{\boldsymbol{F}}$ and $\widetilde{\boldsymbol{H}}$, at each of the 1752 $(= K \times T)$ observations, and then calculate the proportions of regularity violations relative to the total number of observations. The results, presented in the first column of Table 2, indicate that only two (k_2 and r_1) of the six monotonicity conditions are satisfied at all the 1752 observations and that both curvature conditions are violated, with the quasi-convexity in outputs being violated at all observations. We then evaluate the posterior coverage regions of k_n and r_m and of the principal minors of $\widetilde{\boldsymbol{F}}$ and $\widetilde{\boldsymbol{H}}$, again at each of the 1752 observations, and calculate the ratio of the number of observations, where posterior coverage regions span inadmissible values, to the total number of observations (1752). As can be seen in the second column of Table 2, all eight regularity conditions have a positive probability of being violated at some observations. In fact, both of the curvature conditions have a positive probability of being violated at all the 1752 observations.

These violations of monotonicity and curvature in the unconstrained model may lead to perverse conclusions concerning TFP growth. To see this, we also generate the marginal density plots for the shadow input cost shares, ω_n for $n = 1, 2, 3$ in (3), and the shadow output revenue shares, $\tilde{\omega}_m$ for $m = 1, 2, 3$ in (3), from the unconstrained model, evaluated at the mean value of all inputs and outputs in each year. As discussed above, both ω_n and $\tilde{\omega}_m$ are required to be positive and less than one. Due to space limitations, only the marginal densities in 2005 are plotted in Figure 1.1-1.6 — the marginal densities for other years are similar to those in 2005. Clearly, all the three shadow output shares are reasonable, containing no negative values or values larger than one. However, the plot of the shadow input shares shows that the labor share and the capital share may be negative. A negative input share implies that an increase in the use of that input (with all other inputs and outputs held constant) will increase the (measured) productivity of that bank, which is economically implausible. Moreover, Figure 1.2 shows that the shadow input share for funds may be greater than one, implying that an increase in the use of that input (with all other inputs and outputs held constant) will reduce the (measured) productivity of that bank by more than the growth rate of funds, which is again economically implausible.

Since monotonicity and curvature are not attained in the unconstrained model, we follow the procedures specified in Section 4 to impose those constraints on the translog output distance function. Again, we generated a total of 50,000 observations, and then discarded the first 20,000 as a burn-in. The associated estimates of parameters are reported in Table 3, the monotonicity and curvature violations reported in Table 4, and the marginal densities for the shadow input and output shares are plotted in Figure 2.1–2.6. Generally speaking, the constrained model has smaller posterior standard deviations and narrower Bayesian credible intervals in terms of posterior moments for the estimated parameters and shadow revenue and cost shares. This is consistent with Dorfman and McIntosh (2001) and O'Donnell and Coelli (2005) who find that incorporating inequality constraints into the estimation process has the effect of reducing the variances of the estimated marginal pdfs. In addition, Figures 2.1–2.6 show that some densities are asymmetric — for example, those for the funds share, capital share, and non-consumer loans share. Kleit and Terrell (2001) found similar results and suggested that the asymmetry perhaps reflects the fact that the constrained posterior density slices away the portion of the unconstrained posterior density that violates monotonicity and curvature.

As we expected, monotonicity and curvature are satisfied by all measures after monotonicity and curvature are incorporated. In particular, k_n and r_m and the principal minors of $\widetilde{\boldsymbol{F}}$ and $\widetilde{\boldsymbol{H}}$ are correctly signed at all 1752 observations whether they are evaluated by using posterior means or by using posterior coverages. Moreover, the shadow shares are all positive and less than one. In what follows, we will discuss technical efficiency, technological change, returns to scale, and the contributions of each of these components to TFP growth, based on the constrained translog output distance function.

6.2 Consequences of Failure to Impose Theoretical Regularity

Having estimated both the unconstrained and constrained models, in the section we investigate the consequences of failing to impose the theoretical regularity conditions in terms of i) efficiency and productivity rankings of banks, ii) identification of best- and worst-practice banks, and iii) estimates of efficiency and productivity levels.

6.2.1 Rank-Order Correlations of Efficiency and Productivity Growth

As pointed by Bauer *et al.* (1998), identifying the ordering of which financial institutions are more efficient than others is usually more important for regulatory policy decisions than measuring the level of efficiency. Therefore, we first examine whether failure to impose theoretical regularity changes the ranking of individual banks both in terms of efficiency and in terms of productivity growth. For this purpose, we calculate the Spearman rank correlation coefficient between the unconstrained model and the constrained model, for both the case of efficiency ranking and the case of productivity growth ranking — see Table 5. Formally, the Spearman rank correlation coefficient can be written a

$$\rho = 1 - \frac{6\sum_{j=1}^{n_k}(\text{Rank}_{j1} - \text{Rank}_{j2})^2}{n_k(n_k^2 - 1)}, \tag{36}$$

where n_k is the number of banks in the sample, Rank_{j1} is the rank of bank j based on the constrained version of the model, and Rank_{j2} is the rank of the same bank based on the unconstrained version of the model. If $\rho = -1$, there is perfect negative correlation; if $\rho = 1$, there is perfect positive correlation; and if $\rho = 0$, there is no correlation.

We follow Betta and Pietrosanto (2008) and bootstrap each of the 10 Spearman rank correlation coefficients 10000 times. We report the 99% bootstrap confidence intervals in Table 5 (in parentheses); if the confidence interval does not contain unity, the Spearman rank correlation coefficient is significantly different from one at the 1% level. As can be seen from Table 5, none of the 99% bootstrap confidence intervals contain one, suggesting that all the Spearman rank correlation coefficients reported in the table are significantly different from one at the 1% level. In the case of the efficiency ranking of banks (see panel A of Table 5), the rank correlation coefficients range from 0.8210 to 0.9210. This suggests that the imposition of theoretical regularity changes the ranking of banks in terms of efficiency. Moreover, as can be seen in panel B of Table 5, the Spearman rank correlation coefficients for productivity growth are also different than 1, ranging from 0.8254 to 0.9213, suggesting that the imposition of theoretical regularity also changes the productivity growth ranking of banks.

In terms of magnitude, the Spearman rank correlation coefficients in Table 5 are comparable to those reported in earlier studies. For example, Bauer *et al.* (1998), using data on U.S. banks with assets greater than $100 million, over the period from 1977 to 1988, find that the Spearman rank correlation coefficients among the efficiency scores created by different parametric techniques range from 0.484 to 0.976. In particular, they find that the Spearman rank correlation coefficient between the stochastic frontier approach (SFA) and the distribution free approach (DFA) is 0.897, a value comparable in magnitude to those reported here. As is well known, however, SFA and DFA make very different assumptions regarding the distribution of efficiency and also separate inefficiencies from the random error in very different ways — see, for example, Kumbhakar and Lovell (2003, p. 179). In other words, the change in efficiency and productivity ranking caused by failure to impose theoretical regularity conditions is comparable to that caused by switching from one parametric technique (i.e. SFA) to another very different parametric technique (i.e. DFA).

6.2.2 Identification of Best-Practice and Worst-Practice Banks

The identification of the most efficient and least efficient banks is also important for some regulatory purposes. We therefore investigate the consequences of not imposing theoretical regularity in this regard. We calculate the proportion of banks that are identified by the unconstrained model as having efficiency (productivity growth) scores in the least efficient (productive) 25% that are also identified in the bottom quarter by the constrained model — see Table 6.1. We also calculate the proportion of banks that are identified by the unconstrained model as having efficiency (productivity growth) scores in the most efficient (productive) 25% that are also identified in the top quarter by the constrained model — see Table 6.2. We refer to these proportions as 'correspondence proportions' (π). If $\pi = 1$, there is perfect correspondence between the unconstrained and constrained models; and if $\pi = 0$, there is no correspondence.

We also bootstrap (10000 times) each of the 20 correspondence proportions in Tables 6.1 and 6.2 and report the 99% bootstrap confidence intervals in parentheses. As can be seen from these tables, none of the 99% bootstrap confidence intervals contain one, suggesting that the best (or worst) 25% of banks identified by the unconstrained model are different from those identified by the constrained model. In particular, the proportion of banks that are identified by the unconstrained model as having efficiency scores in the least efficient 25% that are also identified in the bottom quarter by the constrained model ranges from 0.6575 to 0.7945 (see panel A of Table 6.1). Also the proportion of banks that are identified by the unconstrained model as having productivity growth scores in the least productive 25% that are also identified in the bottom quarter by the constrained model ranges from 0.6986 to 0.8082 (see panel B of Table 6.1). As can be seen in panel A of Table 6.2, the proportion of banks that are identified by the unconstrained model as having efficiency scores in the most efficient 25% that are also identified in the top quarter by the constrained model ranges from 0.7703 to 0.8514. The proportion of banks that are identified by the unconstrained model as having productivity growth scores in the most productive 25% that are also identified in the top quarter by the constrained model ranges from 0.6757 to 0.8649 (see panel B of Table

6.2).

As with the Spearman rank correlation coefficients in Table 5, the correspondence proportions reported in Tables 6.1 and 6.2 are also comparable (in terms of magnitude) to those among efficiency scores produced by different parametric techniques reported by Bauer *et al.* (1998). For example, they find that the correspondence of best practice banks between SFA and DFA is 0.795, and that the correspondence of worst practice banks between SFA and DFA is 0.790. This shows that failure to impose theoretical regularity conditions will lead to misidentification of best-practice and worst-practice banks, and the extent of this misidentification is comparable to that caused by changing from one parametric technique (i.e. SFA) to another very different parametric technique (i.e. DFA).

6.2.3 Effects on Efficiency and Productivity Growth Estimates

We are also interested in how failure to impose theoretical regularity affects efficiency and productivity growth at the individual bank level. In Table 7 we report on the distribution of the difference in technical efficiency and productivity growth, respectively, between the unconstrained and the constrained model, using the mean difference and the 5th and 95th percentile values.

Regarding efficiency, as can be seen in panel A of Table 7, the mean difference in technical efficiency ranges from -5.64% to -5.46% with an average of -5.54%. That is, compared with the technical efficiency estimates obtained from the constrained model, those obtained from the unconstrained model are 5.54% smaller on average. The decrease in technical efficiency when the theoretical regularity conditions are not imposed is not surprising, considering that the vector of frontier outputs obtained from the constrained model is smaller than that obtained from the unconstrained model, due to the restrictions imposed on the production technology set, $P^t\left(\boldsymbol{x}^t\right)$, by the theoretical regularity conditions. Moreover, in terms of magnitude, an average reduction of 5.54% in technical efficiency estimates caused by failure to impose theoretical regularity conditions is by no means small. For example, taking a representative large bank with the mean value of inputs and outputs in 2005, a reduction of 5.54% in technical efficiency implies that, without changing its technology and inputs, this bank could potentially increase its annual production of securities (y_1), consumer loans (y_2) and non-consumer loans (y_3) by 171.7982, 43.0373, and 284.9830 millions dollars, respectively, by simply increasing its technical efficiency.

As for productivity growth, as can be seen in panel B of Table 7, the mean difference in productivity growth between the unconstrained and constrained model ranges from 0.05% to 0.71%, with an average of 0.41%. The latter figure is not small, considering that the average productivity growth over the sample period based on the constrained model is 1.98% (see Table 11). In other words, failure to impose theoretical regularity conditions leads to an overestimate of average productivity growth by 20.69% (= 0.41%/1.98%). This figure will be even larger if the differences in productivity growth are calculated in absolute value, since negative and positive differences in productivity growth cancel each other out partially when taking the average (this is not a problem for the differences in efficiency, which are all negative). Also, a close examination of the 5th and 95th percentiles (in panel B of Table 7) reveals that, compared with the average difference in productivity growth, the differences in productivity growth for individual banks can be very large. Taking the difference in productivity growth in 2001 as an example, the 95% percentile is 3.14% while the 5% percentile is – 2.42%.

In summary, failure to impose theoretical regularity leads to misleading ranking of banks both in terms of technical efficiency and in terms of productivity growth; misidentification of best and worst practice banks; and mismeaured technical efficiency and productivity growth. For these reasons, in what follows we concentrate on the empirical results from the constrained model.

6.3 Results from the Constrained Model

6.3.1 Technical Efficiency

Table 8.1 reports the estimates of average technical efficiency over the sample period, together with the 90% posterior density regions. The average technical efficiency for each year is evaluated at the mean value of all inputs and outputs in that year. As indicated by the standard deviations and 90% density regions, the estimates of the average technical efficiency are statistically significant for every year over the sample period. The scores of technical efficiency show a high level of efficiency, ranging from 92.43% to 93.41%. Thus, on average, a 7% to 8% proportional increase in outputs can be achieved by solely increasing efficiency, without altering production technology and input usage.

Our estimates of technical efficiency are quite close to those from recent research; see, for example, Stiroh (2000) and Tsionas (2006). Both of these studies employed a translog cost frontier (dual method), rather than a distance frontier (primal method). Thus, one of the differences in efficiency estimates could be due to allocative efficiency. For example, Tsionas used the panel data on 128 large U.S. banks over the period from 1989 to 2000 and found that the average efficiency is 88.9% when a dynamic effect is not considered and 95.5% when a dynamic effect is considered. Further, our technical efficiency estimates show no specific pattern of temporal change. In particular, it starts at 93.41% in 2000, falls to 92.49% in 2001, rebounds slightly in the following two years, falls slightly again in 2004, and picks up again to 92.69% in 2005. This time pattern of technical efficiency means that the change in technical efficiency is not a consistent source of TFP growth.

To get a better understanding of the distribution of technical efficiency across banks, in Table 8.2 we report the minimum and maximum technical efficiency in each year, together with standard deviations, and the 5th and 95th percentile values. The results show that the scores of technical efficiency can differ greatly across banks in all the sample years. Taking the technical efficiencies in 2005 as an example, the highest is 97.63% whereas the lowest is only 35.08%. Despite these extreme cases, the results on standard deviations and the 5th and 95th percentile values show that the vast majority of the banks fall within the range between 84% and 96%.

6.3.2 Returns to Scale

Table 9 summarizes the returns to scale (RTS) estimates, again evaluated at the mean value of all inputs and outputs each year. The standard deviations and 90% density regions indicate that the RTS estimates are statistically significant for every year over the sample period. Clearly, the point estimates of RTS in Table 7 are all greater than one, ranging from 1.037 to 1.056, suggesting that the large commercial banks in the sample exhibit moderate increasing returns to scale. This is consistent with the findings in Bikker and Haaf (2002) and Claessens and Laeven (2003) that the U.S. banking industry is characterized by monopolistic competition.

The presence of moderate increasing returns to scale also has two implications for productivity growth. First, the presence of moderate increasing returns to scale implies that productivity growth will exhibit procyclical behavior to some extent. This is because the contribution of scale economies to productivity growth is positive when the share weighted input aggregate grows over time, but negative when the share weighted input aggregate declines over time, as can be seen from (12). Second, since the economies of scale is moderate in magnitude, the scale effect will not be a consistent significant source of TFP growth. In addition, the presence of moderate increasing returns to scale also implies that the large banks in the U.S. are expected to be engaged in more mergers and acquisitions until the returns to scale are exploited.

6.3.3 Technical Change

Table 10.1 reports technical change rate estimates, again evaluated at the mean value of all inputs and outputs each year. Again, the standard deviations and 90% density regions indicate that the estimates are statistically significant in all years except 2004. On average, the rate of technical change is 2.22% per year. Compared with the estimates of technical efficiency, which show no specific pattern of temporal change, the estimates of the rate of technical change show a declining trend. In particular, the rate of technical change falls consistently from 6.0% in 2000 to -1.79% in 2005. In terms of the output distance function, this means that the frontier is moving outward more slowly over time and even moving inward at the end of the sample period.

Considering the importance of technical change, we also estimated two alternative models to check the robustness of our results regarding the time pattern of technical change. In the first alternative model (Model 1), we treat securities (instead of non-consumer loans) as the numeraire for normalizing the outputs, to see whether the choice of the numeraire has any effect on the time pattern of technical change. In the second alternative model (Model 2), we add an off-balance-sheet variable to see whether the exclusion of non-traditional activities affects the estimated time pattern of the rate of technical change. The estimates of the rate of technical change, together with 90% posterior density regions, from the three alternative models are reported in Table 10.2.

The estimates of the rate of technical change from the first alternative model (Model 1), reported in the first column of Table 10.2, are almost the same as those in Table 10.1 (our standard model), suggesting that the choice of the numeraire has almost no effect on the estimated time pattern of the rate of technical change. When the off-balance-sheet variable is added, the technical change rate estimates change on average by 0.45% in absolute terms. However, as can be clearly seen in the third column of Table 10.2, the time pattern of technical change is still almost the same. As we discussed above, the wide range and imperfect data of the non-traditional activities could introduce more uncertainty regarding the estimates of the rate of technical change. Thus, the second alternative model (Model 2) is not our preferred model.

In summary, the time pattern and (to a lesser degree) magnitude of the rate of technical change estimates are very robust to the different choice of the numeraire output and the inclusion of off-balance-sheet variables.

6.3.4 TFP Growth and Its Components

We now turn to a decomposition of the growth rate of total factor productivity, as shown in Table 11. It should be noted that the first year in the sample period is dropped because we have to difference the technical efficiencies in two consecutive years to obtain efficiency changes. Again, all the estimates are evaluated at the mean values of all inputs and outputs in each year. In addition to the estimates of the three TFP growth components, we also calculate the percentage contribution of each of the three productivity components to total factor productivity growth, shown in brackets in Table 11.

Overall, the results presented in the first column of Table 11 indicate that total factor productivity grew in all years, except the last, at an average annual rate of 1.98%. However, the estimates for total factor productivity growth also exhibit a clear downward trend. In particular, total factor productivity growth is quite impressive in the first three years, in all exceeding 2%. But, it falls almost to zero in 2004 and even turns negative in the last year in the sample. It should be noted that while TFP growth shows a downward trend, TFP level has been increased over the sample period except the last. In particular, if we normalize the productivity level in 2000 to 100, then the productivity level in the last year will be 109.91.

The decomposition of total factor productivity growth in Table 11 identifies the forces that drive its decline. In particular, the estimates for efficiency changes, $-du/dt$, in the second column of Table 11 are rather small in magnitude, averaging only 0.14% per year.

Moreover, they fluctuate around zero, indicating that efficiency change has a minor effect on total factor productivity growth. The small effects of efficiency changes on total factor productivity growth are also reflected in the percentage contribution to total factor productivity growth, reported in Column 3 of Table 11, averaging 7.27% per year. The estimates reported in the fourth column of Table 11 indicate that the scale effect has a moderate positive effect on total factor productivity growth, averaging 0.44% per year. In terms of average percentage contributions, the scale effect is the second largest factor contributing to growth in total factor productivity (22.30%). This is consistent with our estimates of returns to scale, which show moderate economies of scale in large commercial banks in the United States.

Without doubt, the last component, technical change, is the dominant force behind total factor productivity growth. This can be clearly seen from the average annual rate of technical change (of 1.39%) in column 6 of Table 11. The importance of technical change can also be seen from its percentage contribution: it contributes over 75% each year to productivity growth. Further, the technical change estimates show a clear downward trend, accounting for the decline in total factor productivity growth over the sample period.

6.4 Sensitivity Analysis

A possible problem with our estimation of the output distance function is endogeneity. That is, the regressors on the right hand side of equation (14) may not be exogenous. To investigate the robustness of our results to alternative estimation procedures, in this subsection we use instrumental variables.

The variables on the right hand of (14) can be classified into two types of variables: the output ratio variables (i.e. y_m/y_M, $m = 1, \cdots, M-1$) and the input variables. According to Coelli and Perelman (1999), the output ratios are measures of the output mix which are more likely to be exogenous. Schmidt (1988) and Mundlak (1996) also find that, in the context of a production function, the input ratios do not suffer from the endogeneity problem; the basic argument also applies to the output ratios in the transformed output distance function. Thus, the only variables suspected of causing possible endogeneity problems are the input variables. To use instrumental variables for the input variables, we follow the assertion of Griliches (2000, p. 62) that "good instruments are hard to find without the supporting theory that give them a formal role in the model." For the U.S. banking industry, most previous studies find that it is characterized by monopolistic competition — see, for example, Bikker and Haaf (2002) and Claessens and Laeven (2003). Hence, consistent with the theoretical framework of profit maximization in the presence of imperfect competition, input prices and the time trend are chosen as instruments. A similar method of choosing instrumental variables is used in Karagiannis *et al.* (2004) in estimating an input distance function.

The empirical results are summarized in Table 12. A comparison of Tables 11 and 12 reveals that the major conclusions reached in the previous subsection are still valid, although we notice that there are some changes. First, total factor productivity growth still shows a clear downward trend, implying that productivity has been growing at a lower rate. In particular, it has consistently decreased from 0.0491 to 0.033 over the sample period. Second, technical change is still the driving force behind the decline in total factor productivity growth. From the contributions of the three productivity components, we see that technical change is still the dominant force, accounting for 70.32% of the productivity growth on average. With the contributions from the other productivity components being rather small, the consistent decline in technical change (see the last column of Table 12) results in the decline in productivity growth. Third, the estimates of efficiency change and the scale effect when instrumental variables are used are comparable to our earlier estimates. Finally, we also find that the estimates of the contributions of the three productivity components when instrumental variables are used are very similar to our earlier estimates as well. In particular, the average contributions of technical change, scale effect, and efficiency change when instrumental variables are used are 70.32%, 21.94%, and 7.74%, respectively, and they are 70.33%, 22.30%, and 7.27% when instrumental variables are not used. Therefore, our major conclusions in the previous subsection are quite robust to the use of instrumental variables.

6.5 A Comparison with Previous Studies

Unfortunately, previous studies that investigate productivity and efficiency issues of large banks in the United States, use different functional forms, samples periods, and estimation methods than those used in this study. For example, while we apply a translog output distance function to large banks (with assets greater than $1 billion in 2000 U.S. dollars), over the period from 2000 to 2005, and use Bayesian estimation procedures, Bos and Kolari (2005) apply a translog cost frontier to a group of large banks (with assets greater than $1 billion in 1995 U.S. dollars), over the period from 1995 to 1999, and use the maximum likelihood method of estimation. Also, Akhigbe and McNulty (2005) apply a Fourier profit function to a group of large banks (with assets greater than $1 billion in U.S. dollars), for 1995, 1997, 1999 and 2001, using maximum likelihood estimation. These differences in flexible functional forms, samples, and estimation methods make it difficult to provide a meaningful comparison between this study and previous studies and tell whether the differences between our results and those from previous studies are caused by the failure of previous studies to impose theoretical regularity.

However, it is worth providing a comparison between this paper and Feng and Serletis (2009). Both investigate productivity and efficiency issues in the same group of large U.S. banks, over the same sample period, and both impose theoretical regularity, although using different flexible functional forms and different estimation methods. A major difference between them is that Feng and Serletis (2009) find a decline in cost efficiency for large U.S. banks (with assets greater than $1 billion) over the sample period, while this study finds that technical efficiency estimates of those large banks show small temporal variation. There are three potential reasons for this difference. First, different concepts of efficiency are used in these two papers. Feng and Serletis (2009) use cost efficiency, which includes allocative efficiency, whereas this paper uses technical efficiency, which excludes allocative efficiency, in order to keep consistency with the use of the output distance function. Second, the measurement problem in calculating input prices when cost functions are used (as discussed in Section 1) is another potential cause of the difference in the time pattern of the estimates of bank efficiency between these two papers. Third, different approaches to imposing the theoretical regularity conditions are used, with the Bayesian approach used in this study and constrained optimization used in Feng and Serletis (2009). An important difference between the Bayesian and optimization approaches is that in the Bayesian approach the constraints are satisfied for all parameter values where the posterior density is nonzero. That means that all values in probability intervals for parameters will be consistent with the constraints. In the constrained optimization approach, however, only the estimates, not necessarily all values within confidence intervals, will satisfy the constraints. In other words, different approaches to imposing theoretical regularity might lead to different admissible production sets, which in turn lead to different best practice frontiers (both in quantity space). Different best practice frontiers, against which efficiency is measured, will in turn lead to different estimates of bank efficiency.

7 Conclusion

The estimation of output distance functions is gaining increasing popularity in the analysis of bank productivity and efficiency. However, the theoretical regularity conditions (especially those of monotonicity and curvature) required by neoclassical microeconomic theory have been widely ignored in the literature. In this paper, we adopt a Bayesian approach to impose the theoretical regularity conditions on the parameters of a translog output distance function. Implementing the approach involves the use of a Gibbs sampler with data augmentation. A Metropolis-Hastings algorithm is also used within the Gibbs sampler to simulate observations from truncated pdfs. Hence, we provide estimates of technical change, efficiency and returns to scale of large banks in the U.S., subject to theoretical regularity conditions.

Our results confirm that the monotonicity and concavity constrained model yields more accurate and favorable results than an unconstrained model. In particular, shadow revenue and cost shares are well behaved, and the standard deviations are largely reduced. We also find that failure to impose theoretical regularity leads to misleading ranking of banks both in terms of technical efficiency and productivity growth; misidentification of best- and worst-practice banks; and mismeasured technical efficiency and productivity growth. Our results from the constrained model show that total factor productivity grew at an average rate of 1.98% for the large U.S. commercial banks over the sample period. However, the estimates of total factor productivity growth show a clear downward trend and our decomposition of the total factor productivity growth rate indicates that technical change is the driving force that leads to the decline in the total factor productivity growth rate. Our results indicate that returns to scale also have a positive effect on productivity growth, suggesting that the scale effect should be included when examining bank productivity growth.

In estimating technical change, returns to scale, and efficiency in large banks in the United States, we have used a translog output distance function. A locally flexible functional form, the translog is only suitable for samples composed of relatively homogenous firms — for example, only large banks with assets greater than $1 billion are used in this study. In cases where the firms are of widely varying sizes, globally flexible functional forms which can provide greater flexibility will be more appropriate. There are two globally flexible functional forms — the Asymptotically Ideal Model, introduced by Barnett *et al.* (1991), and the Fourier flexible functional form, introduced by Gallant (1982). However, due to the trigonometric terms which are not neoclassical, the Fourier functional forms has been criticized for its possibility of overfitting the data — see, for example, Barnett and Yue (1988). In contrast, with the globally regular Müntz-Szatz series, the AIM model form fits only that part that is globally regular, thus eliminating the risk of overfitting. Therefore, using an AIM output distance function to estimate technical change, returns to scale, and efficiency is an area for potentially productive future research.

References

[1] Aigner, D.J., C.A.K. Lovell, and P. Schmidt. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6 (1977), 21-37.

[2] Akhigbe, A. and J. McNulty. "Profit Efficiency Sources and Differences among Small and Large U.S. Commercial Banks." *Journal of Economics and Finance* 29 (2005), 289-299.

[3] Albert, J.H. and S. Chib. "Computation in Bayesian Econometrics: An Introduction to Markov Chain Monte Carlo." In: *Advances in Econometrics*, Vol 11A: Hill, R.C. (eds.), Greenwich, JAI Press (1996), 3–24.

[4] Atkinson, S.E., C. Cornwell, and O. Honerkamp. "Measuring and Decomposing Productivity Change: Stochastic Distance Function Estimation Versus Data Envelopment Analysis." *Journal of Business and Economic Statistics* 21 (2003), 284-294.

[5] Barnett, W.A. and P. Yue. "Semi-Nonparametric Estimation of the Asymptotically Ideal Model: the AIM Demand System." In *Advances in Econometrics*, Vol VII, Rhodes, G and Fomby, T.B. (eds), Greenwich: JAI Press (1988), 229-252.

[6] Barnett, W.A., J. Geweke, and M. Wolfe. "Semi-nonparametric Bayesian Estimation of the Asymptotically Ideal Production Model." *Journal of Econometrics* 49 (1991), 5-50.

[7] Bauer, P.W., A.N. Berger, G.D. Ferrier, and D.B. Humphrey. "Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Methods." *Journal of Economics and Business* 50 (1998), 85–114.

[8] Berger, A.N. and D.B. Humphrey. "Efficiency of Financial Institutions: International Survey and Directions for Future Research." *European Journal of Operational Research* 98 (1997), 175-212.

[9] Berger, A.N., R.S. Demsetz, and P.E. Strahan. "The Consolidation of the Financial Services Industry: Causes, Consequences, and the Implications for the Future." *Journal of Banking and Finance* 23 (1999), 135-94.

[10] Berger, A.N. and L.J. Mester. "Explaining the Dramatic Changes in the Performance of U.S. Banks: Technological Change, Deregulation, and Dynamic Changes in Competition." *Journal of Financial Intermediation* 12 (2003), 57-95.

[11] Betta, G. and A. Pietrosanto. "A Statistical Approach for Improving the Performance of a Testing Methodology for Measurement Software." *IEEE Transactions on Instrumentation and Measurement* 57 (2008), 1118-1126.

[12] Bikker, J.A. and K. Haaf. "Competition, Concentration and Their Relationship: An Empirical Analysis of the Banking Industry." *Journal of Banking and Finance* 26 (2002), 2191-2214.

[13] Bos, J.W.B. and J.W. Kolari. "Large Bank Efficiency in Europe and the United States: Are There Economic Motivations for Geographic Expansion in Financial Services?" *Journal of Business* 78 (2005), 1555-1592.

[14] Chaffai, M.E., M. Dietsch, and A. Lozano-Vivas. "Technological and Environmental Differences in the European Banking Industries." *Journal of Financial Services Research* 19 (2001), 147-162.

[15] Charnes, A., W.W. Cooper, and E. Rhodes. "Measuring the Efficiency of Decision Making Units." *European Journal of Operations Research* 2 (1978), 429-444.

[16] Claessens S. and L. Laeven. "Financial Development, Property Rights, and Growth." *Journal of Finance* 58 (2003), 2401-2436.

[17] Coelli, T.J. and S. Perelman. "A Comparison of Parametric and Non-parametric Distance Functions: With Application to European Railways." *European Journal of Operational Research* 117 (1999), 326–339.

[18] Denny, M., M. Fuss, and L.Waverman. "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries with an Application to Canadian Telecommunications." In: *Productivity Measurement in Regulated Industries*, T.G. Cowing and R.E. Stevenson (eds.), New York: Academic Press (1981), 179–218.

[19] Diewert, W.E. and T.J. Wales. "Flexible Functional Forms and Global Curvature Conditions." *Econometrica* 55 (1987), 43-68.

[20] Dorfman, J.H. and C.S. McIntosh. "Imposing Inequality Restrictions: Efficiency Gains from Economic Theory." *Economics Letters* 71 (2001), 205-209.

[21] Färe, R. and S. Grosskopf. *Cost and Revenue Constrained Production.* Springer (1994).

[22] Färe, R. and D. Primont. *Multi-Output Production and Duality: Theory and Applications.* Netherlands: Kluwer Academic Publishers (1995).

[23] Feng, G. and A. Serletis. "Productivity Trends in U.S. Manufacturing: Evidence from the NQ and AIM Cost Functions." *Journal of Econometrics* 142 (2008), 281-311

[24] Feng, G. and A. Serletis. "Efficiency and Productivity of the U.S. Banking Industry, 1998-2005: Evidence from the Fourier Cost Function Satisfying Global Regularity Conditions." *Journal of Applied Econometrics* 24 (2009), 105-138.

[25] Fernandez, C., J. Osiewalski, and M.F.J. Steel. "On the Use of Panel Data in Stochastic Frontier Models with Improper Priors." *Journal of Econometrics* 79 (1997), 169-193.

[26] Fuss, M.A. "Productivity Growth in Canadian Telecommunications." *The Canadian Journal of Economics* 27 (1994), 371-392.

[27] Gallant, A.R. "Unbiased Determination of Production Technology." *Journal of Econometrics* 20 (1982), 285-323.

[28] Gallant A.R. and G. Golub. "Imposing Curvature Restrictions on Flexible Functional Forms." *Journal of Econometrics* 26 (1984), 295-321.

[29] Greene W.H. "The Econometric Approach to Efficiency Analysis." In: *The Measurement of Productive Efficiency: Techniques and Applications*. Fried H.O., C.A.K. Lovell, and S.S. Schmidt (eds). Oxford University Press New York (1993), 68-119.

[30] Greene J. and W.P. Heller. "Mathematical Analysis and Convexity with Applications to Economics." in: *Handbook of Mathematical Economics*, Arrow, K.J., and M.D. Intriligator (eds.), North-Holland, Amsterdam (1981), 15-52.

[31] Griliches, Z. *R & D, Education, and Productivity: A Retrospective.* Cambridge, Harvard University Press (2000).

[32] Jones, K.D. and T. Critchfield. "Consolidation in the U.S. Banking Industry: Is the Long, Strange Trip About to End?" *FDIC Banking Review* 17 (2005), 31-61.

[33] Karagiannis, G., P. Midmore, and V. Tzouvelekas. "Parametric Decomposition of Output Growth Using A Stochastic Input Distance Function." *American Journal of Agricultural Economics* 86 (2004), 1044-1057.

[34] Kleit, A. and D. Terrell. "Measuring Potential Efficiency Gains from Deregulation of Electricity Generation: A Bayesian Approach." *Review of Economics and Statistics* 83 (2001), 523-530.

[35] Koetter, M. "Measurement Matters—Alternative Input Price Proxies for Bank Efficiency Analyses." *Journal of Financial Services Research* 30 (2006), 199–227.

[36] Koop, G. and M. Steel. "Bayesian Analysis of Stochastic Frontier Models." In: *A Companion to Theoretical Econometrics*, Baltagi, B. (eds.), MA: Blackwell (2001).

[37] Kumbhakar, S.C. and C.A.K. Lovell. *Stochastic Frontier Analysis*. Cambridge, Cambridge University Press (2003).

[38] Kuosmanen, T. "Duality Theory of Non-convex Technologies." *Journal of Productivity Analysis* 20 (2003), 273–304.

[39] Lovell, C.A.K. "The Decomposition of Malmquist Productivity Indexes." *Journal of Productivity Analysis* 20 (2003), 437–458.

[40] Lovell, C.A.K., S. Richardson, P. Travers, and L.L. Wood. "Resources and Functionings: A New View of Inequality in Australia." In: *Models and Measurement of Welfare and Inequality*, W. Eichhorn (eds.), Berlin: Springer-Verlag Press (1994), 787-807.

[41] Meeusen W. and J. van den Broeck. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review* 18 (1977), 435-444.

[42] Mountain D.C. and H. Thomas. "Factor Price Misspecification in Bank Cost Function Estimation." *Journal of International Financial Markets, Institutions, and Money* 9 (1999), 163–182.

[43] Mundlak, Y. "Production Function Estimation: Reviving the Primal." *Econometrica* 64 (1996), 431-438.

[44] O'Donnell, C.J. and T.J. Coelli. "A Bayesian Approach to Imposing Curvature on Distance Functions." *Journal of Econometrics* 126 (2005), 493-523.

[45] Orea, L. "Parametric Decomposition of a Generalized Malmquist Productivity Index." *Journal of Productivity Analysis* 18 (2002), 5-22.

[46] Ray, S.C. and E. Desli. "Productivity Growth, Technical Progress and Efficiency Change in Industrialized Countries: Comment." *American Economic Review* 87 (1997), 1033-1039.

[47] Schmit, P. "Estimation of Fixed Effect Cobb-Douglas System Using Panel Data." *Journal of Econometrics* 37 (1988), 361-380.

[48] Schoenberg, R. "Constrained Maximum Likelihood." *Computational Economics* 10 (1997), 251–266.

[49] Sealey C. and J. Lindley. "Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions." *Journal of Finance* 32 (1977), 1251-1266.

[50] Shephard, R.W. *Theory of Cost and Production Functions.* Princeton: Princeton University Press (1970).

[51] Stiroh, K.J. "How Did Bank Holding Companies Prosper in the 1990s?" *Journal of Banking and Finance* 24 (2000), 1703-1745.

[52] Terrell, D. "Incorporating Monotonicity and Concavity Conditions in Flexible Functional Forms." *Journal of Applied Econometrics* 11 (1996), 179-194.

[53] Tierney, L. "Markov Chains for Exploring Posterior Distributions (with discussion)" *Annals of Statistics* 22 (1994), 1701-1762.

[54] Tsionas, E.G. "Inference in Dynamic Stochastic Frontier Models." *Journal of Applied Econometrics* 21 (2006), 669-676.

[55] van den Broeck, J., G. Koop, J. Osiewalski, and M. Steel. "Stochastic Frontier Models: A Bayesian Perspective." *Journal of Econometrics* 46 (1994), 39–56.

Table 1

Parameter Estimates From The Unconstrained Model

Variable	Parameter	Estimate	Standard deviation	90% posterior coverage regions
intercept	a_0	0.2060	0.0723	(0.0663, 0.2617)
$\ln x_1$	b_1	−0.0795	0.0418	(−0.1437, − 0.0184)
$\ln x_2$	b_2	−0.9555	0.0357	(−1.0081, − 0.8993)
$\ln x_3$	b_3	−0.0030	0.0276	(−0.0441, 0.0377)
$(\ln x_1)^2$	b_{11}	−0.3669	0.0618	(−0.4754, − 0.2815)
$(\ln x_2)^2$	b_{22}	−0.0268	0.0372	(−0.0815, 0.0273)
$(\ln x_3)^2$	b_{33}	0.09353	0.0281	(0.0499, 0.1352)
$(\ln x_1)(\ln x_2)$	b_{12}	0.2467	0.0399	(0.1906, 0.3114)
$(\ln x_1)(\ln x_3)$	b_{13}	0.1210	0.0344	(0.0719, 0.1782)
$(\ln x_2)(\ln x_3)$	b_{23}	−0.0328	0.0053	(−0.0408, − 0.0247)
$\ln y_1$	a_1	0.3956	0.0209	(0.3635, 0.4261)
$\ln y_2$	a_2	0.1094	0.0102	(0.0942, 0.1246)
$\ln y_3$	a_3	0.4951	0.0202	(0.4651, 0.5260)
$(\ln y_1)^2$	a_{11}	0.0987	0.0231	(0.0584, 0.1296)
$(\ln y_2)^2$	a_{22}	0.0268	0.0039	(0.0207, 0.0326)
$(\ln y_3)^2$	a_{33}	0.1376	0.0218	(0.0997, 0.1680)
$(\ln y_1)(\ln y_2)$	a_{12}	0.0061	0.0066	(−0.0040, 0.0163)
$(\ln y_1)(\ln y_3)$	a_{13}	−0.1047	0.0215	(−0.1342, − 0.0672)
$(\ln y_2)(\ln y_3)$	a_{23}	−0.0328	0.0053	(−0.0408, − 0.0247)
$(\ln x_1)(\ln y_1)$	g_{11}	−0.0211	0.0226	(−0.0565, 0.0125)
$(\ln x_1)(\ln y_2)$	g_{12}	0.0274	0.0132	(0.0080, 0.0479)
$(\ln x_1)(\ln y_3)$	g_{13}	−0.1047	0.0215	(−0.1342, − 0.0672)
$(\ln x_2)(\ln y_1)$	g_{21}	0.0776	0.0196	(0.0483, 0.1083)
$(\ln x_2)(\ln y_2)$	g_{22}	−0.0090	0.0099	(−0.0241, 0.0062)
$(\ln x_2)(\ln y_3)$	g_{23}	−0.0328	0.0053	(−0.0408, − 0.0247)
$(\ln x_3)(\ln y_1)$	g_{31}	−0.0543	0.0156	(−0.0785, − 0.0313)
$(\ln x_3)(\ln y_2)$	g_{32}	−0.0127	0.0082	(−0.0250, − 0.0006)
$(\ln x_3)(\ln y_3)$	g_{33}	0.0671	0.0148	(0.0452, 0.0897)
t	c_t	−0.0867	0.0138	(−0.1073, − 0.0664)
t^2	c_{tt}	0.0163	0.0038	(0.0107, 0.0219)
$(\ln x_1)t$	g_{x1t}	−0.0088	0.0097	(−0.0230, 0.0056)
$(\ln x_2)t$	g_{x2t}	0.0028	0.0079	(−0.0094, 0.0145)
$(\ln x_3)t$	g_{x3t}	0.0053	0.0061	(−0.0036, 0.0145)
$(\ln y_1)t$	g_{y1t}	−0.0229	0.0047	(−0.0300, − 0.0158)
$(\ln y_2)t$	g_{y2t}	0.0002	0.0022	(−0.0031, 0.0034)
$(\ln y_3)t$	g_{y3t}	0.02269	0.0046	(0.0158, 0.0297)

TABLE 2

REGULARITY VIOLATIONS (UNCONSTRAINED MODEL)

Regularity conditions	Regularity violations (at the posterior mean)	pdf > 0 (in inadmissible region)
Monotonicity		
$k_1 \leq 0$	11.59%	89.21%
$k_2 \leq 0$	0%	0.57%
$k_3 \leq 0$	69.29%	98.80%
$r_1 \geq 0$	0%	5.03%
$r_2 \geq 0$	6.51%	42.92%
$r_3 \geq 0$	0.34%	0.74%
Curvature		
All the principal minors of:		
$\widetilde{\boldsymbol{F}}$ are negative, and	100%	100%
$\widetilde{\boldsymbol{H}}$ is positive semidifinite	16.15%	100%

Figure 1. Estimated Distributions of the Shadow Shares from Unconstrained Model Evalauted at Mean Prices in 2005

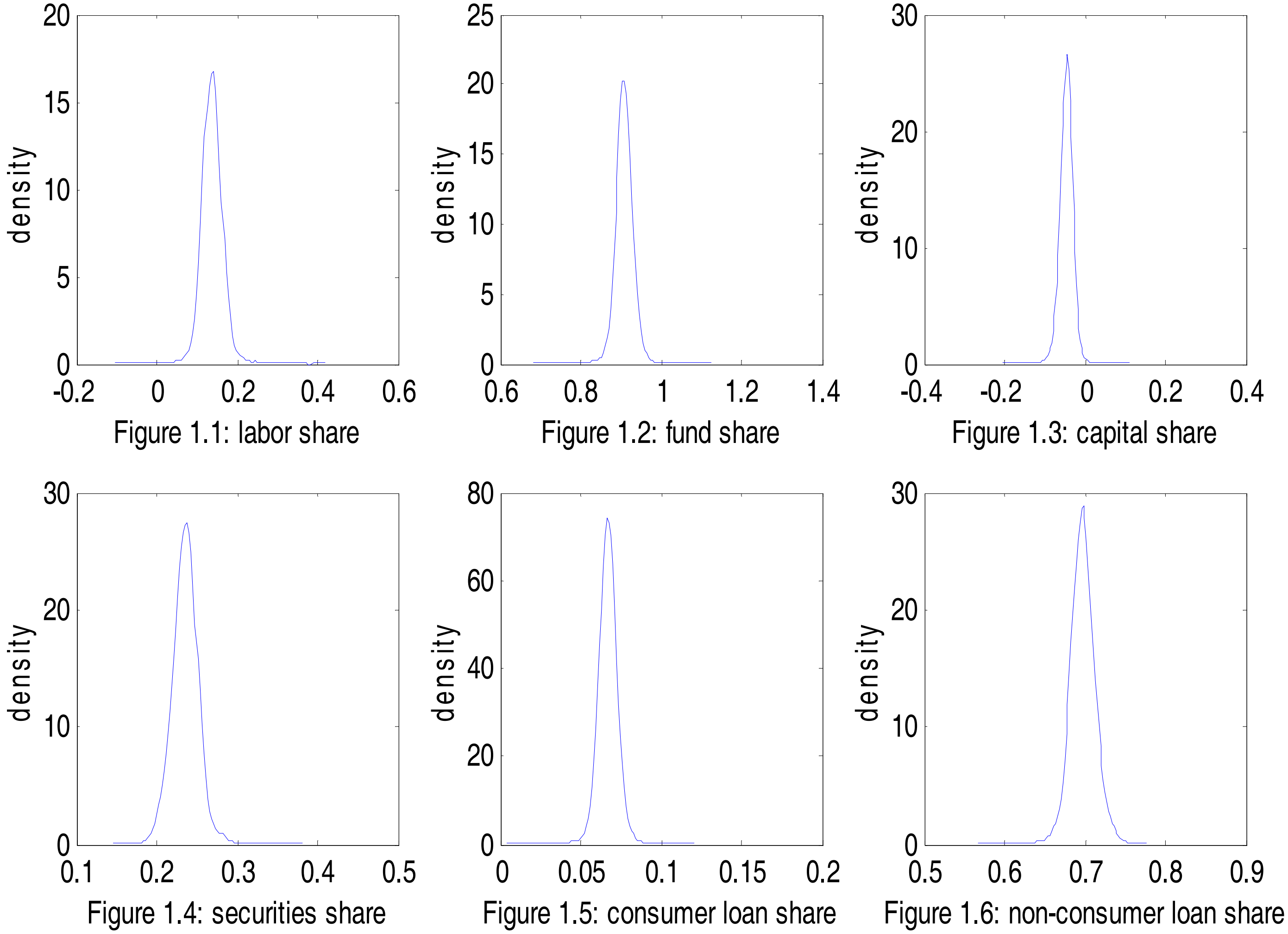

Table 3

Parameter Estimates From The Constrained Model

Variable	Parameter	Estimate	Standard deviation	90% posterior coverage regions
intercept	a_0	0.2548	0.0194	(0.2222, 0.2873)
$\ln x_1$	b_1	−0.1166	0.0223	(−0.1580, −0.0825)
$\ln x_2$	b_2	−0.8705	0.0223	(−0.9094, −0.8363)
$\ln x_3$	b_3	−0.0521	0.0151	(−0.0763, −0.0283)
$(\ln x_1)^2$	b_{11}	−0.0288	0.0112	(−0.0465, −0.0092)
$(\ln x_2)^2$	b_{22}	0.0119	0.0223	(−0.0246, 0.0488)
$(\ln x_3)^2$	b_{33}	0.0076	0.0047	(0.0011, 0.0162)
$(\ln x_1)(\ln x_2)$	b_{12}	0.0140	0.0145	(−0.0105, 0.0370)
$(\ln x_1)(\ln x_3)$	b_{13}	0.0059	0.0042	(−0.0010, 0.0127)
$(\ln x_2)(\ln x_3)$	b_{23}	−0.0243	0.0084	(−0.0386, −0.0112)
$\ln y_1$	a_1	0.3996	0.0169	(0.3741, 0.4301)
$\ln y_2$	a_2	0.1171	0.0059	(0.1069, 0.1264)
$\ln y_3$	a_3	0.4834	0.0171	(0.4524, 0.5098)
$(\ln y_1)^2$	a_{11}	0.0720	0.0076	(0.0590, 0.0837)
$(\ln y_2)^2$	a_{22}	0.0099	0.0007	(0.0087, 0.0111)
$(\ln y_3)^2$	a_{33}	0.0865	0.0054	(0.0776, 0.0951)
$(\ln y_1)(\ln y_2)$	a_{12}	0.0023	0.0023	(−0.0014, 0.0061)
$(\ln y_1)(\ln y_3)$	a_{13}	−0.0743	0.0062	(−0.0842, −0.0639)
$(\ln y_2)(\ln y_3)$	a_{23}	−0.0122	0.0022	(−0.0158, −0.0086)
$(\ln x_1)(\ln y_1)$	g_{11}	−0.0264	0.0107	(−0.0439, −0.0079)
$(\ln x_1)(\ln y_2)$	g_{12}	0.0123	0.0046	(0.0045, 0.0203)
$(\ln x_1)(\ln y_3)$	g_{13}	0.0141	0.0105	(−0.0031, 0.0311)
$(\ln x_2)(\ln y_1)$	g_{21}	0.0582	0.0121	(0.03789, 0.0790)
$(\ln x_2)(\ln y_2)$	g_{22}	0.0064	0.0057	(−0.0042, 0.0152)
$(\ln x_2)(\ln y_3)$	g_{23}	−0.0647	0.0119	(−0.0848, −0.0443)
$(\ln x_3)(\ln y_1)$	g_{31}	−0.0075	0.0032	(−0.0130, −0.0027)
$(\ln x_3)(\ln y_2)$	g_{32}	−0.0023	0.0014	(−0.0046, −0.0002)
$(\ln x_3)(\ln y_3)$	g_{33}	0.0098	0.0036	(0.0041, 0.0158)
t	c_t	−0.0914	0.0120	(−0.1116, −0.0708)
t^2	c_{tt}	0.0183	0.0032	(0.0126, 0.0238)
$(\ln x_1)t$	g_{x1t}	−0.0056	0.0047	(−0.0133, 0.0020)
$(\ln x_2)t$	g_{x2t}	0.0040	0.0046	(−0.0029, 0.0116)
$(\ln x_3)t$	g_{x3t}	0.0010	0.0011	(−0.0009, 0.0028)
$(\ln y_1)t$	g_{y1t}	−0.0158	0.0036	(−0.0219, −0.0098)
$(\ln y_2)t$	g_{y2t}	−0.0021	0.0010	(−0.0037, −0.0003)
$(\ln y_3)t$	g_{y3t}	0.0179	0.0036	(0.0120, 0.0240)

Table 4

Regularity Violations (Constrained Model)

Regularity conditions	Regularity violations (at the posterior mean)	pdf > 0 (in inadmissible region)
Monotonicity		
$k_1 \leq 0$	0%	0%
$k_2 \leq 0$	0%	0%
$k_3 \leq 0$	0%	0%
$r_1 \geq 0$	0%	0%
$r_2 \geq 0$	0%	0%
$r_3 \geq 0$	0%	0%
Curvature		
All the principal minors of		
$\widetilde{\boldsymbol{F}}$ are negative, and	0%	0%
$\widetilde{\boldsymbol{H}}$ is positive semidifinite	0%	0%

Figure 2. Estimated Distributions of the Shadow Shares from Constrained Model Evalauted at Mean Prices in 2005

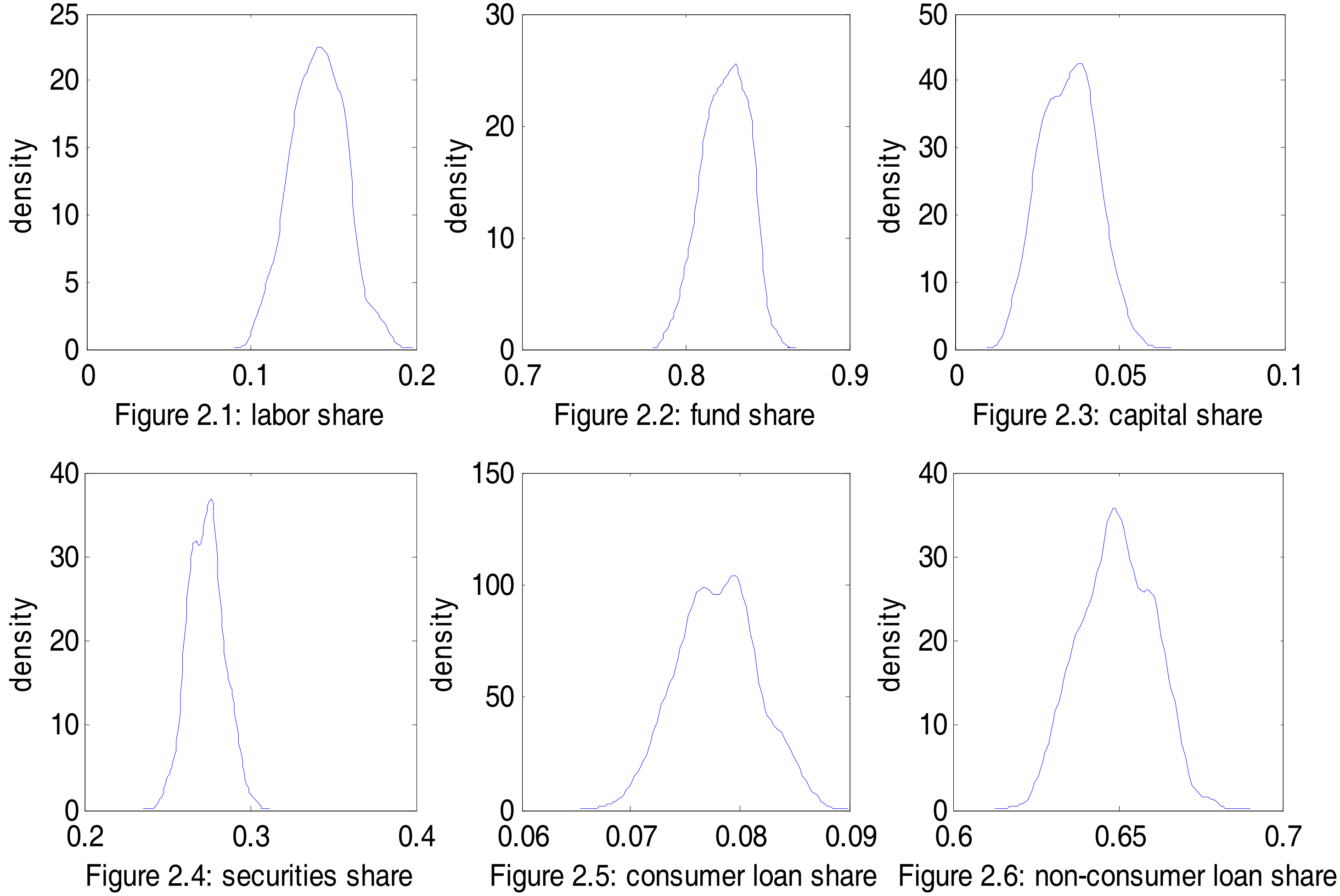

Table 5

Spearman Rank Correlation Coefficients

	A. Efficiency ranking		B. Productivity growth ranking	
Year	Coefficient	99% bootstrap confidence interval	Coefficient	99% bootstrap confidence interval
2001	0.8210	(0.7479, 0.8781)	0.8307	(0.7692, 0.8788)
2002	0.8496	(0.7848, 0.8987)	0.9213	(0.8882, 0.9441)
2003	0.8946	(0.8423, 0.9306)	0.8449	(0.7792, 0.8932)
2004	0.9210	(0.8808, 0.9463)	0.8254	(0.7475, 0.8848)
2005	0.8966	(0.8545, 0.9272)	0.8518	(0.7692, 0.8788)

Note: The Spearman rank correlation coefficient is defined in equation (36).

Table 6.1

Correspondence of Worst Practice Banks
Between Unconstrained and Constrained Models

	A. In terms of efficiency		B. In terms of productivity growth	
Year	correspondence proportion, π	99% bootstrap confidence interval	correspondence proportion, π	99% bootstrap confidence interval
2001	0.6575	(0.5094, 0.7534)	0.7671	(0.6575, 0.8630)
2002	0.6849	(0.5068, 0.7671)	0.7945	(0.6849, 0.8630)
2003	0.7945	(0.6832, 0.8767)	0.6986	(0.5479, 0.7945)
2004	0.7808	(0.6849, 0.8849)	0.8082	(0.7123, 0.9041)
2005	0.7260	(0.5753, 0.8082)	0.7671	(0.6575, 0.8630)

Note: Numbers indicate the proportion of banks that are identified by the unconstrained model as having efficiency (productivity growth) scores in the least efficient (productive) 25% that are also identified in the bottom quarter by the constrained model.

Table 6.2

Correspondence of Best Practice Banks
Between Unconstrained and Constrained Models

	A. In terms of efficiency		B. In terms of productivity growth	
Year	correspondence proportion, π	99% bootstrap confidence interval	correspondence proportion, π	99% bootstrap confidence interval
2001	0.8378	(0.7432, 0.9324)	0.7297	(0.5885, 0.8243)
2002	0.7973	(0.6977, 0.8784)	0.8649	(0.7838, 0.9595)
2003	0.7703	(0.6622, 0.8514)	0.6757	(0.5676, 0.7568)
2004	0.8378	(0.7432, 0.9189)	0.7432	(0.6486, 0.8649)
2005	0.8514	(0.7703, 0.9271)	0.7568	(0.5885, 0.8243)

Note: Numbers indicate the proportion of banks that are identified by the unconstrained model as having efficiency (productivity growth) scores in the most efficient (productive) 25% that are also identified in the top quarter by the constrained model.

Table 7

Differences in Technical Efficiency and Productivity Growth Between Unconstrained and Constrained Models

	A. Efficiency		B. Productivity growth	
Year	Mean difference	5% and 95% percentile	Mean difference	5% and 95% percentile
2001	-0.0546	(-0.0684, -0.0387)	0.0020	(-0.0242, 0.0314)
2002	-0.0564	(-0.0756, -0.0371)	0.0005	(-0.0203, 0.0222)
2003	-0.0548	(-0.0755, -0.0369)	0.0046	(-0.0197, 0.0296)
2004	-0.0553	(-0.0786, -0.0364)	0.0063	(-0.0195, 0.0329)
2005	-0.0562	(-0.0853, -0.0354)	0.0071	(-0.0172, 0.0303)
Average	-0.0554	(-0.0767, -0.0369)	0.0041	(-0.0202, 0.0293)

Notes: Mean difference in efficiency is calculated as the mean of the differences in technical efficiency between the unconstrained and constrained models of all the sample banks. Similarly, mean difference in productivity growth is calculated as the mean of the differences in productivity growth between the unconstrained and constrained models of all the sample banks.

Table 8.1

Average Technical Efficiency

Year	Average technical efficiency	Standard deviation	90% posterior coverage regions
2000	0.9341	0.0048	(0.9259, 0.9418)
2001	0.9249	0.0057	(0.9151, 0.9339)
2002	0.9294	0.0052	(0.9203, 0.9376)
2003	0.9277	0.0054	(0.9183, 0.9361)
2004	0.9243	0.0057	(0.9144, 0.9331)
2005	0.9269	0.0055	(0.9174, 0.9357)

Table 8.2

Distribution of Technical Efficiency Across Banks

Year	Minimum	Maximum	Standard deviation	5% percentile	95% percentile
2000	0.5242	0.9726	0.0335	0.9083	0.9585
2001	0.5245	0.9719	0.0365	0.8882	0.9531
2002	0.4589	0.9789	0.0406	0.8855	0.9616
2003	0.4593	0.9868	0.0441	0.8770	0.9673
2004	0.3717	0.9779	0.0476	0.8700	0.9638
2005	0.3508	0.9763	0.0474	0.8773	0.9603

Table 9

Returns To Scale

Year	Average returns to scale	Standard deviation	90% posterior coverage regions
2000	1.0365	0.0061	(1.0266, 1.0465)
2001	1.0394	0.0047	(1.0315, 1.0474)
2002	1.0413	0.0041	(1.0346, 1.0485)
2003	1.0446	0.0042	(1.0378, 1.0517)
2004	1.0509	0.0047	(1.0430, 1.0583)
2005	1.0560	0.0058	(1.0462, 1.0659)

Table 10.1

Technical Change

Year	Average technical change	Standard deviation	90% posterior coverage regions
2000	0.0684	0.0085	(0.0540, 0.0829)
2001	0.0507	0.0055	(0.0415, 0.0598)
2002	0.0335	0.0030	(0.0282, 0.0383)
2003	0.0153	0.0030	(0.0098, 0.0199)
2004	−0.0051	0.0054	(−0.0143, 0.0040)
2005	−0.0247	0.0083	(−0.0380, −0.0102)

Table 10.2

Technical Change estimates
From Alternative Models

Year	Model 1	Model 2
2000	0.0682 (0.0568, 0.0795)	0.0600 (0.0460, 0.0733)
2001	0.0504 (0.0434, 0.0576)	0.0454 (0.0366, 0.0535)
2002	0.0333 (0.0294, 0.0379)	0.0311 (0.0265, 0.0355)
2003	0.0151 (0.0108, 0.0198)	0.0159 (0.0098, 0.0213)
2004	−0.0053 (−0.0139, 0.0024)	−0.0014 (−0.0115, 0.0092)
2005	−0.0248 (−0.0376, −0.0129)	−0.0179 (−0.0333, −0.0016)

Note: The 90% posterior coverage regions are shown in parentheses.

Table 11

Productivity Change And Its Decomposition

Year	Average productivity change	Efficiency change		Scale effect		Technical change	
		Estimates	Contribution	Estimates	Contribution	Estimates	Contribution
2001	0.0662	0.0092	13.90%	0.0063	9.52%	0.0507	76.59%
	(0.0530, 0.0794)	(0.0013, 0.0172)		(0.0051, 0.0076)		(0.0415, 0.0598)	
2002	0.0311	-0.0045	-14.47%	0.0020	6.43%	0.0335	107.72%
	(0.0211, 0.0409)	(-0.0124, 0.0034)		(0.0017, 0.0024)		(0.0282, 0.0383)	
2003	0.0202	0.0017	8.42%	0.0032	15.84%	0.0153	75.74%
	(0.0107, 0.0296)	(-0.0059, 0.0094)		(0.0027, 0.0037)		(0.0098, 0.0199)	
2004	0.0041	0.0034	82.93%	0.0059	143.90%	-0.0051	-124.39%
	(-0.0087, 0.0166)	(-0.0046, 0.0113)		(0.0050, 0.0067)		(-0.0143, 0.0040)	
2005	-0.0225	-0.0026	11.56%	0.0047	-20.89%	-0.0247	109.78%
	(-0.0395, -0.0049)	(-0.0106, 0.0055)		(0.0039, 0.0056)		(-0.0380, -0.0102)	
Average	0.0198	0.0014	7.27%	0.0044	22.30%	0.0139	70.33%

Notes: The 90% posterior coverage regions are shown in parentheses.

Table 12

Productivity Change And Its Decomposition When Instrumental Variables Are Used

Year	Average productivity change	Efficiency change		Scale effect		Technical change	
		Estimates	Contribution	Estimates	Contribution	Estimates	Contribution
2001	0.0491	0.0021	4.29%	0.0040	8.21%	0.0430	87.50%
	(0.0179, 0.0784)	(-0.0150, 0.0192)		(0.0004, 0.0081)		(0.0204, 0.0645)	
2002	0.0360	0.0007	2.06%	0.0020	5.57%	0.0333	92.36%
	(0.0127, 0.0592)	(-0.0162, 0.0178)		(0.0007, 0.0033)		(0.0192, 0.0472)	
2003	0.0263	0.0008	2.97%	0.0022	8.50%	0.0233	88.53%
	(0.0062, 0.0463)	(-0.0163, 0.0178)		(0.0010, 0.0036)		(0.0145, 0.0333)	
2004	0.0183	0.0032	17.48%	0.0029	16.13%	0.0121	66.39%
	(-0.0039, 0.0407)	(-0.0140, 0.0206)		(0.0012, 0.0052)		(-0.0004, 0.0247)	
2005	0.0033	0.0004	11.89%	0.0023	71.30%	0.0005	16.81%
	(-0.0268, 0.0333)	(-0.0173, 0.0179)		(0.0007, 0.0044)		(-0.0207, 0.0214)	
Average	0.0266	0.0014	7.74%	0.0027	21.94%	0.0224	70.32%

Notes: The 90% posterior coverage regions are shown in parentheses.