How regulatory capital requirement affects banks' productivity: An application on emerging economies banks

Meryem Duygun [a,*], Mohamed Shaban [a], Robin C. Sickles [b], Thomas Weyman-Jones [d]

[a,*] School of Management, University of Leicester, University Road, Leicester, LE1 7RH, UK
[b] Department of Economics, Rice University
[c] School of Business and Economics, Loughborough University

[*] Corresponding author: Tel: +44 116 252 5328; fax: +44 116 252 5515.

E-mail addresses: m.fethi@leicester.ac.uk, m.shaban@leicester.ac.uk, rsickles@rice.edu t.g.weyman-jones@lboro.ac.uk

How regulatory capital requirement affects banks' productivity: An application on emerging economies banks

This version: October 2013

Abstract

This paper presents a novel approach to measure efficiency and productivity decomposition in the banking systems of emerging economies with a special focus on the role of equity capital. We model the requirement to hold levels of a fixed input, i.e. equity, above the long run equilibrium level or, alternatively, to achieve a target equity-asset ratio. To capture the effect of this under-leveraging, we allow the banking system to operate in an uneconomic region of the technology. Productivity decomposition is developed to include exogenous factors such as policy constraints. We use a panel data set of banks in emerging economies during the financial upheaval period of 2005–2008 to analyse these ideas. Results indicate the importance of the capital constraint in the decomposition of productivity.

JEL classification: C23, D24, G21

Keywords: Banking, Efficiency and productivity analysis, Shadow price, Cost function, Regulated capital, Bank capitalization

How regulatory capital requirement affects banks' productivity: An application on emerging economies banks

1. Introduction

When a banking system has undergone a financial shock, there are important lessons to learn from how it reacts, adapts and recovers. These lessons have particularly strong policy implications when most of the developed world is recovering from the financial crisis of 2007–2008. Many emerging economies, however, experienced a number of shocks before this and some began to recover ahead of the developed economies. As a consequence, considerable lessons can be learned from the emerging economies during the last half-decade about financial liberalization, banking system recapitalization and financial crises.

Banking system recapitalization, that is, a greater reliance on equity capital rather than short-term borrowing as a means of providing full loss absorbing capacity for problem loans, is a major preoccupation of policymakers around the world. It is widely believed that a well-capitalized banking system is expected to be less vulnerable to financial crises, whereas an inadequately capitalized banking system is more susceptible to financial shocks (Koutsomanoli-Filippaki et al., 2009). Major recapitalization of the banking systems, however, could impose a resource cost both on the wider economy and on the banking system in particular (Daniel and Jones, 2007)[1].

Our paper attempts to measure this cost and its impact on the banking system. One focus of the research, therefore, will be on measuring the shadow return on equity associated with

[1] In the aftermath of the 2007–2008 financial crisis, this issue has preoccupied regulators; a member of the US Senate Banking Committee asks: "What is the true cost to national economies of higher capital requirements for banks?" Senator Kay Hargan, letter to *The Economist*, June 4, 2010.

capital constraints on the bank balance sheets[2]. A related question is, how has the upheaval in financial markets affected the efficiency and productivity change of banking systems during this period? This paper fills a gap in the literature by departing from the traditional analysis of efficiency and productivity by incorporating regulatory constraints into the cost function. We argue that the balance sheet constraint is a critical aspect to be considered when modelling a banking system cost-minimizing behaviour in order to measure productivity. These ideas are soundly established in the theoretical literature but we wish to develop this theoretical framework into an empirical application. In particular, we aim to measure the productivity cost of changes in the regulatory capital requirements of banks and to relate this to the empirical measurement of the shadow price of equity capital over time and amongst groups of emerging economies.

There is a vast amount of literature on bank efficiency and productivity that examines a number of aspects such as investigating the determinants of efficiency (Canhato and Dermine, 2003; Casu and Molyneux, 2003); ownership (Havrylchyk, 2006; Sturm and Williams, 2004); stock returns and efficiency (Beccalli et al., 2006; Erdem and Erdem, 2008); corporate events and efficiency (Avkiran, 1999; Sherman and Rupert, 2006); regulatory reform, liberalization and efficiency (Brissimis et al., 2008; Fethi et al, 2011; Isik and Hassan, 2003; Tsionas et al., 2003); consolidation and its impact on banks' efficiency (Cuesta and Orea, 2002; Vivas et al. 2011); and comparison of different frontier techniques (Delis et al., 2009)[3]. However, to our knowledge, there is insufficient number of studies that formally consider the relationship between banks' regulated capital and productivity (Fethi et al., 2012).

[2] This shadow return is calculated from the negative of the elasticity of a bank's cost function with respect to the level of equity capital, as shown later in the paper.
[3] Berger and Humphrey (1997) and Fethi and Pasiouras (2010) present detailed reviews of the literature on banking efficiency.

Our paper provides a new direction in the efficiency and productivity literature by exploiting a theoretical feature long overlooked empirically in this strand. The novelty of this paper is owed to constructing a modelling framework that accounts for the impact of the balance sheet constraint (regulatory constraint) on banking production costs. We relax the underlying assumption of the long run cost function by exploiting the envelopment theory and introducing a proxy of capital that is subject to short-run adjustment into the cost function. Our proposed approach is then utilized to obtain the efficiency and productivity decomposition in the banking systems of emerging economies. We further extend the analysis by reflecting our proposed model on the specification of composed error stochastic frontier analysis to derive a productivity decomposition for a panel data set of emerging economy banking systems, where the decomposition includes the impact of the capital constraint.

The paper is organized as follows: Section 2 discusses the theoretical background and the proposed model whilst Section 3 introduces the model specification and data. Section 4 provides analytical discussions on the empirical findings; and the final section concludes.

2. Alternative modelling for the technology and relative efficiency

In this section, we develop a model of banking system activity that takes account of the equity capital requirements. In particular, we look at how the increased capital requirements (compulsory by regulators) may impose additional costs on the efficient allocation of resources. We begin with the parametric frontier dual-cost function, which is based on K variable inputs: $\mathbf{x} = (x_1, \ldots, x_K)$ with input prices: $\mathbf{w} = (w_1, \ldots, w_K)$ and R outputs: $\mathbf{y} = (y_1, \ldots, y_R)$, and an additional input. This input may be either fixed in the short run, or required in a fixed ratio to output, but is variable in the long run. To further clarify, we symbolize this particular input as z_0, with input price w_0. We argue that the interpretation of this fixed input will be critical in a banking industry paradigm, hence it captures the importance of the equity capital level.

We formalize our hypothesis based on the arguments introduced by Braeutigam and Daughety (1983) and Hughes et al. (2001), and we write the long run cost function, with all inputs including z_0 treated as variable, in the form:

$$c(\mathbf{y}, \mathbf{w}, w_0, t) = \min_{\mathbf{x}, z_0} \{\mathbf{w}'\mathbf{x} + w_0 z_0 : F(\mathbf{x}, z_0, \mathbf{y}, t) = 0\} \qquad [1]$$

The efficient boundary of the technology set is represented by a transformation function: $F(\mathbf{x}, z_0, \mathbf{y}, t) = 0$. Assuming weak disposability of the technology implies that the first derivatives, $F_k \equiv \partial F / \partial x_k$, $F_r \equiv \partial F / \partial y_r$, are not restricted in sign. This will allow the model to accommodate both positive and negative shadow prices in the dual-cost function. In that vein, for a banking industry the regulated short run cost function can be modelled in two ways: either by *specifying a fixed level of the critical input equity capital* z_0 is fixed; or alternatively, by *specifying a fixed ratio of the critical input equity capital to a single dimension of output measured as total assets,* $r_0 = z_0 / \mathbf{i}'\mathbf{y} = z_0 / y$. Most of the literature tends to perceive this feature of the envelope theorem application to banking costs through the short run cost function with a fixed equity level. However, we opt to show the relationship between the long run total cost and the short run total cost expressed in regulated equity-asset ratio form. In this case, where the equity capital input z_0 must be held in a regulated or target ratio with output measured as total assets, r_0, the short run cost function is:

$$c(\mathbf{y}, \mathbf{w}, r_0, t) + w_0 z_0 = \min_{\mathbf{x}} \{\mathbf{w}'\mathbf{x} + w_0 z_0 : F(\mathbf{x}, z_0, \mathbf{y}, t) = 0; z_0 = r_0 \mathbf{i}'\mathbf{y} = r_0 y\} \qquad [2]$$

The envelope theorem confirms that long run total cost defines the envelope of short run total cost:

$$c(\mathbf{y}, \mathbf{w}, w_0, t) = \min_{z_0}\{c(\mathbf{y}, \mathbf{w}, r_0, t) + w_0 z_0\} \quad [3]$$

Consequently, the following derivative result holds in the neighbourhood of the optimal ratio of the fixed input, $z_0 = r_0 y$:

$$\partial c(\mathbf{y}, \mathbf{w}, w_0, t)/\partial r_0 = 0 = [\partial c(\mathbf{y}, \mathbf{w}, r_0 *, t)/\partial r_0] + w_0 y \quad [4]$$

Rearranging this last result and expressing it in elasticity form gives the critical interpretation of the shadow price of the target equity capital ratio:

$$-[\partial \ln c(\mathbf{y}, \mathbf{w}, r_0 *, t)/\partial \ln r_0] = (w_0 y)(r_0/C) = (w_0 z_0/C) \quad [5]$$

In other words, the negative log derivative of the short run cost function expresses the shadow share of equity costs to total expenses[4].

There are two particularly important implications in the analysis of banking systems, and these concern the measurement of the shadow price away from equilibrium and the measurement of returns to scale. These implications depend on the nature or choice of the fixed input, either the level of equity capital or equity capital expressed as a ratio to total assets (*equity-asset*). In the *first* case when equity level is involved, we interpret that the negative of the derivative of short run total cost with respect to the equity *level* is the shadow cost of equity. The *second* case is when the model involves the *equity-asset* ratio. In that case we interpret the negative of the derivative of short run total cost with respect to the *equity-asset* ratio as the shadow ratio of equity expenses to total expenses.

In our case, the inclusion of the equity-asset or capital ratio as an explanatory variable in the cost function enables us to examine three possible outcomes that will consequently affect the cost in our model. First, "Over Leverage" banks that are over-leveraged or reliant on debt and under-use equity capital can be expected to show a relatively *low* ratio of equity expenses

[4] In the case where a fixed level of input is the constraint, the corresponding result is that the negative of the derivative of the variable cost function with respect to this fixed input is the input's shadow price.

to total expenses (but with a negative sign on the measured elasticity in the cost function – see equation [5] above). These banks might be engaged in capitalizing themselves, however, with insignificant proportions or relatively very low rates either due to high competition, lucrative opportunities in the loans market, or simply weak accessibility to equity capital. Second, "Active-Capitalizer" banks that are engaged in active recapitalization will show a relatively high ratio of equity expenses to total expenses, but still with a negative sign in [5]. (These types of banks tend to constantly adjust their equity levels to meet the regulatory requirements.) Third, "Excessive-Capitalizer" banks that are far from long run cost minimizing equilibrium, for example because they are undergoing major recapitalization with current equity capital levels well above the long run equilibrium and may be expected to show a significant rise in the ratio of equity expenses to total expenses compared with the long run average when the fitted cost function includes the equity-asset ratio. In the third case, for instance where the fitted cost function is conditioned on the level of equity capital instead of the equity-asset ratio, we will observe a very low possibly severely negative shadow return on equity in the recovery phase from financial crisis. Negative values of the shadow input price or return on the fixed input equity level (corresponding to above average ratio of equity to total expenses) would arise if, for example, the firm was operating in the uneconomic region of the production function.

The sign and magnitude of the shadow return for the equity-asset ratio indeed have an implication on the measurement of returns to scale. Panzar and Willig (1977) derive the following result concerning the inverse of the elasticity of cost with respect to output:

$$E_{\mathbf{cy}}^{-1} = c \Big/ \sum_{r=1}^{r=R} \left(y_r \left(\partial c / \partial y_r \right) \right) = 1 \Big/ \sum_{r=1}^{r=R} \left(\partial \ln c / \partial \ln y_r \right) \qquad [6]$$

Then $E_{\mathbf{cy}}^{-1} < 1$ implies diseconomies of scale (decreasing returns), $E_{\mathbf{cy}}^{-1} = 1$ implies constant returns to scale and $E_{cy}^{-1} > 1$ implies economies of scale (increasing returns). The definition of cost used here, however, is the long run total cost: $c(\mathbf{y}, \mathbf{w}, w_0, t)$, but as Braeutigam and Daughety (1983) demonstrate, close to the optimum level of the fixed input, the short run total cost can be used instead. *The elasticity of scale is measured by adjusting the long run Panzar-Willig estimate by the shadow ratio of equity expenses to total expenses*:

$$E_{c\mathbf{y}}^{-1} \approx \left(1 - \partial \ln C / \partial \ln r_0\right) \Big/ \sum_{r=1}^{r=R} \left(\partial \ln C / \partial \ln y_r\right) \qquad [7]$$

This measures returns to scale at the observed suboptimal level of the fixed input, which may be more appropriate if the industry is expected to remain at a suboptimal allocation of inputs.

We therefore have two possible specifications of the short run total cost function, one using the equity-asset ratio and one using the equity level. We proceed at this point using the equity-asset ratio, but both forms are fitted in the estimation results. The actual cost experienced by the firm is by definition: $C_t \equiv \mathbf{w}'\mathbf{x} + \alpha_0$ where α_0 is expenditure on the fixed input. Consequently, cost efficiency at time *t* is:

$$CE_t = \{c(\mathbf{y}, \mathbf{w}, r_0, t)/C_t\} \in (0,1] \qquad [8]$$

Using $\exp(-u), u \geq 0$ to transform the measure of cost efficiency from the interval $(0,1]$ into a non-negative random variable with support on the non-negative real line $[0,+\infty)$ yields:

$$\ln C_t = \ln c(\mathbf{y}, \mathbf{w}, r_0, t) + u \qquad [9]$$

This function should be homogeneous of degree +1 and concave in input prices. An econometric approach may be adopted by replacing the deterministic kernel of [13] with a fully flexible functional form such as the translog function with an additive idiosyncratic error term v to capture sampling, measurement and specification error. We impose homogeneity by dividing through by one of the input prices, for example w_K, expressing the variables in vector form as:

$$\mathbf{l\tilde{w}} = \left(ln(w_1/w_K) \quad \ldots \quad ln(w_{K-1}/w_K)\right)$$
$$\mathbf{ly} = \left(ln\, y_1 \quad \ldots \quad ln\, y_R\right)$$

and writing the translog approximation with additive error term as $TL(\mathbf{y}, \tilde{\mathbf{w}}, r_0, t) + v$. In the equity-asset ratio specification, these steps give us the following result:

$$\begin{aligned}\ln(C/w_K) = \alpha_0 &+ \boldsymbol{\alpha}'\,\mathbf{ly} + \boldsymbol{\beta}'\,\mathbf{l}\tilde{\mathbf{w}} + \tfrac{1}{2}\mathbf{ly}'\,\mathbf{A}\,\mathbf{ly} + \tfrac{1}{2}\mathbf{l}\tilde{\mathbf{w}}'\,\mathbf{B}\,\mathbf{l}\tilde{\mathbf{w}} + \mathbf{ly}'\,\boldsymbol{\Gamma}\,\mathbf{l}\tilde{\mathbf{w}} + \delta_1 t + \tfrac{1}{2}\delta_2 t^2 + \boldsymbol{\mu}'\,\mathbf{ly}\,t + \boldsymbol{\eta}'\,\mathbf{l}\tilde{\mathbf{w}}\,t \\ &+ \rho_1 \ln r_0 + \tfrac{1}{2}\rho_2 (\ln r_0)^2 + \boldsymbol{\theta}'\mathbf{ly}\ln r_0 + \boldsymbol{\xi}'\mathbf{l}\tilde{\mathbf{w}}\ln r_0 + \omega \ln r_0 t + v + u\end{aligned} \qquad [10]$$

The vectors of elasticity functions (equivalent in the case of the input prices to the share equations by Shephard’s lemma) are derived by differentiating the translog quadratic form:

$$\begin{bmatrix}\boldsymbol{\varepsilon}_y \\ \boldsymbol{\varepsilon}_{\tilde{w}} \\ \varepsilon_t \\ \varepsilon_{r0}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\alpha} & \mathbf{A} & \boldsymbol{\Gamma} & \boldsymbol{\mu} & \boldsymbol{\theta} \\ \boldsymbol{\beta} & \boldsymbol{\Gamma}' & \mathbf{B} & \boldsymbol{\eta} & \boldsymbol{\xi} \\ \delta_1 & \boldsymbol{\mu}' & \boldsymbol{\eta}' & \delta_2 & \omega \\ \rho_1 & \boldsymbol{\theta}' & \boldsymbol{\xi}' & \omega & \rho_2\end{bmatrix} \begin{bmatrix}1 \\ \mathbf{ly} \\ \mathbf{l}\tilde{w} \\ t \\ \ln r_0\end{bmatrix} \qquad [11]$$

This matrix derivative of the translog short run cost function can be used to generate a total factor productivity decomposition.

2.1. Productivity growth decomposition

We derive the total factor productivity index and its decomposition as follows (see Bauer, 1990; Orea, 2002). Differentiating both sides of the cost equation [10] with respect to *t* and rearranging the result, we obtain:

$$E^{-1}\boldsymbol{\varepsilon}_y'\dot{\mathbf{y}} - \mathbf{s}'\dot{\mathbf{x}} = (1 - E/E)\boldsymbol{\varepsilon}_y'\dot{\mathbf{y}} + (\mathbf{s} - \boldsymbol{\varepsilon}_w)'\,\dot{\mathbf{w}} - \varepsilon_t - (du/dt) - \varepsilon_{r0}\dot{r}_0 \qquad [12]$$

In this expression, E^{-1} is the elasticity of scale; $\boldsymbol{\varepsilon}_y$ is the vector of cost elasticity functions with respect to the outputs, with typical element: $\varepsilon_{yr} = \partial \ln c(\mathbf{y}, \mathbf{w}, r_0, t)/\partial \ln y_r$; $\boldsymbol{\varepsilon}_w$ is the vector of cost elasticity functions with respect to the input prices, with typical element: $\varepsilon_{wk} = \partial \ln c(\mathbf{y}, \mathbf{w}, r_0, t)/\partial \ln w_k$; ε_t is the cost elasticity function with respect to the time-based index of technological progress: $\varepsilon_t = \partial \ln c(\mathbf{y}, \mathbf{w}, r_0, t)/\partial t$; (du/dt) is the rate of change of inefficiency; and finally, ε_{r0} is the cost elasticity with respect to the target equity-asset ratio constraint. The left-hand side of this expression is by definition a measure of total factor productivity change with weights that sum to unity, that is, by construction in the case of outputs and by linear homogeneity in the case of inputs. Hence, the right-hand side is a complete decomposition of the total factor productivity index.

The five components of the total factor productivity change on the right-hand side of the equation can therefore be interpreted as follows:

(a) $(1 - E/E)\boldsymbol{\varepsilon}'_y \dot{\mathbf{y}}$: *scale efficiency change*; if $E = 1$ i.e. CRS, there is zero scale efficiency change in the total factor productivity change, *TFPC*, decomposition.

(b) $(\mathbf{s} - \boldsymbol{\varepsilon}_w)' \dot{\mathbf{w}}$: *allocative efficiency change*; if actual input cost shares and optimal input cost shares are equal, there is no potential for allocative efficiency change $\mathbf{s} - \boldsymbol{\varepsilon}_w = \mathbf{0}$.

(c) $-\varepsilon_t$: *technological change*; if the elasticity of cost with respect to time as a proxy for the technological change is negative, $\varepsilon_t < 0$, then this term will raise productivity.

(d) $-(du/dt)$: *cost efficiency change*; if this term, including the sign, is positive then productivity is enhanced by improvements in the technology.

(e) $\varepsilon_{r0}\dot{r}_0$: *regulated equity-asset ratio productivity change*; if this term, including the sign, is positive then productivity is enhanced by relaxation of the equity-asset ratio constraint, and conversely productivity is reduced when the constraint becomes more strongly binding, for example in a recapitalization phase.

It is the last component that allows us to compute the cost of recapitalizing the banking system. If the shadow price or rate of return on equity is positive, then holding higher levels of equity capital or a higher target equity-asset ratio will move the banking system towards a long run equilibrium and will generate a positive impact on productivity growth.

However, if the shadow price or rate of return on equity is negative (i.e. the equity level has a positive coefficient in the fitted cost function), or there is a requirement to hold higher than equilibrium levels of equity capital relative to assets, then this will impose a negative component on productivity growth. This allows us to measure the cost impact of recapitalization by the contribution (negative or positive) of the changes in the equity level or the equity-asset ratio to the measured total factor productivity growth.

The components of total factor productivity change, $T\dot{F}P$, are shown in total differential form; however, we can use them in index number form, as follows:

(a) $\frac{1}{2}\sum_{r}\left[\left(\left(1-E^{t+1}\right)\varepsilon_{yrt+1}/E^{t+1}\right)+\left(\left(1-E^{t}\right)\varepsilon_{yrt}/E^{t}\right)\right]\left(\ln y_{rt+1}-\ln y_{rt}\right)$ is the effect of scale efficiency change.

(b) $\frac{1}{2}\sum_{k}\left[\left(s_{kt+1}-\varepsilon_{wkt+1}\right)+\left(s_{kt}-\varepsilon_{wkt}\right)\right]\left(\ln w_{kt+1}-\ln w_{kt}\right)$ is the effect of the bias in using actual cost share weights instead of optimal cost shares based on shadow prices, i.e. allocative efficiency change.

(c) $-\frac{1}{2}\left[\left(\partial\ln c\left(\mathbf{y},\mathbf{w},z_0,t+1\right)/\partial t\right)+\left(\partial\ln c\left(\mathbf{y},\mathbf{w},z_0 t\right)/\partial t\right)\right]$ is the effect of cost reducing technical progress.

(d) $\left[CE_{t+1}-CE_{t}\right]$ is cost efficiency change.

(e) $-\frac{1}{2}\left[\varepsilon_{r0t+1}+\varepsilon_{r0t}\right]\left(\ln r_{0t+1}-\ln r_{0t}\right)$ is the effect on productivity change of variation in the equity-asset ratio constraint.

3. Methodology and data

The stochastic frontier analysis regression to be estimated, with the error components v representing idiosyncratic error and u representing inefficiency, can be expressed succinctly as follows:

$$\ln\left(C/w_K\right)_{it}=\alpha_0+\mathbf{x}'_{\mathbf{it}}\boldsymbol{\theta}+\varepsilon_{it}\quad;\quad\varepsilon_{it}=v_{it}+u_{it}\quad i=1\ldots N,t=1\ldots T \qquad [13]$$

Here $\mathbf{x}'_{it}$ is a $(K + R + 2)$ vector of explanatory variables representing the input prices, outputs, time and the level of the fixed input equity capital including second order direct and cross product translog expressions. The range of panel data stochastic frontier analysis models reflects different assumptions about the nature of the composed error terms. Because experience suggests that parameter values can be sensitive to the form of the stochastic frontier analysis model that is fitted, we shall use a number of different types of these models. The literature here is immense but we can summarize it briefly as follows.

Within the strict panel data structure, many researchers have followed Schmidt and Sickles (1984) and Pitt and Lee (1981) in adopting a time-invariant model of inefficiency with a short panel; therefore the composed error term is written as: $\varepsilon_{it} = v_{it} + u_i$. The model can be estimated by standard fixed effects using dummy variables (FE-LSDV), standard random effects with generalized least squares (RE-GLS), or by random effects maximum likelihood estimation (RE-MLE), as suggested by Pitt and Lee, if specific distributional assumptions are made, for example the truncated-normal distribution for the inefficiency term.

The RE-GLS and RE-MLE models usually provide very similar results. To incorporate the more general assumption of time-varying inefficiency, two broad approaches are possible. The inefficiency component can be made an explicit function of time: $u_{it} = u_i h(t)$. Battese and Coelli (1992) use an exponential function which is the same across all producers and can be estimated by maximum likelihood with the appropriate distributional assumptions. These methods retain an explicit panel structure.

Firm specific heterogeneity may be incorporated through additional conditioning variables, and a pooled estimation technique based on some form of modified least squares could also be adopted. For example, by making use of the seemingly unrelated regression estimator based on generalized least squares SURE-GLS, we can obtain estimators which are relatively efficient and permit the error terms in the cost share equations to be related to the overall cost equation. This is a generalization which standard stochastic frontier analysis estimators are unable to provide (see Kumbhakar and Lovell, 2000: 156-8).

Finally, Reifschneider and Stevenson (1991), amongst others, suggested the strategy of making specific parameters of the inefficiency density function for u_{it} conditional on time-varying exogenous variables (i.e. conditional mean or conditional heteroscedasticity). Numerous other models in the literature develop variants of these general procedures; for example, the “thick frontier” approach of Berger and Humphrey (1991) splits the sample into quantiles of the dependent variable and estimates average regressions for each quantile; the distribution-free approach of Berger (1993), which is similar in concept to RE-GLS, uses seemingly unrelated regression with generalized least squares (SURE-GLS) applied to each time period separately. Reflecting this discussion, the empirical results in this paper are derived from five broad categories of model. These are summarized in the Appendix table of composed error specifications.

3.1. Data

The data are gathered from several major sources: Bankscope by Bureau van Dijk (2010), the Organisation for Economic Co-operation and Development (OECD) and World Bank databases. The bank data have been reported in $US millions at current prices and market exchange rates. We convert to constant price (year 2000) values by deflating the $US denominated data converted at market exchange rates by the US GDP deflator. Table 1 reports the range of countries and regions used in the sample, while summary statistics for our sample of 485 banks over the period 2005–2008 are reported in Table 2; these indicate the within sample variability of the pre-filtered raw data.

<TABLE 1 HERE>

<TABLE 2 HERE>

The definitions of the key variables in the cost function are standard in the current literature on bank performance (e.g., Bikker and Bos, 2008). They are calculated from the constant price data as follows. Cost, C, is the total operating cost, that is, the sum of interest expenses, salaries and employee benefits and other operating costs. Outputs are: loans, y_1, securities investments, y_2, and off balance sheet total business volume, y_3. The loans variable used is *net loans* after allocating reserves for non-performing loans (NPLs). Equity capital (z_0) is reported separately and the first two outputs, loans, y_1, securities investments, y_2, together account for total assets, (z_1). Input price indices are: the price of labour, w_1, computed as salaries and employee benefits relative to total assets, the price of physical capital, computed as other operating expenses divided by fixed assets, w_2, and the price of funds, computed as interest expenses relative to total assets, w_3. All of these industry variables are sourced from Bureau van Dijk (2010) for each bank and period in the sample, and all have been deflated as above. In addition to these key variables, banking system variables are used along with macroeconomic variables to condition the individual bank cost functions.

Macroeconomic variables are collected from the OECD and World Bank databases and vary through time but are constant across banks. They are measured as percentage rates of change. In this way the banking market is conditioned at the level of the macroeconomy before the beginning of the sample period; then the relative changes in the macroeconomic environment are treated as exogenous shocks. They are measured in differenced form to avoid the spurious correlation problem of entering macroeconomic trending variables in the cost regression. The macroeconomic environmental shocks used in the analysis are as follows:

(a) change in gross domestic product (GDP) at 2000 market prices;

(b) change in GDP at 2000 market prices per head of population.

These reflect the cyclical response to government macroeconomic policy as well as the impact of exogenous shocks from the external economy.

The banking system variables are: Loan loss reserve/Gross loans, Net interest margin, Return on assets, Return on equity, Cost to income ratio, Net loans/Total assets, Net loans/Customer and short term, Funding reserves for impaired loans/NPLs, Non-interest income/Gross revenues, Non-interest expense/Gross expenses, NPL/Gross loans, Reserves for NPL/Gross loans, Reserves for NPL/NPL, and Interbank assets/Interbank liabilities.

All of the data in the fitted regressions are log-mean-corrected, that is, expressed as deviations from the sample means after having been transformed to natural logarithms. This has three advantages: it ensures that the translog function which is an approximation to an arbitrary second order function has the point of approximation at the sample mean; it allows us to check the properties of the fitted translog function at the sample mean by examining the first order estimated coefficients; and it enables computation of the variance of linear functions of the estimated coefficients around the sample mean from the variance-covariance matrix of the regression coefficients.

4. Empirical results: parameter estimates and the shadow price of the equity-asset ratio

Prior to estimation of the models, the data were filtered using the financial ratio rules suggested by Bikker and Bos (2008) together with the addition of a statistical criterion in which we estimated a simple pooled ordinary least squares (OLS) model for the whole sample and dropped observations with a standardized OLS residual exceeding 2 in absolute value. This rule of thumb is approximately equivalent to capturing outliers in the data by an instrumental dummy variable at the 5 percent level of significance. These filters resulted in reducing the sample from 1940 observations to 1869 observations.

Regression results for the first order coefficients in the cost function fitted under different models are shown in Table 3, which presents: (i) the monotonicity effects, that is, elasticity function estimates at the sample mean, and (ii) measures of the presence of inefficiency as a component of the error term and whether the inefficiency is time varying.

<TABLE 3 HERE>

The regression coefficients on the first order terms[5], that is, the cost function elasticities at the sample mean, are relatively consistent across the different econometric specifications. The models all fit well and there are no strong reasons to favour one over another. However, the SURE-GLS model which pools the data without a panel structure finds a negative effect from securities investment while at the same time suggesting that the shadow price of the equity-asset ratio constraint is higher than for other models. The four remaining stochastic frontier analysis models all find a very consistent negative shadow price of about -4.5 to -5 percent on the capital constraint. Interestingly when the equivalent models are estimated with the level of equity capital as the constraint, the shadow return on capital is consistently negative, confirming strongly that these emerging country banks experienced stringent recapitalization during this period. Amongst the four stochastic frontier analysis models the Reifschneider-Stevenson (1991) UHET results indicate the significance of all of the output

[5] There are multiple second order and interaction coefficients too numerous to report here.

variables and have significant and theoretically correct first order elasticity estimates at the sample mean.

The Panzar-Willig estimate of the elasticity of scale at the sample mean and the scale elasticity evaluated out of equilibrium, after adjusting for the regulated equity-assets ratio, are shown in Table 4; they indicate a small degree of increasing returns suggesting scope for some consolidation amongst the banking systems in emerging economies.

<TABLE 4 HERE>

4.1. Empirical results: productivity measurement

In this section of the paper, we use the discrete index number calculation to decompose productivity change during the period encompassing the financial crisis. We could illustrate the impact by using any of the four composed error stochastic frontier analysis models since their coefficients are relatively stable across different approaches. For a number of reasons described above, the Reifschneider-Stevenson model seems to generate the most sensible results and we focus on that model to calculate the productivity decomposition. Table 5 reports the productivity estimates and the component factors for this model; the decomposition covers scale efficiency change, technical change, efficiency change, allocative efficiency change, and constraint relaxation change. The last component illustrates how the requirement to build up a stronger equity-asset ratio during recapitalization may enhance or offset total factor productivity change over the period.

<TABLE 5 HERE>

In Table 5 we see that total factor productivity change in emerging economy banking systems averaged over the sample period has been very slightly negative. The forces driving total factor productivity up have originated in scale efficiency change and allocative efficiency change. Regressive factors have been an apparent loss of technological progress and the impact of the equity-asset constraint. In other words, the need to maintain a certain level of capital has offset the positive forces on total factor productivity change during this critical period. Consistently over the period allocative and scale efficiency change have contributed positively to the performance of banking systems in transition economies. Efficiency change has improved after an initial negative start. Consequently the emerging economies' banking systems have shown signs of resilience while the international financial system has been coping with its recent problems. However, cost performance has been weakened by a failure to take advantage of technological progress and by the need to maintain acceptable equity capital ratios. The capital adequacy constraint has contributed to the weak overall productivity performance.

These results can be seen in more detail when we disaggregate by country groupings in Table 6 to obtain the average productivity change components over time.

<TABLE 6 HERE>

The impact of the capital constraint has been particularly strong for the Middle East (ME) and South East Asia (SEA). The deleveraging implied by a more strongly binding capital constraint and the consequent fact that the shadow return on equity turned negative for these regions has meant that the impact on total factor productivity has been regressive (the growth factor is below one). This allows us to say that this modelling approach provides a direct estimate of the productivity cost of constrained deleveraging activity arising from policy decisions.

5. Conclusions and policy lessons

We have carried out an empirical analysis of the banking systems of a large number of emerging economies during a critical period for the international financial system. In doing this we focused on three aspects of the modelling problem. First, we chose to construct short run balance-sheet-constrained total cost functions for the emerging economy banks. Second, we applied stochastic frontier analysis to these in order to identify sources of variability in economic performance. Third, we were able to derive from the estimated cost functions a decomposition of total factor productivity into: scale efficiency change, allocative efficiency change, technical change, efficiency change and the impact of the equity capital constraint.

We discovered that a variety of time-invariant and time-varying stochastic frontier analysis models produced consistent results for this sample period, but we were able to show that a time varying conditional heteroscedasticity model fitted the data particularly well. Amongst the empirical results that we were able to uncover, we confirm the importance of the equity capital ratio as a constraint on cost minimizing behaviour. This has important policy implications. In the current state of worldwide recovery from the financial crisis, the issue of the recapitalization of the banking system is dominating the policy debate. This has a long run dimension, which is expressed in the question of whether greater reliance on equity capital will raise the long run funding costs of the banks. Policymakers seem relatively optimistic on this issue. However, the equity capital ratio also has a short run dimension: what are the adjustment costs that arise when a banking system recapitalizes? As we indicated at the beginning of the paper, this is an important and unresolved policy problem. This paper has suggested a way of measuring these adjustment costs by examining the role of the equity capital constraint in the determination of total factor productivity of the banking system. Our results suggest that there is a positive adjustment cost. However, it may be relatively small enough not to offset the recognized benefits of moving to a more securely based banking system that uses higher levels of equity capital.

Acknowledgements

We would like to thank the participants of the VI Seminar on Risk, Financial Stability and Banking of the Banco Central do Brasil in São Paulo and our paper's discussant, Emanuel Kohlscheen, for valuable comments and suggestions.

References

Avkiran, N.K., 1999. The evidence on efficiency gains: The role of mergers and the benefits to the public. Journal of Banking and Finance 23, 991–1013.

Battese, G.E., Coelli, T.J., 1992. Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. Journal of Productivity Analysis 3, 153–169.

Battese, G.E., Coelli, T.J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20, 325–332.

Bauer, P.W., 1990. Decomposing TFP growth in the presence of cost inefficiency, nonconstant returns to scale, and technological progress. Journal of Productivity Analysis 1, 287–299.

Beccalli, E., Casu, B., Girardone, C., 2006. Efficiency and stock performance in European banking. Journal of Business Finance and Accounting 33, 245–262.

Berger, A.N., 1993. Distribution-free estimates of efficiency in the US banking industry and tests of standard distributional assumptions. Journal of Productivity Analysis 4 (3) (September), 261–292.

Berger, A.N., Humphrey, D.B., 1991. The dominance of inefficiencies over scale and product mix economies in banking. Journal of Monetary Economics 28, 117–148.

Berger, A.N., Humphrey, D.B., 1997. Efficiency of financial institutions: International survey and directions for future research. European Journal of Operational Research 98, 175–212.

Bikker, J.A., Bos, J.W.B., 2008. Bank Performance: A Theoretical and Empirical Framework for the Analysis of Profitability, Competition and Efficiency. Routledge, London.

Braeutigam, R.R., Daughety, A.F., 1983. On the estimation of returns to scale using variable cost functions. Economics Letters 11, 25–31.

Brissimis, S.N., Delis, M.D., Papanikolaou, N.I., 2008. Exploring the nexus between banking sector reform and performance: Evidence from newly acceded EU countries. Journal of Banking and Finance 32, 2674–2683.

Bureau van Dijk, 2010. Bankscope database. http://www.bvdinfo.com/Home.aspx (accessed November 3, 2012).

Canhoto, A., Dermine, J., 2003. A note on banking efficiency in Portugal: New vs old banks. Journal of Banking and Finance 27, 2087–2098.

Casu, B., Molyneux, P., 2003. A comparative study of efficiency in European banking. Applied Economics 35, 1865–1876.

Cuesta R.A., Orea, L., 2002. Merger and technical efficiency in Spanish savings banks: A stochastic distance function approach. Journal of Banking and Finance 26, 2231–2247.

Daniel, B.C., Jones, J.B., 2007. Financial liberalization and banking crises in emerging economies. Journal of International Economics 72, 202–221.

Delis, M.D., Koutsomanoli-Fillipaki, A., Staikouras, C.K., Gerogiannaki, K., 2009. Evaluating cost and profit efficiency: A comparison of parametric and nonparametric methodologies. Applied Financial Economics 19, 191–202.

Erdem, C., Erdem, M.S., 2008. Turkish banking efficiency and its relation to stock performance. Applied Economics Letters 15, 207–211.

Fethi, M. Duygun, Pasiouras, F., 2010. Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. European Journal of Operational Research 204 (2), 189–198.

Fethi, M. Duygun, Shaban, M., Weyman-Jones, T., 2011. Liberalization, privatization and the productivity of Egyptian banks: A non-parametric approach. The Service Industries Journal 31, 1143–1163.

Fethi, M. Duygun, Shaban, M., Weyman-Jones, T., 2012. Turkish banking re-capitalization and the financial crisis: An efficiency and productivity analysis. Emerging Markets, Finance and Trade, Nov 2012 Supplement 5, Vol. 48, p76-90.

Havrylchyk, O., 2006. Efficiency of the Polish banking industry: Foreign versus domestic banks. Journal of Banking and Finance 30, 1975–1996.

Hughes, J.P., Mester, L.J., Moon, C.-G., 2001. Are scale economies in banking elusive or illusive? Evidence obtained by incorporating capital structure and risk taking into models of bank production. Journal of Banking and Finance 25, 2169–2208.

Isik, I., Hassan, M.K., 2003. Financial deregulation and total factor productivity change: An empirical study of Turkish commercial banks. Journal of Banking and Finance 27, 1455–1485.

Koutsomanoli-Filippaki, A., Margaritis, D., Staikouras, C., 2009. Efficiency and productivity growth in the banking industry of central and Eastern Europe. Journal of Banking and Finance 33, 557–567.

Kumbhakar, S., Lovell, C.A.K., 2000. Stochastic Frontier Analysis. Cambridge University Press, Cambridge.

Lozano-Vivas, A., Kumbhakar, S.C., Fethi, M.D., Shaban, M., 2011. Consolidation in the European banking industry: How effective is it? Journal of Productivity Analysis 36, 247–261.

Orea, L., 2002. A generalised parametric Malmquist productivity index. Journal of Productivity Analysis 18, 5–22.

Panzar, J., Willig, R., 1977. Economies of scale in multi-output production. Quarterly Journal of Economics 8 (August), 48l–493.

Pitt, M.M., Lee, L.-F., 1981. Measurement and sources of technical inefficiency in the Indonesian weaving industry. Journal of Development Economics 9, 43–64.

Reifschneider, D., Stevenson, R., 1991. Systematic departures from the frontier: A framework for the analysis of firm efficiency. International Economic Review 32 (3) (August), 715–723.

Schmidt, P., Sickles, R.C., 1984. Production frontiers and panel data. Journal of Business and Economic Statistics 2, 367–374.

Sherman, H.D., Rupert, T.J., 2006. Do bank mergers have hidden or foregone value? Realized and unrealized operating synergies in one bank merger. European Journal of Operational Research 168, 253–268.

Sturm, J.-E., Williams, B., 2004. Foreign bank entry, deregulation and bank efficiency: Lessons from the Australian experience. Journal of Banking and Finance 28, 1775–1799.

Tsionas, E.G., Lolos, S.E.G., Christopoulos, D.K., 2003. The performance of the Greek banking system in view of the EMU: Results from a non-parametric approach. Economic Modelling 20, 571–592.

APPENDIX

Table of composed error specifications

Schmidt-Sickles (1984) fixed effects, SSFE	Panel, time invariant LSDV fixed effects to measure inefficiency; can permit inefficiency and RHS variables to be correlated	Shift frontier by maximal effect to measure efficiency $CE_i = \exp\left[-\left(\hat{\alpha}_i - \min\left(\hat{\alpha}_j\right)\right)\right]$
Pitt-Lee (1981) time invariant, PL	Panel, time invariant MLE, normal and half normal errors	Measure efficiency $CE_i = E\left(\exp\left[-u_i \middle\| \tilde{e}_{it}\right]\right)$
Battese-Coelli (1992) time varying BC92	Panel, time varying inefficiency same across all firms, MLE, normal and truncated normal errors	Measure efficiency $CE_{it} = E\left(\exp\left[-u_i \exp\left(-\eta\left(t-T\right)\right) \middle\| \tilde{e}_{it}\right]\right)$
Reifschneider-Stevenson (1991), UHET	Pooled, time varying inefficiency differs across all firms, MLE, normal and truncated normal errors with conditional heteroscedasticity	Measure efficiency $CE_{it} = E\left(\exp\left[-u_{it} \middle\| \tilde{e}_{it}\right]\right)$
Seemingly unrelated system, SURE-CGLS	Pooled, constrained system, time varying inefficiency, no distributional assumptions; estimate cost function and share equations together	Shift frontier by minimum residual to measure efficiency $CE_{it} = \exp\left[-\left(e_{it} - \min\left(e_{jt}\right)\right)\right]$

Table 1

Countries where the sampled banking firms are located

No.	Country name	2005	2006	2007	2008
1	ARGENTINA	14	14	14	14
2	BAHRAIN	6	6	6	6
3	BELARUS	10	10	10	10
4	BOLIVIA	8	8	8	8
5	BRAZIL	47	47	47	47
6	BULGARIA	12	12	12	12
7	CHINA, PEOPLE'S REP.	14	14	14	14
8	COSTA RICA	15	15	15	15
9	CROATIA	17	17	17	17
10	CZECH REPUBLIC	9	9	9	9
11	GEORGIA, REP. OF	8	8	8	8
12	GREECE	13	13	13	13
13	HONG KONG	11	11	11	11
14	HUNGARY	7	7	7	7
15	INDIA	43	43	43	43
16	INDONESIA	6	6	6	6
17	ISRAEL	10	10	10	10
18	JORDAN	10	10	10	10
19	KOREA, REP. OF	15	15	15	15
20	LATVIA	17	17	17	17
21	LITHUANIA	6	6	6	6
22	PERU	9	9	9	9
23	PHILIPPINES	20	20	20	20
24	POLAND	17	17	17	17
25	ROMANIA	17	17	17	17
26	SLOVAKIA	10	10	10	10
27	SLOVENIA	12	12	12	12
28	SOUTH AFRICA	8	8	8	8
29	TAIWAN	13	13	13	13
30	THAILAND	16	16	16	16
31	TURKEY	12	12	12	12
32	UKRAINE	26	26	26	26
33	UNITED ARAB EMIRATES	11	11	11	11
34	VENEZUELA	16	16	16	16
	Total	485	485	485	485

Table 2

Summary data on core variables prior to sample filtering

Variable	Number in unfiltered sample	Mean	Std. Dev.	Min	Max
Loans	1940	89.88604	339.6181	0.0002287	5272.131
Securities and investments	1940	40.29417	210.9409	0.000	3657.231
Off balance sheet income	1940	57.23229	224.4006	0.000	3342.204
Total assets	1940	161.4529	669.0114	0.1085498	11596.22
Deposits and short-term funding	1940	126.5258	576.3351	0.0176429	10547.89
Interest expenses	1940	4.555784	14.12974	0.0003471	211
Personnel expenses	1940	1.439621	4.573288	0.0022202	63.28918
Other operating expenses	1940	1.496821	4.347391	0.0006781	46.88668
Equity-asset ratio (%)	1940	11.72962	8.703521	0.102	86.24

$US million at year 2000 prices except where otherwise stated.

Table 3

First order regression coefficients of cost function variables

Variable	SSFE	PL	BC92	UHET	SURE_GLS
Core outputs, input prices, time and cost function constraint variables					
Loans	0.938***	0.957***	0.957***	0.805***	1.006***
Securities	0.005	0.001	0.003	0.145***	-0.035***
Off balance sheet	0.007**	0.010***	0.008***	0.017***	0.011***
Funding price	0.052***	0.059***	0.059***	0.064***	0.235***
Capital price	0.571***	0.554***	0.555***	0.448***	0.543***
Time	0.025***	0.024***	0.054***	0.020***	0.011
Equity-asset ratio	-0.049***	-0.049***	-0.041***	-0.044***	-0.099***
Z-variables used to condition the cost frontier or the inefficiency estimates					
Net loans/total assets	-0.01422***	-0.01520***	-0.01460***	-0.00704	-0.01768***
Net loans/deposits and short-term funds	-0.00339***	-0.00360***	-0.00380***	-0.07951***	-0.00181***
Liquid assets/deposits and short-term funds	-0.00066***	-0.00084***	-0.00070***	0.01689***	-0.00015
Reserves for impaired loans/ non-performing loans	-0.00001	-0.00002***	-0.00001**	-0.00003	0.00001
Non-interest expenses/gross revenues	0.00028	0.00033*	0.00034*	-0.00114	0.00115***
Non-performing loans/gross loans	0.00066	0.00049	0.00049	-0.00199	0.00150**
Non-performing loans/gross loans relative to the average for the country	0.00196	0.00333***	0.00247*	0.11730***	0.00146
Equity-asset ratio relative to the average for the country	0.00384	0.00562***	0.00335*	0.10639***	0.00760***
Per capita GDP growth rate	-0.00036	0.00004	-0.00068	0.02777	-0.00005
Mu		0.18958***	0.17306***		
Eta			0.14049***		
Time				0.29306**	
Model statistics					
F value	2150.00				
chi-square		272000.00	262000.00	200000.00	
sigma_u	0.15	0.12	0.09	* conditional on z-variables above	no u component
sigma_v	0.06	0.06	0.06	0.15	

* p<0.05
** p<0.01
*** p<0.001
where p = probability-value significance level.

Table 4

Estimated elasticity of scale at the sample mean

Sample mean values	SSFE	PL	BC92	UHET	SURE_GLS
Panzar-Willig elasticity of scale	1.053	1.033	1.033	1.034	1.018
Adjusted elasticity of scale	1.105	1.084	1.075	1.079	1.120

Table 5

Total factor productivity change and its components for the whole sample

Year	Scale	Allocative	Technical	Capital constraint	Efficiency	Total factor productivity
2005	1.000	1.000	1.000	1.000	1.000	1.000
2006	1.003	1.020	0.991	0.987	0.994	0.994
2007	1.003	1.012	0.979	0.999	0.999	0.992
2008	1.001	1.033	0.964	1.004	1.007	1.007
Mean over time	1.002	1.016	0.983	0.997	1.000	0.998

Table 6

Productivity change components by region

TFP component	Africa	CA	CEE	ME	SA	SEA	WE
Scale efficiency change	1.003	1.007	1.011	1.004	1.008	1.003	1.009
Allocative efficiency change	1.042	1.043	1.032	1.004	1.035	1.022	1.042
Technical change	0.972	0.981	0.975	0.983	0.991	0.977	0.970
Constraint efficiency change	1.006	1.000	1.001	0.996	1.000	0.997	1.008
Technical efficiency change	1.003	1.019	1.021	1.016	1.008	1.009	1.002
Total factor productivity change	1.025	1.065	1.039	1.000	1.041	1.005	1.029