2010

Banks' efficiency and productivity analysis using the Hicks-Moorsteen approach: a case study of Iran

Amir Arjomandi
University of Wollongong, amira@uow.edu.au

Abbas Valadkhani
University of Wollongong, abbas@uow.edu.au

University of Wollongong
Economics Working Paper Series 2010

http://www.uow.edu.au/commerce/econ/wpapers.html

Banks' Efficiency and Productivity Analysis Using the Hicks-Moorsteen Approach: A Case Study of Iran

Amir Arjomandi

and

Abbas Valadkhani

School of Economics
University of Wollongong

WP 11-10
October 2010

Banks' Efficiency and Productivity Analysis Using the Hicks-Moorsteen Approach: A Case Study of Iran

Amir Arjomandi and Abbas Valadkhani

Abstract

This study is the first to use the Hicks-Moorsteen TFP index developed by O'Donnell (2008, 2009, 2010c) to analyse efficiency and productivity changes in the banking system. The advantage of this approach over the popular Malmquist productivity index is that it is free from any assumptions concerning firm optimising behaviour, the structure of markets, or returns to scale. The effects of Iranian government regulations launched in 2005 on the Iranian banking industry are investigated through an analysis of performance over the period 2003-2008 assuming variable returns to scale. The results obtained show that although the Iranian banking industry has been inefficient over the entire period of the study, the industry's technical efficiency level - which had improved over the period 2003-2006 - deteriorated considerably after the regulatory changes were introduced. The industry experienced its highest negative efficiency growth in 2006 which was 43% and became more mix inefficient after 2005, with a considerably negative productivity change after 2007. Overall, changes of production possibility set and scale efficiency changes exerted dominant effects on productivity changes.

Keywords Regulation, Productivity, Banking, Data envelopment analysis, Malmquist index, Hicks-Moorsteen index

1. Introduction

Over the last decade the Iranian banking industry has undergone substantial change due to factors such as liberalization, increased government regulation and technological advances, all of which have resulted in an extensive restructuring of the industry. Changes in policy have affected both public banks (which are government-owned banks including commercial and specialised banks) and private banks in Iran. The former have been the most successful in acquiring market share, and it is mainly due to this reason that private banks did not join the market until after 2001. However, it seems that public banks were more noticeably affected by the Iranian government regulatory initiatives launched in 2005, which obliged all banks to considerably reduce deposit and loan interest rates. The government also imposed differing interest rates and conditions on public and private banks, for example, an obligation on public banks to assign higher priority in their lending operations to areas such as advanced technology projects, small and medium enterprises, and housing projects for low income earners. As a result, the level of non-performing loans (NPLs) of public banks increased dramatically after 2006. According to the Central Bank of Iran (CBI 2007), the annual growth rate of public banks' NPLs was less than 30% before 2005 but increased to a staggering 129% in 2006. The highest share of NPLs being attributed to the manufacturing and mining (20.1%) and construction (19.5%) sectors (CBI, 2007). For these reasons, in particular, this study investigates the effect of government policies on the productivity of the Iranian banking industry.

Fethi and Pasiouras (2010), in their comprehensive survey of 196 bank performance studies, reveal that of those studies where estimates of total factor productivity growth are obtained, almost all employ a DEA[1]-type Malmquist index. This finding demonstrates that the Malmquist index has widespread use in examining total factor productivity growth (see also Sturm and Williams, 2004; Coelli and Rao, 2005; Chen and Lin, 2007; Mukherjee et al., 2001; and Sufian, 2006). Initially, Caves et al. (1982) introduced the Malmquist productivity index as a theoretical index. Färe et al. (1992) later merged Farrell's (1957) measurement of efficiency with Caves et al.'s (1982) measurement of productivity to develop a new Malmquist index of productivity change. Färe et al. (1992) subsequently demonstrate that the resulting total factor productivity (TFP) indices could be decomposed into efficiency change and technical change components. Färe et al. (1994) further decompose the efficiency change into pure technical efficiency change and changes in scale efficiency, a development which results in the Malmquist index becoming widely popular as an empirical index of productivity changes.

However, despite extensive literature on the Malmquist index and its evident popularity as a measure of productivity change, the pros and cons of constant returns to scale (CRS) to estimate Malmquist indices have been extensively discussed. Grifell-Tatje and Lovell (1995) demonstrate that with non-constant returns to scale the Malmquist productivity index does not precisely measure productivity change. They suggest that the bias is systematic and relies on the magnitude of scale economies. Coelli and Rao (2005) maintain the importance of imposing CRS upon any technology that is used for the estimation of distance functions for the calculation of a Malmquist TFP index, applicable to both firm-level and aggregate data; without CRS the result may incorrectly measure TFP gains or losses arising from scale economies.

In contrast however, Ray and Desli (1997) and Wheelock and Wilson (1999) argue that the decomposition of the Malmquist index performed by Färe et al. (1994) is not reliable. Wheelock and Wilson (1999) demonstrate that when a firm's location (from one period to another) has not changed, and scale efficiency change is entirely due to a shift in the variable returns to scale (VRS) estimate of technology, there appears no resulting technical change under CRS. They thus conclude that under such circumstances the CRS estimate of technology is statistically inconsistent.

To avoid these problems O'Donnell (2008) proposed a new way to decompose multiplicatively complete TFP indices into a measure of technical change and various measures of efficiency change, without any assumptions concerning firm optimising behaviour, the structure of markets, or returns to scale for a multiple-input multiple-output case. According to O'Donnell (2008), any TFP index that represents the ratio of aggregate output to aggregate input is said to be multiplicatively complete, where completeness is an essential requirement for an economically-meaningful decomposition of the TFP change. He further demonstrates that the group of complete TFP indices includes Fisher, Konus, Törnqvist, and Hicks-Moorsteen indices, but not the popular Malmquist index of Caves, et al. (1982). Apart from special cases such as constant returns to scale, O'Donnell (2008) states that the Malmquist index is a biased measure of TFP change. Consequently, the popular Färe et al. (1994) decomposition of the Malmquist index also generally leads to unreliable estimates of technical change and/or efficiency change.

In this study, therefore, the Hicks-Moorsteen TFP index (O'Donnell, 2008, 2009, 2010c) is employed to analyse productivity changes of Iranian banks. The remainder of this paper is structured as follows: Section 2 provides the basic idea of multiplicatively completeness and presents the Hicks-Moorsteen TFP index. Section 3 shows the method utilised by O'Donnell (2008) to decompose a multiplicatively complete TFP. It presents a simple two dimensional geometric representation of TFP for a multiple-input multiple-output firm. Section 4 describes how measures of technical, scale and mix efficiency can be defined in relation to quantity aggregates and brings all of these concepts together to show that the multiplicatively-complete TFP index is capable of being decomposed into different implicit measures of technical change and technical efficiency change, in addition to measures of mix and scale efficiency change. Section 5 explains the data employed in the paper and Section 6 discusses the results, followed by some concluding remarks in section 7.

2. Hicks-Moorsteen TFP index

In the case of a multiple-input multiple-output firm, O'Donnell (2008) uses the usual definition of total factor productivity following Jorgenson and Grilliches (1967), and Good et al. (1997); $TFP_{nt} = Y_{nt}/X_{nt}$, where TFP_{nt} indicates the TFP of firm *n* in the period *t*, $Y_{nt} \equiv Y(y_{nt})$, and $X_{nt} \equiv X(x_{nt})$ that Y_{nt} and X_{nt} are aggregate output and aggregate input respectively. This definition allows one to define TFP changes as the ratio of an output quantity index to an input quantity index (a ratio of an output growth to an input growth). Index numbers formed in this way are referred to as multiplicatively complete indexes.

Hicks-Moorsteen TFP index is the only multiplicatively-complete index that can be computed without price data, and has not previously been used to analyse a country's banking system. The Hicks-Moorsteen TFP is actually a ratio of Malmquist output and input quantity indexes, so named because Diewert (1992, p. 240) attributes its origins to Hicks (1961) and Moorsteen (1961). Although Caves et al. (1982) advocated the application of Malmquist indexes, they did not apply ratios of these indexes to develop a complete TFP index in the role of an aggregate output to an aggregate input ratio. Their indexes are complete if and only if the technology is of a restrictive form (O'Donnell, 2008, p.10). The Hicks-Moorsteen TFP index operates as follows:

$$TFP_{HM}^{t,t+1} = \left(\frac{D_o^{t+1}(x^{t+1}, y^{t+1}) D_o^t(x^t, y^{t+1})}{D_o^{t+1}(x^{t+1}, y^t) D_o^t(x^t, y^t)} \frac{D_I^{t+1}(x^t, y^{t+1}) D_I^t(x^t, y^t)}{D_I^{t+1}(x^{t+1}, y^{t+1}) D_I^t(x^{t+1}, y^t)} \right)^{1/2} \quad (1)$$

Where $D_o(x, y)$ and $D_I(x, y)$ are output and input distance functions, respectively, defined by Shephard (1953) as:

$$D_O^T(x, y) = \min\{\delta > 0 : (x, y/\delta) \in P\}, \text{ and} \quad (2)$$

$$D_I^T(x, y) = \max\{\rho > 0 : (x/\rho, y) \in P\}, \quad (3)$$

where *P* denotes the period-T production possibilities set. Using DEA, one is able to calculate these distance functions. O'Donnell (2009) develops a DEA methodology for computing and decomposing the Hicks-Moorsteen TFP index (for a complete explanation of the linear programmes see O'Donnell 2009 and 2010c)[2].

3. A simple two dimensional geometric representation of TFP

To demonstrate that every multiplicatively-complete TFP index is able to be decomposed into a measure of technical change and various measures of efficiency change, the TFP of a multiple-input multiple-output firm in two-dimensional aggregate quantity space is used as an example. In Figure 1, TFP_{nt} is given by the slope of the line that passes through the origin (0,0) and point *A*. In the same way, TFP_{ms} which shows the TFP of firm *m* in period *s*, is given by the slope of the line through the origin and point *Z*. The angles between the horizontal axis and the lines passing through points *A* and *Z* are lower-case *a* and *z* respectively. So,

$$TFP_{nt} = \frac{Y_{nt}}{X_{nt}} = slope\ OA = \tan a \quad (4)$$

$$TFP_{ms} = \frac{Y_{ms}}{X_{ms}} = slope\ OZ = \tan z \quad (5)$$

$$TFP_{ms,nt} = slope\ OA / slope\ OZ = \tan a / \tan z \quad (6)$$

where $TFP_{ms,nt}$ is a TFP index which measures TFP change between the two firms *n* and *m* in periods *s* and *t*, respectively. Thus, any multiplicatively-complete TFP index can be written as the ratio of (tangent) functions of angles in aggregate quantity space. For instance, assume *e* denotes the angle between the horizontal axis and the line passing through the origin and any non-negative point like *E*. Subsequently, it is obvious that the change in TFP between firm *m* and firm *n* can be decomposed as:

$$TFP_{ms,nt} = \tan a / \tan z = (\tan a / \tan e)(\tan e / \tan z) \quad (7)$$

Based on this structure, an unbounded number of points, like *E*, can be used to produce a decomposition of a multiplicatively-complete TFP index.

4. The components of TFP change

O'Donnell (2008) used this approach to provide further insights into the relationships between aggregate quantities and to conceptualise different alternative components of TFP change; measures of technical change and various measures of efficiency change; pure technical efficiency, mix efficiency, scale efficiency, residual scale efficiency and residua mix efficiency. The author mapped multiple-input multiple-output production points into aggregate quantity space; see Figure 2. This figure presents such a mapping for feasible input-output combinations represented by points A, C and V. In this Figure the curved line which passes through points *B* and *C* denotes the frontier of a restricted production possibilities set. It is named restricted since it only includes input and output aggregate vectors, which can be written as scalar multiples of x_t and y_t. In Figure 2, Firm A can boost its aggregate output (and consequently its TFP) by expanding outputs until it achieves point *C*. Hence, the vertical distance from point *A* to point *C* shows the measure of output-oriented technical efficiency (OTE), and can be defined as:

$$OTE_t = \frac{Y_t}{\bar{Y}_t} = \frac{\tan a}{\tan c}. \tag{8}$$

where $\bar{Y}_t$ is the maximum aggregate output that is technically feasible when using x_t to generate a scalar multiple of y_t. Accordingly, TFP of firm A, and maximum TFP possible (holding the input vector and output mix fixed) can be defined as $Y_t / X_t = \tan a$ and $\bar{Y}_t / X_t = \tan c$, respectively.

It is obvious from Figure 2 that any enhancements in technical efficiency imply expansions in TFP, however the TFP of Firm A is not maximized by shifting to the technically efficient point *C*. When the input-output mixes are held fixed, firm A can maximize its TFP by shifting to a point where a line through the origin is tangent to the restricted production possibilities frontier. This point is denoted as point *D* in Figure 2, and named as the point of mix-invariant optimal scale (MIOS) by O'Donnell (2008). Subsequently, pure scale efficiency is a measure of the difference between TFP at *C*, which is the technically efficient point, and TFP at *D* that is the point of MIOS. The term "pure" is used since input and output mixes are being held fixed, thus the change in TFP is a pure scale effect. The vertical distance from point *C* to point *S* represents the measure of output-oriented scale efficiency (OSE) and can be written as:

$$OSE = \frac{\bar{Y}_t / X_t}{\tilde{Y}_t / \tilde{X}_t} = \frac{\tan c}{\tan d} \tag{9}$$

where $\tilde{X}_t$ and $\tilde{Y}_t$ denote the aggregate input and output quantities at the MIOS point.

The efficiency measures discussed so far have been defined relating to a restricted production frontier. Removing restrictions on input and/or output mix causes an upward shift in the production possibilities set. The frontier of this developed production possibilities set is named as unrestricted production frontier which encloses a restricted boundary of the type shown in Figure 2. If the restrictions on output mix are relaxed, Firm A is able to expand more aggregate output compared with point *C* and move vertically to point *V* in Figure 2. In view of that, O'Donnell (2008) defined the mix efficiency measure as a difference between TFP at a technically efficient point on the mix-restricted frontier and TFP at a point on the unrestricted frontier. Hence, the pure output-oriented mix efficiency (OME) is defined as:

$$OME_t = \frac{\bar{Y}_t}{\hat{Y}_t} = \frac{\bar{Y}_t / X_t}{\hat{Y}_t / X_t} = \frac{\tan c}{\tan v} \tag{10}$$

where $\hat{Y}_t$ is the maximum aggregate output which is feasible when using x_t to produce any output vector.

It is obvious that any improvement in technical and mix efficiency implies enhancements of TFP. However the TFP of firm A is not maximized by shifting to the technically- and mix-efficient point *V*. More exactly, its TFP will be maximized only by moving to the point E where a straight line through the origin is tangent to the unrestricted production possibilities frontier. Point *E* is named as the point of maximum productivity. O'Donnell (2008) defined residual scale efficiency measure as the difference between the TFP amount at point *V* and TFP amount at point *E*. The vertical distance from point *V* to point *H* represents measure of residual output-oriented scale efficiency (ROSE):

$$ROSE_t = \frac{\hat{Y}_t / X_t}{Y^*_t / X^*_t} = \frac{\tan v}{\tan e} \tag{11}$$

where Y_t^* and X_t^* are the aggregate output and input quantities at point E. Hence, TFP_t^* is defined as the maximum TFP possible using any technically feasible inputs and outputs, and is depicted as $Y_t^* / X_t^* = \tan e$. O'Donnell (2008) also defined residual mix efficiency (RME) measure which can be calculated as the difference between TFP at the mix-invariant optimal scale point and TFP at the maximum productivity point:

$$RME_t = \frac{\tilde{Y}_t / \tilde{X}_t}{Y_t^* / X_t^*} = \frac{\tan d}{\tan e}. \tag{12}$$

This difference can be represented in Figure 2 as a movement from point *D* to point *E* (or the vertical distance between points *S* and *H*).

According to the definitions provided above, it can be concluded that:

$$\text{TFP efficiency } = TFPE_t = \frac{TFP_t}{TFP_t^*} = \frac{\tan a}{\tan e} = \frac{\tan a}{\tan c}\frac{\tan c}{\tan v}\frac{\tan v}{\tan e}. \tag{13}$$

This is an output-oriented measure of TFP efficiency that calculates the proportionate expansion in TFP as the firm moves the entire way from point *A* to point *E* (see Figure 2). As can be seen in Figure 2 there are many pathways from *A* to *E*. Thus there are many ways to decompose TFP efficiency. Pathway ACVE is used for $TFPE_t$. Thus, another way that can be traced by this firm is ACDE, so TFP efficiency can also be defined as:

$$TFPE_t = \frac{\tan a}{\tan e} = \frac{\tan a}{\tan c}\frac{\tan c}{\tan d}\frac{\tan d}{\tan e}. \tag{14}$$

In relation to the efficiency measures defined in this section, Equations (8) to (12), these two output-oriented decompositions can be defined as:

$$TFPE_t = \frac{TFP_t}{TFP_t^*} = \frac{\frac{Y_t}{X_t}}{\frac{Y_t^*}{X_t^*}} = OTE_t \times OME_t \times ROSE_t \tag{15}$$

$$TFPE_t = \frac{TFP_t}{TFP_t^*} = \frac{\frac{Y_t}{X_t}}{\frac{Y_t^*}{X_t^*}} = OTE_t \times OSE_t \times RME_t. \tag{16}$$

These two decompositions can be used as a foundation of an output-oriented decomposition of a multiplicatively complete TFP index, and can be rephrased as:

$$TFP_t = TFP_t^* \times (OTE_t \times OME_t \times ROSE_t) \tag{17}$$

$$TFP_t = TFP_t^* \times (OTE_t \times OSE_t \times RME_t) \tag{18}$$

A similar equation can be written for any other firm like *m* in period *s*. Accordingly, the index number that compares the TFP of firm *n* in period *t* with the TFP of firm *m* in period *s* is defined as:

$$TFP_{ms,nt} = \frac{TFP_{nt}}{TFP_{ms}} = \underbrace{\left(\frac{TFP_t^*}{TFP_s^*}\right)}_{\text{Technical change}} \times \underbrace{\left(\frac{OTE_{nt}}{OTE_{ms}} \times \frac{OME_{nt}}{OME_{ms}} \times \frac{ROSE_{nt}}{ROSE_{ms}}\right)}_{\text{Technical efficiency change}} \tag{19}$$

$$TFP_{ms,nt} = \frac{TFP_{nt}}{TFP_{ms}} = \underbrace{\left(\frac{TFP_t^*}{TFP_s^*}\right)}_{\text{Technical change}} \times \underbrace{\left(\frac{OTE_{nt}}{OTE_{ms}} \times \frac{OSE_{nt}}{OSE_{ms}} \times \frac{RME_{nt}}{RME_{ms}}\right)}_{\text{Technical efficiency change}}. \tag{20}$$

The first parentheses on the right-hand sides of these equations are measures of technical changes since they measure the difference between the maximum TFP possible using the technology feasible in period *t*, and the maximum TFP possible using the technology feasible in period *s*. Thus, the industry experiences technical improvement or decline as TFP_t^* / TFP_s^* is greater than or less than 1. In Figure 2, TFP_t^* / TFP_s^* measures the change in the slope of the line which passes through point *E*. On the contrary, in the decomposition of the Malmquist TFP index, Färe et al. (1994) compute the change in the slope of the line passing through point *D*. Hence, O'Donnell (2008) state that this technical change includes a mixed effect and characteristically differs from firm to firm. The second ratios in parentheses on the right-hand sides are understandable measures of technical efficiency change, (residual) mix efficiency change and (residual) scale efficiency change. Equation (20) is applied in this study to analyse different components of the technical efficiency change.

5. The data

To facilitate measurement of productivity changes, we initially had to specify sets of inputs and outputs for the banks in our sample. However, there being no consensus as to how to specify inputs and outputs, in this study we employ the intermediation approach to focus on bank services. Under this approach, banks are viewed as financial intermediaries with outputs measured in local currency, and with labour, capital, and different funding sources as inputs. This approach has a number of variants; *asset*, *value-added* and *user cost* views. Sealy and Lindley (1977) focus on the banks' role as financial intermediaries between depositors and the final users of bank *assets*. They classify deposits and other liabilities, plus real resources (labour and capital), as inputs, and only bank assets, such as loans, as outputs. Berger et al. (1987) classify loans and all types of deposits as "important" outputs since these balance sheet categories contribute to bank *value added*, whilst labour, capital, and purchased funds are classified as inputs. On the other hand, Aly et al. (1990) and Hancock (1991) implement a *user-cost* framework to determine whether a financial product is an input or an output based on its net contribution to bank revenue. Utilising this approach, a bank asset can be categorised as an output if the financial return on the asset goes above the opportunity cost of the investment, and a liability can be categorised as an output if the financial cost of the liability is less than its opportunity cost.

As our measurement of productivity relys on a mutually exclusive distinction between inputs and outputs, following Aly et al. (1990), as well as Wheelock and Wilson (1999) and Burgess and Wilson (1995), we classify inputs and outputs on the basis of the user cost approach. We include three inputs: labour (x_1) measured by the number of full-time equivalent employees on the payroll at the end of each period, physical capital (x_2) measured by the book value of premises and fixed assets, and purchased funds (x_3) including all time and savings deposits and other borrowed funds (not including demand deposits). We include three outputs: total demand deposits (y_1), public sector loans (y_2) including loans for agriculture, manufacturing, mining and services, and non-public loans (y_3). All data were obtained from Iran's Central Bank archives (CBI, 2005; and CBI, 2008). We consider all but three banks operating in the Iranian banking industry, as these three were not homogenous in input and output mixes. In all, we have used balanced panel data for 14 banks over 6 years (2003-2008). All estimates were made by means of DPIN software written by O'Donnell (2010a).

6. Empirical results

As the Hicks-Moorsteen is a distance-based index, DEA methodology developed by O'Donnell (2009; and 2010c) is applied for estimating the distances under VRS. The interpretation is straightforward. An efficiency estimate equal to unity indicates that the bank lies on the boundary of the production set, and, accordingly, is (relatively) efficient. An estimate below unity indicates that the bank is positioned under the frontier and is (relatively) inefficient. The estimates of output-oriented efficiency levels are reported in Table 1, and categorised into four groups; commercial banks, specialised banks, private banks and mean efficiency for the banking industry over the period 2003-2008[3]. Columns 1 and 2 of Table 1 show the different categories of the banks and years 2003 through 2008, respectively. Columns 3-5 list the measures of pure technical efficiency, scale efficiency and mix efficiency, respectively, for each year.[4]

Table 1. Measures of output-oriented technical efficiency (OTE), scale efficiency (OSE) and mix efficiency (OME)

Banks	Year	OTE	OSE	OME
Commercial Banks (Public)	2003	0.8905	0.9454	0.9379
	2004	0.9821	0.9736	0.9896
	2005	0.9820	0.9775	0.9804
	2006	0.9928	0.9397	0.9650
	2007	0.9950	0.6366	0.9532
	2008	0.9349	0.8806	0.9629
Specialized Banks (Public)	2003	1.0000	1.0000	0.9648
	2004	0.9263	0.9194	0.9078
	2005	0.9548	0.8851	0.9211
	2006	0.9911	0.8351	0.9105
	2007	0.9846	0.7420	0.8844
	2008	1.0000	0.8386	0.9030
Private Banks	2003	0.7949	0.9876	0.9502
	2004	0.9364	0.9383	0.9681
	2005	1.0000	0.9333	1.0000
	2006	0.9897	0.9527	0.9831
	2007	0.8971	0.9336	0.9016
	2008	0.8806	0.8684	0.9122
The Banking Industry	2003	0.8951	0.9777	0.9510
	2004	0.9482	0.9438	0.9552
	2005	0.9789	0.9319	0.9671
	2006	0.9912	0.9091	0.9528
	2007	0.9589	0.7707	0.9130
	2008	0.9385	0.8625	0.9260

Note: Efficiency estimates equal to unity indicate that the bank-group is the most efficient, and estimates below unity indicate that the bank-group is inefficient.
Source: Authors' calculations.

Table 1 reveals that, as a whole, the industry is inefficient over the entire period, however, levels of the public banks suggest pure efficiency improves over 2006-2008, whereas the mean of the private banks' pure efficiency levels declines considerably over this period. These level changes coincided with the program of major banking reform initiated in 2005 by then the newly-elected government; public banks were obliged to provide more direct facilities to less privileged areas and to provide lower interest rates and banking services compared to those of private banks. It may be argued that due to the large expansion of public banks' advances on the non-public sector, public banks became more purely efficient than private banks. On the other hand, considerably lower pure efficiency of private banks after 2005 can be attributable to their poor management of deposits which increased considerably due to the different banking rates, increase of public confidence in private banks, and the low attractiveness of investment in other markets (2007).[5] Table 1 also shows that, on average, Iranian banks become highly scale and mix inefficient after regulations were imposed, and scale inefficiency became a major problem for the industry. These weak levels of banks scale efficiency and mix efficiency can be attributed to inefficient scale size and the lack of independence of the banks in terms of managing their inputs-outputs, respectively.

Table 2. Total Factor Productivity changes and its various components assuming VRS

Banks	Period	ΔTFP	ΔTech	ΔEff	ΔOTE	ΔROSE	ΔOME
Commercial Banks (Public)	2003/2004	0.7656	0.8252	0.9209	1.1259	0.7734	1.0576
	2004/2005	1.0206	1.4253	0.7133	0.9999	0.7201	0.9908
	2005/2006	1.1901	1.0605	1.1234	1.0130	1.1266	0.9843
	2006/2007	1.1417	2.2734	0.5039	1.0023	0.5093	0.9870
	2007/2008	0.8179	0.8432	0.9765	0.9387	1.0254	1.0146
Specialized Banks (Public)	2003/2004	0.8762	0.8252	1.0597	0.9263	1.2225	0.9358
	2004/2005	1.1186	1.4253	0.7820	1.0404	0.7362	1.0209
	2005/2006	0.9110	1.0605	0.8553	1.0443	0.8319	0.9846
	2006/2007	1.8700	2.2734	0.8104	0.9934	0.8464	0.9638
	2007/2008	0.9682	0.8432	1.1448	1.0162	1.0971	1.0269
Private Banks	2003/2004	0.9065	0.8252	1.1298	1.2447	0.8877	1.0226
	2004/2005	1.0733	1.4253	0.7830	1.0854	0.6959	1.0366
	2005/2006	1.1838	1.0605	1.1107	0.9897	1.1417	0.9831
	2006/2007	0.9530	2.2734	0.4290	0.9078	0.5147	0.9182
	2007/2008	0.9633	0.8432	1.1437	0.9720	1.1582	1.0159
The Banking Industry	2003/2004	0.8494	0.8252	1.0619	1.0989	0.9612	1.0053
	2004/2005	1.0708	1.4253	0.7595	1.0419	0.7174	1.0161
	2005/2006	1.0950	1.0605	1.0327	1.0157	1.0334	0.9840
	2006/2007	1.3215	2.2734	0.5771	0.9678	0.6235	0.9563
	2007/2008	0.9164	0.8432	1.0873	0.9756	1.0935	1.0191

Note: ΔTFP = ΔTech×ΔEff, and ΔEff = ΔOTE×ΔROSE×ΔOME.

Source: Authors' calculations.

Table 2 lists measures of the banks' total factor productivity changes (ΔTFP) and its components, technical change (ΔTech) and efficiency change (ΔEff), in the four categories over five pairs of years between 2004 and 2008. The table also presents components of the ΔTech; changes in output-oriented pure technical efficiency (ΔOTE), residual scale efficiency (ΔROSE) and mix efficiency (ΔOME). Estimated values greater than unity indicate an improvement in the measures, and estimated values less than unity indicate a deterioration in these measures.

Table 2 shows technical changes (ΔTech) are the same for each group of banks in any period, indicating that banks have access to the same production possibilities set. Thus, all banks will be affected equally by expansions or contractions in the production possibilities set. A change in the production possibilities set (ΔTech) can be attributable to any changes in the environment. Thus, it will capture the effect of technological change as well as the longer term effects of government regulations and central bank policies. In 2004-2005, 2005-2006, and 2006-2007, the industry's estimated ΔTech was greater than unity, suggesting an overall technological progress in the industry. This is most probably due to the technological advances in the banking industry, which commenced in 2004, such as increased numbers of automated teller machines (ATM), credit cards, debit cards, and online-branches, as well as the increased pressure on commercial banks to expand credit in 2006. Despite this, the industry shows a large decrease in technical change for the period 2007-2008, which coincided with a substantial rise in the public banks' NPLs.

A general comparison of the different indexes presented in Table 2 reveals that the important components of Iranian banking TFP changes have been technical changes and changes in residual output-oriented scale efficiency (ROSE). There are two periods when the industry experienced a significant deterioration of ΔTFP: the period 2003-2004 when ΔTFP worsened by 16% (ΔTFP=84%), and the period 2007-2008 when ΔTFP exacerbated by 9% (ΔTFP=91%). Each of these periods was associated with a significant fall in the technical changes. Commercial banks, which are the largest banks in Iran, experienced the lowest level of scale efficiency changes (high negative changes) over almost all periods. Their scale efficiency rate during 2006-2007 was considerably negative (ΔROSE=0.5093) but this measure increased to 1.0254 over 2007-2008. These variations coincided with decisions made by a number of the largest commercial banks (*e.g. National Bank*, *Bank Saderat* and *Bank Sepah*) to reduce the number of bank branches and staff. Conversely, specialised banks and private banks improved their rate of scale efficiency growth from -16% (ROSE=0.84) to +9% (ROSE=1.09) and from -49% (ROSE=0.51) to +15% (ROSE=1.15), respectively, by increasing the number of their branches and employees over the same period.

In terms of output mix efficiency (OME), every bank group experienced negative changes over the periods 2005-2006 and 2006-2007, reflecting bank problems with the resource allocation in the post-regulation era when interest rates and the allocation of direct lending facilities were regulated. Hence, mix efficiency levels of the industry worsened by 2% during 2005-2006 (ΔOME=0.98) and by 5% during 2006-2007 (ΔOME=0.95). During 2006-2007 and 2007-2008 ΔOTE in all bank groups was low but private banks had the worst performance with a negative growth of 10% (ΔOTE=0.90) and 3% (ΔOTE=0.97), respectively. Consequently the industry, on average, showed negative changes in technical efficiency by 4% (ΔOTE=0.96) and 3% (ΔOTE=0.97) over these periods, respectively.

In general, the results in Tables 1 and 2 indicate that while government regulations may have resulted in large advances in the production possibilities set over time, the state regulatory measures exacerbated scale inefficiencies. In 2006-2007 commercial banks, specialised banks and private banks experienced extensive inefficiency changes (ΔEff) by 50%, 19% and 58%, respectively. However, the technology advances of banks offset the increase in efficiency changes (which is due to negative changes of scale, mix and pure technical efficiencies) over this period. Hence, public banks showed positive productivity changes and private banks showed only a 5% decrease in their productivity growth. On the other hand, over the period 2007-2008 large increases in the scale efficiency of banks did not offset the large reduction in production possibilities, and, on average, the banks' productivity deteriorated considerably through this period.

7. Conclusions

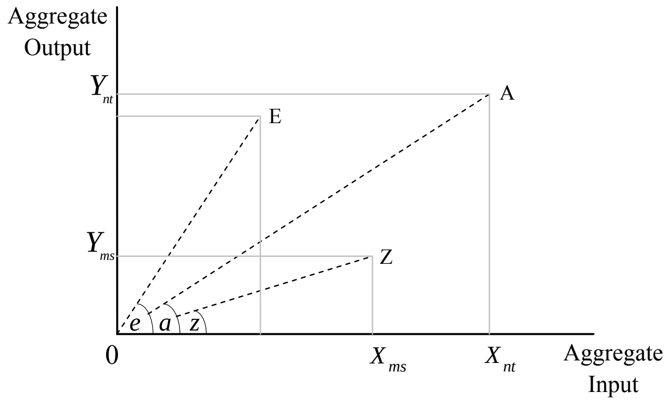
This paper has employed the Hicks-Moorsteen TFP index developed by O'Donnell (2008, 2009, 2010c) to analyse efficiency and productivity changes for the first time in a banking context. We investigate the effects of Iranian government regulations, launched in 2005, on technical efficiency and productivity changes in the Iranian banking industry over the period 2003-2008. Four different components of productivity change were estimated; i.e. technical changes, changes in pure technical efficiency, changes in scale efficiency, and changes in mix efficiency. Different efficiency measures were also computed.

Based on our results, it appears that, although the industry has been inefficient over the entire period of the study, the industry's technical efficiency has improved overall over the period 2003-2006, and deteriorated considerably soon after the regulatory changes were introduced in 2005. The efficiency level of public banks, in particular specialized banks, increased considerably after 2005 which is likely because of this reason that these banks, by virtue of undertaking most of the government borrowing programs, could generate significant advances from this source and thus tend to be more efficient under intermediation approach. Public banks' productivity changes show the same fluctuations as technical changes and the extent of productivity changes declined significantly and became negative after 2007. Private banks experienced negative productivity changes after 2006 despite the fact that they were not obliged to follow government guidelines for lending. In general, it can be concluded that the pure efficiency, mix efficiency and productivity of the industry have been affected considerably after introduction of regulations, and scale inefficiency has been a major problem for Iranian banks. Hence, there is significant room for improvement in Iranian banks in terms of scale efficiency and mix efficiency. Also, it seems that government control of the public banks has tended to limit the incentives and ability of managers to allocate their resources efficiently and to operate on an efficient scale.

It can, therefore, be suggested that the privatization of the banking industry should be expedited, and that government intervention should be reduced to boost the efficiency and productivity of banks in Iran. In addition, one may argue that the lacklustre performance of the private banks was mainly due to a considerable rise in deposits after the regulations were imposed, and that scale inefficiency was attributable to the lack of institutional growth. For future study, there is one technical problem with DEA that should be addressed; DEA does not have any statistical foundation, hence it is not possible to make inferences about DEA scores. One possible solution would be to use the bootstrap simulation method defined by Simar and Wilson (1998, 2000). This allows us to determine the statistical properties of the non-parametric estimators in the multi-input and multi-output case, and hence enabling the construction of confidence intervals for DEA efficiency scores.

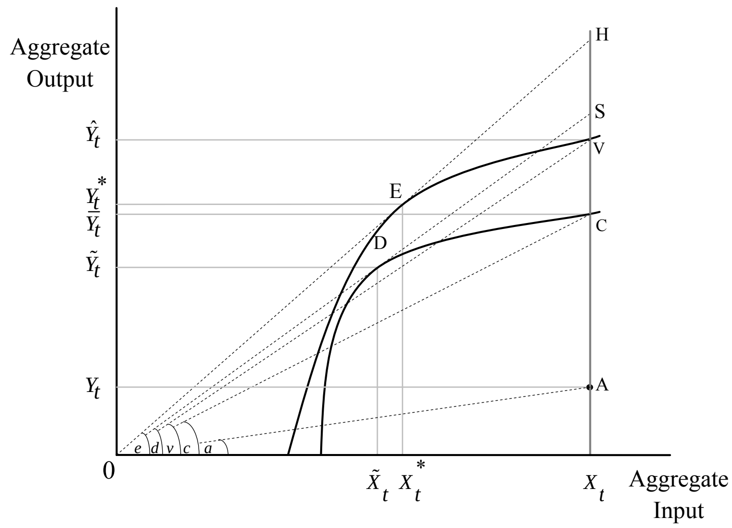
The major findings of this paper can be summarised as follows: First, overall the Iranian banking industry was inefficient during the period 2003-2008. Second, with the introduction of government regulation in 2005, the industry witnessed immediately its highest negative efficiency growth of 43% (ΔEff=0.57). Third, while the state ownership of public banks helped to reduce the extent of inefficiency of commercial banks by providing banking services to the government-specified areas. But the lack of independence of specialized banks from government controls led to their considerable mix inefficiency particularly after the regulatory measured introduced in 2005. Finally, Iranian banks need to focus on optimizing their scale size so as to improve their economies of scale and overcome their substantial scale inefficiencies.

Figure 1: Total factor productivity change



Source: O'Donnell (2008, p.25), edited by the authors.

Figure 2: Output-oriented decompositions of TFP efficiency



Source: O'Donnell (2009, p.20), edited by the authors.

References

Aly, H.Y., Grabowski, R., Pasurka, C. and Rangan, N. (1990), "Technical, scale, and allocative efficiencies in U.S. banking: an empirical investigation", *The Review of Economics and Statistics*, Vol. 72, pp. 211-218.

Berger, A.N., Hanweck, G.A. and Humphrey, D.B. (1987), "Competitive viability in banking: scale, scope, and product mixeconomies", *Journal of Monetary Economics*, Vol. 20, pp. 501-520.

Burgess, J.F. and Wilson, P.W. (1995), "Decomposing hospital productivity changes, 1985-1988: a nonparametric Malmquist approach", *The Journal of Productivity Analysis*, Vol. 6, pp. 343-363.

Caves, D.W., Christensen, L.R. and Diewert, W.E. (1982), "The economic theory of index numbers and the measurement of input, output, and productivity", *Econometrica*, Vol. 50 No. 6, pp. 1393-1414.

CBI (2005), *Iranian Banks Performance Report 2005*, Central Bank of Iran, Tehran.

CBI (2007), *Annual Report 2007*, Central Bank of Iran, Tehran.

CBI (2008), *Iranian Banks Performance Report 2008*, Central Bank of Iran, Tehran.

Chen, Y. and Lin, C. (2007), "Empirical study on the efficiency analysis of Australian banks", *Banks and Bank Systems*, Vol. 2 No. 4, pp. 38-49.

Coelli, T.J. and Rao, D.S.P. (2005), "Total factor productivity growth in agriculture: a Malmquist index analysis of 93 countries, 1980-2000", *Agricultural Economics*, Vol. 32 No. s1, pp. 115-134.

Diewert, W.E. (1992), "Fisher ideal output, input, and productivity indexes revisited", *Journal of Productivity Analysis*, Vol. 3, pp. 211-248.

Färe, R., Grosskopf, S., Lindgren, B. and Roos, P. (1992), "Productivity changes in Swedish pharmacies 1980-1989: a nonparametric Malmquist approach", *Journal of Productivity Analysis*, Vol. 3 No. 3, pp. 85-101.

Färe, R., Grosskopf, S., Norris, M. and Zhang, Z. (1994), "Productivity growth, technical progress, and efficiency change in industrialized countries", *American Economic Review*, Vol. 84 No. 1, pp. 66-83.

Farrell, M.J. (1957), "The measurement of productive efficiency", *Journal of the Royal Statistical Society* Vol. Series A (general) l20 No. 3, pp. 229-253.

Fethi, M.D. and Pasiouras, F. (2010), "Assessing bank performance with operational research and artificial intelligence techniques: a survey", *European Journal of Operational Research*, Vol. 204, pp. 189-198.

Good, D., Nadiri, M.I. and Sickles, R. (1997), "Index number and factor demand approaches to the estimation of productivity", in Pesaran, H. and Schmidt, P. (eds.), *Handbook of Applied Econometrics*, Basil Blackwell, Oxford, pp. 14-80.

Grifell-Tatje, E. and Lovell, C.A.K. (1995), "A note on the Malmquist productivity index", *Economics Letters*, Vol. 47 No. 2, pp. 169-175.

Hancock, D. (1991), *A Theory of Production for the Financial Firm*, Kluwer Academic Publishers, Boston.

Hicks, J. R. (1961), "Measurement of capital in relation to the measurement of other economic aggregates", in Lutz, F.A. and Hague, D.C. (eds.), *Theory of Capital*, Macmillan, London.

Jorgenson, D.W. and Grilliches, Z. (1967), "The explanation of productivity change", *The Review of Economic Studies*, Vol. 34 No. 3, pp. 249-283.

Moorsteen, R.H. (1961), "On measuring productive potential and relative efficiency", *The Quarterly Journal of Economics*, Vol. 75 No. 3, pp. 151-167.

Mukherjee, K., Ray, S.C. and Miller, S.M. (2001), "Productivity growth in large US commercial banks: the initial post-deregulation experience", *Journal of Banking and Finance*, Vol. 25 No. 5, pp. 913-939.

O'Donnell, C.J. (2008), "An aggregate quantity-price framework for measuring and decomposing productivity and profitability change", working paper, Centre for Efficiency and Productivity Analysis, University of Queensland, 13 November.

O'Donnell, C.J. (2009), "Measuring and decomposing agricultural productivity and profitability change", working paper, Centre for Efficiency and Productivity Analysis, University of Queensland, 11 February.

O'Donnell, C.J. (2010a), "DPIN version 1.0: a program for decomposing productivity index numbers", working paper, Centre for Efficiency and Productivity Analysis, University of Queensland, March.

O'Donnell, C.J. (2010b), "Measuring and decomposing agricultural productivity and profitability change", *Australian Journal of Agricultural and Resource Economics*: in press.

O'Donnell, C.J. (2010c), "Nonparametric estimates of the components of productivity and profitability change in U.S. agriculture", working paper, Centre for Efficiency and Productivity Analysis, University of Queensland, 5 July.

Ray, S.C. and Desli, E. (1997), "Productivity growth, technical progress, and efficiency change in industrialized countries: comment", *The American Economic Review*, Vol. 87 No. 50, pp. 1033-1039.

Sealy, C.W. and Lindley, J. (1977), "Inputs, outputs and a theory of production and cost at depository financial institutions", *Journal of Finance*, Vol. 32, pp. 1252-1266.

Shephard, R.W. (1953), *Cost and Production Functions*, Princeton University Press, Princeton.

Simar, L. and Wilson, P.W. (1998), "Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models", *Management Science*, Vol. 44 No. 1, pp. 49–61.

Simar, L. and Wilson, P.W. (2000), "Statistical inference in nonparametric frontier models: The state of the art", *Journal of Productivity Analysis*, Vol. 13 No. 1, pp. 49-78.

Sturm, J. and Williams, B. (2004), "Foreign bank entry, deregulation and bank efficiency: lessons from the Australian experience", *Journal of Banking and Finance*, Vol. 28 No. 7, pp. 1775-1799.

Sufian, F. (2006), "Post crisis productivity and efficiency of commercial banks: empirical evidence from Malaysia", *International Journal of Economic Research*, Vol. 3 No. 2, pp. 143-163.

Wheelock, D.C. and Wilson, P.W. (1999), "Technical progress, inefficiency and productivity change in U.S banking 1984-1993", *Journal of Money, Credit and Banking*, Vol. 31 No. 2, pp. 212-234.

Notes:

[1] Data Envelopment Analysis (DEA) is one of the most popular non-parametric approaches in the literature that has been used widely in frontier efficiency and productivity methods.

[2] O'Donnell's (2009) paper is also in press in Australian Journal of Agricultural and Resource Economics (O'Donnell 2010b).

[3] Results for all years are available from the authors upon request.

[4] As methods for estimating residual mix and residual scale efficiency levels are not presently available, hence, we only could provide estimates of pure technical efficiency, scale efficiency and mix efficiency.

[5] The ratio of Private Banks' deposits on Total deposits in the banking system increased considerably from 7% in 2004 to 23.8% in 2008.