Malmquist indexes using a geometric distance function (GDF). Application to a sample of Portuguese bank branches

Maria Conceição A. Silva Portela[1] ·
Emmanuel Thanassoulis

Keywords Efficiency · Malmquist Indexes · Bank Branches

Abstract Traditional approaches to calculate total factor productivity (TFP) change through Malmquist indexes rely on distance functions. In this paper we show that the use of distance functions as a means to calculate TFP change may introduce some bias in the analysis, and therefore we propose a procedure that calculates TFP change through observed values only. Our total TFP change is then decomposed into efficiency change, technological change, and a residual effect. This decomposition makes use of a non-oriented measure in order to avoid problems associated with the traditional use of radial oriented measures, especially when variable returns to scale technologies are to be compared. The proposed approach is applied in this paper to a sample of Portuguese bank branches.

Introduction

Productivity change has been a topic of interest since the earlier developments on this topic by Caves et al. (1982) on Malmquist productivity indexes. Earlier in 1978 the work of Charnes et al. (1978) provided a straightforward way to measure efficiency through linear programming models, and since then this framework (best known as Data Envelopment Analysis—DEA) has been applied to the measurement of productivity change through Malmquist indexes. Malmquist indexes, using DEA efficiency measures calculated in relation to a constant returns to scale (CRS) technology, are argued to be equivalent to a total factor productivity (TFP) index (see e.g. Färe et al. 1994, 1998) This is easily proved for single input/output technologies, but for multiple input/output technologies the calculated TFP Malmquist index has some problems. In this paper we refer in particular to problems arising from the fact that TFP is measured through the comparison of technical efficiency measures that require the definition of a reference technology. Indeed, the TFP Malmquist index computes productivity change between two observed points, say **a** and **b**, by finding one or more reference points relative to which the technical efficiency of **a** and

[1] The authors acknowledge the financial support of the Portuguese Foundation for Science and Technology, and the European Social Fund. The contents of the paper are the responsibility of the authors.

M. C. A. Silva Portela (✉)
Universidade Católica Portuguesa,
Centro Regional do Porto,
R. Diogo Botelho, 1327, 4169-005 Porto, Portugal
E-mail: csilva@porto.ucp.pt

E. Thanassoulis
Aston Business School,
Aston Triangle, B4 7ET
Birmingham, U.K.
E-mail: e.thanassoulis@aston.ac.uk

b is assessed. Productivity change is then inferred from changes in technical efficiency. We argue in this paper that TFP should be measured by comparing directly the points **a** and **b** rather than using references that might not be the same for each point.

This paper proposes, therefore, a novel way to compute TFP using observed values only, which does not require any specifications about the technology on which points such as **a** and **b** operate. The proposed TFP measure is then decomposed into efficiency change (EFCH), technological change (THCH), and a residual effect (RES) which reflects scale and allocative shifts. This decomposition obviously requires assumptions about the technology under which production units operate. In our TFP decomposition we try to account for some problems with existing methodologies such as the approach of Färe et al. (1994), (FGNZ throughout) and the approach of Ray and Desli (1997) (RD throughout). Both approaches calculate the TFP Malmquist productivity index in the same way (through radial efficiency scores calculated in relation to a CRS technology), but they decompose it differently. In the FGNZ approach the THCH component is calculated with reference to a CRS frontier, while in the RD approach it is calculated with reference to a VRS (variable returns to scale) frontier. Therefore, the FGNZ approach has the advantage of measuring changes in "maximal average product" (Färe et al. 1997b), but it has the disadvantage of not accounting for changes in the VRS technology. This might be a serious drawback if there are strong reasons to believe that the true technology is indeed VRS (see e.g. Balk 2001) The RD approach tries to solve the problem of the FGNZ approach by specifying a THCH component that is defined in relation to a VRS technology. This, however, may result in some DEA models being infeasible when assessments involve cross-period data (Bjurek 1996), meaning that for some units no THCH component can be calculated.

Both the FGNZ and the RD approaches are based on radial efficiency measures that are oriented either towards input contraction or output expansion. This provides different results concerning some components of productivity change depending on the model orientation. The orientation of DEA models is in some cases a given since some inputs and outputs are not under the control of production units. However, in many cases (for example in the assessment of bank branches as we undertake in this paper) at least some inputs and some outputs are under the control of production units and in such cases non-oriented models (i.e. models that allow for simultaneous changes in inputs and outputs towards the efficient frontier) might be used instead. The use of non-oriented efficiency measures solves the problem of sensitivity of the solution to the model's orientation, while at the same time solving the computational problems inherent to the RD approach. Examples of non-oriented efficiency measures that have been used in this context are the directional distance function used by Chambers et al. (1996) and Chung et al. (1997), and the hyperbolic efficiency measure used by Zofio and Lovell (2001).

Another problem of the FGNZ and RD approaches to calculating Malmquist indexes is that they rely on radial measures that do not account for slacks. If slacks are important sources of inefficiency, then the resulting Malmquist indexes may be based on biased measures of efficiency that do not fully reflect the distance between observed values and targets. Some authors have addressed this problem and solved it through the use of non-radial efficiency measures (note that non-radial efficiency measures are not necessarily non-oriented, though the reverse is true). For example, Grifell-Tatjé. (1998) developed a quasi-Malmquist productivity index that tries to overcome this problem (see also Førsund 1998, who criticise this paper), and Thrall (2000) developed an efficiency measure (based on a weighted additive model) that can be used in the computation of Malmquist type indexes.

The main contribution of this paper is, therefore, the development of a TFP change measure that is based on observed values only, and its decomposition through an efficiency measure that can evaluate both input and output changes (ICH and OCH) concurrently, that can consider variable returns to scale, and that can account for all the sources of inefficiency. The measure of efficiency that we will use in this paper was first proposed in Portela and Thanassoulis (2002) and is called geometric distance function (GDF) (see also Portela and Thanassoulis 2005). In the next section we briefly introduce the GDF efficiency measure and then we move on to point out the problems that may happen when the traditional approaches to calculate TFP are used. In Section "Malmquist type indexes based on the GDP" we show how the GDF measure can be used to calculate a TFP index, based on observed data only, and to decompose it into EFCH, THCH, and a residual component. In Section "Application to bank branches" we apply the developed procedure to a sample of bank branches, and the last section concludes this paper.

Geometric distance function

Let the vector $\mathbf{x} = (x_1, \ldots, x_m) \in R_+^m$ correspond to inputs used to produce an output vector $\mathbf{y} = (y_1, \ldots, y_s) \in R_+^s$ in a technology involving n production units. Consider that, for each production unit, efficient input and output levels are known and are equal to $(\mathbf{x}^*, \mathbf{y}^*) = ((x_1^*, \ldots, x_m^*), (y_1^*, \ldots, y_s^*))$. We denote these efficient levels of production by targets, as they correspond to projections of each observation on the Pareto-efficient frontier. The procedure used to arrive at such target points in not material to our approach and so the reader may assume that they are calculated by some known DEA model.

The GDF as first defined in Portela and Thanassoulis (2002), assumes the form shown in (1), where θ_i represents the ratio between a target input and an observed input i (x_i^*/x_i) and β_r represents the ratio between a target output and an observed output r (y_r^*/y_r).

$$(\text{GDF}) = \frac{(\Pi_i \theta_i)^{1/m}}{(\Pi_r \beta_r)^{1/s}}. \tag{1}$$

The GDF is defined in (1) as the ratio between the geometric mean of inputs' contraction towards target levels, and the geometric mean of outputs' expansion towards target levels. The geometric mean is, therefore, the way used to aggregate varying expansion and contraction factors towards target levels on the production frontier. As target levels can be calculated using any known procedure, the GDF is in fact a general measure that encompasses other existing measures in the literature. Note for example that if an input oriented DEA model was used to calculate target levels, then output targets (y_r^*) would equal observed levels (y_r) and therefore $\beta_r = 1$. At the same time all input targets (x_i^*) would equal the observed level of input times the radial efficiency measure $(x_i^* = \theta x_i)$ meaning that GDF in (1) would reduce to the traditional radial input efficiency score (θ).

When both inputs and outputs are allowed to change towards the efficient frontier, the GDF is a non-oriented measure that incorporates both input contraction and output expansion towards that frontier. It can also incorporate all the sources of inefficiency as long as target levels used in (1) are Pareto-efficient.

Problems with traditional ways of calculating TFP

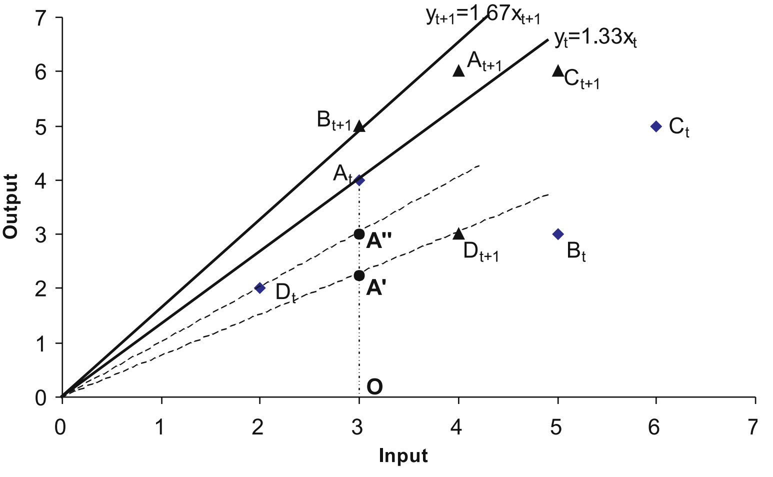
The FGNZ and RD approaches use radial efficiency measures calculated in relation to CRS frontiers to calculate Malmquist TFP indexes. Careful reflection shows, however, that this is just a means to an end since the computation of productivity change only requires the comparison of observed values in two different time periods. The computation of efficiency measures can be dispensed with for this purpose.

Consider the single input/output case where a measure of productivity change from period t to $t+1$ is given by the ratio P_{t+1}/P_t, where $P_t = y_t/x_t$ in each time period t. This productivity change measure is put forward by most authors that analyse productivity change and is free of controversy. Graphically P_{t+1}/P_t corresponds to the distance between the rays that pass through a given observation in period t and $t+1$ (see Fig. 1). The ray with the largest slope in each time period is the CRS frontier of that period (as it is associated with highest productivity).

Obviously the distance between the rays that pass through, for example, points D_{t+1} and D_t in Fig. 1 can be alternatively calculated with reference to another ray. Taking this reference as being the CRS frontier of period t defined by unit A_t, we have that the distance between the rays that pass through D_{t+1} and A_t divided by the distance between the rays that pass through D_t and A_t is equal to the distance between the rays that pass through D_{t+1} and D_t (in the graph this is the same as to say $\frac{OA'}{OA''} = (OA'/OA_t)/(OA''/OA_t)$). In a generalisation of the foregoing illustration existing approaches use distance functions defined in relation to CRS technologies to calculate productivity change indexes for the general case of multiple inputs/multiple outputs. Such distance functions are of the type introduced by Farrell (1957), which are usually operationalised through DEA (Charnes et al. 1978) (note that in Fig. 1 the ratios OA'/OA_t and OA''/OA_t are the Farrell output efficiency measures of units A' and A'', respectively).

Consider now a measure $\gamma_{j_t}^t$ indicating the radial efficiency of unit j as observed in period t and assessed in relation to the technology of period t (superscript). A Malmquist productivity index, M_j^t, is usually computed as $\gamma_{j_{t+1}}^t/\gamma_{j_t}^t$, when the reference is the t frontier. Obviously the reference technology could also have been $t+1$, which would result in $M_j^{t+1} = \gamma_{j_{t+1}}^{t+1}/\gamma_{j_t}^{t+1}$. The values of these two Malmquist indexes may differ

Fig. 1 One input/output example



and, as such, Färe et al. (1994) consider the geometric mean of both as the Malmquist TFP index as shown in (2).

$$M_j = \left(\frac{\gamma^t_{j_{t+1}}}{\gamma^t_{j_t}} \times \frac{\gamma^{t+1}_{j_{t+1}}}{\gamma^{t+1}_{j_t}}\right)^{(1/2)}. \tag{2}$$

Note that productivity change, as shown in Fig. 1, is not dependent on efficiency or functional form of the efficient frontier as defined in DEA. The use of distance functions is just a means to operationalise the concept for the multiple input/output case. This approach relies, however, on efficiency being calculated in relation to a unique referent line or plane. This necessarily happens in the single input/output case as the ray presenting maximum productivity in each time period is unique. If the referent hyperplane is not the same for observations in t and $t+1$, then the Malmquist index as defined in (2) is just an approximation for true productivity change and not a real measure of productivity change. In the multiple input/output case, CRS technologies are defined by a cone that has multiple facets, and projections on this cone may happen on any of its facets. This means that the referent hyperplane, or facet, is not necessarily the same for every two observations in t and $t+1$ between which productivity change is to be measured.

To illustrate the above, consider the example in Table 1, where 5 units producing one output (y) from 2 inputs (x_1 and x_2) are considered.

In Table 1 we also show the growth in partial productivity between periods t and $t+1$. That is, calculating the partial productivity of the output in relation to input 1, y/x_1, and the partial productivity of the output in relation to input 2, y/x_2, for each time period, the ratio $\Delta \mathbf{y}/\mathbf{x_i} = y_{t+1}/x_{i_{t+1}}/y_t/x_{i_t}$ shows partial productivity growth of output in relation to each input i. Inspecting these ratios in Table 1, it is clear that units 1, 3 and 5 increased their productivity from t to $t+1$, while the productivity of unit 4 decreased in the same period. Note also, that unit 5 shows the highest productivity increase from t to $t+1$ since the partial productivity growth ratios seen jointly are the highest that can be found. If we now apply (2) to calculate productivity change the results are as shown in Table 2, where M_j is the geometric mean of M^t_j and M^{t+1}_j.

These results show some contradiction to what was expected from the partial productivity ratios, especially

Table 1 Illustrative example

Unit	Period t			Period $t+1$			Growth	
	y	x_1	x_2	y	x_1	x_2	$\Delta(\mathbf{y}/\mathbf{x_1})$	$\Delta(\mathbf{y}/\mathbf{x_2})$
Unit 1	12	5	13	22	8	14	1.146	1.702
Unit 2	14	16	12	12	12	11	1.143	0.935
Unit 3	26	16	26	26	8	25	2	1.081
Unit 4	26	17	15	20	15	14	0.872	0.824
Unit 5	8	12	14	8	6	10	2	1.4

Table 2 Malmquist results for illustrative example

Unit	$\gamma^t_{j_t}$	$\gamma^{t+1}_{j_{t+1}}$	$\gamma^{t+1}_{j_t}$	$\gamma^t_{j_{t+1}}$	M^t_j	M^{t+1}_j	M_j
Unit 1	1.000	1.000	0.784	1.396	1.397	1.275	1.334
Unit 2	0.673	0.694	0.742	0.646	0.961	0.935	0.947
Unit 3	0.852	1.000	0.636	1.354	1.588	1.571	1.579
Unit 4	1.000	0.909	1.103	0.857	0.857	0.824	0.840
Unit 5	0.398	0.509	0.363	0.692	1.740	1.400	1.559

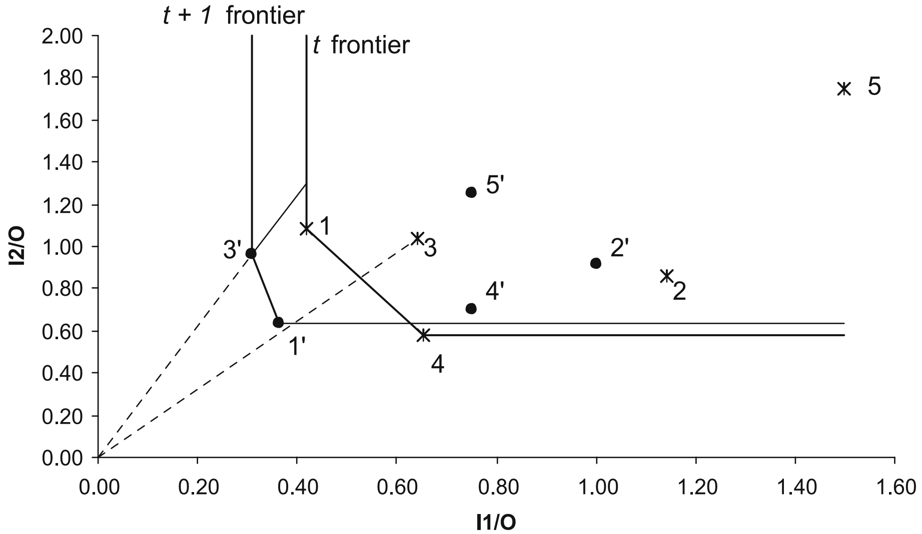


Fig. 2 Illustrative 2 Inputs 1 output example

because unit 5 does not have the highest Malmquist index as one would expect. At the same time, while it is clear that unit 4 exhibited a productivity decrease (and the Malmquist index correctly identifies this decrease), it is not clear that unit 2 also had a productivity decline. In fact a guess on the productivity change of this unit would more likely be a productivity increase, because the growth on $\Delta(\mathbf{y}/\mathbf{x_1})$ is higher than the decline in $\Delta(\mathbf{y}/\mathbf{x_2})$.

The reasons for the above behaviour of the Malmquist TFP index can be better explained trough Fig. 2, where observations in $t+1$ are represented by dots and observations in t are represented by crosses.

It is clear in Fig. 2 that the hyperplane against which efficiency is measured is not necessarily the same for an observation in t and $t+1$. For example, when unit 3 is evaluated in relation to the period t frontier it happens that unit 3 as observed in t is projected on the hyperplane defined by units 1 and 4, but unit 3 observed in $t+1$ is projected on the hyperplane of the t frontier defined only by unit 1 (facet where free disposability applies, or weakly efficient facet). The same happens with other units in Fig. 2.

Note also that the Malmquist index of units 2, 4, and 5 as evaluated in relation to the $t+1$ frontier (M_j^{t+1}) in Table 2 is exactly equal to the partial productivity change of input 2 ($\Delta(\mathbf{y}/\mathbf{x_2})$) in Table 1. This means that when productivity change for these units is evaluated in relation to the $t+1$ frontier one of the inputs (in this case input 1) is completely neglected in the analysis. Such a result is due to projections on the 'flat' part of the frontier of $t+1$ in Fig. 2, that satisfy free disposability of input 1. This fact strengthens what was previously said about the importance of using efficiency measures that account for all sources of inefficiency. The radial efficiency measures shown in Table 2 do not account for the slacks on input 1 that exist for units 2, 4, and 5 when these are projected on the $t+1$ frontier and, therefore, the Malmquist indexes based on these measures cannot account for productivity changes in input 1 but only in input 2.

In summary, the non-existence of a single referent hyperplane against which efficiency is measured for the same unit in different time periods, causes biased results on Malmquist TFP indexes that are based on such measures. In the next section we propose a GDF based approach that attempts to resolve some of the problems identified here.

Malmquist type indexes based on the GDF

The GDF measure defined in (1) has a double role in this paper. On the one hand it is used to calculate efficiency measures that are non-oriented and account for all sources of inefficiency, and on the other hand it is used to calculate a TFP index based on observed values only, without reference to a technology frontier. This TFP is then decomposed into three components, namely EFCH, THCH, and a RES in the way shown in (3).

$$\text{TFP} = \text{EFCH} \times \text{THCH} \times \text{RES}. \quad (3)$$

The way each of the above terms is computed through the GDF is presented next.

Calculating TFP

As noted earlier, TFP change is defined for the single input/output case as a ratio between output to input ratios in two different time periods ($\frac{y_{t+1}/x_{t+1}}{y_t/x_t}$), or alternatively as a ratio between output change and input change from t to $t+1$ ($\frac{y_{t+1}/y_t}{x_{t+1}/x_t}$). When there are multiple inputs and/or outputs the use of such ratios calls for the aggregation of the varying output and/or various input change. This obviously implies the definition of an aggregation formula like a simple or weighted average. Traditional index number approaches (e.g. Py 1990) provide such formulae, where factor prices are usually the aggregation instrument. The best known examples of such indexes are the Laspeyres, Paasche and Fisher indexes. Without price information the aggregation of amounts expressed in different units of measurement is not easy. However, if we express input and output values in terms of a rate of change from one period to the other, then aggregation of rates without using factor prices, and making use of observed values only is possible. The GDF provides a way of aggregating meaningfully input and output change rates from one period to the other that can be interpreted as a TFP index. (Note that by ignoring factor prices, we are calculating TFP change from a technological perspective and not from an economic/allocative perspective. This is analogous to well established Malmquist index approaches (such as the FGNZ approach) that also ignore factor prices.)

Although the GDF has been originally proposed in Portela and Thanasoulis (2002) as a way to measure efficiency, it can be adapted to the present context to calculate productivity change. This is shown in (4), where the input/output levels considered are not observed versus target as in (1) but observed in t versus observed in $t+1$.

$$\text{TFP} - \text{GDF}\,(x_t,\, y_t,\, x_{t+1},\, y_{t+1}) = \frac{\left(\Pi_r \frac{y_{r_{t+1}}}{y_{r_t}}\right)^{1/s}}{(\Pi_i \frac{x_{i_{t+1}}}{x_{i_t}})^{1/m}}. \quad (4)$$

For the single input/output case is is easy to see that (4) corresponds to a TFP index. In the multiple input/output case the GDF is a ratio between a geometric mean of output growth and a geometric mean of input growth, which is in fact a TFP index.[2]

Table 3 TFP results for illustrative example based on the GDF

	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
TFP-GDF	1.3967	1.034	1.4999	0.84765	1.6733

If we apply (4) to the illustrative example shown in the previous section the values are those shown in Table 3.

Productivity growth is identified for all units except unit 4 as expected. Note also that unit 5 is now the unit that offers the highest productivity growth, exactly as one would expect from the knowledge of the partial productivity ratios calculated previously (see Table 1). The GDF seems therefore a good alternative to calculate TFP change, having the advantage of relying only on observed values and making, therefore, no assumptions about the technology.

EFCH and THCH Components

The GDF in (4) used to calculate TFP is not an efficiency measure as it does not account for distances between observed and target levels but between two points observed in different time periods. In this sense the calculation of TFP does not require any assumptions regarding the technological specification of the frontier. However, when the GDF is used to calculate the EFCH and THCH components of TFP such assumptions about the technology are required.

Consider a measure $\text{GDF}^t\,(x_t,\, y_t)$, as calculated through (1), representing the efficiency measure of the input/output vector $(\mathbf{x}, \mathbf{y})$ as observed in period t and projected against technology of period t (superscript). A Malmquist type index based on the GDF is given by (5).

$$\text{MGDF}=\left(\frac{\text{GDF}^t(y_{t+1},x_{t+1})}{\text{GDF}^t(y_t,x_t)}\times\frac{\text{GDF}^{t+1}(y_{t+1},x_{t+1})}{\text{GDF}^{t+1}(y_t,x_t)}\right)^{\frac{1}{2}}. \quad (5)$$

[2] See e.g. Diwert and Nakamura (2003, p. 148) who defined TFP as the ratio between a measure of output growth and input growth in the multiple input/output case, however defining differently each aggregate measure of growth.

As in other approaches to computing a Malmquist index, the index in (5) can be decomposed in EFCH and THCH as shown in (6).

$$\mathrm{MGDF} = \frac{\mathrm{GDF}^{t+1}(y_{t+1}, x_{t+1})}{\mathrm{GDF}^{t}(y_t, x_t)} \times \left[\frac{\mathrm{GDF}^{t}(y_{t+1}, x_{t+1})}{\mathrm{GDF}^{t+1}(y_{t+1}, x_{t+1})} \times \frac{\mathrm{GDF}^{t}(y_t, x_t)}{\mathrm{GDF}^{t+1}(y_t, x_t)} \right]^{1/2}. \tag{6}$$

That is,

$$\mathrm{MGDF} = \mathrm{EFCH} \times \mathrm{THCH}. \tag{7}$$

As the GDF is a general measure, the above decomposition is also general and encompasses as special cases other decomposition approaches in the literature. For example when input or output oriented CRS models are used to calculate target points implicit in the GDF, the above reduces to the FGNZ approach. Note, however, that MGDF is not necessarily equal to TFP as it is usually measured in the literature. We consider that MGDF is simply the product of EFCH and THCH. TFP as calculated in the previous section includes these components but may also include another component as will become clearer later.

Assume now a single input/output case, where technical efficient projections of each observation are identified by the superscript $*t$ if the projection lies on the t frontier and by $*t+1$ if the projection lies on the $t+1$ frontier. In this case, each of the MGDF components assumes the form shown in (8).

$$\mathrm{EFCH} = \frac{\dfrac{x_{t+1}^{*t+1}/x_{t+1}}{y_{t+1}^{*t+1}/y_{t+1}}}{\dfrac{x_t^{*t}/x_t}{y_t^{*t}/y_t}}, \quad \mathrm{THCH} = \left(\frac{\dfrac{x_{t+1}^{*t}}{y_{t+1}^{*t}}}{\dfrac{x_{t+1}^{*t+1}}{y_{t+1}^{*t+1}}} \times \frac{\dfrac{x_t^{*t}}{y_t^{*t}}}{\dfrac{x_t^{*t+1}}{y_t^{*t+1}}} \right)^{\frac{1}{2}}. \tag{8}$$

The EFCH component in (8) [and (6)] is interpreted in the usual way, i.e. when it is higher than one the efficiency of observation in $t+1$ evaluated in relation to the $t+1$ frontier (measured for the single input output case as $\frac{x_{t+1}^{*t+1}/x_{t+1}}{y_{t+1}^{*t+1}/y_{t+1}}$) is higher than the efficiency of observation in t evaluated in relation to the t frontier, and therefore there was an efficiency increase from t to $t+1$. (When EFCH is lower than one there is an efficiency decrease in moving from t to $t+1$). In the same way a THCH component higher than one means technological progress and a THCH component lower than one means technological regress. Note that the technical change component may be re-organised so that we have a product of ICH and OCH. That is $\mathrm{THCH} = \left(\frac{x_{t+1}^{*t}}{x_{t+1}^{*t+1}} \times \frac{x_t^{*t}}{x_t^{*t+1}}\right)^{\frac{1}{2}} \times \left(\frac{y_{t+1}^{*t+1}}{y_{t+1}^{*t}} \times \frac{y_t^{*t+1}}{y_t^{*t}}\right)^{\frac{1}{2}}$. An ICH factor greater than 1 means that the frontier at t has higher inputs than the frontier at $t+1$. That is, there was an improvement (decrease) in inputs in moving from t to $t+1$. If the OCH is higher than 1, it means that outputs in $t+1$ are higher than outputs in t, which also means an improvement in outputs in moving from t to $t+1$. So progress from t to $t+1$ happens when both input and OCH are greater than 1. Obviously one may have movements in different directions and in this case the resulting technological progress or regress will depend on which factor dominates the other. Note that the input and output change components of THCH are closely related to the input and output scale bias defined in Färe et al. (1997a) (see also Färe et al. 1998, 2001) In Appendix A we relate the input and output bias components of Färe et al. (1997a) with those identified by our approach.

In the multiple input/output case the above technological and EFCH components are calculated as shown in (9), where again THCH change is the product of ICH and OCH.

$$\mathrm{EFCH} = \frac{\dfrac{\left(\Pi_i \theta_{i_{t+1}}^{t+1}\right)^{1/m}}{\left(\Pi_r \beta_{r_{t+1}}^{t+1}\right)^{1/s}}}{\dfrac{\left(\Pi_i \theta_{i_t}^{t}\right)^{1/m}}{\left(\Pi_r \beta_{r_t}^{t}\right)^{1/s}}},$$

$$\mathrm{THCH} = \left(\left(\Pi_i \frac{x_{i_{t+1}}^{*t}}{x_{i_{t+1}}^{*t+1}} \times \Pi_i \frac{x_{i_t}^{*t}}{x_{i_t}^{*t+1}}\right)^{\frac{1}{m}} \right)^{\frac{1}{2}} \times \left(\left(\Pi_r \frac{y_{r_{t+1}}^{*t+1}}{y_{r_{t+1}}^{*t}} \times \Pi_r \frac{y_{r_t}^{*t+1}}{y_{r_t}^{*t}}\right)^{\frac{1}{s}} \right)^{\frac{1}{2}} \tag{9}$$

The EFCH and THCH components in (9) include those existing in the literature, though being more general because they can handle situations where non-oriented models are used to calculate target levels. If both inputs and outputs change towards the technical efficient frontier, then the ratios considered in (9) account simultaneously for these changes. These ratios can be calculated both when targets lie on a CRS frontier or on a VRS frontier. We shall use, however, only the latter technological specification for reasons that will become clearer in the next section.

Residual Effect

The MGDF in (5) can alternatively be decomposed as shown in (10), where it equals the product of a TFP index as calculated through the GDF (see Section "Calculating TFP") and a residual component that is scale related.

$$\mathrm{MGDF} = \frac{\left(\Pi_r \frac{y_{r_{t+1}}}{y_{r_t}}\right)^{1/s}}{\left(\Pi_i \frac{x_{i_{t+1}}}{x_{i_t}}\right)^{1/m}} \times \left(\frac{\left(\Pi_r \frac{y^{*t}_{r_t}}{y^{*t}_{r_{t+1}}}\right)^{1/s}}{\left(\Pi_i \frac{x^{*t}_{i_t}}{x^{*t}_{i_{t+1}}}\right)^{1/m}} \times \frac{\left(\Pi_r \frac{y^{*t+1}_{r_t}}{y^{*t+1}_{r_{t+1}}}\right)^{1/s}}{\left(\Pi_i \frac{x^{*t+1}_{i_t}}{x^{*t+1}_{i_{t+1}}}\right)^{1/m}} \right)^{\frac{1}{2}} \quad (10)$$

Note that it is TFP that one wants to decompose, and therefore the above is better expressed as (11).

$$\mathrm{TFP} = \mathrm{MGDF} \times \left(\frac{\left(\Pi_i \frac{x^{*t}_{i_t}}{x^{*t}_{i_{t+1}}}\right)^{1/m}}{\left(\Pi_r \frac{y^{*t}_{r_t}}{y^{*t}_{r_{t+1}}}\right)^{1/s}} \times \frac{\left(\Pi_i \frac{x^{*t+1}_{i_t}}{x^{*t+1}_{i_{t+1}}}\right)^{1/m}}{\left(\Pi_r \frac{y^{*t+1}_{r_t}}{y^{*t+1}_{r_{t+1}}}\right)^{1/s}} \right)^{\frac{1}{2}} \quad (11)$$

To see that the square root in (11) is scale related, consider the single input/output case, where (11) reduces to (12).

$$\mathrm{TFP} = \mathrm{MGDF} \times \left(\frac{\frac{x^{*t}_t}{x^{*t}_{t+1}}}{\frac{y^{*t}_t}{y^{*t}_{t+1}}} \times \frac{\frac{x^{*t+1}_t}{x^{*t+1}_{t+1}}}{\frac{y^{*t+1}_t}{y^{*t+1}_{t+1}}} \right)^{\frac{1}{2}} \quad (12)$$

The second term of this decomposition compares changes between input and output targets along the t and the $t+1$ frontier. As it is arbitrary to measure these changes on the t or on the $t+1$ frontier the geometric mean between both is taken in (12). As all the points considered in the square root in (12) are efficient points, the movements between these points (on each frontier) can only reflect the exploitation of scale economies or changes in the mix of operations.

The TFP as calculated through the GDF approach decomposes, therefore, in MGDF (which includes a THCH and EFCH components) and in a residual component (RES) that is scale related. Note that if input and output targets were calculated in relation to a CRS technology, then in (12) one would have TFP = MGDF as the residual component would equal 1.[3] On the other hand, if target points are calculated in relation to a VRS technology, then the above decomposition in (12) is equivalent to the RD approach, where the residual component in (12) is equal to the RD scale effect. In Appendix B this equivalence is proved for the single input/output case. For the multiple input/output case, the RD and the GDF approaches yield different TFP and RES components, but can yield the same EFCH and THCH components when the same efficiency models are used in both approaches (to calculate efficiency scores in the RD model and target levels in the GDF model).

Interpreting the RD scale change factor is not easy (e.g. see (Lovell (2001) and Ray (2000)), since it is not a straightforward ratio of scale efficiency in two different periods (as happens to be the case in the FGNZ approach). However, it is not clear that the scale related component of productivity change should reflect changes in scale efficiency. For example, Lovell (2001) points out that the scale component of productivity change should reflect the influence of scale economies on productivity change rather than changes in scale efficiency. The author further points out that this contribution of scale economies to productivity change is provided by the scale component of the RD approach, whereas the contribution of the scale EFCH of the FGNZ approach to explain scale economies is unclear. Given that our residual component is related to the RD approach its interpretation in terms of scale effects is not straightforward, especially for technologies involving multiple inputs/outputs.

In order to shed some light on the interpretation of our residual component, note that in the single input/output case the square root in (12) can be alternatively written as: $\left(\frac{\frac{y^{*t}_{t+1}}{x^{*t}_{t+1}}}{\frac{y^{*t}_t}{x^{*t}_t}} \times \frac{\frac{y^{*t+1}_{t+1}}{x^{*t+1}_{t+1}}}{\frac{y^{*t+1}_t}{x^{*t+1}_t}} \right)^{\frac{1}{2}}$. Considering the production frontier at period t and assuming that $x^{*t}_{t+1} > x^{*t}_t$, we have that if $[y^{*t}_{t+1}/x^{*t}_{t+1}]/[y^{*t}_t/x^{*t}_t]$ is greater than 1 then the production function exhibits increasing returns to scale; if this ratio is equal to 1 we have CRS,

[3] Assume for example an input oriented CRS model, where target outputs correspond to observed outputs whatever the frontier where the observed unit is projected. This means that $y^{*t}_t = y^{*t+1}_t = y_t$, and $y^{*t}_{t+1} = y^{*t+1}_{t+1} = y_{t+1}$. At the same time target inputs are given by the function of the ray that passes through the origin and point (x_t, y_t). Let this function be given by $y_t = a_t x_t$ for period t and $y_{t+1} = a_{t+1} x_{t+1}$ for period $t+1$. Replacing this in the second term of (12) we have $\left(\frac{\frac{y_t/a_t}{y_{t+1}/a_t}}{\frac{y_t}{y_{t+1}}} \times \frac{\frac{y_t/a_{t+1}}{y_{t+1}/a_{t+1}}}{\frac{y_t}{y_{t+1}}} \right)^{\frac{1}{2}}$, which equals 1.

and if it is lower than 1 we have decreasing returns (see Diewert and Nakamura 2003, who also put forward this interpretation). Each term of the above geometric mean can therefore be interpreted as containing information on the returns to scale properties of the production frontier.

Though attractive this interpretation it may have some problems. Note, for example, that the above interpretation implies $x_{t+1}^{*t} > x_t^{*t}$. If $x_{t+1}^{*t} < x_t^{*t}$, then a ratio of two input–output coefficients lower than 1 would indicate increasing returns while a value higher than 1 would indicate decreasing returns. Therefore the interpretation of values higher or lower than 1 is conditional to the relationship between input levels at the two points being compared. Another difficulty relates to the fact that the RES component is an aggregate measure of returns to scale on both the t and $t+1$ frontiers. While a component on each frontier can be interpreted in the way suggested by Diwert and Nakamura (2003) a geometric mean of RTS on the t and $t+1$ frontier seems to be lacking an easy interpretation. For the multiple input/output case difficulties are even greater because movements along each production frontier may reflect, apart from scale effects, also mix effects. This means that in this case the interpretation of this factor becomes even more complicated. As one of the referees to this paper pointed out, the residual term can also be a host of factors that are not considered in the input/output specification such as workforce motivation, leadership style, etc. In fact, such factors may be one of the reasons for TFP change, but the decomposition of TFP into its efficiency and THCH components cannot account for them and the residual term may be capturing their effect. It is not our aim in this paper to further analyse the scale change component of the GDF measure. Suffice it to say at this stage that it reflects a residual that accounts for differences between TFP and a Malmquist index calculated in relation to a VRS technology. When the RES is greater than 1 it means that TFP change benefits from a positive influence of the exploitation of scale economies, whereas when it is lower than 1 this influence on TFP is negative.

Application to bank branches

The GDF approach is now applied to a sample of bank branches. The input and output factors used in the efficiency assessment are identified in Table 4.

Table 4 Inputs and outputs used to assess efficiency in month t

Inputs	Outputs
Number of staff [Staff](t)	Value current accounts [Curracc] (t)
Supply costs [Supplycosts] (t)	Value other resources [Othress](t)
	Value credit bank [Credb](t)
	Value credit associates [Credass](t)

These input–output variables were selected to reflect the cost sources on the input side and the revenue sources on the output side. This makes our model, consistent with the intermediation approach to analysing bank branches' efficiency (see for example Oral and Yolalan 1990; Athanassopoulos 1997; Berger et al. 1997, for bank branch studies that also adopted this perspective).

The outputs in Table 4 capture the major sources of revenue to the bank branch: interest revenue from the management of the various products the bank has to offer, namely, current accounts, other resources (which includes term deposit accounts, emigrant accounts, investment funds, savings insurance, etc.), and credit. The bank under analysis distinguishes between two types of credit: directly by the bank and by associates. The former consists of all types of credit that the bank itself can provide, while the latter consists of special types of credit that the bank provides through some associate companies (like leasing or factoring credit). The models used to measure technical efficiency are non-oriented, permitting simultaneous input contraction and output expansion. This is compatible with seeking profit maximisation which could involve non-oriented movements in this manner rather than oriented cost minimisation or output maximisation.

Technical efficiency analysis over time

The GDF Malmquist procedure developed here is applied to our data set which covers the period from March to December 2001 on a month by month basis. Technical efficiencies for each input–output bundle of a branch in a given month are computed both relative to the efficient frontier of that month and relative to the efficient frontier of a previous or succeeding month as our approach dictates. The model used to obtain all

Table 5 Malmquist GDF index results

Period	GDF_t	GDF_{t+1}	$MGDF_V$	EFCH	THCH	ICH	OCH	RES	TFP
M–A	0.6950	0.7380	1.0575	1.1415	0.9697	0.9851	0.9857	0.9952	1.0193
A–My	0.7380	0.7063	0.9420	0.9722	0.9894	1.0014	0.9890	1.0825	0.9952
My–J	0.7063	0.7852	1.0021	1.1646	0.8908	0.9764	0.9129	0.9995	0.9930
J–Jy	0.7807	0.7894	1.0265	1.0265	1.0138	1.0133	1.0011	0.9964	1.0158
Jy–Au	0.7894	0.7823	1.0581	1.0015	1.0705	1.0081	1.0637	1.0306	1.0843
Au–S	0.7823	0.7338	0.9044	0.9471	0.9935	0.9837	1.0097	1.0260	0.9234
S–Oc	0.8545	0.8223	1.0131	0.9653	1.0523	1.0120	1.0408	1.0029	1.0121
Oc–N	0.8223	0.7599	0.6562	0.9311	0.7139	0.7185	1.0048	0.9897	0.6467
N–D	0.7599	0.7443	1.1999	0.9981	1.2278	1.0774	1.1466	1.0002	1.1921

target levels is model (C) shown in Appendix C.1 The GDF efficiency measure was then used to reflect the distance between the data of the branch in a given month and the various sets of targets derived on the foregoing frontiers.

Our data consist of 57 branches for each month from March to June. In July one branch closed and the sample was left with 56 units. In October four more branches closed and the sample was left with 52 units. Table 5 shows the average values for each of the TFP components.

Average efficiency (GDF_t and GDF_{t+1} in Table 5) is high with a lower value of 69.5% in March and a maximum of 85.45% in September. The EFCH of bank branches or the catching up effect is not stable during the period of analysis with progress happening from March to April, regress from April to May, and progress again from May to June. After June EFCH started deteriorating and values fell below unity after August, though from November to December there was a slight improvement.

As far as the THCH component is concerned, Fig. 3 depicts the values in Table 5.

Technological regress happened from March to June, followed by a slight progress. The most severe regress shows from October to November followed by strong progress from November to December. The THCH recorded reflects changes to input and output levels as shown in Table 5 (see also Fig. 3). OCH is the main factor underlying THCH in all months except November 2001. In that month it is input rise that accounts for the technological regress from October to November, since in this period outputs showed a slight rise (see Fig. 3) but obviously not enough to compensate for the higher input levels.

TFP change as measured by our approach, which does not use a reference technology but relies on observed values only, is generally close to 1 from March to August. From August to September average productivity decreases considerably, to increase again from September to October. The decrease in TFP in September seems to be mostly due to a decrease in efficiency rather than technological regress. Nevertheless, the highest decrease in productivity occurs between October and November, though in this case the factor that mostly explains this decrease is technological regress brought about especially by input rise. It is worth noting that there is a large increase in supply costs in November. In December both inputs and outputs exhibit gains and therefore TFP increased greatly from November to December. Note that RES do not seem to contribute much to productivity change as on average values are very close to 1. This means that on average TFP is not much affected by the exploitation of scale economies.

Our results so far focus on two successive months at a time giving that a short term view of progress or regress. It would be useful to complement this analysis with a longer period as noted for example by Berg et al. (1992). Such an analysis will use a fixed base period and ascertain progress or regress from that point in time on. Another reason for choosing a base period approach relates to the use of monthly data, and to the fact that THCH requires, in principle, longer periods than a month to take place.

Using March as the base period[4] the results of our GDF procedure are shown in Table 6.

[4] March was chosen as the base period because it is the first month of analysis. Any other month could have been chosen as the base period. Since our intention here is just to compare a base period approach to the standard chain approach, the period that is taken as a base is not crucial for our conclusions. Therefore we did not perform any sensitivity analysis regarding the impact of the base period chosen in the conclusions reached.

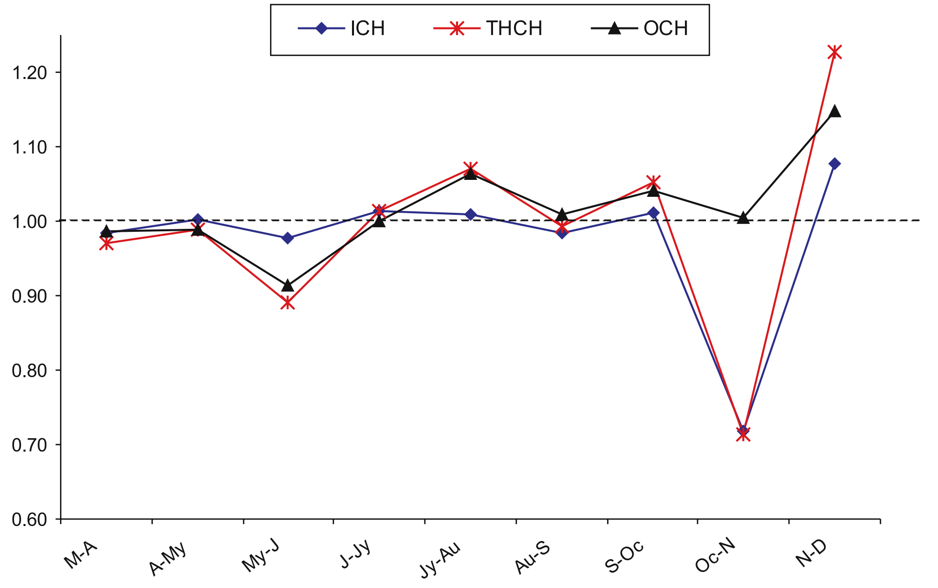


Fig. 3 Technological change and its components

Table 6 Malmquist GDF index results with base period

Period	GDF_t	GDF_{t+1}	$MGDF_V$	EFCH	THCH	ICH	OCH	RES	TFP
M–A	0.6950	0.7380	1.0741	1.1415	0.9882	0.9822	1.0152	0.9915	1.0193
M–My	0.6950	0.7063	0.9916	1.0398	0.9682	0.9705	1.0078	1.0703	1.0130
M–J	0.6950	0.7852	0.9598	1.2209	0.8403	0.9489	0.8947	1.0867	1.0050
M–Jy	0.7037	0.7894	0.9891	1.1963	0.8797	0.9646	0.9174	1.0669	1.0176
M–Au	0.7037	0.7823	1.1369	1.1696	1.0132	0.9833	1.0384	0.9944	1.1024
M–S	0.7037	0.7338	0.9865	1.0641	0.9343	0.9531	0.9929	1.0723	1.0176
M–Oc	0.7929	0.8223	0.9897	1.0680	0.9391	0.9566	0.9954	1.0637	1.0322
M–N	0.7929	0.7599	0.6481	0.9854	0.6653	0.6397	1.0502	1.0493	0.6642
M–D	0.7929	0.7443	0.7777	0.9592	0.8193	0.7228	1.1514	1.0418	0.7891

Results from a base period analysis provide new insights concerning changes in the various components of the TFP index. Figure 4 is an aid to understanding the values in Table 6.

EFCH (see left graph in Fig. 4) is mostly decreasing except in June, but values are still above one in most months, except in the last two months (November and December). This means that efficiency is higher than in March for most months but it is decreasing and in the last two months average efficiency falls below that in March. These conclusions are not particularly different from those obtained before, meaning that bank branches are not on average catching up with frontier movements especially after June.

In terms of THCH (see right graph in Fig. 4) the pattern is very similar to that shown when we did not use a fixed base period, but values are almost always below one. This means that between March and December we mostly have technological regress the only exception being the month of August. In Fig. 4 we can see that technological regress is mostly due to input deterioration (increase) rather than output fall. In fact OCH only falls well below 1 in June and July, showing a marked increasing trend after October. From both, the fixed base and ‘moving base’ approaches, we can conclude that the best month in terms of technological progress was August (and not December as we might have thought from the moving base approach) and that the worst month in terms of technological progress was November. In August technological progress is mostly due to progress in outputs, while in November regress is mostly due to regress in inputs. One possible reason

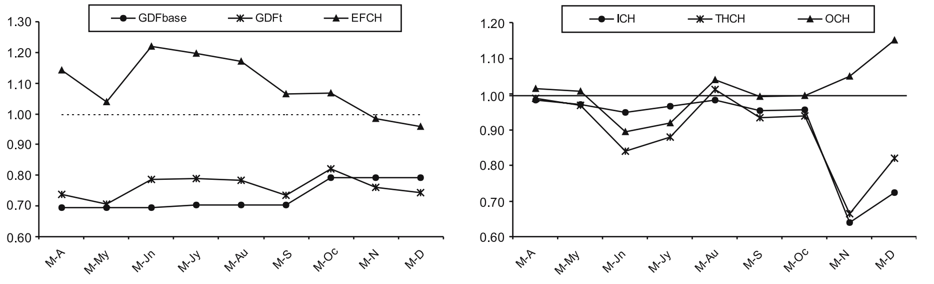


Fig. 4 Technological, and EFCH

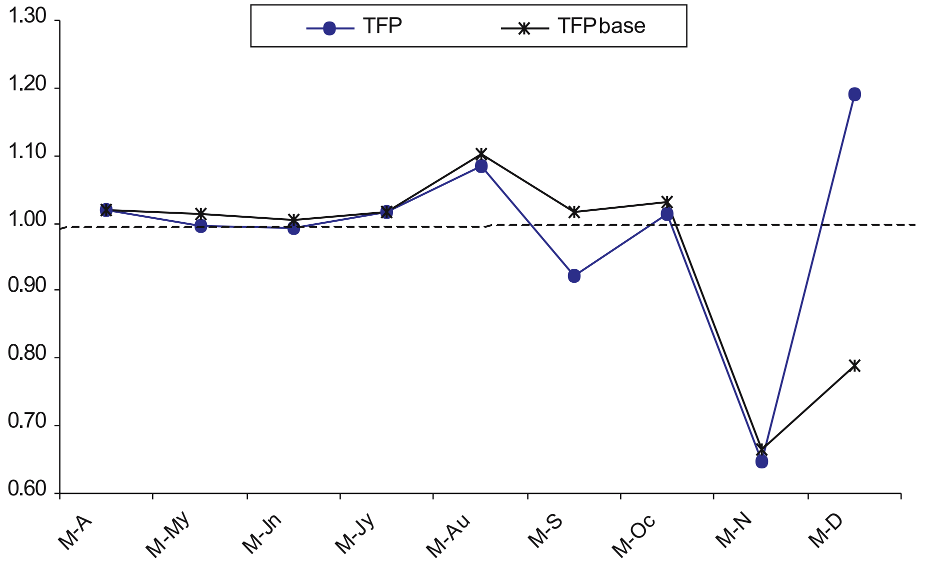


Fig. 5 TFP measures

for August being a good month in terms of technological progress seems to be related to the big flow of emigrants holidaying in Portugal in this month, which raises the business volume of bank branches on emigrant accounts.

So far as TFP is concerned it is mostly around 1 with a big decrease in November that is partly recovered in December, though productivity in December is still well below the productivity levels shown in March. Again the RES does not seem to contribute much to TFP change. Comparing the levels of TFP when we use the fixed and the moving base approaches we can see, in Fig. 5, that the results are very similar except for the recovery in December which is not so marked in the fixed base approach.

In fact, in the moving base approach the December values look good only because they are compared with the worst situation (namely November). In summary, from both approaches, one can say that the best month is August and the worst is November. The productivity recovery in December is noticeable when compared to November levels but still very far from the levels with which bank branches started with in March.

TFP shows, therefore, stability before July 2001 but an oscillatory behaviour after July. In view of the inputs and outputs used in the computation of TFP, this oscillatory behaviour basically means that change in outputs (namely the value of debit and credit accounts) is not being followed by a proportional change in inputs (operational costs). In November this fact is particularly noticeable, since the increase in operational costs (mainly due to the increase in supply costs) was not followed by a similar improvement of outputs, meaning that productivity levels decreased considerably in relation to initial levels observed in March.

Comparison of the GDF approach with earlier existing approaches

In order to ascertain the degree to which the GDF approach provides different results from earlier existing approaches we calculated Malmquist indexes for our banking data this time using the FGNZ and RD approaches noted earlier. The various components of TFP change for some chosen units (for the period of March to April 2001) and based on all three approaches are shown in Table 7.

Differences between the FGNZ and the RD approaches occur only for the THCH and scale change components. For most branches (see Table 7) the values are very close but for branches B1 and B43 differences are substantial. Note that in some cases we could not compute the THCH component for the RD approach due to infeasible models when assessing observations in one period in respect to a VRS frontier of a different period. This problem is not very important in most cases except for the computation of THCH from October to November, where the assessment of branches in October against the November frontier through VRS DEA models is for most branches infeasible (47 out of 52 bank branches could not be assessed). Considering comparisons with the GDF approach we have cases where TFP is not very different but components are (e.g. B18 shows a similar TFP under the GDF and the FGNZ/RD approaches, but whereas the FGNZ/RD indicate that the TFP decrease is mostly due to EFCH the GDF assigns most of this decrease to residual components). Further, there are cases where the TFP shows different directions of change. For example B1 and B21 exhibit a decline in TFP under the FGNZ/RD approaches but an improvement in TFP under the GDF approach. In other cases the direction is the same but values may vary slightly. For example B43 shows an improvement in TFP under all approaches, and the direction of all the components is the same for the RD and GDF approaches though there are some differences in the values.

If we compute partial productivity ratios, as we did in Section "Problems with traditional ways of calculating TFP", values are as shown in Table 8 for units B1 and B21. These partial productivity ratios show a clear productivity growth for branch B21 (since all the ratios are above 1 except the productivity ratio of current accounts per staff), and they also seem to indicate a productivity growth for branch B1. For this branch there are more partial productivity ratios above unity than below unity, and the growth observed in the ratio of credit associates per staff and supply costs seems to compensate for the decline observed in the ratios that fell below unity. These results seem to indicate that the GDF approach reflects reality better than the FGNZ or the RD approaches that attributed a TFP decline to branches B1 and B21. Note that the GDF TFP change is in fact the geometric mean of the ratios shown in Table 8.

Considering differences between the average values computed through the GDF and the FGNZ and RD approaches these are not very marked as can be seen from the graphs in Fig. 6.

In these graphs it is clear that the average behaviour of TFP is coincident for all approaches, the same happening for its THCH component. The main differences concern the EFCH component and the residual/scale component, although even in these cases movements happen mostly on the same direction, though with different magnitudes. What this comparisons seem to suggest is that when one is interested in analysing overall changes in productivity and its components the

Table 7 Comparison of GDF approach with standard approaches for 3 units

Approach	EFCH	THCH	RES	TFP
B1 FGNZ	1.0735	1.0033	0.9229	0.994
B1 RD	1.0735	0.8801	1.052	0.994
B1 GDF	1.2933	0.7453	1.0934	1.0538
B18 FGNZ	0.7905	1.0192	1.0097	0.8135
B18 RD	0.7905	1.0129	1.016	0.8135
B18 GDF	1.059	1.0669	0.7797	0.881
B21 FGNZ	0.9339	1.0383	0.9947	0.9645
B21 RD	0.9339	1.0333	0.9994	0.9645
B21 GDF	0.7219	1.1229	1.2842	1.0411
B43 FGNZ	3.6601	1.0018	0.3629	1.3307
B43 RD	3.6601	0.7915	0.4593	1.3307
B43 GDF	4.8286	0.5702	0.4357	1.1997

Table 8 Partial productivity ratios for units B1 and B21

	Curracc	Othress	Credb	Credass
$\Delta \frac{y}{\text{staff}}$ B1	1.006037	1.01952	1.00901	1.28180
$\Delta \frac{y}{\text{Suppcost}}$ B1	0.969989	0.982987	0.972852	1.235870
$\Delta \frac{y}{\text{staff}}$ B21	0.95735	1.045611	1.003646	1.029835
$\Delta \frac{y}{\text{Suppcost}}$ B21	1.020077	1.114121	1.06941	1.097311

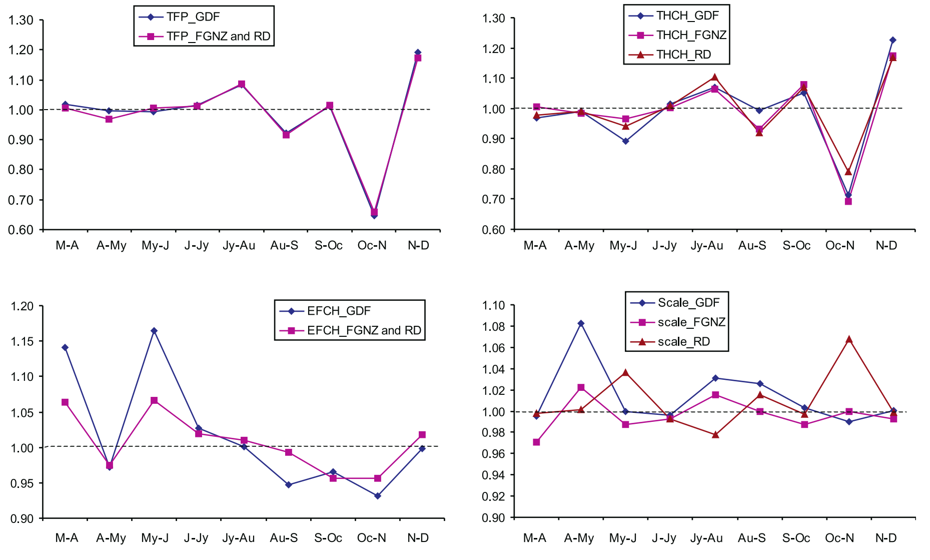


Fig. 6 TFP and its components—comparison between approaches

approach one chooses to calculate TFP and to decompose it is not of utmost importance, since on average they tend to agree on the main findings. However, when one is interested in analysing productivity change of individual units and understand the sources of this productivity change, then the approach one chooses to use clearly can make a difference in the conclusions reached and in the courses of action one decides to take. Therefore, the GDF approach is clearly a contribution for assessing productivity change of individual units and to analyse the sources of this change given its substantial advantages over other approaches as it was highlighted throughout this paper.

Conclusion

This paper draws attention for some limitations of current approaches to calculate Malmquist indexes, and attempts to resolve them through the use of a GDF approach. The GDF is used here with two purposes. (i) To calculate a TFP measure based on observed values only, and (ii) to calculate measures of technical efficiency that are non-oriented and account for all the sources of inefficiency. The latter use of the GDF solves the problem of infeasibility of some DEA models when VRS technologies are used, and resolves the ambiguity resulting from the use of oriented models that yield conflicting information depending on whether the models used are input or output oriented. The former use of the GDF to calculate TFP is consistent with the single input/output case, where it is widely accepted that a ratio of productivity at two different points in time reflects productivity change. Such ratios are based on observed values only, and do not require any assumptions regarding the form of the production frontier. In the multiple input/output case the most usual procedure to calculate TFP change is through Malmquist indexes that use efficiency measures calculated in relation to CRS technologies. We demonstrate through an example how this procedure may provide biased results due to changes in the reference hyperplane used when the unit's data changes over time.

Since in our approach a reference technology is used to decompose the TFP index (but not to compute the TFP index itself), one could argue that the THCH, EFCH, and RES components could suffer bias due to projection on different facets of a frontier as the data of the unit changes over time. This issue is not, however, a problem because the interest in computing these components is in fact to analyse differences in reference boundaries over time. Indeed, the definition of EFCH and THCH (see (6)) implies a comparison of a given observation against different time period frontiers in order to ascertain differences between these frontiers. So the issue of the facet being different is not important since this will generally be so unless the frontiers in time t and $t+1$ are exactly the same. However, the definition of the Malmquist TFP index itself (rather than its components–see (5)) implies a comparison of two observations (of the same unit in different time periods) against the same time period frontier, and therefore when different facets of the same boundary are used in this comparison the index itself can be biased.

We have applied our GDF approach to a sample of Portuguese bank branches. Both a fixed and a moving base period were used to analyse productivity change of this sample of bank branches between March and December of 2001. A comparison between the GDF approach and standard approaches is also provided in this paper, showing that there is good agreement between the approaches at the general level over the collection of branches but there can be differences at unit level. We demonstrate that in certain cases the GDF approach reflects better the underlying changes in productivity as they can be deduced using partial productivity ratios.

Appendix A

According to Färe et al. (1988, 2001) technical change is the product of input biased technical change (IBTECH), output biased technical change (OBTECH) and magnitude of technical change (MATECH), all defined in (A.1).

$$\begin{aligned}
\text{IBTECH} &= \left(\frac{D^{t+1}(x_t, y_t)}{D^t(x_t, y_t)} \times \frac{D^t(x_{t+1}, y_t)}{D^{t+1}(x_{t+1}, y_t)}\right)^{\frac{1}{2}} \\
\text{OBTECH} &= \left(\frac{D^t(x_{t+1}, y_{t+1})}{D^{t+1}(x_{t+1}, y_{t+1})} \times \frac{D^{t+1}(x_{t+1}, y_t)}{D^t(x_{t+1}, y_t)}\right)^{\frac{1}{2}} \\
\text{MATECH} &= \frac{D^t(x_t, y_t)}{D^{t+1}(x_t, y_t)}
\end{aligned} \tag{A.1}$$

Consider that the distance functions in (A.1) are CRS input oriented and can be expressed as $D^t(x_t, y_t) = \frac{x_t^{*t}/x_t}{y_t^{*t}/y_t}$ for the single input/output case (note that Färe et al. (2001) defined the components in (A.1) for output oriented measures). We can, therefore, write the above (A.1) equivalently as (A.2).

$$\begin{aligned}
\text{IBTECH} &= \left(\frac{x_t^{*t+1}}{x_t^{*t}} \times \frac{x_{t+1}^{*t}}{x_{t+1}^{*t+1}}\right)^{\frac{1}{2}} \\
\text{OBTECH} &= \left(\frac{y_{t+1}^{*t+1}}{y_{t+1}^{*t}} \times \frac{y_t^{*t}}{y_t^{*t+1}}\right)^{\frac{1}{2}} \\
\text{MATECH} &= \frac{x_t^{*t}/y_t^{*t}}{x_t^{*t+1}/y_t^{*t+1}}
\end{aligned} \tag{A.2}$$

Note that the IBTECH and the OBTECH are very similar to our input and output change components except that in each case one of the ratios is inverted. This inversion results from the existence of the magnitude of technical change component. In fact when multiplying all the three components in (A.2) we get (A.3), which is equivalent to our product of input change and output change components of technological change.

$$\begin{aligned}
&\left(\left(\frac{x_t^{*t+1}}{x_t^{*t}} \times \frac{x_{t+1}^{*t}}{x_{t+1}^{*t+1}}\right) \times \left(\frac{y_{t+1}^{*t+1}}{y_{t+1}^{*t}} \times \frac{y_t^{*t}}{y_t^{*t+1}}\right) \right.\\
&\left. \times \left(\frac{x_t^{*t}/y_t^{*t}}{x_t^{*t+1}/y_t^{*t+1}}\right)^2\right)^{\frac{1}{2}} \Leftrightarrow \left(\left(\frac{x_{t+1}^{*t}}{x_{t+1}^{*t+1}} \times \frac{x_t^{*t}}{x_t^{*t+1}}\right)\right. \\
&\left. \times \left(\frac{y_{t+1}^{*t+1}}{y_{t+1}^{*t}} \times \frac{y_t^{*t+1}}{y_t^{*t}}\right)\right)^{\frac{1}{2}}
\end{aligned} \tag{A.3}$$

Appendix B—Equivalence between the GDF and the RD Approaches

The equivalence between the residual component in (12) and the RD scale component is proved next for the single input/output case. When input oriented efficiency measures are calculated both in relation to a CRS (c) technology and in relation to a VRS (v) technology, then the scale component of the RD approach is defined by (B.1), where $\gamma_t^{t(r)} = x_t^{*t(r)}/x_t$, being $x_t^{*t(r)}$ the target input in frontier r that can be either c or v.

$$\begin{aligned}
&Scale_{RD} \\
&= \left(\frac{\gamma_{t+1}^{t(c)}/\gamma_{t+1}^{t(v)}}{\gamma_t^{t(c)}/\gamma_t^{t(v)}} \times \frac{\gamma_{t+1}^{t+1(c)}/\gamma_{t+1}^{t+1(v)}}{\gamma_t^{t+1(c)}/\gamma_t^{t+1(v)}}\right)^{1/2} \\
&= \left(\frac{x_{t+1}^{*t(c)}/x_{t+1}^{*t(v)}}{x_t^{*t(c)}/x_t^{*t(v)}} \times \frac{x_{t+1}^{*t+1(c)}/x_{t+1}^{*t+1(v)}}{x_t^{*t+1(c)}/x_t^{*t+1(v)}}\right)^{1/2}
\end{aligned} \tag{B.1}$$

Consider the CRS technology defined by $y = bx$, then we can replace projections on c by $x_{t+1}^{*t(c)} = y_{t+1}/b_t$, $x_t^{*t(c)} = y_t/b_t$, $x_{t+1}^{*t+1(c)} = y_{t+1}/b_{t+1}$, and $x_t^{*t+1(c)} = y_t/b_{t+1}$, which results in (B.1) being equivalent to (B.2).

$$Scale_{RD} = \left(\frac{(y_{t+1}/b_t)/x_{t+1}^{*t(v)}}{(y_t/b_t)/x_t^{*t(v)}} \times \frac{(y_{t+1}/b_{t+1})/x_{t+1}^{*t+1(v)}}{(y_t/b_{t+1})/x_t^{*t+1(v)}} \right)^{1/2} = \left(\frac{\frac{x_t^{*t(v)}}{x_{t+1}^{*t(v)}}}{\frac{y_t}{y_{t+1}}} \times \frac{\frac{x_t^{*t+1(v)}}{x_{t+1}^{*t+1(v)}}}{\frac{y_t}{y_{t+1}}} \right)^{\frac{1}{2}} \quad \text{(B.2)}$$

The latter expression in (B.2) is exactly equivalent to the scale component in (12) since in input oriented measures target outputs are equal to observed outputs. The GDF residual component is, therefore, equal to the scale component of the RD approach, having the advantage of not requiring the computation of efficiency scores in relation to a CRS technology. The GDF approach uses, instead of projections on the CRS frontier, output relationships on the VRS frontier to account for scale effects.

Appendix C—Measuring technical efficiency

The GDF efficiency measure in (1) was designed to be used *a posteriori* after targets have been computed. It can, nevertheless, be used as the objective function of a DEA model where all inputs and outputs are allowed to change by different proportions. In our implementation we simplified the GDF measure in the objective function of the DEA model used, which is shown in (C.1).

$$\text{EFF}_o = \min \left\{ \frac{\theta}{\beta} \mid \sum_{j \in E} \lambda_j \, y_{rj} \geq \beta y_{ro} \ (a), \quad \sum_{j \in E} \lambda_j \, x_{ij} \leq \theta x_{io} \ (b), \quad \sum_{j \in E} \lambda_j = 1 \ (c), \quad \lambda_j, \ \geq \ 0 \ (d), \quad [0 \leq \theta \leq 1, \text{ and } \beta \geq 1] \ (e) \text{ or } [\theta \geq 1, \text{ and } 0 \leq \beta \leq 1] \ (f) \right\} \quad \text{(C.1)}$$

We use in (C.1) an equiproportional factor associated with expanding outputs, and a different equiproportional factor associated with contracting inputs. Though the objective function of (C.1) provides an efficiency score, we do not use it as the final efficiency measure. Instead Pareto-efficient targets resulting from model (C.1) are used for calculating the GDF technical efficiency measure for each unit. We assure that Pareto-efficient targets result from the linear combination of the λs in (C.1) by restricting the reference set to Pareto-efficient units (units in the set E).

The last set of constraints (e and f) in (C.1) ensure that the right direction (expansion or contraction) is followed by inputs and outputs. When units are assessed in relation to a frontier containing observations of the same time period, only constraints (e) are activated. When the frontier relates to a different time period then two things may happen for an observation: either it lies below the frontier (and then constraints (e) are activated) or it lies above the frontier (and constraints (f) are activated). This was easily programmed using GAMS.

References

Athanassopoulos AD (1997) Service quality and operating efficiency synergies for management control in the provision of financial services: evidence from Greek bank branches. Eur J Operat Res 98:300–313

Balk BM (2001) Scale efficiency and productivity change. J Prod Anal 15(3):159–183

Berg SA, Førsund, FR Jansen, ES (1992) Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980–89. Scand J Econ 94:S211–S228

Berger AN, Leusner JH, Mingoi JJ (1997), The efficiency of bank branches. J Monet Econ 40(1):141–162

Bjurek H (1996) The Malmquist total factor productivity index. Scand J Econ 98(2):303–314

Caves D, Chistensen LR, Diewert W (1982) The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica 50(6):1393–1414

Chambers RG, Färe R, Grosskopf S (1996) Productivity growth in APEC countries. Pacific Econ Rev 1(3):181–190

Charnes A, Cooper WW, Rhodes E (1978) Measuring efficiency of decision making units. Eur J Operat Res 2:429–444.

Chung Y, Färe R, Grosskopf S (1997), Productivity and undesirable outputs: a 30 directional distance function approach. J Environ Manage 51(3):229–240

Diewert ER, Nakamura AO, (2003) Index number concepts, measures and decompositions of productivity growth. J Product Anal 19(2/3):127–159

Farrell MJ (1957) The measurement of productive efficiency. J Roy Stat Soc Series A, general 120 (part 3):253–281

Färe R, Grifell-Tatje E, Grosskopf S, Lovell CAK (1997a) Biased technical change and malmquist productivity index. Scand J Econ 99(1):119–128

Färe R, Grosskopf S, Lee W-F (2001) Productivity and technical change: the case of Taiwan. Appl Econ 33(15):1911–1925

Färe R, Grosskopf S, Norris M, (1997b), Productivity growth, technical progress and efficiency changes in industrialized countries: reply'. Am Econ Rev 87(5): 1040–1043

Färe R, Grosskopf S, Norris M, Zhang, Z. (1994), Productivity growth, technical progress and efficiency changes in industrialised countries. Am Econ Rev 84:66–83

Färe R, Grosskopf S, Roos P (1998) Malmquist productivity indexes: a survey of theory and practice. In: Färe R, Grosskopf S, Russell RR (eds) Index numbers: essays in honour of Sten Malmquist. Kluwer Academic Publishers, pp. 127–190

Førsund FR (1998) The rise and fall of slacks: comments on quasi-Malmquist productivity indices. J Prod Anal 10:21–34

Grifell-Tatjé, E, Lovell CAK, Pastor JT, (1998), A quasi-Malmquist productivity index. J Prod Anal 10:7–20

Lovell CAK (2001) The decomposition of Malmquist productivity indexes. Paper presented at the 7th European Workshop on Efficiency and Productivity Analysis, Oviedo, Spain, 25–29 September

Oral M, Yolalan R (1990) An Empirical Study on measuring operating efficiency and profitability of bank branches. Eur J Oper Res 46(3):282–294

Portela MCS, Thanassoulis E (2002), Profit efficiency in DEA. Aston Business School Research Paper RP0206, ISBN 185449502X, University of Aston, Aston Triangle, Birmingham B4 7ET, UK

Portela MCS, Thanassoulis E (2005) Profitability of a sample of Portuguese bank branches and its decomposition into technical and allocative components. Eur J Oper Res 162(3): 850–866

Py B (1990) Statistique Descriptive, 3e Édition. Economica, Paris

Ray SC (2001) On an extended decomposition of the Malmquist productivity index. Paper presented at the 7th European Workshop on Efficiency and Productivity Analysis, Oviedo, Spain, September 25–29

Ray SC, Desli E (1997), Productivity growth, technical progress and efficiency changes in industrialized countries: comment. Am Econ Rev 87(5):1033–1039

Thrall RM (2000) Measures in DEA with an application to the Malmquist index. J Prod Anal 13(2):125–137

Zofio JL, Lovell CAK (2001) Graph efficiency and productivity measures: an application to U.S agriculture. Appl Econ 33(11):1433–1442