# Appendix A: Variations of the Malmquist Index

This appendix provides a more detailed formal presentation of the methods to compute the productivity change. Specifically, the definition of the Malmquist index is based on the technology set at a particular period in time $t$,

$$\mathcal{T}_t = \left\{(\boldsymbol{x}_t, \boldsymbol{y}_t) \in \mathbb{R}^{p+q} \,:\, \boldsymbol{x}_t \geq \mathbf{0} \text{ can produce } \boldsymbol{y}_t \geq \mathbf{0}\right\}, \tag{A1}$$

with $\boldsymbol{x}_t \geq \mathbf{0}$ as the vector of $p$ inputs used to produce the vector of $q$ outputs $\boldsymbol{y}_t \geq \mathbf{0}$ of a particular country in period $t$. For later reference this technology set is referred to as the contemporaneous technology set. This technology set is the basis to define the output-oriented distance function (Shephard (1970)) of the input-output combination of a country in period $s$ to the frontier function in period $t$ by

$$D_t(\boldsymbol{x}_s, \boldsymbol{y}_s) = \inf\{\theta \,:\, (\boldsymbol{x}_s, \boldsymbol{y}_s/\theta) \in \mathcal{T}_t\}. \tag{A2}$$

This distance function measures the reciprocal of the maximum enhancement of the outputs of period $s$ needed so that the input-output combination $(\boldsymbol{x}_s, \boldsymbol{y}_s/\theta)$ remains in the technology set of period $t$, $\mathcal{T}_t$. In the case of the same-period distances (with $s = t$) the largest possible value of the distance function is equal to one for input-output combinations on the boundary of the technology set (the frontier function). For input-output combinations below the frontier function the distance function is smaller than one, implying that the output vector can be enhanced by the largest factor $1/\theta$ (smallest $\theta$) to reach the frontier function. Taking the reciprocal is required to obtain an efficiency measure with the largest value normalized to one for the efficient countries on the frontier and values lower than one for the inefficient countries below the frontier. In the case of the mixed-period distance functions (with $s \neq t$) values of the distance function larger than one may arise when the input-output combination of a country in period $s$ is located outside of the technology set of a different period $t$.

For the actual computation with real data there exist linear programming formulations related to the distance functions. Let $\boldsymbol{x}_{it}$ and $\boldsymbol{y}_{it}$ denote the input and output data, respectively, of country $i = 1, ..., n$ at time $t = 1, ..., T$. Correspondingly, $\boldsymbol{X}_t = (\boldsymbol{x}_{1t} \vdots ... \vdots \boldsymbol{x}_{nt})$ is the $p \times n$ matrix containing the data for the $p$ inputs of all $n$ countries at time $t$ and $\boldsymbol{Y}_t = (\boldsymbol{y}_{1t} \vdots ... \vdots \boldsymbol{y}_{nt})$ is the $q \times n$ matrix containing the data for the $q$ outputs for the entire country sample at time $t$ (here '$\vdots$' indicates side-by-side stacking of the vectors). Under the assumption of constant returns to scale (CRS) the distance function can be computed by solving the DEA linear programming problem

$$D_t^c(\boldsymbol{x}_{is}, \boldsymbol{y}_{is}) = \min\{\theta \,:\, \boldsymbol{X}_t\boldsymbol{\lambda} \leq \boldsymbol{x}_{is},\, \boldsymbol{Y}_t\boldsymbol{\lambda} \geq \boldsymbol{y}_{is}/\theta,\, \boldsymbol{\lambda} \geq \mathbf{0}\} \tag{A3}$$

for country $i \in \{1, ..., n\}$ (Charnes et al. (1978)). The values in the $n$-vector $\boldsymbol{\lambda}$ are computed jointly with $\theta$ as the solution of the linear programming problem and give the weights of the sample countries for constructing the projection point of the observation $(\boldsymbol{x}_s, \boldsymbol{y}_s)$ on the frontier function of period $t$. This leads to a frontier function constructed as a piece-wise linear envelopment of the input-output combinations in the sample. Therefore, the method does not rely on a particular functional form such as a production function, giving the method its nonparametric character.

## Basic Malmquist Index

Based on the four distance functions $D_t^c(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})$, $D_{t+1}^c(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})$, $D_{t+1}^c(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})$ and $D_t^c(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})$, the Malmquist index $MI$ for the change of total factor productivity between periods $t$ and $t+1$ under CRS can be stated for country $i \in \{1, ..., n\}$ as

$$MI_{t,t+1} = \left[ \frac{D_t^c(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})}{D_t^c(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})} \times \frac{D_{t+1}^c(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})}{D_{t+1}^c(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})} \right]^{1/2}. \tag{A4}$$

This index is defined as the geometric mean of two ratios of distance functions for the input-output combinations of the periods $t$ and $t+1$ with respect to the frontier function of period $t$ in the case of the first ratio and with respect to the frontier function of period $t+1$ in the case of the second ratio. The first ratio is larger (smaller) than one if the input-output combination of a country in period $t+1$ is closer to (farther from) the frontier of period $t$ than it was in period $t$. With technological progress input-output combinations may be positioned above that frontier. The second ratio makes the same comparison with respect to the frontier function of period $t+1$. Since there is no reason to prefer the frontier function of period $t$ or that of period $t+1$ as the benchmark for the measurement the geometric mean of both is taken.

Defining the Malmquist index in this way has the advantage that the productivity change can be decomposed into two meaningful factors as shown by Färe at al. (1994), i.e.

$$\begin{aligned} MI_{t,t+1} &= \frac{D_{t+1}^c(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})}{D_t^c(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})} \times \left[ \frac{D_t^c(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})}{D_{t+1}^c(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})} \frac{D_t^c(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})}{D_{t+1}^c(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})} \right]^{1/2} \\ &= EC_{t,t+1} \times TC_{t,t+1}. \end{aligned} \tag{A5}$$

The first factor is the efficiency change component $EC$ which is larger (smaller) than one if the input-output combination of the country in period $t+1$ is closer to (farther from) the frontier function of this period than it was in period $t$ with respect to the frontier function of that period. This component can be interpreted as indicating catching up or falling behind of a country with respect to the frontier function between two periods $t$ and $t+1$. The second factor is the technology change component $TC$ which is larger (smaller) than one if the part of the piece-wise linear frontier function pertaining to the country under evaluation is shifting forwards (backwards).[1]

As defined above, the Malmquist index and its components are all measured under CRS. Wheelock and Wilson (1999), among others, provide a refinement of this decomposition which takes account of the possibility that the true technology may exhibit variable returns to scale (VRS). This refinement permits a four-factor decomposition of the Malmquist index with further decomposing the efficiency change and technology change components under CRS into efficiency change and technology change component under VRS and the respective scale change components. This requires the computation of the distance functions also under a VRS technology which can be achieved by solving the modified linear programming problem

$$D_t^v(\boldsymbol{x}_{is}, \boldsymbol{y}_{is}) = \min\{\theta \,:\, \boldsymbol{X}_t\boldsymbol{\lambda} \leq \boldsymbol{x}_{is},\, \boldsymbol{Y}_t\boldsymbol{\lambda} \geq \boldsymbol{y}_{is}/\theta,\, \boldsymbol{1}'\boldsymbol{\lambda} = 1,\, \boldsymbol{\lambda} \geq \boldsymbol{0}\} \tag{A6}$$

[1]See Färe et al. (1994, p. 70) for a graphical illustration.

for country $i \in \{1, ..., n\}$ (Banker et al. (1984)) with the additional constraint $\mathbf{1}'\boldsymbol{\lambda} = 1$ which forces the weights to add up to unity.

Following Wheelock and Wilson (1999), the Malmquist index can be decomposed as

$$\begin{aligned} MI_{t,t+1} &= \frac{D^v_{t+1}(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})}{D^v_t(\boldsymbol{x}_t, \boldsymbol{y}_t)} \times \left[ \frac{D^v_t(\boldsymbol{x}_t, \boldsymbol{y}_t)}{D^v_{t+1}(\boldsymbol{x}_t, \boldsymbol{y}_t)} \frac{D^v_t(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})}{D^v_{t+1}(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})} \right]^{1/2} \\ & \times \frac{D^c_{t+1}(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})/D^v_{t+1}(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})}{D^c_t(\boldsymbol{x}_t, \boldsymbol{y}_t)/D^v_t(\boldsymbol{x}_t, \boldsymbol{y}_t)} \\ & \times \left[ \frac{D^c_t(\boldsymbol{x}_t, \boldsymbol{y}_t)/D^v_t(\boldsymbol{x}_t, \boldsymbol{y}_t)}{D^c_{t+1}(\boldsymbol{x}_t, \boldsymbol{y}_t)/D^v_{t+1}(\boldsymbol{x}_t, \boldsymbol{y}_t)} \frac{D^c_t(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})/D^v_t(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})}{D^c_{t+1}(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})/D^v_{t+1}(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})} \right]^{1/2} \\ &= PEC \times PTC \times SEC \times STC. \end{aligned} \tag{A7}$$

Therein, the first component ($PEC$) represents the pure efficiency change and is interpreted analogously to $EC$ but now with respect to the VRS technology instead of the CRS technology. The second component ($PTC$) represents pure technology change component and is interpreted analogously to $TC$ but now also with respect to the VRS technology. The third and fourth components ($SEC$ and $STC$) are the scale-related components scale efficiency change and scale technology change.[2]

While the basic Malmquist index under CRS has the disadvantage that the CRS technology is more restrictive and therefore less realistic, the variant of Wheelock and Wilson has two other deficiencies. First, the solvability of the mixed-period linear programming problems (with $t \neq s$) is guaranteed only under CRS, they need not have a solution under VRS. This infeasibility problem always affects the technology change component under VRS and therefore the whole productivity index for a subset of the sample countries. Excluding these countries from the sample provides no resolution since other countries would then be affected by this problem. It is not much recognized that this inevitably leads to a sample-selection problem since the drop-outs from the sample are not random but systematically related to their position in the input-output space. Second, the results of the scale-related components arising in the four-factor decomposition (scale efficiency change and scale technology change) are difficult to interpret as the related discussion in Wheelock and Wilson (1999) shows.

For the subsequent empirical application we build a Malmquist index under VRS by just putting together the pure efficiency change and pure technology change components resulting the following basic variant

$$MI^v_{t,t+1} = \frac{D^v_{t+1}(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})}{D^v_t(\boldsymbol{x}_t, \boldsymbol{y}_t)} \times \left[ \frac{D^v_t(\boldsymbol{x}_t, \boldsymbol{y}_t)}{D^v_{t+1}(\boldsymbol{x}_t, \boldsymbol{y}_t)} \frac{D^v_t(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})}{D^v_{t+1}(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})} \right]^{1/2}. \tag{A8}$$

In this way we circumvent the difficult interpretation of the scale-related components. We accept the infeasibility problem for this index variant and take the opportunity to evaluate its importance in the context of the empirical application by a comparison with two other variations of the Malmquist index which are explained next.

[2] See Diewert and Fox (2017) and Zofio (2007) for overviews of alternative decompositions of the Malmquist index.

## Global and Biennial Malmquist Index

One practical way out of the infeasibility problem is the idea of the global Malmquist productivity index put forward by Pastor and Lovell (2005). This proposal relies on modified distance functions which are computed with respect to a "global" technology set in which the observations from all countries *and* all periods are pooled together.

Formally, the idea relies on two different specifications of the technology set. The first is the contemporaneous benchmark technology $\mathcal{T}_t$ consisting of the production possibilities in a particular period $t$ as defined above in (A1). This is the technology set we considered so far for defining the distance function (A2) which may be computed under CRS (A3) or VRS (A6). The second specification of the technology set is the global benchmark technology $\mathcal{T}_G$ proposed by Pastor and Lovell (2005). This is defined as the convex envelope of the union of all contemporaneous technology sets $\mathcal{T}_1 \cup ... \cup \mathcal{T}_T$ spanning the entire time period under investigation. The corresponding global distance function is defined as

$$D_G(\boldsymbol{x}_s, \boldsymbol{y}_s) = \inf\{\theta \,:\, (\boldsymbol{x}_s, \boldsymbol{y}_s/\theta) \in \mathcal{T}_G\}. \tag{A9}$$

and can be computed for a particular country $i$ by substituting the $p \times nT$ matrix $(\boldsymbol{X}_1 \vdots ... \vdots \boldsymbol{X}_T)$ for $\boldsymbol{X}_t$ and the $q \times nT$ matrix $(\boldsymbol{Y}_1 \vdots ... \vdots \boldsymbol{Y}_T)$ for $\boldsymbol{Y}_t$ in the linear programming problems (A3) for CRS or (A6) for VRS. Now, the dots '$\vdots$' simply mean stacking matrices side-by-side which implies a simple pooling of all observations from all periods.

Productivity change between periods $t$ and $t+1$ can be defined with respect to the global benchmark technology as the ratio of the global distance function evaluated at the observation of period $t+1$ and the global distance function evaluated at the observation of period $t$ (for some country $i \in \{1, ..., n\}$) leading to the global Malmquist index[3]

$$GMI_{t,t+1} = \frac{D_G(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})}{D_G(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})}. \tag{A10}$$

The index is larger (smaller) than one if the country $i$ is more (less) efficient in period $t+1$ than in period $t$, in both periods evaluated with respect to the global technology set. Since the index is based on the global technology, it avoids taking the geometric mean with two different base periods. This construction endows the index with the property of circularity as emphasized by Pastor and Lovell (2005), which is a desirable feature for index numbers in general and is not given for the basic Malmquist index. Moreover, all linear programming problems involved in the computation of the index and its components are guaranteed to be feasible also under variable returns to scale.

Like the basic Malmquist index discussed above the global Malmquist index can be decomposed into two components

[3]It is important to recognize that the distance functions have no indication of constant and variable returns to scale (the superscripts '$c$' and '$v$' are missing) since this index can always be computed under both constant and variable returns to scale.

$$GMI_{t,t+1} = \frac{D_{t+1}(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})}{D_t(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})} \times \frac{D_G(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})/D_{t+1}(\boldsymbol{x}_{i,t+1}, \boldsymbol{y}_{i,t+1})}{D_G(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})/D_t(\boldsymbol{x}_{it}, \boldsymbol{y}_{it})} \qquad \text{(A11)}$$
$$= EC_{t,t+1} \times BPC_{t,t+1}.$$

The first component of (A11) represents efficiency change $EC$ with the same interpretation as before. The second component depends on the so-called best-practice gap $D_G(\cdot)/D_t(\cdot)$ between the global and the contemporaneous technology frontier. Technological change is measured by the change of this best-practice gap between the two time periods $t$ and $t+1$. Thus, here $BPC > 1$ ($BPC < 1$) represents an improvement (deterioration) of the technology.

A potential problem with the global index, as emphasized by Afsharian and Ahn (2015), is that the global frontier may consist of observations from different time periods implying that convex combinations of observations across rather different time periods are feasible. This problem is mitigated by a variation of the global Malmquist index, the biennial Malmquist productivity index as proposed by Pastor et al. (2011). The biennial index provides a compromise of the basic and the global index which amounts to modify the specification by restricting the global technology to just the two periods $t$ and $t+1$ over which the productivity change is to be measured. This means that $\mathcal{T}_G$ is here the convex union only of $\mathcal{T}_t$ and $\mathcal{T}_{t+1}$ and the matrices for the inputs and the outputs used in the linear programming problems are constructed by just stacking the two single-period matrices side-by-side, i.e. substituting $(\boldsymbol{X}_t \vdots \boldsymbol{X}_{t+1})$ for $\boldsymbol{X}_t$ and $(\boldsymbol{Y}_t \vdots \boldsymbol{Y}_{t+1})$ for $\boldsymbol{Y}_t$ in (A3) or (A6). This is sufficient to guarantee the computational feasibility under both CRS and VRS. The circularity property of the global index is lost for the biennial variant, however.[4]

In the presentation of the results we indicate in the figure headings for which variant of the index (basic, global or biennial) the results are computed. For clarity and simplicity, the global index is just abbreviated by $MI$ and the components are abbreviated by $EC$ and $TC$ (with the latter actually meaning the $BPC$ component in (A11) in the case of the global and biennial index variants). All linear programs in the papers are computed under VRS.

## Additional References Not Cited in the Main Text

Afsharian, M. & Ahn, H. (2015). The overall Malmquist index: A new approach for measuring productivity changes over time. *Annals of Operations Research*, 226, 1–27.

Diewert, W.E. & Fox, K.J. (2017). Decomposing productivity indexes into explanatory factors. *European Journal of Operational Research*, 256, 275–291.

Shephard, R.W. (1970). *Theory of cost and production functions.* Princeton: Princeton University Press.

Wheelock, D.C. & Wilson, P.W. (1999). Technical progress, inefficiency, and productivity change in U.S. banking, 1984-1993. *Journal of Money, Credit, and Banking*, 31, 212–234.

[4] A major motivation for the biennial variant put forward by Pastor et al. (2011) was that this index can be more easily updated when new time periods arise whereas the global index must be recomputed anew for all periods. Regarding the very fast software solutions available for solving linear programming problems this aspect appears less crucial.

Zofio, J.L. (2007). Malmquist productivity index decompositions: A unifying framework. *Applied Economics*, 39, 2371–2387.