Bremen Energy Working Papers

Paperseries No. 29

Roland Meyer, Gert Brunekreeft, and George Elias

TOTEX Malmquist Index for RPI-X Regulation: Does it Correctly Estimate the True Frontier Shift?

July 2019

Jacobs University Bremen
Bremen Energy Research (BER)

Editors:

Prof. Dr. Gert Brunekreeft

Dr. Roland Meyer

Jacobs University Bremen

Bremen Energy Research (BER)

Campus Ring 1 / South Hall

28759 Bremen

www.jacobs-university.de/

https://bremen-energy-research.de/

JACOBS UNIVERSITY

Contact:

Dr. Roland Meyer

Tel. +49 (0) 421 - 200-4869

E-mail ro.meyer@jacobs-university.de

Suggested citing:
Meyer, R., Brunekreeft, G., and Elias, G. (2019). TOTEX Malmquist Index for RPI-X Regulation: Does it Correctly Estimate the True Frontier Shift?, *Bremen Energy Working Papers* No. 29, Jacobs University Bremen.

TOTEX Malmquist index for RPI-X regulation: does it correctly estimate the true frontier shift?

Roland Meyer[*], Gert Brunekreeft[**] and George Elias*[**]

[*]corresponding author
Jacobs University Bremen, Germany
ro.meyer@jacobs-university.de

Phone: +49(0)421 200 4869

[**]Jacobs University Bremen, Germany
[***] Polynomics, Olten, Switzerland

July 30, 2019

This paper is based on a study for Netze BW. The authors would like to thank Netze BW and especially Eva Deuchert for valuable contributions and comments. Furthermore, we are grateful for useful comments at the international IAEE conference 2018 in Groningen.

Abstract

The X in RPI-X regulation aims to adjust price or revenue allowances of regulated firms to changes in total factor productivity and input prices. If calculated correctly, both terms together correspond to the change in efficient costs which can be determined by applying a cost Malmquist index. However, regulators typically lack the required data on input quantities and prices. As an alternative, regulation authorities may apply a TOTEX Malmquist index to measure the total cost change. We study under which conditions this total cost change correctly estimates the true efficient cost change. We find that the TOTEX Malmquist index provides an undistorted estimate at least under two conditions, namely if (1) the frontier firms identified in the benchmarking procedure are *either* fully efficient, *or* if their degree of inefficiency remains constant over time, and (2) if input prices *either* stay constant *or* change by the same proportion for all firms.

JEL classification: L51, D24

Key words: *price regulation, Malmquist index, total factor productivity, data envelopment analysis*

1 Introduction

The basic formula of price-based monopoly regulation is known as RPI-X (Littlechild, 1983). This type of regulation aims to set efficiency incentives by decoupling output prices temporarily from actual cost levels of the regulated firms. During a predefined regulatory period, prices or revenues are adjusted yearly according to RPI-X. While the first term, the retail price index RPI, is published in official statistics, the critical task for the regulator is to determine X. But what is X and how can it be accurately measured?

RPI-X-regulation aims to mimic competition (Bernstein and Sappington 1999): the zero-profit condition requires that the change in regulated prices is equal to the change in input price minus the change in total factor productivity. The main important point of X is the estimate of the expected productivity increase: the frontier shift.[1]

A common regulatory practice to calculate the productivity changes relies on Törnqvist quantity indices (Törnqvist 1936; Coelli et al. 2005), while the input price changes are determined separately based on statistical price index data. The weakness of the Törnqvist index, however, is that it cannot distinguish between a *frontier shift* and *catch-up* effects. The former reflects the change in productivity, and the latter captures changes in technical or allocative efficiency of individual firms as compared to the efficiency frontier. If some firms have increased their efficiency over time – which RPI-X aims to achieve – the Törnqvist index will overestimate the past productivity change and thereby impose an excessive productivity target for the near future.

As an alternative, the Malmquist index (Malmquist 1953; Coelli et al. 2005) addresses precisely this distinction between frontier shift and catch-up effects by applying benchmarking techniques like data envelopment analysis. The Malmquist index is a well-established and widely used method for productivity analysis but is hardly used for monopoly regulation[2]. Literature gets completely silent in case regulators do not have data on input prices and quantities but only total cost data (TOTEX).

This paper examines whether the Malmquist index can be usefully applied in setting the X-factor in RPI-X regulation of monopoly. The focus will be on a *TOTEX Malmquist index*, which can be applied if the regulator only has TOTEX information. In this case, neither the traditional *production* Malmquist index (Malmquist 1953), nor the *cost* Malmquist index (Maniadakis and Thanassoulis 2004) are

[1] To be more precise, in RPI-X-regulation, regulators use the general rate of inflation (RPI), instead of the sector specific difference between input price change and change in total factor productivity. Accordingly, X is a correction term for RPI that captures the deviations between the sectoral and overall developments. As we will explain further below, X consists of four terms.

[2] The energy regulator in Germany, Bundesnetzagentur (BNetzA), has started to apply the Malmquist Index for the regulation of the energy networks in 2018 and 2019 for gas and electricity distribution and transmission network operators, respectively.

applicable. This paper derives the TOTEX Malmquist index[3] and analyses its theoretical accuracy in determining the productivity and input price change in combination. Which information on the frontier shift gets lost, if we do not have input price and quantity data but only TOTEX? We identify two sets of assumptions under which no biases occur. The first case addresses technical and allocative inefficiency. The TOTEX Malmquist index is undistorted if the efficiency frontier is set by firms which are *either* technically and allocatively efficient in both periods *or* if inefficiencies of these frontier firms stay constant over time. The second case considers firms facing different input prices. The TOTEX Malmquist index is undistorted if input prices *either* stay constant *or* change by the same proportion for all firms.

The remainder of this paper is organized as follows. Section 2 reviews the economic target of price-based regulation and derives the relevant components of RPI-X. The analysis shows that the term RPI-X is theoretically equivalent to the (unobserved) efficient cost change of a regulated firm. Section 3 defines and explains the three variants of the Malmquist index: while the production Malmquist index only measures the technical frontier shift, the cost Malmquist index captures the cost frontier shift, which is exactly the efficient cost change representing RPI-X. Lastly, the TOTEX Malmquist approximates the cost frontier shift. Section 4 analyses the accuracy of the TOTEX Malmquist and derives assumptions, under which it is unbiased. Section 5 concludes.

2 What is X in RPI-X-regulation?

The aim of monopoly regulation is to prevent companies from exploiting their market position as natural monopolies to earn excessive revenues at the expense of their consumers. The regulatory goal is to mimic competition by setting prices at an efficient level that ensures cost recovery for the regulated company while providing the right incentives both for cost efficient operation and investments. This in turn is a trade-off for the regulatory design (see e.g. Brunekreeft and Meyer 2016; Borrmann and Brunekreeft 2011). The traditional cost-based approach of regulation is well suited to achieve cost recovery and investments by adjusting the prices to the firm's actual cost level; but it is also known for setting poor efficiency incentives and may (theoretically) result in gold-plating, if the regulated rate of return is set too high (Averch and Johnson 1962). In contrast, modern regulatory regimes are price-based approaches that define price or revenue allowances in advance for a predefined regulatory period of typically 4 to 5 years. By decoupling prices from actual costs, which in absence of competition are not necessarily efficient, firms should be given incentives for cost reductions. The difficulty, however, is to acquire the information needed to determine the adequate price path in advance. Besides uncertainty about

---

[3] See also Polynomics and Jacobs University Bremen (2016).

future productivity and input price changes, the main problem is asymmetric information about the firms' technology: the regulator is less well informed about the relevant cost factors than the regulated firm (Schmalensee 1989; Beesley and Littlechild 1989).[4] Most of the currently applied price-cap regimes are based on an "RPI-X" formula, which goes back to the seminal paper of Littlechild (1983). The two terms in the simplified RPI-X formula refer to the retail price index (RPI) and a correction term (X). As derived below, the X denotes the difference between the expected future productivity growth and the price changes of the regulated sector as the deviations from total economy. Both terms are to be determined by the regulator in advance for the forthcoming regulatory period and prescribe the price path of the regulated firm independent of its actual cost.

But what is the economic rationale behind RPI-X? We follow the approach set out by Bernstein and Sappington (1999), albeit in a slightly different way of exposition.[5] Denoting the regulated sector with superscript R, the prescribed price path is given by

$$P_t^R = P_{t-1}^R(RPI - X) = P_{t-1}^R(1 + \Delta RPI - X), \tag{1}$$

where ΔRPI is the change in the retail price index, reflecting the total economy's output price increase. Written in index terms this gives

$$\Delta P_t^R = \frac{P_t^R}{P_{t-1}^R} - 1 = \Delta RPI - X, \tag{2}$$

where ΔP_t^R denotes the change in the output price index of the regulated sector.

Assuming perfect competition and constant returns to scale for the total economy, the term ΔRPI, depicting the change in the retail price index of the total economy, is equal to the change in input prices (Δw^T) minus the change in total factor productivity (ΔTFP^T) of the total economy (denoted by superscript T):

$$\Delta RPI = \Delta w^T - \Delta TFP^T. \tag{3}$$

Moreover, if the zero-profit condition holds, the difference between changes in input prices and total factor productivity will be equal to the change in efficient cost for a constant output, which we denote with ΔECC^T:

$$\Delta ECC^T = \Delta w^T - \Delta TFP^T. \tag{4}$$

---

[4] A good overview is provided in Joskow (2014).

[5] Our exposition is very short and concentrates only on those parts which we need further in the paper. The interested reader will find the comprehensive derivation in Bernstein and Sappington (1999).

As regulation aims to mimic competition, the regulated output price change (P_t^R) should follow the same rule. Hence, prices should adjust according to sectoral input price changes minus total factor productivity change according to the following regulatory formula:

$$\Delta P^R = \Delta w^R - \Delta TFP^R = \Delta ECC^R, \tag{5}$$

To relate this rule to the RPI-X formula, we bring together the parts for the total economy (T) and the regulated sector (R). Subtracting equation (3) from (5) gives:

$$\Delta P^R - \Delta RPI = [\Delta w^R - \Delta w^T] - [\Delta TFP^R - \Delta TFP^T], \tag{6}$$

or

$$\Delta P^R = \Delta RPI - X \tag{7}$$

with

$$X = (\Delta TFP^R - \Delta TFP^T) + (\Delta w^T - \Delta w^R). \tag{8}$$

Hence, the RPI-X formula (2) says that the change of the regulated price is equal to the rate of inflation minus the X-factor, which according to equation (8) is a combination of total factor productivity and input prices changes of the total economy and the regulated sector. In other words, the X is a correction term for the RPI, as it corrects for the deviation of productivity and input price changes of the total economy from the regulated sector.

The careful reader will quickly see that when putting (8) back into the RPI-X formula (2), and using equation (3), the terms for the total economy (RPI, Δw^T and ΔTFP^T) vanish, and equation (5) is all that remains and, in fact, all that is needed for regulation. The use of RPI is entirely due to the regulatory practice, as input price data is difficult to collect (Bernstein and Sappington 1999), which we will not pursue here. From the perspective of regulatory practice, it follows however that the X-factor defined as above is by and large a correction factor: strictly speaking, we would only be interested in ΔTFP^R and Δw^R while all the other terms are there because the RPI-X formula applies RPI instead of Δw^R. For our purposes in this paper, we concentrate completely and only on the regulated firms (R) and from now on ignore the terms for total economy (T). Hence, for ease of notation, we will drop the superscript R and focus on the simplified regulatory formula

$$\Delta P = \Delta w - \Delta TFP = \Delta ECC. \tag{9}$$

2.1 Törnquist and Malmquist

The regulatory challenge is to determine the two terms Δw and ΔTFP in the regulatory formula (9), most notably the last term: ΔTFP. The commonly applied methodology to determine ΔTFP is the "Törnqvist index", named after Törnqvist (1936).[6] It calculates the historical long-run development of total factor productivity (TFP) defined as output divided by input; usually, prices are used as weighting factors to sum various inputs and outputs. If the output index grows faster than the input index, TFP increases and reverse. The crucial difficulty with this approach is an identification problem: the Törnqvist index cannot distinguish between a *frontier shift* and *catch-up* effects. Typically, the regulated sector contains several monopoly firms which are regulated under the same regulatory regime. The firms typically differ in their performance: some firms will be efficient (i.e. operate at the efficiency frontier) and some firms will lag behind as compared to the efficiency frontier. If an inefficient firm reduces its relative inefficiency through time, we call this a catch-up effect. In contrast, if the efficient firms at the frontier improve efficiency, we call this a frontier shift. The frontier shift is the change in sector productivity and is the primary target of the X-factor as defined above. The catch-up factors are important but should not be part of the X-factor as defined above; instead they should be regulated differently.[7] In case of catch-up effects, the Törnqvist index will overestimate the frontier shift and thereby ΔTFP.

To address precisely this problem, the Malmquist index (named after Malmquist, 1953) serves as an alternative to the Törnqvist index. The Malmquist index, which will be explained in detail below, disentangles the efficiency effects into catch-up effects and the frontier shift. There are different variations of the Malmquist index. The *productivity Malmquist index* (PMI), which only applies quantity and no price data, calculates technological changes and, hence, ΔTFP. An important innovation was made with the *cost Malmquist index* (CMI) by Maniadakis and Thanassoulis (2004); this approach assumes the availability of data on both input quantities and input prices. Provided these data are in fact available, this method can be used to calculate the term $\Delta w - \Delta TFP$, which is the efficient cost change ΔECC required according to the regulatory formula (9) above. In practice, however, regulators often do not have separate data on input prices and input quantities. Instead, regulators must rely on total cost data (TOTEX). To this end, we analyse a *TOTEX Malmquist index* (TMI) as a third alternative. The TMI is basically the traditional PMI, where TOTEX instead of quantity data are used as input. We know from Maniadakis and Thanassoulis (2004) that the CMI can accurately isolate the cost frontier shift ΔECC. But how accurate is

[6] See e.g. see Coelli et al. (2005).
[7] For this reason, regulatory practice distinguishes between "individual X-factors" reflecting the relative inefficiencies and catch-up factors on the one hand, and the "general X-factor", reflecting the overall productivity change as in the frontier shift on the other hand. For an overview of international practices to treat inefficiencies in regulation see e.g. Jamasb and Pollitt (2000; 2004).

the approximation of ΔECC by applying a TMI? In this paper, we study the distortions which result if TOTEX data is applied instead of data on input prices and quantities.

3 Malmquist index approach

3.1 The traditional production Malmquist index (PMI)

The production Malmquist index (PMI) is a means of productivity analysis first introduced by Caves, Christensen and Diewert (1982). It is based on a production function approach, and measures efficiency values as radial distances of a firm's actual input or output quantities to the efficiency frontier. Figure 1 illustrates the input distance function approach for the case of one output (Y_t) and two-inputs, x_t^1 and x_t^2 over two periods t_0 and t_1.

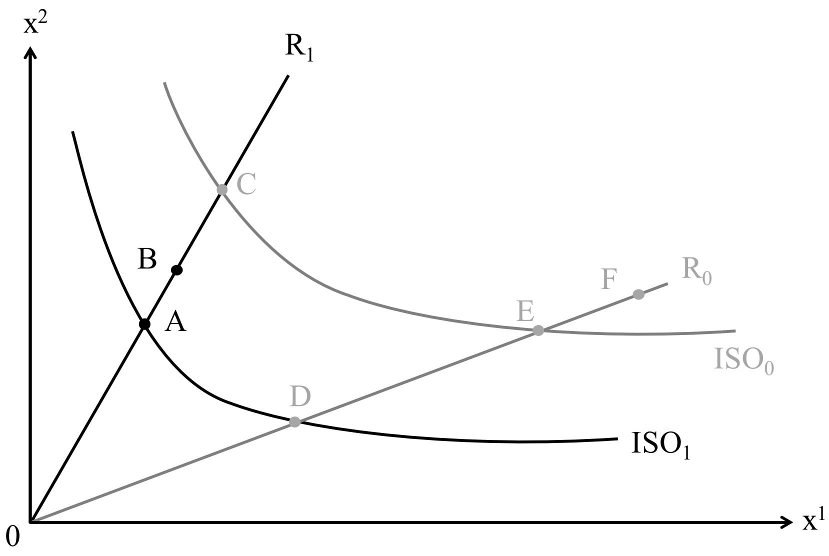

Figure 1: Illustration of the production Malmquist index

Assume a firm using input combinations F (in t_0) and B (in t_1) to produce an output *Y* under constant returns to scale. Technically efficient input combinations are given by the production function, which is represented by the isoquants ISO_0 and ISO_1.

The technical efficiency (TE) refers to the quantity of inputs used by the firm in comparison to the production frontier. TE can be measured by the radial distances of F and B to the technically efficient input combinations E (in t_0) and A (in t_1) along the lines of origin. TE results as 0E/0F and 0A/0B. In distance functions with $\mathbf{x_t}$ denoting the vector of input quantities, the term TE reads

$$TE_t(\mathbf{x_t}, Y_t) = \frac{1}{d_t(\mathbf{x_t}, Y_t)}. \tag{10}$$

In this representation, the distance function d_t gives the radial distance of actual to efficient input use for given outputs and is defined as

$$d_t(\mathbf{x_t}, Y_t) = \max_{\rho}\{\rho: (\mathbf{x_t}/\rho) \epsilon L(Y_t)\}, \tag{11}$$

with $L(Y_t)$ being the input requirement set that contains all input vectors that can produce output Y_t at a given technology. The input distance function is thus the largest factor by which input quantities can be divided and one is still able to produce a given output Y_t.

For the purpose of this paper, we are interested in productivity change over time, i.e. between periods t_0 and t_1. Färe et al. (1994) show that the Malmquist index can be used to decompose a firm's efficiency change over time into *technical efficiency change (TEC)* and *technical change (TC)*. As both periods t_0 and t_1 can be used as reference technology, and the choice of either would be arbitrary, the PMI is typically written in geometric means for t_0 and t_1:

$$PMI = \left[\left(\frac{0B/0C}{0F/0E}\right) \cdot \left(\frac{0B/0A}{0F/0D}\right)\right]^{1/2} = \underbrace{\frac{0B/0A}{0F/0E}}_{TEC} \cdot \underbrace{\left[\left(\frac{0B/0C}{0B/0A}\right) \cdot \left(\frac{0F/0E}{0F/0D}\right)\right]^{1/2}}_{TC} \tag{12}$$

The term TEC denotes the catch-up effect, measuring how the individual firm has increased its technical efficiency relative to the production frontier. TC represents the frontier shift indicating the change in sectoral productivity over time. The focus of this paper is on the latter term, which describes the technological change for the industry and eliminates the individual efficiency changes.

In distance function notation, TC can be written as

$$\begin{aligned} \mathrm{TC} &= \left[\left(\frac{d_0(\mathbf{x_1}, Y_1)}{d_1(\mathbf{x_1}, Y_1)}\right) \cdot \left(\frac{d_0(\mathbf{x_0}, Y_0)}{d_1(\mathbf{x_0}, Y_0)}\right)\right]^{1/2} \\ &= \left[\left(\frac{TE_1(\mathbf{x_1}, Y_1)}{TE_0(\mathbf{x_1}, Y_1)}\right) \cdot \left(\frac{TE_1(\mathbf{x_0}, Y_0)}{TE_0(\mathbf{x_0}, Y_0)}\right)\right]^{1/2} \end{aligned} \tag{13}$$

Assuming there are at least some technically efficient firms in the sample, a benchmarking methodology like data envelopment analysis (see e.g. Färe et al. 1985) can be used to calculate the technical change (TC), which is exactly the inverse of the change in technical efficiency (TE) that corresponds to the TFP change.[8] Hence, writing these terms in differences yields $\Delta\mathrm{TC} = -\Delta\mathrm{TFP}$. The second term in the regulatory formula (9), the input price change Δw, must be determined separately.

[8] While TFP is defined as a positive productivity change, TC is defined negatively as the reduction of input uses in period 1 compared to period 0.

3.2 The cost Malmquist index (CMI)

The traditional PMI only identifies the total productivity change, but not the price change. Maniadakis and Thanassoulis (2004) derive a cost Malmquist index (CMI) by introducing input price information to the calculations. Figure 2 illustrates the CMI graphically.

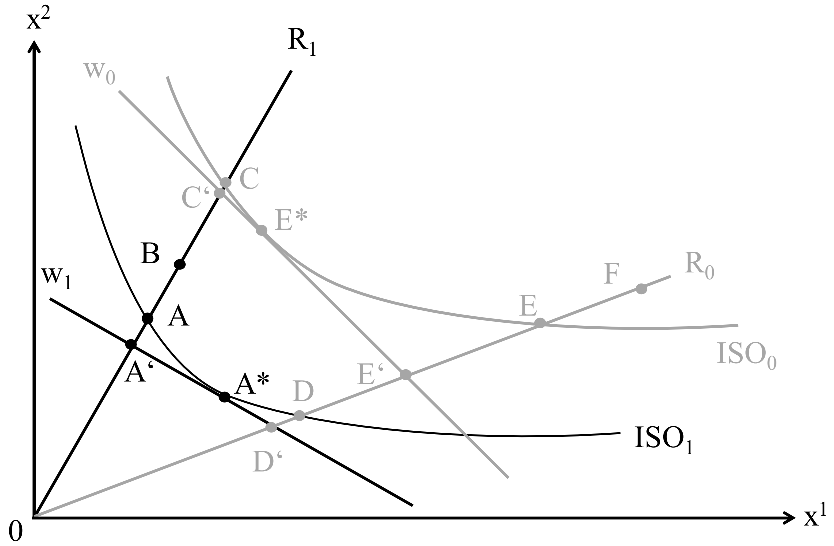

Figure 2 Illustration of the cost Malmquist index

Given relative input price w_0 and w_1, cost minimal input use is given by the tangential points E* (in t_0) and A* (in t_1) between the isoquants and the iso-cost lines w_0 and w_1, respectively.

Cost efficiency (CE) compares the firm's individual costs based on input combinations and input prices (F and B) with the efficient costs (E* and A*). Hence, the firm's cost efficiency is given by the distances 0E'/0F and 0A'/0B in Figure 2. Defining a linearly homogeneous cost function $C_t(\mathbf{w_t}, Y_t)$ subject to the cost minimization approach for given output Y_t, input price vector $\mathbf{w_t}$, and technology $y(\mathbf{x_t})$

$$C_t(\mathbf{w_t}, Y_t) = min\, \mathbf{w_t}\mathbf{x_t} \;\; s.t.\, y(\mathbf{x_t}) \geq Y_t, \tag{14}$$

Cost efficiency is measured as

$$CE\big(\mathbf{w_t}, \mathbf{x_t}, C_t(\mathbf{w_t}, Y_t)\big) = \frac{C_t(\mathbf{w_t}, Y_t)}{\mathbf{w_t}\mathbf{x_t}} \tag{15}$$

Input prices add another dimension to the efficiency analyses, namely *allocative efficiency (AE)*. AE measures the efficiency of relative input quantities for given input prices. In distance measures, AE corresponds to the ratios 0E'/0E and 0A'/0A in Figure 2, respectively.

Allocative efficiency is exactly the difference between cost efficiency and technical efficiency. Using equations (10) and (15), AE can be stated in cost and distance notation as

$$AE\big(\mathbf{w_t}, \mathbf{x_t}, C_t(\mathbf{w_t}, Y_t)\big) = \frac{d_t(\mathbf{x_t}, Y_t) C_t(\mathbf{w_t}, Y_t)}{\mathbf{w_t x_t}} \tag{16}$$

Maniadakis and Thanassoulis (2004) derive the CMI to decompose cost efficiency changes over time. Writing in cost notation and using the geometric mean of both periods, the CMI can be stated as follows (Maniadakis and Thanassoulis, 2004):

$$\begin{aligned} CMI &= \left[\frac{C_0(\mathbf{w_0}, Y_0)/\mathbf{w_0 x_0}}{C_0(\mathbf{w_0}, Y_1)/\mathbf{w_0 x_1}} \times \frac{C_1(\mathbf{w_1}, Y_0)/\mathbf{w_1 x_0}}{C_1(\mathbf{w_1}, Y_1)/\mathbf{w_1 x_1}}\right]^{1/2} \\ &= \underbrace{\left[\frac{C_0(\mathbf{w_0}, Y_0)/\mathbf{w_0 x_0}}{C_1(\mathbf{w_1}, Y_1)/\mathbf{w_1 x_1}}\right]}_{=OEC} \\ &\times \underbrace{\left[\frac{C_1(\mathbf{w_1}, Y_1)/\mathbf{w_1}\boldsymbol{x_1}}{C_0(\mathbf{w_0}, Y_1)/\mathbf{w_0}\boldsymbol{x_1}} \times \frac{C_1(\mathbf{w_1}, Y_0)/\mathbf{w_1}\boldsymbol{x_0}}{C_0(\mathbf{w_0}, Y_0)/\mathbf{w_0}\boldsymbol{x_0}}\right]^{1/2}}_{=CTC} \end{aligned} \tag{17}$$

with:

OEC: overall efficiency change

CTC: cost technical change.

In this representation, the term $C_0(\mathbf{w_0}, Y_0)/\mathbf{w_0}\boldsymbol{x_0}$, for instance, is the cost efficiency of a firm with costs $\mathbf{w_0}\boldsymbol{x_0}$ compared to efficient costs $C_0(\mathbf{w_0}, Y_0)$ in period t_0, resulting from the cost minimization approach (14).

The decomposition of CMI separates the catch-up effect, which in cost terms is the overall efficiency change (OEC), from the cost frontier shift, which corresponds to the cost technical change (CTC). In this decomposition, OEC denotes the change in technical and allocative efficiency with respect to the efficiency frontier: the catch-up effect. The term CTC covers both the change in production technology (technological progress) and the cost effects caused by changing input prices. We focus on CTC, which turns out to include the efficient cost change ECC needed for regulation.

A further decomposition of CTC yields

$$\begin{aligned} CTC &= \underbrace{\left[\frac{C_1(\mathbf{w_0}, Y_1)}{C_0(\mathbf{w_0}, Y_1)} \times \frac{C_1(\mathbf{w_0}, Y_0)}{C_0(\mathbf{w_0}, Y_0)}\right]^{1/2}}_{=TC} \times \underbrace{\left[\frac{C_1(\mathbf{w_1}, Y_1)}{C_1(\mathbf{w_0}, Y_1)} \times \frac{C_1(\mathbf{w_1}, Y_0)}{C_1(\mathbf{w_0}, Y_0)}\right]^{1/2}}_{\equiv \mathrm{w}} \\ &\times \underbrace{\left[\frac{\boldsymbol{w_0 x_1}}{\boldsymbol{w_1 x_1}} \times \frac{\boldsymbol{w_0 x_0}}{\boldsymbol{w_1 x_0}}\right]^{1/2}}_{\equiv \text{bias}} \end{aligned} \tag{18}$$

We denote the second term in (18) as the input price index (**w**), as it is the geometric mean of the input price change with the efficient input quantities in both periods as weights. This is the definition of the well-known Fisher index (Fisher 1922):

$$\mathrm{w} = \left[\frac{C_1(\mathbf{w_1}, Y_1)}{C_1(\mathbf{w_0}, Y_1)} \times \frac{C_1(\mathbf{w_1}, Y_0)}{C_1(\mathbf{w_0}, Y_0)}\right]^{1/2} = \left[\frac{\mathbf{w_1 x_1^*}(\mathbf{w_1}, Y_1)}{\mathbf{w_0 x_1^*}(\mathbf{w_1}, Y_1)} \times \frac{\mathbf{w_1 x_0^*}(\mathbf{w_0}, Y_0)}{\mathbf{w_0 x_0^*}(\mathbf{w_0}, Y_0)}\right]^{1/2} \tag{19}$$

Note that multiplying the first two terms of CTC (TC and w) yields the ratio of efficient costs of both periods, which is the efficient cost change (ECC) written in geometric means:

$$TC \times \mathrm{w} = \left[\frac{C_1(\mathbf{w_1}, Y_1)}{C_0(\mathbf{w_0}, Y_1)} \times \frac{C_1(\mathbf{w_1}, Y_0)}{C_0(\mathbf{w_0}, Y_0)}\right]^{1/2} = ECC. \tag{20}$$

Writing the cost ratio ECC in differences gives

$$\Delta ECC = \Delta TC + \Delta \mathrm{w}. \tag{21}$$

As in case of the production Malmquist, TC is the inverse of total factor productivity change ($\Delta\mathrm{TC} = -\Delta\mathrm{TFP}$). Hence, equation (21) is the regulatory price adjustment according to equation (9):

$$\Delta ECC = \Delta w - \Delta TFP, \tag{22}$$

The question is how ECC can be calculated in practice. Using benchmarking techniques like data envelopment analysis (DEA), CTC as stated in (17) results as a combination of four cost efficiency scores:

$$CTC = \left[\frac{CE\big(\mathbf{w_1}, \mathbf{x_1}, C_1(\mathbf{w_1}, Y_1)\big)}{CE\big(\mathbf{w_0}, \mathbf{x_1}, C_0(\mathbf{w_0}, Y_0)\big)} \times \frac{CE\big(\mathbf{w_1}, \mathbf{x_0}, C_1(\mathbf{w_1}, Y_1)\big)}{CE\big(\mathbf{w_0}, \mathbf{x_0}, C_0(\mathbf{w_0}, Y_0)\big)}\right]^{1/2}. \tag{23}$$

According to (18), however, the calculation of ECC requires a correction of CTC for a bias-term:

$$ECC = \frac{CTC}{bias} = CTC \times \left[\frac{\boldsymbol{w_1 x_1}}{\boldsymbol{w_0 x_1}} \times \frac{\boldsymbol{w_1 x_0}}{\boldsymbol{w_0 x_0}}\right]^{1/2} \tag{24}$$

Both the application of the production and the cost Malmquist index require that the sample contains firms that are technical efficient to reveal the true efficiency frontier. The CMI, however, is more demanding regarding data availability: it requires both input quantities and prices for both periods. As equation (23) reveals, CTC involves cross efficiency scores from combinations of inputs and outputs from different periods.[9] These data are also needed to correct for the bias term in (24). This term is related to potential allocative inefficiency of the sample firms, which is ignored in the production Malmquist index. As equation (24) shows, the bias term is a distorted price index that in contrast to the Fisher index (19) uses (potentially) allocatively inefficient input quantities $\mathbf{x_t}$ as weights. A correction for this allocative bias also requires input quantity and price information.

3.3 The TOTEX Malmquist index (TMI)

A regulator usually lacks separate price and quantity information on the regulated firms, but instead works with total costs (TOTEX). This section presents a TOTEX Malmquist index (TMI) as an alternative way to approximate the efficient cost change needed for regulation. We define the TMI as a classical PMI with the difference that the firms' total costs $C_t(\mathbf{w_t}, Y_t)$ are used as the only input. Applying the distance function approach analogously to (11), gives the following distance functions:

$$\hat{d}_t(C_t, Y_t) = \max_{\rho}\{\rho: (C_t/\rho)\epsilon F(Y_t)\} \tag{25}$$

This leads to an alternative definition of cost efficiency,

$$\widehat{CE}_t(C_t, Y_t) = \frac{1}{\hat{d}_t(C_t, Y_t)}, \tag{26}$$

where $\widehat{CE}_t(C_t, Y_t)$ is strictly speaking technical efficiency based on aggregated total costs (i.e. TOTEX) instead of physical input quantities.

The respective TOTEX Malmquist index

$$TMI = \left[\left(\frac{\hat{d}_0(C_1, Y_1)}{\hat{d}_0(C_0, Y_0)}\right) \cdot \left(\frac{\hat{d}_1(C_1, Y_1)}{\hat{d}_1(C_0, Y_0)}\right)\right]^{1/2} \tag{27}$$

can be decomposed as follows:

$$TMI = \frac{\hat{d}_1(C_1, Y_1)}{\hat{d}_0(C_0, Y_0)} \cdot \underbrace{\left[\left(\frac{\hat{d}_0(C_1, Y_1)}{\hat{d}_1(C_1, Y_1)}\right) \cdot \left(\frac{\hat{d}_0(C_0, Y_0)}{\hat{d}_1(C_0, Y_0)}\right)\right]^{1/2}}_{=TCC}, \tag{28}$$

TCC denotes the total cost change. Writing TCC in efficiency scores gives

---

[9] For instance, the efficiency value CE($\mathbf{w_0}$,$\mathbf{x_1}$, $\mathbf{w_0 x_0}^*$) results as a firm's benchmark using its actual inputs from period 1 but prices from period 0 against all remaining firms with inputs and prices from period 0. Hence, calculating CTC requires separate information on price and quantity data for each observation period.

$$TCC = \left[\left(\frac{\widehat{CE}_1(C_1, Y_1)}{\widehat{CE}_0(C_1, Y_1)} \right) \cdot \left(\frac{\widehat{CE}_1(C_0, Y_0)}{\widehat{CE}_0(C_0, Y_0)} \right) \right]^{1/2}. \tag{29}$$

In this representation, TCC is used as an approximation of the efficient cost change ECC. The difference of the two approaches is that the CMI uses the optimal input quantities as weights for input prices, while the TMI uses observed input quantities, as there is no explicit distinction between prices and quantities.

TCC consists of four efficiency scores, which can be calculated with DEA, for instance. The difference to the original cost Malmquist approach by Maniadakis and Thanassoulis (2004) is that only total cost is used for the calculations. Formally, TCC is a classical production Malmquist is based on aggregated TOTEX as the only input instead of physical units.

That means, however, that any information on allocative efficiency is ignored – as it is done in the traditional PMI approach. Given that a regulator is neither able to directly observe efficient costs nor does he have separate input price and quantity information on the firms in the sample, the question is in which cases a TOTEX-based benchmarking approach is still an unbiased approximation of the efficient cost change ECC. Section 4 analyses TCC with respect to possible distortions.

4 Discussion of the TOTEX Malmquist

The total cost change (TCC) derived from the TOTEX Malmquist index may be used by regulators as an approximation of the efficient cost change ECC. The problem of real-world regulation is that the regulator lacks detailed price and quantity data. The TMI allows the regulator to make use of the total costs data (TOTEX). Moreover, the technical change and the input price changes can be calculated in combination; a separate information on prices and quantities is not necessary.

However, this simplification in data requirement comes at a cost: the quality of the Malmquist calculation critically depends on the data for a limited number of firms, which are identified as the efficient firms in the benchmarking procedure. We derive the conditions under which the TOTEX Malmquist leads to an unbiased calculation of the efficient cost change ECC. Two scenarios are considered. First, we discuss the implications of allocative inefficiency. Second, we analyse the case that firms face different input prices. The two propositions below outline the conditions under which a TOTEX Malmquist leads to an unbiased approximation of the cost frontier shift ECC, and hence "Δw – ΔTFP".

Case 1: impact of allocative inefficiency on TCC

First, we derive the relationship between TCC and ECC by rewriting TCC according to (28) in cost notation:

$$TCC = \left[\left(\frac{\widehat{CE}_1(C_1,Y_1)}{\widehat{CE}_0(C_1,Y_1)}\right)\cdot\left(\frac{\widehat{CE}_1(C_0,Y_0)}{\widehat{CE}_0(C_0,Y_0)}\right)\right]^{1/2}$$

$$= \left[\frac{C_1^F(\mathbf{w_1},Y_1)/\mathbf{w_1x_1}}{C_0^F(\mathbf{w_0},Y_0)/\mathbf{w_1x_1}}\times\frac{C_1^F(\mathbf{w_1},Y_1)/\mathbf{w_0x_0}}{C_0^F(\mathbf{w_0},Y_0)/\mathbf{w_0x_0}}\right]^{1/2} \tag{30}$$

$$= \underbrace{\frac{\mathbf{w_1x_1^*}(\mathbf{w_1},Y_1)}{\mathbf{w_0x_0^*}(\mathbf{w_0},Y_0)}}_{ECC}\times\underbrace{\frac{\mathbf{w_0x_0^*}(\mathbf{w_0},Y_0)/\mathbf{w_0x_0^F}}{\mathbf{w_1x_1^*}(\mathbf{w_1},Y_1)/\mathbf{w_1x_1^F}}}_{OEC^F}.$$

In equation (30) C_0^F and C_1^F denote the observed cost frontiers identified in the benchmarking process. Hence, cost efficiency $\widehat{CE}_1(C_1,Y_1)$, for instance, measures the distance of a firm's total cost $\mathbf{w_tx_t}$ to the observed cost frontier $C_1^F(\mathbf{w_1},Y_1)$, which is not necessarily the true efficient costs $C_1(\mathbf{w_1},Y_1) = \mathbf{w_1x_1^*}(\mathbf{w_1},Y_1)$.

The last term in (30) represents a possible distortion between TCC and ECC. This term is a ratio of efficiency values, giving the overall efficiency change of the frontier firms with respect to the unobserved true cost function (OEC^F). No distortion exists if OEC^F is equal to one. We formalize this in the following proposition.

Proposition 1: Assuming the same technology and input prices for all firms, TCC is an undistorted measure for ECC if

a) the efficiency frontier in benchmarking is set by firms which are technically (i.e. operate on the true frontier) and allocatively efficient in both periods, *or* if
b) technical and allocative inefficiency of the frontier firms stay constant over time.

Proof:

$$TCC = ECC \Leftrightarrow \frac{\mathbf{w_0x_0^*}(\mathbf{w_0},Y_0)/\mathbf{w_0x_0^F}}{\mathbf{w_1x_1^*}(\mathbf{w_1},Y_1)/\mathbf{w_1x_1^F}} = 1$$

$$\Leftrightarrow \left[\frac{\mathbf{w_0x_0^*}(\widetilde{\mathbf{w_0^F}},\bar{Y})}{\mathbf{w_1x_1^*}(\widetilde{\mathbf{w_1^F}},\bar{Y})}\right]\times\left[\frac{\mathbf{w_1x_1^*}(\widetilde{\mathbf{w_1^F}},\bar{Y})}{\mathbf{w_0x_0^*}(\widetilde{\mathbf{w_0^F}},\bar{Y})}\right]\times\left[\frac{\mathbf{w_0x_0^*}(\mathbf{w_0},Y_0)/\mathbf{w_0x_0^F}}{\mathbf{w_1x_1^*}(\mathbf{w_1},Y_1)/\mathbf{w_1x_1^F}}\right] = 1 \tag{31}$$

$$\Leftrightarrow \underbrace{\left[\frac{\mathbf{w_0 x_0^*}/\mathbf{w_0 x_0^*}(\widetilde{\mathbf{w_0^F}}, Y_0)}{\mathbf{w_1 x_1^*}/\mathbf{w_1 x_1^*}(\widetilde{\mathbf{w_1^F}}, Y_1)}\right]}_{AEC^F} \times \underbrace{\left[\frac{\mathbf{w_0 x_0^*}(\widetilde{\mathbf{w_0^F}}, Y_0)/\mathbf{w_0 x_0^F}}{\mathbf{w_1 x_1^*}(\widetilde{\mathbf{w_1^F}}, Y_1)/\mathbf{w_1 x_1^F}}\right]}_{TEC^F} = 1.$$

With

$\mathbf{w_t}$	Vector of input prices in t
$\widetilde{\mathbf{w_t^F}}$	Vector of distorted input prices of frontier firms in t,
$\mathbf{x_t^*}(\mathbf{w_t}, Y_t)$	Vector of efficient input quantities in t for input prices $\mathbf{w_t}$ and output Y_t
$\mathbf{x_t^*}(\widetilde{\mathbf{w_t^F}}, Y_t)$	Vector of efficient input quantities in t for distorted prices $\widetilde{\mathbf{w_t^F}}$ and output Y_t

Allocative inefficiency may be stated as if the frontier firms optimize their inputs with respect to distorted input prices $\widetilde{\mathbf{w_t^F}}$ instead of the true prices $\mathbf{w_t}$. Equation (31) relates the condition for TCC being an undistorted measure for ECC to the frontier firms' change in allocative (AEC^F) and technical efficiency (TEC^F) over time. If the frontier firms are fully efficient in both periods, proposition 1a is obviously fulfilled, because allocative and technical catch-up effects are zero.

If the frontier firms in the benchmarking dataset are inefficient, however, i.e. there is either technical or allocative inefficiency, then the total cost change (TCC) will only be undistorted if both inefficiencies are constant over time (proposition 1b).[10] As the cost Malmquist does include data on input prices, it can correctly isolate the change of allocative efficiency, and as a result can correctly calculate the frontier shift. The TOTEX Malmquist does not have input price data, and wrongly assigns changes in technical or allocative efficiency to the frontier shift, thereby distorting the true frontier shift.

The result is important in a wider regulatory context: according to proposition 1, a TOTEX Malmquist is unbiased if the frontier firms are both technically and allocatively efficient or if inefficiency stays constant over time. The assumption of technical efficiency may be reasonable, if incentive regulation has been in place long enough to have already incentivized cost reductions as it intends to do. At least some best-practice firms may operate fairly close to technical efficiency. Allocative efficiency appears to be more critical, as it would require an instant adjustment of inputs to changes of relative input prices.

---

[10] Alternatively, the condition also holds if changes in technical and allocative inefficiency exactly compensate each other, which would be a rather unlikely coincidence, however.

Especially long-lived capital assets in infrastructure (e.g. electricity networks) cannot be adjusted to price changes in the short term but are stranded costs (see e.g. Baumol and Sidak 1994). Therefore, even the most efficient firms may not be able to achieve full allocative efficiency. In this case, the TOTEX Malmquist will still be unbiased if the frontier firms' allocative inefficiency does not change between the two observation periods, which seems plausible only in case of constant input price ratios.

How large possible distortions are remains an empirical question. Two things should be noted, however. First, the efficiency requirements only apply to the best-practice firms in the sample. Any possible catch-up effects (technical or allocative) of firms not defining the frontier will not distort TCC. Second, any possible distortion of the TOTEX Malmquist also applies to the Törnqvist index based on average sector data. By methodology, the Törnqvist index does not distinguish between catch-up effects and frontier shift and will always consider catch-up effects (including allocative efficiency) incorrectly as productivity change.

Case 2: impact of different input prices on TCC

We extend the above analysis for the possibility that firms face different (not observable) input prices. As a consequence, even cost-efficient firms may turn out being inefficient as they are benchmarked against firms with lower input prices. The focus is again on best-practice firms. Assume a group of frontier firms F that, due to different input prices, face lower efficient costs than other efficient firms in the benchmarking:

$$\mathbf{w_t^F}\mathbf{x_t^{*F}}(\mathbf{w}_t^F, Y_t) \leq \mathbf{w_t^F}\mathbf{x_t^*}(\mathbf{w_t}, Y_t) \leq \mathbf{w_t}\mathbf{x_t^*}(\mathbf{w_t}, Y_t).$$

Any firm with less favorable input prices faces a true (unobservable) cost frontier $\mathbf{w_t}\mathbf{x_t^*}$, but the observed frontier is set by $\mathbf{w_t^F}\mathbf{x_t^{*F}}$. The condition, under which the TOTEX Malmquist still provides an undistorted measure for the true cost change can be states as follows:

$$TCC = \underbrace{\frac{\mathbf{w_1}\mathbf{x_1^*}(\mathbf{w_1}, Y_1)}{\mathbf{w_0}\mathbf{x_0^*}(\mathbf{w_0}, Y_0)}}_{ECC} \times \underbrace{\frac{\mathbf{w_0}\mathbf{x_0^*}(\mathbf{w_0}, Y_0)/\mathbf{w_0^F}\mathbf{x_0^F}}{\mathbf{w_1}\mathbf{x_1^*}(\mathbf{w_1}, Y_1)/\mathbf{w_1^F}\mathbf{x_1^F}}}_{OEC^F}. \tag{32}$$

Proposition 2: Assuming technical and allocative efficiency of all firms, but differences in input prices with

$$\mathbf{w_t^F}\mathbf{x_t^{*F}}(\mathbf{w}_t^F, Y_t) \leq \mathbf{w_t^F}\mathbf{x_t^*}(\mathbf{w_t}, Y_t) \leq \mathbf{w_t}\mathbf{x_t^*}(\mathbf{w_t}, Y_t).$$

Then TCC is an undistorted measure for ECC if all factor prices differ by the same proportion λ and the difference in factor prices stays constant over time:

$$\lambda_t \mathbf{w_t^F} = \mathbf{w_t}.$$

Proof: Assume that the cost function is linearly homogenous in w:

$$C_{\mathrm{t}}(\mathbf{w_t}, Y_{\mathrm{t}}) = C_{\mathrm{t}}(\lambda_t \mathbf{w_t^F}, Y_{\mathrm{t}}) = \lambda_t C_{\mathrm{t}}(\mathbf{w_t^F}, Y_{\mathrm{t}}).$$

Then the distortion term according to (30) is equal to 1 if the following condition holds:

$$\frac{C_0(\mathbf{w_0}, Y_0)/C_0(\mathbf{w_0^F}, Y_0)}{C_1(\mathbf{w_1}, Y_1)/C_1(\mathbf{w_1^F}, Y_1)} = \frac{\lambda_0 C_0(\mathbf{w_0^F}, Y_0)/C_0(\mathbf{w_0^F}, Y_0)}{\lambda_1 C_1(\mathbf{w_1^F}, Y_1)/C_1(\mathbf{w_1^F}, Y_1)} = 1 \leftrightarrow \lambda_t = \bar{\lambda}\, \forall\, t.$$

Proposition 2 states that in case of groups of firms facing different input prices, TCC is unbiased only if input prices for both groups change by the same proportion, i.e. if the ratio of the price index between the groups remains constant over time. What does this mean for regulatory practice? The problem of firms facing different prices is that the frontier shifts (in cost terms) may be different, even if the technical change is the same. A regulatory regime may cover two structurally or geographically different areas; assume now that wages increase faster in one area than in the other. The TOTEX Malmquist index cannot identify this difference because data on the input prices are not available. Ignoring price differences in benchmarking may result in distorted calculations of the frontier shift. Proposition 2 shows that TCC will only be an unbiased approximation of the correct frontier shift if prices change proportionally between the groups of firms. How large possible distortions are is an empirical question that is beyond the scope of this paper.

5 Conclusions

This paper examines a topical issue in the regulation of monopoly. Modern forms of regulation apply the so-called RPI-X regulation, where X reflects *inter alia* estimates of the expected development of total productivity and input price changes. The regulatory challenge is how to determine X. The usual practice is to use the Törnqvist index, which, however, cannot distinguish between the frontier shift itself and catch-ups to the frontier. As an alternative, the Malmquist index makes such a distinction.

We build upon the cost Malmquist index, which extends the productivity Malmquist index by including input prices. By doing so it can give an accurate measure of the efficient cost change (i.e. an indicator of the cost frontier shift, combining quantity and price effects). The cost Malmquist index requires that data on input prices and quantities are available. However, usually regulators only have total cost (TOTEX) data. For this case, a TOTEX Malmquist index is derived. This is basically a production Malmquist index with TOTEX as the only input.

The main aim of this paper is to analyse potential biases in the determination of the frontier shift when a TOTEX Malmquist is applied instead of a cost Malmquist. Under which conditions does the

TOTEX Malmquist calculate the correct frontier shift? We consider two cases. The first case addresses technical and allocative inefficiency. The TOTEX Malmquist index is undistorted if the efficiency frontier is set by firms which are *either* technically and allocatively efficient in both periods *or* if inefficiencies of these frontier firms stay constant over time. The second case considers firms facing different input prices. The TOTEX Malmquist index is undistorted if input prices *either* stay constant *or* change by the same proportion for all firms.

From the viewpoint of regulatory practice, we consider allocative inefficiency a more relevant issue, as input quantities may not be quickly adjusted to changing input prices. It should be noted, however, that all conditions apply to frontier firms only: inefficiencies and differing input prices of firms not defining the frontier in the benchmarking procedure do not result in distortion of the TOTEX Malmquist index. However, it is important to be aware of possible distortions, as they may have significant consequences for the regulated companies.

References

Averch, H. and L. L. Johnson (1962). Behavior of the firm under regulatory constraint, *American Economic Review* 52, 1053–69.

Baumol, W. J., and Sidak, J. G. (1994). Stranded costs. *Harvard Journal of Law and Public Policy*, 18(3), 835–49.

Beesley, M. E., and Littlechild, S. C. (1989). The regulation of privatized monopolies in the United Kingdom. *The RAND Journal of Economics*, 454–472.

Bernstein J.I. and Sappington, D.E. (1999). Setting the X factor in price-cap regulation plans, *Journal of Regulatory Economics*, 16(1), 5–25.

Borrmann, J., and Brunekreeft, G. (2011). The effect of monopoly regulation on the timing of investment. *Bremen Energy Working Papers* No. 09, May 2011.

Brunekreeft, G., and Meyer, R. (2016). Anreizregulierung bei Stromverteilnetzen: Effizienz versus Investitionen. *Perspektiven der Wirtschaftspolitik*, 17(2), 172–187.

BNetzA, 2018a, *Beschluss im Verwaltungsverfahren nach §29 Abs. 1 EnWG i.V.m. §32 Abs. 1 Nr. 2a i.V.m. §9 Abs. 3 ARegV*, February 21, 2018.

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*, 1393–1414.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., and Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.

Färe R., Grosskopf S., Lindgren B., and Roos P. (1994). Productivity Developments in Swedish Hospitals: A Malmquist Output Index Approach. In: *Data Envelopment Analysis: Theory, Methodology, and Applications*. Springer, Dordrecht.

Färe, R., Grosskopf, S., Logan, J., and Lovell, C. K. (1985). Measuring efficiency in production: with an application to electric utilities. *The Measurement of Efficiency of Production*, 185–214.

Fisher, I. (1922). *The making of index numbers: a study of their varieties, tests, and reliability (No. 1).* Houghton Mifflin.

Jamasb, T., and Pollitt, M. (2004). Benchmarking and regulation of electricity transmission and distribution utilities: lessons from international experience, *Cambridge Working Papers in Economics (CWPE)*, University of Cambridge.

https://doi.org/10.17863/CAM.5398.

Jamasb, T., and Pollitt, M. (2000). Benchmarking and regulation: international electricity experience. *Utilities policy*, 9(3), 107–130.

Joskow, P. L. (2014). Incentive regulation in theory and practice: electricity distribution and transmission networks. In: *Economic Regulation and Its Reform: What Have We Learned?* University of Chicago Press, 291–344.

Littlechild, S. C. (1983). *Regulation of British Telecommunications' profitability: report to the Secretary of State*, February 1983. Department of Industry.

Malmquist, S. (1953). Index Numbers and Indifference Surfaces, *Trabajos de Estatistica*, Vol. 4, 209–242.

Maniadakis, N., and Thanassoulis, E. (2004). A cost Malmquist productivity index, *European Journal of Operational Research*, 154(2), 396–409.

Polynomics and Jacobs University Bremen (2016). *Die Ermittlung des technologischen Fortschritts anhand von Unternehmensdaten – Der Einsatz der Malmquist-Methode im deutschen Regulierungsrahmen*. Final report. Study conducted for Netze BW GmbH, August 24, 2016.

Schmalensee, R. (1989). Good Regulatory Regimes, *Rand Journal of Economics*, 20:3, 417–436.

Törnqvist, L. (1936). The Bank of Finland's consumption price index, *Bank of Finland Monthly Bulletin*, No. 10(1936), 1–8.

WIK (2017). *Gutachten zur Bestimmung des generellen sektoralen Produktivitätsfaktors*, Überarbeitete Version nach Eingang der Stellungnahmen [in German]. Study Bundesnetzagentur, July 10, 2017.