

COST AND PRODUCTIVITY

E. Grifell-Tatjé

Departament d'Economia de l'Empresa

Universitat Autònoma de Barcelona

08193 Bellaterra (Barcelona), SPAIN

emili.grifell@uab.es

C. A. K. Lovell

Department of Economics

University of Georgia

Athens, Ga 30602, USA

knox@terry.uga.edu

This is a postprint version of the article that was published in

Grifell-Tatjé, E. and C.A.K. Lovell (2000), "Cost and Productivity," *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.

DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

Abstract

We develop an analytical model capable of decomposing both intertemporal and multilateral cost variation. We begin by attributing cost variation to a price effect and a quantity effect. We then decompose the quantity effect into a productivity effect and an activity effect. The productivity effect in turn decomposes into a cost efficiency effect and, in the intertemporal context, a technical change effect. We also show how the intertemporal and multilateral cost decompositions can be implemented, using linear programming techniques. These techniques offer certain advantages over conventional econometric techniques whenever a substantial portion of cost variation is due to variation in cost efficiency. We illustrate the two cost decompositions with a pair of benchmarking exercises based on a panel of 93 US electric power generating companies, in which variation in cost efficiency does play a key role.

Keywords: Cost, productivity, benchmarking, electricity generation

JEL Codes: D24, M2

COST AND PRODUCTIVITY*

1. Introduction

Operating cost varies through time and across producers of even a reasonably homogeneous product. In such circumstances it is important for managers to be able to attribute both intertemporal and multilateral cost variation to its causal factors. In this paper we develop an analytical model capable of attributing cost variation to a price effect, which identifies cost variation attributable exclusively to variation in input prices, and a quantity effect, which identifies cost variation attributable solely to variation in input use. We continue by decomposing the quantity effect into an activity effect, which identifies cost variation attributable to variation in cost-efficient input use with given technology, and a productivity effect, which identifies cost variation attributable to variation in cost efficiency and, in an intertemporal context, to variation in the structure of production technology. Finally we decompose variation in cost efficiency into variation in productive efficiency and variation in the efficiency with which inputs are allocated in light of their prices. Thus we decompose cost variation into five sources in the intertemporal context, and into four sources in the multilateral context. These decompositions provide a useful framework within which to conduct a cost-based benchmarking exercise, either through time or across producers.

Our cost decomposition model is new, although it does have precedents. It can be considered as an attempt to merge the concept of cost variance analysis in the cost accounting literature with the concept of a cost frontier in the economics literature. In the economics literature both neoclassical theory and modern duality theory express minimum cost as a function of output quantities, input prices, and perhaps also a time trend. Thus the potential is there to use economic theory to enrich cost variance analysis. Indeed econometric models based on this framework, from Nerlove (1963) to the present day, have attributed cost variation to input price variation, scale economies and technical change. However these models have imposed a symmetric error structure on the equation(s) to be estimated, and have been unable to incorporate variation in cost efficiency as a source of cost variation. More recent stochastic cost frontier models, e.g., Greene (1980) and Stevenson (1980), incorporate a composed error structure and do allow for variation in cost efficiency. However apart from the restrictive Cobb-Douglas case of Schmidt and Lovell (1979), these stochastic cost frontier models have been unable to decompose cost efficiency into its technical and allocative components, the purely analytical results of Kumbhakar (1997) notwithstanding. Fortunately an older linear programming - based approach initiated by Afriat (1972) and Hanoch and Rothschild (1972) is capable of generating a complete cost variation decomposition. However it was not used for this purpose by these authors, who developed linear programs to test consistency of data with various regularity conditions and behavioral hypotheses. Our objective in this paper is to exploit economic theory, and to apply linear programming techniques, to achieve an empirically tractable attribution of cost variation to its economic sources.

The paper is organized as follows. In Section 2 we summarize a standard analytical framework of production and cost. In Section 3 we exploit this analytical framework to

Grifell-Tatjé, E. and C.A.K. Lovell (2000), “Cost and Productivity,” *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.

DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

decompose intertemporal cost change into five sources. In Section 4 we show how to implement the intertemporal cost change decomposition, using linear programming techniques. In Section 5 we show how the analysis developed in Sections 3 and 4 simplifies in the multilateral context, and we show how the multilateral analysis can be employed to conduct a benchmarking exercise against the performance of a target producer. In Section 6 we illustrate the cost decomposition techniques with a pair of benchmarking exercises based on a panel of 93 US electric power generating companies. Their costs vary substantially, both through time and across companies, and we are interested in decomposing these cost variations into their sources.

2. The Analytical Framework

We consider a producer using N resources represented by the input quantity vector $x = (x_1, \dots, x_N) \geq 0$ to produce M products represented by the output quantity vector $y = (y_1, \dots, y_M) \geq 0$. The producer pays resource prices represented by the input price vector $w = (w_1, \dots, w_N) > 0$. The producer's observed cost in period t is given by $c^t = w^{tT} x^t = \sum_{n=1}^N w_n^t x_n^t$. Since we are interested in decomposing cost change from one period to the next, $(c^{t+1} - c^t)$, we are not concerned with the revenue side of the producer's activities, and so we ignore the prices at which it sells its products.

We begin by defining the structure of period t technology in terms of its *input sets*

$$L^t(y^t) = \{x^t: (x^t \text{ can produce } y^t)\}, \quad t = 1, \dots, T. \quad (1)$$

Input sets are assumed to be closed and convex, and to satisfy strong disposability [$x^t \in L^t(y^t) \Rightarrow x'^t \in L^t(y^t)$, $x'^t \geq x^t$]. The lower boundaries of input sets are their *input isoquants*, which are defined as

$$I^t(y^t) = \{x^t: x^t \in L^t(y^t), \lambda x^t \notin L^t(y^t), \lambda < 1\}, \quad t = 1, \dots, T. \quad (2)$$

Although an input vector x^t must belong to its contemporaneous input set $L^t(y^t)$, it need not necessarily belong to its input isoquant $I^t(y^t)$. The Malmquist (1953) - Shephard (1953) *input distance function* provides a radial measure of the distance from an input vector to an input isoquant, and is defined as

$$D^t(y^t, x^t) = \max \{\theta: x^t/\theta \in L^t(y^t)\}, \quad t = 1, \dots, T. \quad (3)$$

$D^t(y^t, x^t) \geq 1$ since $x^t \in L^t(y^t)$, and $D^t(y^t, x^t) = 1 \Leftrightarrow x^t \in I^t(y^t)$. Dual to an input distance function is a *cost frontier*, which describes the minimum cost required to produce output vector y^t with input price vector w^t and with period t technology in place. The cost frontier is defined as

$$c^t(y^t, w^t) = \min_x \{(w^{tT} x): x \in L^t(y^t)\} = \min_x \{(w^{tT} x): D^t(y^t, x) \geq 1\}, \quad t = 1, \dots, T. \quad (4)$$

$c^t(y^t, w^t)$ is nondecreasing in y^t , and nondecreasing, concave and homogeneous of degree +1 in w^t . Since $c^t(y^t, w^t)$ is defined as a minimum concept, $c^t \geq c^t(y^t, w^t)$, with equality holding if, and only if, the producer is both technically and allocatively efficient.

We also use an adjacent-period input set $L^{t+1}(y^t)$, which describes the set of input vectors capable of producing output vector y^t with period $t+1$ technology, its adjacent-period isoquant $I^{t+1}(y^t)$, and the corresponding adjacent-period input distance function $D^{t+1}(y^t, x^t)$. However since (y^t, x^t) may not be feasible with period $t+1$ technology, $x^t \notin L^{t+1}(y^t)$ is possible, and it follows that $D^{t+1}(y^t, x^t) \geq 1$. Finally, an adjacent period cost frontier dual to the adjacent period input distance function is defined as $c^{t+1}(y^t, w) = \min_x \{(w^T x) : x \in L^{t+1}(y^t)\} = \min_x \{(w^T x) : D^{t+1}(y^t, x) \geq 1\}$, where w can be either w^t or w^{t+1} . Each of these adjacent-period representations of technology is defined in a similar manner to the contemporaneous representations in (1) - (4).

3. The Intertemporal Cost Change Decomposition

We are now prepared to decompose a producer’s cost change between period t and period $t+1$ into five distinct sources. We begin by decomposing cost change resulting from a movement from (x^t, w^t, y^t) to $(x^{t+1}, w^{t+1}, y^{t+1})$ into two basic components: a *price effect* reflecting the impact on cost of changes in input prices from w^t to w^{t+1} , holding input usage fixed, and a *quantity effect* reflecting the impact on cost of changes in input usage and output production from (x^t, y^t) to (x^{t+1}, y^{t+1}) , holding input prices fixed. A preliminary issue concerns how to hold input usage fixed in the price effect and how to hold input prices fixed in the quantity effect.

One possible decomposition of cost change into a price effect and a quantity effect is

$$c^{t+1} - c^t = (w^{t+1} - w^t)^T x^{t+1} + w^{tT} (x^{t+1} - x^t), \quad (5)$$

and another is

$$c^{t+1} - c^t = (w^{t+1} - w^t)^T x^t + w^{t+1T} (x^{t+1} - x^t). \quad (6)$$

In (5) the price effect is a Paasche type of input price index (fixing input usage at x^{t+1}) and the quantity effect is a Laspeyres type of input quantity index (fixing input prices at w^t), both expressed in difference rather than ratio form. The roles are reversed in (6), in which the price effect is a Laspeyres type of input price index (fixing input usage at x^t) and the quantity effect is a Paasche type of input quantity index (fixing input prices at w^{t+1}), both expressed in difference rather than ratio form. In light of the well-known deficiencies of Paasche and Laspeyres types of indexes, Bennet (1920) recommended a decomposition of cost change by means of the arithmetic mean of (5) and (6). This procedure generates Fisher type arithmetic mean input usage weights in the price index and Fisher type arithmetic mean input price weights in the quantity index. More recently Diewert (1998) has shown that Bennet's input price and quantity indexes satisfy a number of desirable axioms, not all of which are satisfied by (5) or (6).¹ Consequently we base our cost change decomposition on Bennet's approach, which leads to

Proposition 1: The cost change between period t and period $t+1$ decomposes as

$$c^{t+1} - c^t = w^{t+1T} x^{t+1} - w^{tT} x^t$$

$$\begin{aligned}
 &= (1/2)(x^t + x^{t+1})^T(w^{t+1} - w^t) && \text{price effect} \\
 &+ (1/2)(w^t + w^{t+1})^T(x^{t+1} - x^t) && \text{quantity effect}
 \end{aligned}$$

Both the price effect and the quantity effect can be computed directly from the data, with no analysis being required. However a decomposition of the quantity effect requires economic analysis. We begin by decomposing the quantity effect into two components: a *productivity effect* reflecting the combined impact on cost of a change in the cost efficiency of resource use and of a change in the structure of production technology, and an *activity effect* reflecting the impact on cost of a change in the scale of the producer’s operation.

Proposition 2: The **quantity effect** between period t and period t+1 decomposes as

$$\begin{aligned}
 (1/2)(w^t + w^{t+1})^T(x^{t+1} - x^t) &= \\
 (1/2)(w^t + w^{t+1})^T[(x^{t+1} - x^B) - (x^t - x^E)] &&& \text{productivity effect} \\
 + (1/2)(w^t + w^{t+1})^T(x^B - x^E) &&& \text{activity effect}
 \end{aligned}$$

Figure 1 illustrates the decomposition of the quantity effect into a productivity effect and an activity effect. Two cost frontiers are drawn, with $c^{t+1}(y, w^{t+1})$ located beneath $c^t(y, w^t)$, on the assumptions that technical progress has occurred between period t and period t+1, and that the impact of $w^{t+1} \geq w^t$ does not swamp the impact of technical progress, although neither of these assumptions is necessary in the analysis. Observed period t cost $c^t > c^t(y, w^t)$, and observed period t+1 cost $c^{t+1} > c^{t+1}(y, w^{t+1})$, the assumption being that resource use is inefficient in each period. Finally, eliminating the price effect in order to focus on the quantity effect leads to an upward displacement of c^t to $c^{*t} = (1/2)(x^t + x^{t+1})^T w^t$ and to a downward displacement of c^{t+1} to $c^{*t+1} = (1/2)(x^t + x^{t+1})^T w^{t+1}$, the assumption being that $x^{t+1} \geq x^t$.

With the price effect out of the way, we can focus on decomposing the quantity effect. We begin by noting that x^B is a cost minimizing input vector for (y^{t+1}, w^{t+1}) and period t+1 technology, and so $x^B = \nabla_{w^{t+1}} c^{t+1}(y^{t+1}, w^{t+1})$. Similarly x^E is a cost minimizing input vector for (y^t, w^{t+1}) and period t+1 technology, and so $x^E = \nabla_{w^{t+1}} c^{t+1}(y^t, w^{t+1})$. Consequently the productivity effect measures the excess cost of operating above the period t+1 cost frontier $c^{t+1}(y, w^{t+1})$ in period t+1, less the excess cost of operating above the period t+1 cost frontier $c^{t+1}(y, w^{t+1})$ in period t, with both excess costs being evaluated at the arithmetic mean of period t and period t+1 input prices. The productivity effect contributes to a reduction (an increase) in cost from period t to period t+1 according as $[(1/2)(w^t + w^{t+1})^T(x^t - x^E)] > (<)$ $[(1/2)(w^t + w^{t+1})^T(x^{t+1} - x^B)]$.

The activity effect measures the cost change between period t and period t+1 resulting from a change in the scale of the business from $x^E \in I^{t+1}(y^t)$ to $x^B \in I^{t+1}(y^{t+1})$, with both excess costs being evaluated at the arithmetic mean of period t and period t+1 input prices. The activity effect contributes to a reduction (an increase) in cost from period t to period t+1 according as $[(1/2)(w^t + w^{t+1})^T(x^B - x^E)] < (>) 0$.²

It should be apparent from Figure 1 that the productivity effect has cost efficiency change and technical change components. Accordingly we next decompose the productivity effect into a *technical efficiency* effect, an *allocative efficiency* effect and a *technical change effect*.

Proposition 3: The productivity effect between period t and period t+1 decomposes as

$$\begin{aligned}
 (1/2)(w^t + w^{t+1})^T[(x^{t+1} - x^B) - (x^t - x^E)] = \\
 (1/2)(w^t + w^{t+1})^T[(x^{t+1} - x^D) - (x^t - x^C)] & \quad \text{technical efficiency effect} \\
 + (1/2)(w^t + w^{t+1})^T[(x^D - x^B) - (x^C - x^A)] & \quad \text{allocative efficiency effect} \\
 + (1/2)(w^t + w^{t+1})^T(x^E - x^A) & \quad \text{technical change effect}
 \end{aligned}$$

The decomposition of the productivity effect is illustrated in Figure 2. x^t is technically inefficient, since $x^C = \theta^t x^t \in I^t(y^t)$, where $\theta^t = [D^t(y^t, x^t)]^{-1} < 1$. In addition, x^C is allocatively inefficient for w^t , employing too much x_1 and too little x_2 , relative to the cost-efficient input vector x^A . x^{t+1} is also technically inefficient, since $x^D = \theta^{t+1} x^{t+1} \in I^{t+1}(y^{t+1})$, where $\theta^{t+1} = [D^{t+1}(y^{t+1}, x^{t+1})]^{-1} < 1$. In addition, x^D is allocatively inefficient for w^{t+1} , employing too much x_1 and too little x_2 relative to the cost-efficient input vector x^B . The technical efficiency effect measures the impact on cost (evaluated at the arithmetic mean of period t and period t+1 input prices) of any change in technical efficiency between period t and period t+1, and appears in Figure 2 and the second line of Proposition 3 as $(1/2)(w^t + w^{t+1})^T[(x^{t+1} - x^D) - (x^t - x^C)]$. The allocative efficiency effect measures the impact on cost (evaluated at the arithmetic mean of period t and period t+1 input prices) of any change in allocative efficiency between period t and period t+1, and appears in Figure 2 and the third line of Proposition 3 as $(1/2)(w^t + w^{t+1})^T[(x^D - x^B) - (x^C - x^A)]$. The sum of the two efficiency effects appears in Figure 1 as $[(x^{t+1} - x^B) - (x^t - x^A)]$, both differences being evaluated at the arithmetic mean of period t and period t+1 input prices.

We now turn to the impact on cost of technical change. In Figure 2 the cost-efficient input vector $x^A \in I^t(y^t)$, and so $w^{tT} x^A = c^t(y^t, w^t)$. The cost-efficient input vector $x^E \in I^{t+1}(y^t)$, and so $w^{t+1T} x^E = c^{t+1}(y^t, w^{t+1})$. Thus the movement from x^A to x^E represents the impact on cost of technical progress. Since x^E is not necessarily a radial contraction of x^A , technical progress is not restricted to be input-neutral, and the nature of the bias can be determined by comparing the input mix at x^E with the input mix at x^A . In addition, since x^A and x^E are cost-efficient for different input price vectors, we evaluate the cost saving impact of technical progress at the arithmetic mean of period t and period t+1 input prices. Thus the technical change effect appears in Figures 1 and 2 and the final line of Proposition 3 as $(1/2)(w^t + w^{t+1})^T(x^E - x^A)$.

4. Implementing the Intertemporal Cost Change Decomposition

The *measurement* of intertemporal cost change requires information on (y^t, x^t, w^t) and $(y^{t+1}, x^{t+1}, w^{t+1})$, and all of these data are (potentially) observable. However the *decomposition*

of intertemporal cost change requires that the unobserved input quantity vectors $(x^A, x^B, x^C, x^D, x^E)$ be identified so that Propositions 2 and 3 can be implemented. We now show how to use linear programming techniques to calculate each of the unobserved input quantity vectors. We assume that we have a sample consisting of I producers, indexed $i = 1, \dots, I$, each being observed over a sequence of T time periods, indexed $t = 1, \dots, T$. In the following analysis we decompose cost change for producer “ o ”, where “ o ” is a member of I .³

Consider first the calculation of the technical efficiency effect in Proposition 3. This requires the identification of the input quantity vectors x^C in period t and x^D in period $t+1$. Since x^C is a radial contraction of x^t , $x^C = \theta^C x^t = x^t/D^t(y^t, x^t)$, with $\theta^C \leq 1$. The scalar θ^C is determined as the solution to the linear programming problem

$$[D^t(y^{ot}, x^{ot})]^{-1} = \min_{\theta^C, \lambda^t} \theta^C \quad (7)$$

$$\text{subject to} \quad \theta^C x_n^{ot} \geq \sum_{i=1}^I \lambda_i^t x_n^{it}, \quad n = 1, \dots, N$$

$$\sum_{i=1}^I \lambda_i^t y_m^{it} \geq y_m^{ot}, \quad m = 1, \dots, M$$

$$\lambda_i^t \geq 0, \quad i = 1, \dots, I$$

$$\sum_{i=1}^I \lambda_i^t = 1.$$

The constraints of this program guarantee that $x^{oC} = \theta^C x^{ot} \in I^t(y^{ot})$, as it does in Figure 2.

Since x^D is a radial contraction of x^{t+1} , $x^D = \theta^D x^{t+1} = x^{t+1}/D^{t+1}(y^{t+1}, x^{t+1})$, with $\theta^D \leq 1$. The scalar θ^D is determined as the solution to a linear programming problem identical in form to (7), with period t data $(x^{it}, y^{it}, x^{ot}, y^{ot})$ being replaced with period $t+1$ data $(x^{it+1}, y^{it+1}, x^{ot+1}, y^{ot+1})$. The constraints of this program guarantee that $x^{oD} = \theta^D x^{ot+1} \in I^{t+1}(y^{ot+1})$, as it does in Figure 2.

Consider next the calculation of the allocative efficiency effect in Proposition 3. This requires the identification of the input quantity vectors x^A in period t and x^B in period $t+1$. Since x^A minimizes the cost of producing outputs y^t when input prices are w^t and period t technology is in place, x^A can be identified as the solution to the linear programming problem

$$c^t(y^{ot}, w^{ot}) = \min_{x, \lambda^t} w^{ot} x \quad (8)$$

$$\text{subject to} \quad x_n \geq \sum_{i=1}^I \lambda_i^t x_n^{it}, \quad n = 1, \dots, N$$

$$\sum_{i=1}^I \lambda_i^t y_m^{it} \geq y_m^{ot}, \quad m = 1, \dots, M$$

$$\lambda_i^t \geq 0, \quad i = 1, \dots, I$$

$$\sum_{i=1}^I \lambda_i^t = 1.$$

Grifell-Tatjé, E. and C.A.K. Lovell (2000), “Cost and Productivity,” *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.

DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

The constraints of this program guarantee that $w^{otT}x^{oA} = c^t(y^{ot}, w^{ot})$, as it does in Figures 1 and 2.

Since x^B minimizes the cost of producing outputs y^{t+1} when input prices are w^{t+1} and period $t+1$ technology is in place, x^B can be identified as the solution to a linear programming problem identical in form to (8), with period t data $(x^{it}, y^{it}, y^{ot}, w^{ot})$ being replaced with period $t+1$ data $(x^{it+1}, y^{it+1}, y^{ot+1}, w^{ot+1})$. The constraints of this program guarantee that $w^{ot+1T}x^{oB} = c^{t+1}(y^{ot+1}, w^{ot+1})$, as it does in Figures 1 and 2.

Finally we identify the input quantity vector x^E which, together with x^B and x^A , enables us to identify both the activity effect in Proposition 2 and the technical change effect in Proposition 3. Since x^E minimizes the cost of producing output y^t when input prices are w^{t+1} and period $t+1$ technology is in place, x^E can be identified as the solution to a linear programming problem identical in form to (8), with period t data $(x^{it}, y^{it}, y^{ot}, w^{ot})$ being replaced with mixed-period data $(x^{it+1}, y^{it+1}, y^{ot}, w^{ot+1})$. The constraints of this program guarantee that $w^{ot+1T}x^E = c^{t+1}(y^{ot}, w^{ot+1})$, as it does in Figures 1 and 2.

Solving each of the five linear programming problems of forms (7) and (8) I times, once for each producer in the sample, generates all the information required to identify the five unobserved input quantity vectors $(x^A, x^B, x^C, x^D, x^E)$. When combined with the observed input quantity vectors (x^t, x^{t+1}) , and with the observed input price vectors (w^t, w^{t+1}) , the observed cost change $(c^{t+1} - c^t)$ can be decomposed into its five sources by means of Propositions 1-3.

5. Conducting a Cross - Sectional Benchmarking Exercise

Businesses frequently evaluate their performance by benchmarking against some target performance. Although such a benchmarking exercise is typically conducted on the basis of key financial ratios, such as return on assets, it can perhaps more usefully be conducted on the basis of the *economic determinants*, rather than the *financial consequences*, of performance variation. We now show how the techniques developed in Sections 3 and 4 can be adapted for use in a cross-sectional framework for the purpose of conducting a benchmarking exercise.

Suppose that we observe a sample of I producers, indexed $i = 1, \dots, I$ as before, but during a single time period, so we dispense with the time index. The previous cost change decomposition exercise now becomes a *cost gap* or *cost variance* decomposition exercise, the objective being to decompose the gap between the cost of a producer, c^o , and the cost of a benchmark target, c^* . Although the benchmark can be determined in a number of ways, we consider two possibilities. The first possibility is to define the benchmark as the producer having the lowest unit cost (c/y) in the sample. We refer to this model as the “low-cost benchmark” model. This model is easy to implement in the single output case, although it requires the construction of an output quantity index in the multiple output case. The second possibility is to define the benchmark as a producer identified as being cost-efficient, being located on rather than above the cost frontier $c(y, w)$. We refer to this model as the “cost-efficient benchmark” model. This model is easy to implement even in the multiple

Grifell-Tatjé, E. and C.A.K. Lovell (2000), “Cost and Productivity,” *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.

DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

output case. In both models we associate the benchmarking producer “^o” with period t, and the benchmark target producer “^{*}” with period t+1, in Figures 1 and 2.

5.1 The Low-Cost Benchmark Model

As in Section 3, the cost frontier $c(y,w)$ provides an analytical foundation for the cost gap analysis. In contrast to Section 3, we have only one time period and one cost frontier, and so the technical change effect disappears with the time index.⁴ In Figures 1 and 2 the absence of technical change means that $x^E = x^A$. The benchmarking producer has observed cost $w^{oT}x^o \geq c(y^o, w^o)$, and the low-cost benchmark has observed cost $w^{*T}x^* \geq c(y^*, w^*)$. The low-cost benchmark is not necessarily cost-efficient, since its low cost may be due to relatively low input prices. We are interested in the cost gap $c^o - c^* = (w^{oT}x^o - w^{*T}x^*)$. Adaptation of Propositions 1 - 3 suggests that the cost gap may be due to any of four factors:

i) the *price effect* $[(1/2)(x^o + x^*)^T(w^* - w^o)]$ measures that part of the cost gap attributable to input price differences between the two producers;

ii) the *technical efficiency effect* $[(1/2)(w^o + w^*)^T(x^* - x^D) - (1/2)(w^o + w^*)^T(x^o - x^C)]$ measures that part of the cost gap attributable to the difference between the technical inefficiency of the two producers;

iii) the *allocative efficiency effect* $[(1/2)(w^o + w^*)^T(x^D - x^B) - (1/2)(w^o + w^*)^T(x^C - x^A)]$ measures that part of the cost gap attributable to the difference between the allocative inefficiency of the two producers;

iv) the *activity effect* $[(1/2)(w^o + w^*)^T(x^B - x^A)]$ measures that part of the cost gap attributable to difference between the cost-efficient size of the two producers.

The cost gap analysis is based on observed data (y^o, x^o, w^o) and (y^*, x^*, w^*) , and also on unobserved data (x^D, x^B) and (x^C, x^A) . The four unobserved input quantity vectors can be retrieved by solving a pair of atemporal versions of the two linear programming problems (7) and (8) in Section 4. The corresponding unobserved costs are then obtained by multiplying each respective input quantity vector by $(1/2)(w^o + w^*)$. In this manner a complete decomposition of the cost gap is obtained, for each benchmarking producer and the low-cost benchmark.

5.2 The Cost-Efficient Benchmark Model

The analysis simplifies considerably when the benchmark is cost-efficient. This is because, as can readily be determined from Figures 1 and 2, the benchmark has $x^* = x^D = x^B$. In this case the *price effect* is unchanged from its version in Section 5.1, but the *technical efficiency effect* becomes $[(1/2)(w^o + w^*)^T(x^o - x^C)]$, the *allocative efficiency effect* becomes $[(1/2)(w^o + w^*)^T(x^C - x^A)]$, and the *activity effect* becomes $[(1/2)(w^o + w^*)^T(x^* - x^A)]$. In this model the observed data remain (y^o, x^o, w^o) and (y^*, x^*, w^*) , but the unobserved data which must be retrieved reduce to (x^C, x^A) . Consequently only two linear programming problems need to be solved: atemporal versions of (7) and (8). Substituting the solutions to these problems into the four effects generates a complete decomposition of the cost gap between each benchmarking producer and the cost-efficient benchmark.

6. An Empirical Application to US Electric Power Generation

Our data set consists of a balanced panel of 93 problem-free investor-owned electric power generating companies observed in 1977, 1982, 1987 and 1992. The variables include total generating cost (current dollars), a single output (net volume, in Mwh, of electricity generated), and quantities and nominal prices of three inputs, labor, fuel and capital. The quantity of labor is the total (full time plus one half of part time) number of employees engaged in generation activities, and the price of labor is calculated as the ratio of total labor cost to the total number of employees. The quantity of fuel is measured in MBTU of coal, oil, natural gas and nuclear fuel, and the price of fuel is the delivered cost (in cents) per MBTU. The quantity and price of capital are measured as multilateral Törnqvist indexes of the quantities and prices of capital services employed in various generation activities.

The data underlying the analysis are summarized in Table 1, which reports annual means and standard deviations of all variables used in the analysis. The first striking feature of the data is the pattern of intertemporal variation in mean generating cost. Mean generating cost more than doubled from 1977 to 1982, despite a mere one percent increase in mean electricity generation. Mean generating cost increases moderated from 1982 to 1987 (23%), and from 1987 to 1992 (5%), and were much more in line with mean electricity generation increases during these periods (17% and 7% respectively). There is thus an intertemporal cost variation story to be told, and not all of the explanation lies with variation in mean fuel prices, which varied wildly as a consequence of both political and economic developments in international petroleum markets (the oil price shocks of the 1970s), and domestic regulatory developments (the 1977 Amendment to the 1970 Clean Air Act).

The second striking feature of the data is the large variation in generating cost across firms within each year. In each of the four years of the panel the standard deviation of generating cost is virtually the same as the mean generating cost itself. It appears that much of this cost variation is attributable to size variation, since in each year the standard deviations of output and the three inputs are nearly as large as the means of output and the three inputs, although there remains plenty of room for alternative explanations, and thus for interesting benchmarking exercises.

The third feature of the data is the considerable amount of input substitution that has occurred in the industry. The mean capital/fuel and capital/labor ratios both increased over six-fold during the period, whereas the mean labor/fuel ratio, after increasing by 15% from 1977 to 1982, declined thereafter and increased by barely four percent during the entire period. The extent to which these substitution patterns were driven by input price changes or by changing regulatory requirements is a matter of speculation, although our cost decomposition will shed light on the question.

6.1 The Intertemporal Cost Change Decomposition

We begin by running the data through the sequence of five linear programming problems specified in (7) and (8). We then insert calculated values of the unobserved input quantity vectors $(x^A, x^B, x^C, x^D, x^E)$ into the intertemporal cost change decompositions in

Grifell-Tatjé, E. and C.A.K. Lovell (2000), "Cost and Productivity," *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.

DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

Propositions 1 - 3. This enables us to decompose intertemporal cost change for each utility. Means and standard deviations of calculated cost change and its sources appear in Table 2.

Between 1977 and 1982 cost increased by an average of \$344.2 million, or by an average of 150%. Three effects contributed the lion's share of the mean cost increase. Over 40% of the cost increase was attributable to the price effect, which resulted from a 64% increase in the input price index. Nearly one third of the cost increase was attributable to the technical change effect, which resulted from a 48% increase in cost-efficient input usage ($x^A - x^E$). The technical change effect provides a measure of the cost of a deterioration in best practice performance. This deterioration was likely a combination of rate-base padding in response to rate of return regulation and investment in pollution abatement capital equipment in response to environmental regulation.⁵ As a result, the mean capital input nearly quintupled from 1977 to 1982. Nearly one fourth of the cost increase was attributable to the technical efficiency effect, which resulted from a 35% increase in the divergence between actual and technically efficient input use $[(x^{t+1} - x^D) - (x^t - x^C)]$.⁶ The allocative efficiency effect made a very small contribution to the cost increase; this is evident from Table 1, where it is clear that utilities substituted capital for fuel and, to a lesser extent, for labor, in response to changing input price ratios.⁷ The activity effect made virtually no contribution to the cost increase; this can be explained by the fact that mean output barely changed, and so cost-efficient input use ($x^B - x^E$) barely changed (despite the fact that actual input use increased substantially).

The nature of the cost change and its decomposition changed dramatically during the second and third periods. Cost increases abated, due primarily to declines in the input price index driven by substantial declines in fuel prices. Mean cost efficiency also improved between 1982 and 1987, and declined only marginally between 1987 and 1992.⁸ The primary driver of cost increases in the two latter periods was the activity effect, reflecting increases in cost-efficient input usage in response to steady increases in electricity demand. Finally, the technical change effect contributed to cost increases through 1987, reflecting the continuing installation of costly pollution abatement capital equipment, although this process seems to have been completed by 1987.⁹

6.2 A Pair of Multilateral Benchmarking Exercises

We now conduct a pair of benchmarking exercises using 1992 data, one against the low-cost utility and the other against a cost-efficient utility. The former exercise requires identifying the low-cost utility, which in 1992 was Kentucky Power, with an average generating cost of \$18/Mwh, less than half the sample mean of \$43/Mwh. We then ran the 1992 data through four linear programming problems (7) and (8) and inserted the solutions into Propositions 1 - 3 to obtain a decomposition of the cost gap between each benchmarking utility and the low-cost benchmark. Results of this exercise are summarized in Table 3, which reports means and standard deviations of the cost gap and its four components.

Somewhat surprisingly, the low-cost utility serves as a useful benchmark. Less than 20% of the enormous mean cost gap of \$619 million can be attributed to the price effect; the benchmark is low-cost only partly because it faces relatively low input prices. The main reason the benchmark is low-cost is its relatively small size; 42% of the mean cost gap is attributable to the activity effect. Although benchmarking utilities can learn nothing of value

Grifell-Tatjé, E. and C.A.K. Lovell (2000), "Cost and Productivity," *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.

DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

from the benchmark's small size, and little from its favorable input price structure, the benchmark's superior cost efficiency does provide ample opportunity for learning. Nearly 40% of the mean cost gap is attributable to the technical and allocative efficiency effects; since these effects are presumably under the control of management, significant cost savings are available to benchmarking utilities.

The latter exercise requires identifying a cost-efficient utility against which to benchmark. Program (8) identified a total of six cost-efficient utilities in 1992, and so we somewhat arbitrarily selected the largest, Texas Utilities Electric, as the cost-efficient benchmark. We then ran the 1992 data through the two linear programming problems (7) and (8) and inserted the solutions into Propositions 1 - 3 to obtain a decomposition of the cost gap between each benchmarking utility and the cost-efficient benchmark. The results of this exercise are summarized in Table 4, which reports means and standard deviations of the cost gap and its four components.

The cost-efficient utility also serves as a useful benchmark, but in a different way. The large negative mean cost gap is due primarily to the fact that the benchmark is large, and partly to the fact that it faces relatively high input prices. Once again, benchmarking utilities have little if anything to learn from these components of the mean cost gap. However the unfavorable mean cost efficiency effect is over four times as large as the favorable mean price effect, once again suggesting ample opportunity for benchmarking utilities to find significant cost savings through waste elimination and resource reallocation.

7. Summary and Conclusions

We have developed a theory-based decomposition of variation in operating cost, both through time and across producers. This decomposition draws from, and extends, approaches developed in the linear programming literature. A virtue of this decomposition is that it distinguishes sources of cost variation which are arguably beyond management's control from those which are deserving of management's attention. The decomposition is based on observed and unobserved input quantity vectors, and we have shown how to derive the unobserved input quantity vectors from the observed input quantity vectors.

We have applied our cost decomposition analysis to a panel of US electric power generating companies, and we found substantial variation in operating cost, both through time and across producers. In an intertemporal cost decomposition exercise we identified two effects as having been the primary cost drivers: the input price effect and the technical change effect, the latter of which we associated with regulatory change. In a pair of multilateral benchmarking exercises we attributed a significant portion of the mean cost gap with the superior cost efficiency of both the low-cost benchmark and the cost-efficient benchmark. This finding suggests that benchmarking can be a profitable exercise for utilities, since not all of the mean cost gap is due to exogenous factors beyond the control of their managements.

Footnotes

* A previous version of this paper was presented at the Fifth European Workshop on Efficiency and Productivity Analysis held at the Royal Veterinary and Agricultural University, Copenhagen, Denmark, October 9-11, 1997. We are grateful to Comisión del Sistema Eléctrico Nacional (CSEN) and DGICYT PB94-0708, DGES PB95-0616 and Generalitat de Catalunya 19965GR 00113 for generous financial support, to Dr. Herbert Thompson for providing the data used in this study, to Julie Millington and Yih Pin Tang for their exceptional research assistance, and to two referees whose comments we found unusually helpful.

1. Just as a Fisher index is defined as a geometric mean of Paasche and Laspeyres indexes when the latter are expressed in ratio form, a Bennet index is defined as an arithmetic mean of Paasche and Laspeyres indexes when the latter are expressed in difference form. In addition, just as a Fisher index is a superlative index, Diewert (1998) calls a Bennet index a "pseudosuperlative" index because it has the same first and second partial derivatives as a superlative index around an equal price and quantity vector.

2. When $M = 1$ the nature of scale economies along the path from x^E to x^B can be inferred from the ratio $R = [(1/2)(w^t + w^{t+1})^T(x^B/y^{t+1}) / (1/2)(w^t + w^{t+1})^T(x^E/y^t)]$. Returns to scale are locally decreasing, constant or increasing along this path according as $R \geq < 1$. If $M > 1$ y^t must be replaced with an output quantity index.

3. These linear programming techniques were developed by Banker et al (1984) and Färe et al (1985).

4. As a referee has correctly noted, different firms can have different technologies in a given year, so in principle a technical change effect is possible. As a practical matter, however, it is not possible to construct different cost frontiers for different firms in a given year, since there is only one observation on each firm in a given year.

5. Gollop and Roberts (1983) found similar regulatory impacts on a sample of 56 utilities during the 1973-1979 period. They found that environmental regulations were capital-using, and forced utilities to substitute high-cost low-sulfur fuel for low-cost high-sulfur fuel. Both effects had a dampening impact on growth rates during the period. More recently Yaisawarng and Klein (1994) found sulfur dioxide controls to have dampened productivity growth in the 1980s.

6. Although the divergence between actual and technically efficient input use increased, technical efficiency itself actually improved, from a 1977 mean of 0.623 to a 1982 mean of 0.644. This apparent anomaly is due to the fact that the technical efficiency effect is based on differences, while technical efficiency is a ratio.

7. Many studies (e.g., Gollop and Karlson (1978), Atkinson and Halvorsen (1980)) have found evidence of substantial allocative inefficiency during the 1970s, and have attributed these inefficiencies to rate base padding and the fuel adjustment mechanisms. Our finding

Grifell-Tatjé, E. and C.A.K. Lovell (2000), “Cost and Productivity,” *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.

DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

suggests that the degree of input misallocation worsened only marginally during the 1977-1982 period.

8. The post-1982 improvement in cost efficiency was no doubt driven in part by the Public Utilities Regulatory Policies Act of 1978, which enhanced competition in electricity generation.

9. The regulatory factors we have associated with technical regress were costly, contributing nearly one third of the mean cost increase between 1977 and 1982, and contributing nearly 150% of the mean cost increase between 1982 and 1987. They were also biased with respect to their input requirements. Cost-efficient capital/labor ratios increased fivefold, and cost-efficient capital/fuel ratios increased fourfold, on average between 1977 and 1982. The biases diminished in magnitude, and changed directions, between 1982 and 1987.

Grifell-Tatjé, E. and C.A.K. Lovell (2000), "Cost and Productivity," *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.

DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

References

Afriat, S. N. (1972), "Efficiency Estimation of Production Functions," *International Economic Review* 13:3 (October), 568-98.

Atkinson, S. E., and R. Halvorsen (1980), "A Test of Relative and Absolute Price Efficiency in Regulated Utilities," *Review of Economics and Statistics* 62:1 (February), 81-88.

Banker, R. D., A. Charnes and W. W. Cooper (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science* 30:9 (September), 1078-92.

Bennet, T. L. (1920), "The Theory of Measurement of Changes in Cost of Living," *Journal of the Royal Statistical Society* 83, 455-62.

Diewert, W. E. (1998), "Index Number Theory Using Differences Rather than Ratios," Discussion Paper No. 98-10, Department of Economics, University of British Columbia, Vancouver V6T 1Z1, Canada.

Färe, R., S. Grosskopf and C. A. K. Lovell (1985), *The Measurement of Efficiency of Production*. Boston: Kluwer-Nijhoff Publishing.

Gollop, F. M., and S. H. Karlson (1978), "The Impact of the Fuel Adjustment Mechanism on Economic Efficiency," *Review of Economics and Statistics* 60:4 (November), 574-84.

Gollop, F. M., and M. J. Roberts (1983), "Environmental Regulations and Productivity Growth: The Case of Fossil-Fueled Electric Power Generation," *Journal of Political Economy* 91:4 (August), 654-74.

Greene, W. (1980), "Maximum Likelihood Estimation of Econometric Frontier Functions," *Journal of Econometrics* 13:1 (May), 27-56.

Grifell-Tatjé, E., and C. A. K. Lovell (1999), "Profits and Productivity," *Management Science* 45:9 (September).

Hanoch, G., and M. Rothschild (1972), "Testing the Assumptions of Production Theory: A Nonparametric Approach," *Journal of Political Economy* 80:2 (March/April), 256-75.

Kumbhakar, S. C. (1997), "Modelling Allocative Inefficiency in a Translog Cost Function and Cost Share Equations: An Exact Relationship," *Journal of Econometrics* 76:1/2 (January/February), 351-56.

Malmquist, S. (1953), "Index Numbers and Indifference Surfaces," *Trabajos de Estadística* 4, 209-42.

Grifell-Tatjé, E. and C.A.K. Lovell (2000), "Cost and Productivity," *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.

DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

Nerlove, M. (1963), "Returns to Scale in Electricity Supply," in C. Christ et al., eds., *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*. Stanford, CA: Stanford University Press.

Schmidt, P., and C. A. K. Lovell (1979), "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers," *Journal of Econometrics* 9:3 (February), 343-66.

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

Stevenson, R. E. (1980), "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics* 13:1 (May), 58-66.

Yaisawarng, S., and J. D. Klein (1994), "The Effects of Sulfur Dioxide Controls on Productivity Change in the U.S. Electric Power Industry," *Review of Economics and Statistics* 76:3 (August), 447-60.

Grifell-Tatjé, E. and C.A.K. Lovell (2000), “Cost and Productivity,” *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.
 DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

		1977	1982	1987	1992
Total Cost (\$ million)	Mean	229.79	573.98	707.01	744.70
	SD	237.99	545.71	689.12	760.06
Output (million Mwh)	Mean	13.70	13.86	16.18	17.27
	SD	13.11	12.98	15.16	16.25
Labor (x₁) (#)	Mean	1,373	1,797	1,995	2,021
	SD	1,412	1,751	2,115	2,206
Fuel (x₂) (000 MBTU)	Mean	1,288.29	1,440.61	1,677.43	1,823.54
	SD	1,186.28	1,430.77	1,538.72	1,643.46
Capital (x₃)	Mean	44,756	211,622	371,041	396,386
	SD	41,807	183,817	385,629	442,892
w₁ (\$)	Mean	14,208	22,594	31,233	39,949
	SD	2,508	3,845	5,384	8,399
w₂ (cents)	Mean	125	226	162	144
	SD	49	113	46	44
w₃	Average	1,028	1,004	1,002	1,005
	SD	214	84	86	109

Table 1. Data Summary

Grifell-Tatjé, E. and C.A.K. Lovell (2000), “Cost and Productivity,” *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.

DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

		1982-1977		1987-1982		1992-1987	
		\$ million	%	\$ million	%	\$ million	%
Cost Change	Mean	344.2	149.79	133.0	23.18	37.7	5.33
	SD	355.1		318.5		181.5	
Price Effect	Mean	147.8	64.26	-101.0	-17.59	-14.2	-2.00
	SD	205.0		238.2		46.1	
Activity Effect	Mean	0.4	0.15	84.4	14.69	54.0	7.63
	SD	45.1		190.4		225.9	
Technical Efficiency Effect	Mean	81.4	35.50	-40.8	-7.12	43.1	6.09
	SD	142.3		226.5		145.9	
Allocative Efficiency Effect	Mean	4.6	1.95	-11.2	-1.95	-39.2	-5.54
	SD	69.5		237.4		192.5	
Technical Change Effect	Mean	110.0	47.93	201.6	35.11	-6.1	-0.86
	SD	88.7		315.1		116.0	

Table 2. Intertemporal Cost Change Decomposition

		\$ million	%
Cost Gap	Mean	619.3	494.08
	SD	760.1	
Price Effect	Mean	115.8	92.37
	SD	154.0	
Activity Effect	Mean	259.2	206.75
	SD	526.3	
Technical Efficiency Effect	Mean	164.5	131.20
	SD	169.1	
Allocative Efficiency Effect	Mean	79.9	63.77
	SD	113.1	

Table 3. 1992 Cost Gap Decomposition Using the Low-Cost Benchmark

		\$ million	%
Cost Gap	Mean	-2,524.0	-77.22
	SD	760.1	
Price Effect	Mean	-74.7	-2.28
	SD	107.6	
Activity Effect	Mean	-2,764.6	-84.58
	SD	679.1	
Technical Efficiency Effect	Mean	214.5	6.56
	SD	218.5	
Allocative Efficiency Effect	Mean	100.8	3.08
	SD	141.6	

Table 4. 1992 Cost Gap Decomposition Using a Cost-Efficient Benchmark

Grifell-Tatjé, E. and C.A.K. Lovell (2000), “Cost and Productivity,” *Managerial and Decision Economics* vol. 21, issue 1, pages 19 – 30.
DOI: 10.1002/1099-1468(200001/02)21:1<19::AID-MDE962>3.0.CO;2-7

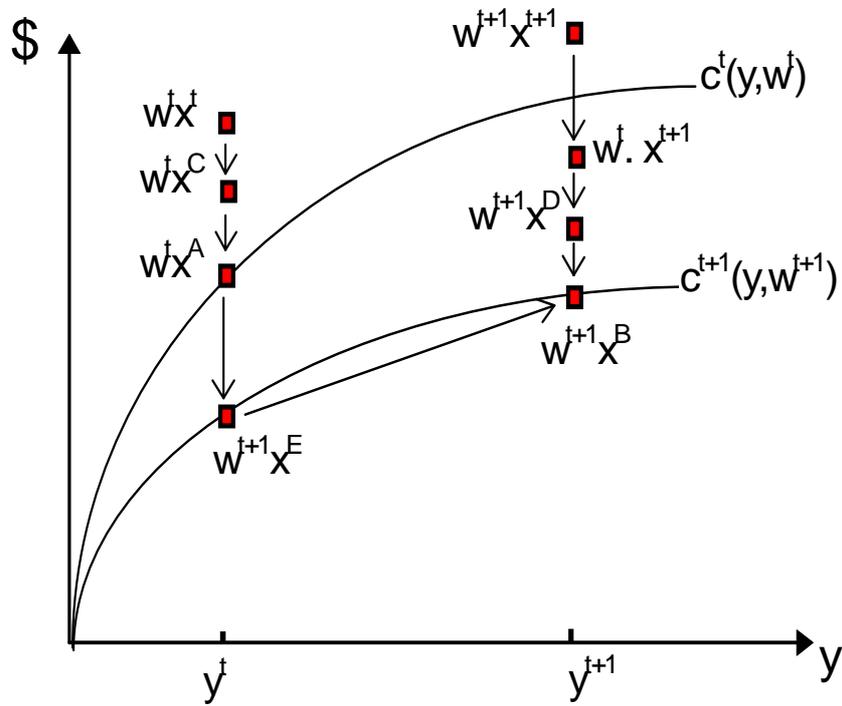


Figure 1

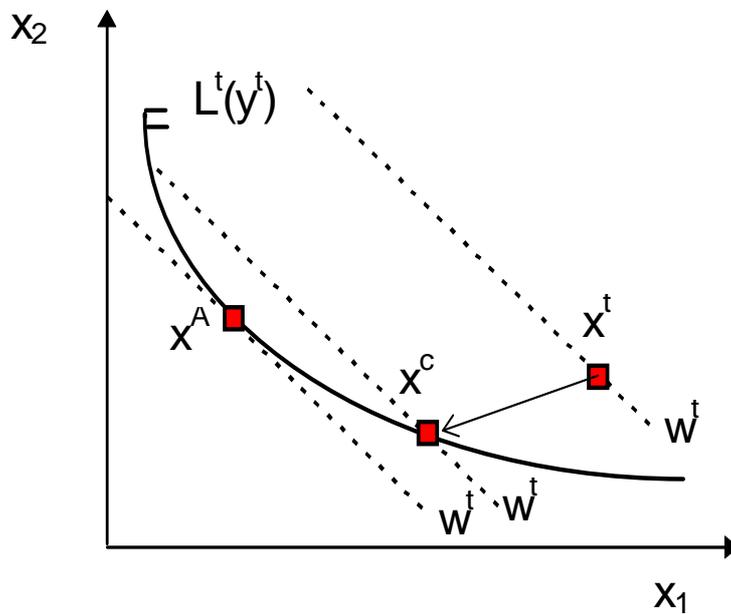


Figure 2

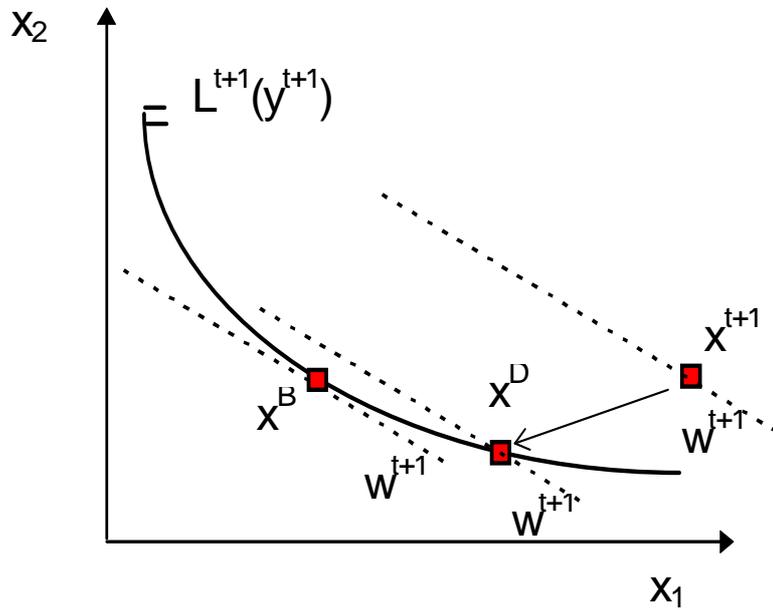


Figure 3

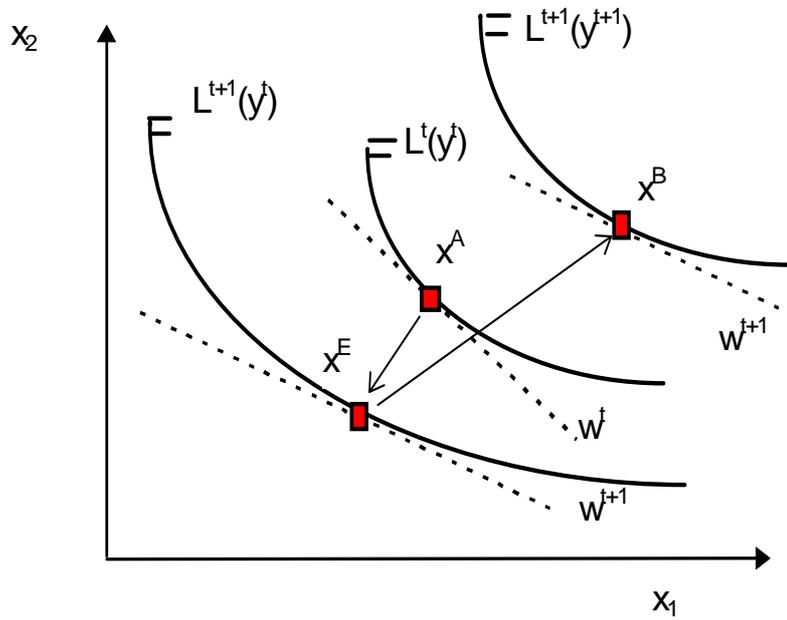


Figure 4