

**On the determinants of cost performance and the decomposition  
of returns to scale measures in the presence of quasi-fixed inputs.  
A comment on Morrison and Schwartz (1996) and related work**

**Angel de la Fuente\***

**Instituto de Análisis Económico (CSIC)**

**September 1999**

**UAB-IAE WP 445.00**

**Abstract**

In this note I amend and clarify the framework for cost analysis developed and applied by Morrison and Schwartz (M&S) in a series of papers. I propose a consistent decomposition of a cost-based measure of short-run returns to scale and develop a version of M&S's dynamic decomposition that makes transparent its nature and usefulness as a cost-accounting device.

---

\* I would like to thank Xavier Vives for his useful comments and suggestions. This research has been financed in part by the Spanish Ministry of Education and Culture through DGICYT grants no. PB95-0130 and SEC99-1189

*Address* Instituto de Análisis Económico (CSIC), Campus de la Universidad Autónoma de Barcelona, 08193, Bellaterra, Barcelona, Spain. *Tel:* 93-580-6612. *Fax:* 93-580-1452. *E-Mail address:* delafuente@cc.uab.es.

## 1.- Introduction

In a number of papers, Morrison and Schwartz (M&S) and various co-authors<sup>1</sup> have developed and applied a framework for empirical cost analysis that attempts to disentangle the separate effects of technical progress, returns to scale and the existence of quasi-fixed and unpaid inputs on the evolution of costs over time and on their responsiveness to output. In this note I will argue that, while the approach they propose is potentially quite useful, it needs clarification, and that some of their theoretical results are incorrect and must be amended.

Morrison and Schwartz's theoretical contribution can be largely summarized by two formulas or decompositions, one static and one dynamic, that are then combined in a way that can be used to gain considerable insight into the determinants of cost performance. The usefulness of the dynamic formula, however, is far from apparent on a first reading. As stated in their work, M&S's formula shows how to go from an incorrect measure of technical change to the rate of cost reduction induced by (properly measured) technical progress at constant levels of output and factor stocks (and constant input prices). The authors start out from a measure of total factor productivity (TFP) growth computed using Solow's (1957) growth accounting formula in a context where it does not hold, and then add to it a series of terms that correct for violations of Solow's assumptions and take us from TFP growth to the induced cost reduction. At first sight, the exercise seems pointless, as the correct rate of technical change could have been computed directly using the parameters they estimate in the empirical part of the paper. It is only when one realizes, with extremely little help from the authors, that the incorrect TFP growth measure delivered by Solow's formula happens to be equal to the rate of reduction of unit costs at constant factor prices that the decomposition begins to make sense. The correction factors for the biases in Solow's formula then acquire economic meaning as they turn out to capture the contributions of non-constant returns to scale and quasi-fixed factors to the firm's cost performance. The presentation, however, seems to be carefully designed to hide this crucial fact and is made unnecessarily obscure by the authors' loose use of language and by their failure to define their terms with precision.<sup>2</sup>

M&S's second formula is a decomposition of a cost-based measure of short-run returns to scale into a long-run or technological measure of returns to scale and two additional factors that are meant to capture the distortions caused by the existence of quasi-fixed inputs. Inserted into the previous formula, this "result" allows us to isolate a pure scale factor in the evolution of costs, which captures the constraint imposed by technology as such, and to formalize the intuition that it is important for cost performance whether or not the stocks of quasi-fixed factors "keep up" with output. Unfortunately, the formula is incorrect, although it can be salvaged as we will see later. The proposed

---

<sup>1</sup> See among others Morrison and Schwarz (1994, 1996) and Morrison and Siegel (1999).

<sup>2</sup> The authors use the terms "productivity" and "technical change" to refer indistinctly to cost performance and to correct and incorrect measures of TFP growth, and carefully avoid explicit definitions of the relevant variables, giving instead various expressions for their computation that may or may not be equal for conceptually different variables depending on technological and fixity assumptions.

decomposition involves elasticities of functions that are never defined. If one attempts to define them in what seems to be the natural way in view of the informal discussion, they may fail to exist. And when they do exist, the proposed expression collapses because the short- and long-term cost elasticities are the same when the firm is on its long-run cost schedule. The authors are aware of this last fact and seem to have in mind a situation in which the firm is off the long-run cost curve, without realizing that in this case the proposed formula does not apply, as there is no systematic relationship between the slopes of the short and long-run cost curves away from their tangency.

In the remainder of this note I will provide a derivation of M&S's dynamic decomposition that will make apparent its nature and usefulness as a "cost accounting" device, and propose an amended version of their static formula that captures their intuition in a consistent manner. The paper is organized as follows. In Section 2, I will consider the short- and long-run cost-minimization problems faced by a firm in the presence of two quasi-fixed inputs, one of which is public capital provided free of charge by the government, and collect some useful results. Building on these results, in Section 3 I will discuss the shortcomings of M&S's static decomposition and suggest one possible way to get around them. Section 4 then analyzes the determinants of the evolution of costs over time. I derive a version of M&S's dynamic formula by computing the rate of reduction of unit costs at constant factor prices, and show that this last variable coincides with the growth rate of the standard Solow residual. Section 5 concludes.

## 2.- Some results on cost functions

Consider a firm that operates in competitive factor markets and is endowed with a production function of the form

$$(1) Y = F(K, L, P, t)$$

where  $Y$  is output,  $K$  the stock of private capital,  $L$  the level of employment,  $P$  the stock of public capital and  $t$  time, which is included as an argument of the production function to capture exogenous technical progress. It will be assumed that  $K$  is fixed in the short run and that  $P$  is an unpaid input provided free of charge by the government. The firm's short-run cost minimization problem can be written:

$$(P.S) C(w, r, K, P, Y, \lambda, t) = G(w, K, P, Y, \lambda, t) + rK = \\ = \min_L \{wL \text{ s.t. } F(\lambda K, \lambda L, P, t) = Y, \text{ and } K, P \text{ given}\} + rK$$

where  $G()$  is the variable cost function,  $w$  the wage rate and  $r$  the rental rate or user cost of capital. Notice that we have introduced an auxiliary scale parameter,  $\lambda$ , that will be used to relate cost elasticities to output elasticities. All the computations will be made under the assumption that  $\lambda = 1$ .

The Lagrangian associated with the minimization of variable costs is given by

$$\mathcal{L} = wL + \mu[Y - F(\lambda K, \lambda L, P, t)]$$

and the first-order condition for an optimum requires that

$$(2) w = \mu F_L.$$

By the envelope theorem, it follows that the marginal cost is given by the multiplier,  $\mu$ :

$$(3) \frac{\partial C}{\partial Y} = \frac{\partial G}{\partial Y} = \frac{\partial \mathcal{E}}{\partial Y} = \mu.$$

Using this expression, the elasticity of total cost with respect to output is given by the ratio of marginal to average cost,

$$(4) \varepsilon_{C,Y} \equiv \frac{\partial C}{\partial Y} \frac{Y}{C} = \frac{\mu}{C/Y},$$

where the symbol " $\equiv$ " is used to emphasize that the first part of (4) is a definition.

The output elasticity of the cost function is often used as a measure of returns to scale, as cost responsiveness to changes in output will depend inversely on output's responsiveness to changes in factor inputs. To relate  $\varepsilon_{C,Y}$  to the relevant "primal" production function elasticities, we can proceed as follows. Using the first-order condition (2) and equation (4), we have

$$(5) \frac{\partial C}{\partial \lambda} \frac{1}{C} = \frac{-1}{C} (\mu F_K K + \mu F_L L) = \frac{-1}{C} (\mu F_K K + wL) = -\frac{wL}{C} - \frac{\mu}{C/Y} \frac{F_K K}{Y} = - (S_L + \varepsilon_{C,Y} \varepsilon_{Y,K})$$

where  $S_L \equiv wL/C$  is the share of labour in total cost and  $\varepsilon_{Y,K} \equiv F_K K/Y$  the elasticity of output with respect to labour input. Setting the second and last terms in (5) equal to each other and operating, we have

$$S_L + \varepsilon_{C,Y} \varepsilon_{Y,K} = \frac{\mu}{C} (F_K K + F_L L) = \frac{\mu}{C/Y} \frac{F_K K + F_L L}{Y} = \varepsilon_{C,Y} (\varepsilon_{Y,K} + \varepsilon_{Y,L})$$

from where

$$(6) \varepsilon_{C,Y} = \frac{S_L}{\varepsilon_{Y,L}}.$$

Hence, the output elasticity of the short-run cost function depends on the inverse of the elasticity of output with respect to labour (which determines the elasticity of the variable cost function) and the share of variable costs in total costs.

We will now define and compute a series of other variables that will play a role in the derivation of M&S's formulae and relate them to their "primal" counterparts. We define the shadow prices  $Z_k$  and  $Z_p$  of the fixed inputs as their marginal contributions to the reduction of variable costs. Using the envelope theorem again, we have

$$(7) Z_k \equiv -\frac{\partial G}{\partial K} = -\frac{\partial \mathcal{E}}{\partial K} = \mu F_K$$

$$(8) Z_p \equiv -\frac{\partial G}{\partial P} = -\frac{\partial \mathcal{E}}{\partial P} = \mu F_P.$$

By analogy with the share of labour in total cost, we define the shadow shares of the fixed inputs by

$$(9) S_k^* \equiv \frac{Z_k K}{C} \quad \text{and} \quad S_p^* \equiv \frac{Z_p P}{C}$$

and observe that they are related to the relevant output elasticities in the same way as  $S_L$ :

$$(10) \varepsilon_{Y,K} \equiv F_K \frac{K}{Y} = \frac{\mu F_K K}{C} \frac{C/Y}{\mu} = \frac{Z_k K}{C} \frac{C/Y}{\mu} \equiv \frac{S_k^*}{\varepsilon_{C,Y}}$$

$$(11) \varepsilon_{Y,P} \equiv F_P \frac{P}{Y} = \frac{\mu F_P P}{C} \frac{C/Y}{\mu} = \frac{Z_p P}{C} \frac{C/Y}{\mu} \equiv \frac{S_p^*}{\varepsilon_{C,Y}}$$

where we have made use of (4), and (7)-(9).

Consider now the long-run cost function obtained by allowing the firm to optimize over the stock of private capital taking the stock of public capital as given:

$$(P.L) C^L(w, r, Y, P, \lambda, t) = \min_K C(w, r, K, P, Y, \lambda, t) = \min_K \{G(w, K, P, Y, \lambda, t) + rK\}$$

The first-order condition for this problem,

$$(12) \frac{\partial C}{\partial K} = \frac{\partial \mathcal{E}}{\partial K} + r = -\mu F_K + r = 0$$

(where  $\mu$  is still the multiplier associated with the minimization of labour costs in (P.S)), implicitly defines a long-term demand for private capital,  $K^L(w, r, Y, P, t)$ .

By the envelope theorem, the cost elasticity of the long-run cost function is equal to the cost elasticity of the short run cost function tangent to it at each point,

$$(13) \varepsilon_{C,Y}^L \equiv \frac{\partial C^L}{\partial Y} \frac{Y}{C^L} = \frac{\partial C}{\partial Y} \frac{Y}{C} = \frac{\mu}{C/Y} = \varepsilon_{C,Y}$$

When the firm is on the long-run cost curve, however, this elasticity reflects the degree of returns to scale in capital and labour. Notice that now

$$\frac{\partial C^L}{\partial \lambda} \frac{1}{C^L} = \frac{\partial C}{\partial \lambda} \frac{1}{C} = \frac{-1}{C} (\mu F_K K + \mu F_L L) = \frac{-1}{C} (rK + wL) = -1$$

where we have made use of the first-order conditions for capital and labour, (2) and (12). Rearranging terms, we can write

$$1 = \frac{\mu}{C/Y} \frac{F_K K + F_L L}{Y} = \varepsilon_{C,Y}^L (\varepsilon_{Y,K} + \varepsilon_{Y,L})$$

from where<sup>3</sup>

$$(14) \varepsilon_{C,Y}^L = \frac{1}{\varepsilon_{Y,K} + \varepsilon_{Y,L}}$$

Observe that the elasticity of output with respect to infrastructure capital,  $\varepsilon_{Y,P}$ , does not enter this expression because the firm is not free to adjust this input at will, even in the long run. Hence, the output elasticity of the long-run cost curve will in general be larger than the inverse of the overall returns to scale in production, given by by

$$(15) \varepsilon_{Y,K} + \varepsilon_{Y,L} + \varepsilon_{Y,P} \equiv \rho$$

### 3.- Decomposing returns to scale and cost elasticities

One of the issues Morrison and Schwartz address has to do with the determinants of the responsiveness of short-run costs to changes in output. Intuition suggests that fixities (associated with both internal and external inputs) imply that costs will rise with output faster than required by the long-term technological constraint imposed by the overall returns to scale displayed by the

<sup>3</sup> Notice that expression (6) above for  $\varepsilon_{C,Y}$  must still hold, so it must be the case that, when the firm is on its long-term cost schedule

$$\frac{1}{\varepsilon_{Y,K} + \varepsilon_{Y,L}} = \frac{S_L}{\varepsilon_{Y,L}} \quad \text{or,} \quad S_L = \frac{\varepsilon_{Y,L}}{\varepsilon_{Y,K} + \varepsilon_{Y,L}}$$

i.e. each input's share in costs is proportional to its output elasticity when the firm is in long-run equilibrium.

production function --i.e. that short-term returns to scale will typically be lower than long-term returns. To formalize this intuition and serve as the basis for the empirical measurement of the various forces at play, the authors seek a meaningful decomposition of the relevant short-term cost elasticity into factors that reflect purely technological considerations and the impact of each of the quasi-fixed inputs they consider.

Their approach, which must be largely inferred from the discussion as the authors provide no details, is to proceed backwards from a hypothetical long-term cost function which allows for the "optimal" adjustment of quasi-fixed inputs in a rather imprecise way. Formally, their procedure seems to involve the introduction of two functions  $K^0(Y, \dots)$  and  $P^0(Y, \dots)$  that describe notional long-term or "optimal" demands for the quasi-fixed inputs and their substitution into the short-term cost function to derive a hypothetical "optimal" cost function,

$$(16) C^0(Y) = C[K^0(Y), P^0(Y), Y],$$

free of the effects of fixities (notice that I am omitting all arguments other than  $Y$ ). Applying the chain rule to (16), the elasticity of the optimal cost function is given by

$$(17) \epsilon_{C,Y}^0 = \frac{\partial C}{\partial Y} \frac{Y}{C} + \left( \frac{\partial C}{\partial K} \frac{K}{C} \right) \left( \frac{\partial K^0}{\partial Y} \frac{Y}{K} \right) + \left( \frac{\partial C}{\partial P} \frac{P}{C} \right) \left( \frac{\partial P^0}{\partial Y} \frac{Y}{P} \right) \equiv \epsilon_{C,Y} + \epsilon_{C,K} \epsilon_{K,Y}^0 + \epsilon_{C,P} \epsilon_{P,Y}^0$$

where  $\epsilon_{K,Y}^0$  and  $\epsilon_{P,Y}^0$  are the output elasticities of the "optimal" demands for quasi-fixed inputs and  $\epsilon_{C,Y}$  the output elasticity of the short-run cost function. This expression is then solved for  $\epsilon_{C,Y}$  to obtain the desired decomposition:

$$(18) \epsilon_{C,Y} = \epsilon_{C,Y}^0 - \epsilon_{C,K} \epsilon_{K,Y}^0 - \epsilon_{C,P} \epsilon_{P,Y}^0$$

where  $\epsilon_{C,Y}^0$  is interpreted as a cost-based measure of long-term returns to scale.

The main problem with this construction is that the authors never bother to make clear the nature of the optimal demands for quasi-fixed inputs, or to show that  $C^0(Y)$  does indeed capture cost behaviour in the absence of input fixities. In the case of private capital, it seems reasonable to assume that what they have in mind is the long-term cost minimization problem we have discussed in the previous section, so that  $K^0(Y)$  is in fact the function we have called  $K^L(w, r, Y, P, t)$  above. The definition of  $P^0(Y)$  raises more difficult questions, but let us put them aside for a moment and consider the case where  $P$  either does not enter the production function or is treated as an exogenous variable.

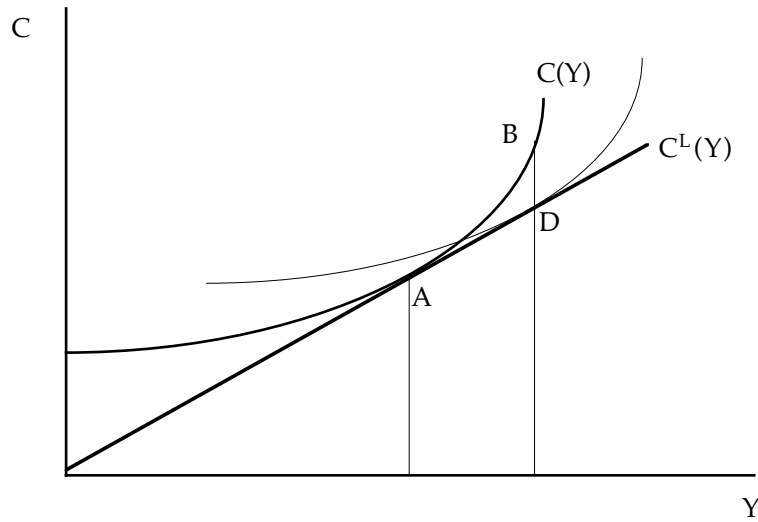
Then, the last term of (18) disappears and the relevant elasticities are those of the long-term cost and capital demand functions defined in (P.L), leaving us with the following simplified version of M&S's decomposition

$$(19) \epsilon_{C,Y} = \epsilon_{C,Y}^L - \epsilon_{C,K} \epsilon_{K,Y}^L.$$

If we return to the results of the previous section, a problem becomes immediately apparent. The first-order condition for optimal capital given in (12) implies that  $\epsilon_{C,K} = 0$ , leaving us with  $\epsilon_{C,Y}^L = \epsilon_{C,Y}$ , a result that can also be obtained directly by applying the envelope theorem in the computation of  $\epsilon_{C,Y}^L$  as in (13). Hence, the decomposition collapses because of the tangency of the short- and long-run cost

curves. The authors are indeed aware of this, but they seem to want to apply (19) to situations where the firm is off its long-run cost schedule (e.g. to point B in Figure 1 below). It is clear, however, that this cannot be done, as the  $\epsilon_{C,Y}$  that appears in (19) would then be the elasticity of the short-run cost curve tangent to  $C^L()$  at point D, and not that of the relevant short-run schedule through B. In other words, (19) is only valid at tangency points. Away from them, there is no known systematic relationship between the slopes or elasticities of the two curves.

**Figure 1: Short and long-run cost curves**



Let us now return to the definition of  $P^0(Y)$ . The relevant sections of the different M&S papers on the topic suggest two possible ways to define a notional long-term demand for public capital. The first one is in terms of an artificial cost-minimization problem in which the firm selects its stock of private capital so as to minimize its total costs and some hypothetical planner determines the stock of public capital with the same objective. That is,

$$(P.L') C^0(w, r, Y, t) = \min_{K,P} C(w, r, K, P, Y, t) = \min_P C^L(w, r, Y, P, t).$$

This approach is extremely problematic. First, because it is likely that an optimal solution to (P.L') will not exist. Since  $P$  is a free input, the optimal demand for it involves setting its marginal product equal to zero, which will be impossible for finite values of  $P$  for many standard production functions. Second, because if the solution does exist we run into the same problem we have just discussed: the decomposition collapses when the firm is on its long-term cost schedule and is not applicable when it is not. And third, because this formulation would be inconsistent with the fact that  $P$  is an exogenous variable, and not a decision variable for the firm.

In fact, the same can be said of just about any definition of  $P^0()$ . M&S are aware of the problem and try to wiggle out of the tight corner into which they have backed themselves by the curious expedient of labeling the derivatives involved in the calculation of  $\epsilon_{p,Y}^0$  as "definitional" rather than "behavioral" (1994, p. 253). In practice, what M&S propose in this second approach to the construction of  $P^0()$  is to

define the elasticity of the optimal demand for public capital,  $\varepsilon_{P,Y}^0$ , as the inverse of the elasticity of output with respect to P,  $1/\varepsilon_{Y,P}$ . This would have to mean that  $P^0()$  is defined by the requirement that P be set in such a way that the desired level of output is produced, given the values of the remaining inputs, i.e. that  $P^0$  is such that

$$(20) F(K, L, P) = Y$$

so that indeed  $(\partial P^0/\partial Y) = 1/(\partial F/\partial P)$ .

This may be interpreted in two alternative ways. In the first interpretation,  $P^0()$  describes a decision rule for the determination of P which is known by the firm when it solves its own long-term cost-minimization problem. In this case, the firm can induce the planner to supply the quantity of public capital it desires by choosing K and L appropriately, and we are back to (P.L') with the firm choosing P, K and L to minimize total private costs subject to the production constraint. Alternatively, we can think of applying  $P^0()$  ex-post, once the values of K and L have been chosen by the firm. But in that case (20) does not define a function, for it then holds identically for any value of P: since the firm chooses K and L as functions of P in order to satisfy a production constraint that coincides with (20), any existing value of P will satisfy this condition ex-post. Hence, we cannot use (20) to define the desired function as it now becomes the identity

$$F[K^L(P, Y), L^L(P, Y), P] \equiv Y$$

where  $K^L()$  and  $L^L()$  are the long-term factor demands obtained from the solution of (P.L), an expression that holds for all P by definition.

### *An alternative decomposition*

In conclusion, the decomposition of the short-run cost elasticity with respect to output proposed by M&S is unworkable because it relies on a concept of long-term demands for quasi-fixed inputs that is left unspecified by the authors and that, as far as I can tell, cannot be made precise in a sensible way. Since the intuition behind the proposed formula is quite appealing, however, it would be desirable to construct a consistent decomposition that preserves it.

One workable way to proceed is as follows. First, we define an artificial long-term cost elasticity constructed to capture the desired technological constraint; then, we relate it to the output elasticity of the short-run cost function and check whether the difference between these two variables can be given a sensible interpretation in terms of the effects of input fixities. By analogy with (14), a natural measure of the technological constraint on cost performance would be the inverse of the total returns to scale,

$$\frac{1}{\rho} = \frac{1}{\varepsilon_{Y,K} + \varepsilon_{Y,L} + \varepsilon_{Y,P}},$$

which would indeed be the output elasticity of the long-term cost function if the firm could choose P at a given user cost per unit. To relate this variable to the cost-based measure of short-run returns to scale given by  $\varepsilon_{C,Y}$ , notice that, recalling equations (6), (10) and (11) above,  $1/\rho$  can be written in the form



$$(21) \frac{1}{\rho} = \frac{\varepsilon_{C,Y}}{S_k^* + S_p^* + S_L}.$$

Hence, the desired decomposition would be of the form,

$$(22) \varepsilon_{C,Y} = \frac{S_k^* + S_p^* + S_L}{\rho} = \frac{S_k^* + S_p^* + 1 - S_k}{\rho} = \frac{1}{\rho} + \frac{S_k^* - S_k}{\rho} + \frac{S_p^*}{\rho},$$

and its last two terms can indeed be interpreted as measures of the impact of input fixities. As intuition suggests, the output elasticity of short-run costs will tend to be larger than required by the purely technological constraint represented by  $\rho$  due to the existence of quasi-fixed inputs. The contribution of private capital to this expression will be positive whenever its stock is below the optimal long-run level (i.e. whenever  $S_k^* < S_k$  or equivalently the shadow price  $Z_k = \mu F_k$  is higher than the market price,  $r$ ). Since public capital is an unpaid (and uncontrollable) input, its shadow price is presumably always positive and therefore higher than its (zero) market price, making the existing stock suboptimally low from the firm's perspective and its contribution to the elasticity of short-run costs positive and directly proportional to its "scarcity," as measured by its shadow share.

#### 4.- Technical progress and cost performance

We now turn to the measurement of technical progress and the analysis of the determinants of cost performance. Following the standard usage, we will define the rate of technical progress or TFP growth as the rate of growth of output that would follow simply from the passage of time with fixed factor stocks, that is,

$$(23) \varepsilon_{Y,t} \equiv \frac{F_t}{Y}$$

where the subscript indicates a partial derivative. Given information on the rates of input and output growth and the values (or estimates) of the relevant output elasticities,  $\varepsilon_{Y,t}$  can be computed as a residual by the following procedure. Differentiating the production function totally with respect to time, we have

$$Y = F_k K + F_L L + F_p P + F_t \Rightarrow F_t = Y - F_k K - F_L L - F_p P$$

from where

$$(24) \varepsilon_{Y,t} \equiv \frac{F_t}{Y} = \frac{Y}{Y} - \frac{F_k K}{Y} \frac{K}{K} - \frac{F_L L}{Y} \frac{L}{L} - \frac{F_p P}{Y} \frac{P}{P} = Y' - \varepsilon_{Y,K} K' - \varepsilon_{Y,L} L' - \varepsilon_{Y,P} P'$$

where the circumflex accent denotes a rate of growth (e.g.  $K' = \frac{K}{K}$ ).

The closest dual counterpart of the rate of TFP growth is the rate of cost reduction induced by technical progress for constant values of output and fixed input prices and quantities,

$$(25) -\varepsilon_{C,t} \equiv -\frac{\partial C}{\partial t} \frac{1}{C} = -\frac{\partial G}{\partial t} \frac{1}{C} = -\frac{\partial E}{\partial t} \frac{1}{C} = \frac{\mu F_t}{C}.$$

Using our previous results, it is easily seen that this dual measure of technical progress is related to TFP growth by the following expression

$$(26) -\varepsilon_{C,t} = \frac{F_t}{Y} \frac{\mu}{C/Y} = \varepsilon_{Y,t} \varepsilon_{C,Y}.$$

Hence,  $-\varepsilon_{C,t}$ , which is what M&S (1996) call the "corrected rate of technological change" ( $-\varepsilon_{C,t}^T$  in their notation)<sup>4</sup>, reflects not only TFP growth per se but also the degree of returns to scale in the short run. As is well known (e.g. Ohta (1974)), the two measures coincide when we have constant returns on private inputs and instantaneous adjustment, so that  $\varepsilon_{C,Y} = 1$ .

### *A cost accounting formula*

A natural measure of cost performance over time is given by the rate of reduction of unit costs at constant factor prices. It is clear that, holding  $w$  and  $r$  constant, the evolution of unit costs,

$$c(w, r, K, P, Y, t) \equiv \frac{C(w, r, K, P, Y, t)}{Y},$$

depends not only on the rate of technical progress and on the degree of short-run returns to scale (which determine  $C_t$ ) but also on the rates of accumulation of quasi-fixed inputs and on the growth rate of output. The size of this last effect, moreover, will depend on the degree of returns to scale.

We will now derive a cost accounting decomposition that is analogous to M&S's dynamic formula except in that it is written explicitly in terms of the rate of reduction of unit costs at constant factor prices, defined as

$$(27) \eta_{C,t} \equiv - \frac{\dot{c}}{c} \Big|_{w,r} = - \left( \frac{\dot{C}}{C} \Big|_{w,r} - \dot{Y} \right).$$

Holding  $w$  and  $r$  constant, the growth rate of total costs is given by

$$\begin{aligned} \frac{\dot{C}}{C} \Big|_{w,r} &\equiv \frac{dC[w, r, K(t), P(t), Y(t), t]}{dt} \frac{1}{C} = \frac{d\{G[w, r, K(t), P(t), Y(t), t] + rK(t)\}}{dt} \frac{1}{C} = \\ &= \left( \frac{\partial G}{\partial K} \dot{K} + \frac{\partial G}{\partial P} \dot{P} + \frac{\partial G}{\partial Y} \dot{Y} + \frac{\partial G}{\partial t} + r\dot{K} \right) \frac{1}{C} = \left( -Z_k \dot{K} - Z_p \dot{P} + r\dot{K} \right) \frac{1}{C} + \frac{\partial G}{\partial Y} \frac{Y}{C} \frac{\dot{Y}}{Y} + \frac{\partial G}{\partial t} \frac{1}{C} \\ &= \left( -\frac{Z_k K}{C} \frac{\dot{K}}{K} - \frac{Z_p P}{C} \frac{\dot{P}}{P} + \frac{rK}{C} \frac{\dot{K}}{K} \right) + \varepsilon_{C,Y} \dot{Y} + \varepsilon_{C,t} \\ &= \left( -S_k^* \dot{K} - S_p^* \dot{P} + S_k \dot{K} \right) + \varepsilon_{C,Y} \dot{Y} + \varepsilon_{C,t} \\ &= \varepsilon_{C,t} + \varepsilon_{C,Y} \dot{Y} - (S_k^* - S_k) \dot{K} - S_p^* \dot{P}. \end{aligned}$$

Substituting this expression in (27), we arrive finally at

$$(28) \eta_{C,t} = \dot{Y} - \frac{\dot{C}}{C} \Big|_{w,r} = -\varepsilon_{C,t} + (1 - \varepsilon_{C,Y}) \dot{Y} + (S_k^* - S_k) \dot{K} + S_p^* \dot{P}$$

In addition to the effects of technical progress, captured by  $-\varepsilon_{C,t}$ , this expression captures the impact on cost performance of changes in scale and the accumulation of fixed factors. For a given rate of technical progress, unit costs decrease with infrastructure investment and with the accumulation of

<sup>4</sup> Since the authors never define this variable, we have to show that the formula for it given in equation (2) in their (1996) paper holds with our definition. This is done below. Notice that my equation (31) is equation (2) in p. 1099 of M&S (1996), i.e. that my  $-\varepsilon_{C,t}$  and M&S's are the same thing.

private capital whenever the shadow value of this last input exceeds its user cost. Output growth will increase unit costs whenever there are decreasing returns to scale in the short run, i.e. when  $\epsilon_{C,Y} > 1$ .

Following M&S, we can go one step further. Notice that the returns to scale term that appears in (28) depends on the short-run output elasticity of costs, which reflects not only the technological constraint dictated by the total returns to scale displayed by the production function but also the increase in the cost responsiveness of output due to the existence of quasi-fixed inputs. We can use the static formula derived in the previous section to further decompose this term into a pure technological effect, reflecting total returns to scale, and terms that capture the short run fixity of public and private capital. Substituting (22) into (28), we have:

$$(29) \eta_{C,t} = -\epsilon_{C,t} + \left(1 - \frac{1}{\rho} - \frac{S_k^* - S_k}{\rho} - \frac{S_p^*}{\rho}\right) \dot{Y} + (S_k^* - S_k)\dot{K} + S_p^*\dot{P}$$

$$= -\epsilon_{C,t} + \left(1 - \frac{1}{\rho}\right) \dot{Y} + (S_k^* - S_k) \left(\dot{K} - \frac{\dot{Y}}{\rho}\right) + S_p^* \left(\dot{P} - \frac{\dot{Y}}{\rho}\right)$$

Notice that the last two terms in (22) can be conveniently combined with the corresponding terms of (28) to provide measures of the "net effect" of the behaviour of quasi-fixed factors on cost performance. If both private and public capital "keep up with output" (i.e. grow at the rate  $\dot{Y}/\rho$ ), their net effects are zero and costs vary only in response to technical progress and a purely technological scale effect. Faster accumulation of either input will reduce unit costs whenever the relevant shadow price exceeds the corresponding market price.

#### *The biases in Solow's formula and M&S's dynamic decomposition*

The expression we have derived above for the computation of the rate of technical progress (equation (24)) is written in terms of output elasticities rather than input shares in output or in total cost. Solow (1957) observed, however, that under certain assumptions (perfectly competitive markets, constant returns to scale, cost minimization in all inputs at each point in time and absence of unpaid factors) all these variables coincide, allowing us to compute the rate of technical progress without explicit knowledge of the production function. In the absence of estimates of the relevant production parameters, it has become standard practice to use Solow's formula to obtain an approximate rate of TFP growth. In the present context, the application of this procedure (using input shares in total cost) would yield the following "Solow" estimate of the rate of technical progress

$$(30) \hat{\epsilon}_{Y,t} = \dot{Y} - S_k \dot{K} - S_L \dot{L}.$$

Unlike equation (24), M&S's dynamic formula is written in terms of the Solow residual,  $\hat{\epsilon}_{Y,t}$ . As stated in M&S (1996), it relates this variable to the rate of cost reduction induced by technical progress,  $-\epsilon_{C,t}$ . This relationship can be easily derived using our previous results. Substituting (6), (10) and (11) into the "primal" formula for TFP growth given in equation (24) we obtain an expression for the rate of technical change similar to Solow's formula but written in terms of the shadow cost shares of K and P with a correction for short-term returns to scale:

$$(31) \varepsilon_{Y,t} = \dot{Y} - \frac{S_k^*}{\varepsilon_{C,Y}} \dot{K} - \frac{S_L}{\varepsilon_{C,Y}} \dot{L} - \frac{S_p^*}{\varepsilon_{C,Y}} \dot{P}.$$

Substituting this expression into equation (26), we have

$$(32) -\varepsilon_{C,t} = \varepsilon_{C,Y} \varepsilon_{Y,t} = \varepsilon_{C,Y} \dot{Y} - S_k^* \dot{K} - S_L \dot{L} - S_p^* \dot{P}.$$

We can now rearrange terms to rewrite (32) in terms of the Solow residual. We have

$$\begin{aligned} (33) -\varepsilon_{C,t} &= (1 + \varepsilon_{C,Y} - 1) \dot{Y} - (S_k + S_k^* - S_k) \dot{K} - S_L \dot{L} - S_p^* \dot{P} \\ &= (\dot{Y} - S_k \dot{K} - S_L \dot{L}) + (\varepsilon_{C,Y} - 1) \dot{Y} - (S_k^* - S_k) \dot{K} - S_p^* \dot{P} \\ &= \xi_{Y,t} + (\varepsilon_{C,Y} - 1) \dot{Y} - (S_k^* - S_k) \dot{K} - S_p^* \dot{P} \end{aligned}$$

which is the first version of M&S's dynamic formula (given in the first part of equation (4) in M&S (1996)). Indeed, this expression can be (roughly) interpreted in terms of "the error biases in traditional [PTF growth] measures" (M&S 1994, p. 255), but this is of limited interest in itself. It is unclear, in particular, why we may want to compute the Solow residual when we have or can recover estimates of the parameters required to compute TFP growth correctly.

The problem disappears ex-post once we observe that the Solow residual turns out to be equal to the rate of reduction of unit costs at constant factor prices, i.e. that  $\xi_{Y,t} = \eta_{C,t}$  as can be seen by comparing equations (33) and (28). Hence, M&S's formula does indeed provide insight on the immediate sources of cost reductions. This is not, however, readily apparent from the discussion. Citing Ohta (1974), M&S (1994, p. 254) assert the equality under constant returns to scale and instantaneous adjustment of three, conceptually quite different quantities: the rate of cost reduction induced by technical change ( $-\varepsilon_{C,t}$  in our notation), the rate of reduction of unit costs at constant factor prices ( $\eta_{C,t}$ ) and the Solow residual ( $\xi_{Y,t}$ ). When these assumptions fail, however, the three quantities may conceivably differ, and it is not clear which of them is being referred to in the subsequent discussion. The context seems to indicate that the symbol  $-\varepsilon_{C,t}$  in M&S (1994) is meant as the Solow residual, as the discussion refers repeatedly to the "standard" or "traditional" measure of multifactor productivity. This is even clearer in M&S (1996) where  $-\varepsilon_{C,t}$  is replaced by a different symbol that explicitly refers to the Solow measure of TFP growth. In neither case is the link between the Solow residual and the rate of cost reduction (which may or may not be known by the authors) brought to the attention of the reader.

## 5.- Conclusion

Unit costs vary over time in response to technical progress and factor accumulation in a way that is sensitive to returns to scale and to the presence of quasi-fixed and external inputs. Building on well-known duality concepts and results, Morrison and Schwartz have developed a framework that can be used to conceptually isolate the separate effects of these factors and to measure their impact on cost performance. Their analysis, however, suffers from two important shortcomings. The first is that one key aspect of it, involving the decomposition of a cost-side measure of short-run returns to scale, is either incorrect or inconsistent. The proposed expression involves derivatives that either necessarily vanish or do not exist because the underlying functions are not well defined. The second problem is

that the exposition is extremely misleading. It seems, in fact, purposefully designed to hide the real interest of the exercise (the analysis of the determinants of cost behaviour) by presenting the second formula as a recipe for correcting the biases in a patently incorrect measure of PTF growth whose link with the problem at hand is far from obvious. In this note I have attempted to correct both of these problems while preserving the spirit of M&S's analysis by i) developing a consistent decomposition of a measure of short-run returns to scale and ii) providing a direct derivation of their dynamic formula that makes transparent its usefulness as a device for cost analysis.

### References

- Morrison, C. and Schwartz A. (1994). "Distinguishing external from internal scale effects: the case of public infrastructure." *Journal of Productivity Analysis* 5, pp. 249-70.
- Morrison, C. and Schwartz A. (1996). "State infrastructure and productive performance." *American Economic Review* 86(5), pp. 1095-1111.
- Morrison, C. and D. Siegel (1999). "Scale economies and industry agglomeration externalities: a dynamic cost function approach." *American Economic Review* 89(1), pp. 272-90
- Ohta, M. (1974). "A note on the duality between production and cost functions: Rate of return to scale and rate of technical progress." *Economic Studies Quarterly* 25(3), pp. 63-5.
- Solow, R. (1957). "Technical change and the aggregate production function." *Review of Economics and Statistics* 39, pp. 312-320.