Experience, Learning, and Returns to Scale

UTD AUTHOR(S): Daniel G. Arce M

©2014 Southern Economic Association
Experience, Learning, and Returns to Scale

Daniel G. Arce*

The experience curve is a tool for forecasting future decreases in average cost as a function of cumulative output/volume. The extent of an experience effect has profound implications for both pricing strategy and the focus on market share as a managerial objective. At the same time, the underlying sources of the experience effect are not well understood. This article demonstrates that, as commonly measured, experience effects are aggregated with the effects of increasing returns to scale. This implies that standard experience curve estimates are misspecified because they suffer from an omitted variable bias. Strategic implications of the experience-scale link are discussed.

JEL Classification: D2, L2, L1, M11

1. Introduction

The experience (or learning) curve captures the statistical relationship between the cumulative volume of production and declining average (variable) costs. Beginning with the Boston Consulting Group (hereafter BGC 1968) it remains a centerpiece of the strategy toolbox of major business consulting practices (Stern and Deimler 2006; Gottfredson and Schaubert 2008; Kiechel 2010; Thompson 2012). Indeed, capturing market share is legitimized as a managerial objective specifically because of the cost efficiencies that arise from the experience curve. The associated rationale is as follows: If cost is a function of accumulated experience/volume, then profit is a function of sustained market share. Market share therefore has a value directly reflected in relative costs.

From early on, economists as well have recognized the importance of experience as a cost-reducing phenomenon (e.g., Hirsch 1956; Alchian 1958) but tend to instead emphasize the effects of economies of scale for cost reduction and producing below capacity under increasing returns to scale as a potential entry-deterring strategy. At the same time, increasing returns to scale is viewed as one of the contributing factors to cost reductions associated with experience curve efficiencies (uniquely so in BGC 1968), along with other factors such as learning-by-doing, decreasing costs of capital as the firm grows, technological advances, and decreasing average fixed costs, to name but a few. This begs the question, How separate are learning/experience effects from increasing returns to scale?

This article investigates the formal relationship between the experience curve (as it is mostly commonly specified in the empirical literature in cost accounting, economics, marketing, and strategic management) and increasing returns to scale, which—to my knowledge—has not
been characterized until now. Various studies have derived learning/experience effects from underlying production functions (e.g., Rosen 1972; Womer 1981; Berndt 1991). Such analyses necessarily start with assumptions about returns to scale as defined by the effect that proportional changes in inputs have on output. By contrast, this analysis begins with the experience curve itself and asks, what does this assumption about the functional form of the effect of experience on cost—as measured by the cumulative volume of production—imply for returns to scale? Note that the cost function associated with the experience curve is not a cost function in the usual sense; that is, it does not parse out costs in terms of input prices and production quantities. It is instead an expression of cost as a function of the firm’s initial average cost, the cumulative volume, and the learning elasticity. Specifically, this version of the experience curve is viewed as a purely statistical relationship between cumulative output and average cost (Oi 1967; Hall and Howell 1985; Johnson 1987; Liao 1988; Alberts 1989; Sinclair, Klepper, and Cohen 2000; Zwan and Wene 2011; Thompson 2012). This has two implications for my analysis: (i) it is not possible to employ cost-benefit duality to back out the firm’s production technology because the experience curve is not expressed in terms of input prices and quantities (i.e., it is not formally derived from a cost-minimization exercise); therefore, (ii) I employ an alternative but equivalent measure of returns to scale based on the elasticity of total cost, which is the ratio of marginal cost to average cost. For example, under this measure increasing returns to scale occur when the average costs for a production period are decreasing owing to marginal costs that are below average costs. By contrast, learning/experience effects are regarded as shifters of the average cost curve, resulting in lower average costs for all possible returns to scale. Yet I show that under standard assumptions the experience effect is consistent only with increasing returns to scale. This is a novel finding given that the experience curve was originally meant to capture a statistical relationship only. Moreover, it implies that learning/experience curve analyses based on this statistical relationship are misspecified; as a portion of the cost reductions associated with increases in cumulative volume are instead attributable to increasing returns to scale.

The article proceeds as follows. The next section presents the (statistical/empirical) experience curve, where average cost is given as a log-linear function of cumulative volume. The following section describes various measures of marginal cost that stem from the experience curve. In the penultimate section it is shown that the experience effect occurs jointly with increasing returns to scale. In the final section I discuss the strategic implications of the experience-scale relationship and provide brief concluding remarks.

2. The Experience Curve Defined

This article focuses on the experience curve, which measures the reduction in average cost as cumulative volume increases. By contrast, the learning curve measures the reduction in average variable costs (typically labor and/or other factor costs) as cumulative volume increases.¹ There is also an alternative experience curve often found in the cost engineering

¹ These definitions are consistent with the literature in strategic management and managerial economics, which differentiate between the experience and learning curves in terms of whether fixed costs are included or not, respectively. This is decidedly different from differentiating between learning, which is a flow, and experience, which is a stock.
literature (known as the Crawford model) that relates the decline in incremental costs to cumulative volume. Again, the present analysis is concerned with the experience curve where average cost is the dependent variable (known as the Wright model). The existence of an experience effect translates into a pricing strategy that deviates from current-period (myopic) profit maximization in order to attain increased market share that leads to future cost reductions. This strategy is known as \textit{racing down one’s experience curve}.

The effect of experience on cost reduction is traditionally quantified through the following relationships between cumulative output and average cost. Specifically, let \( v \) = the firm’s experience (accumulated output/volume), which is defined by the following identity:

\[
v = \sum_{t=1}^{T-1} q_t,
\]

where: \( q_t \) is the level of output in period \( t \). Furthermore, \( AC_1 \) = the (initial) average cost of the 1st unit and \( AC_v \) = average cost after \( v \) cumulative units of output.

The Wright version of the experience curve relationship is given by either of the following equations:

\[
TC_v = AC_1 \cdot v^{1-\lambda} \quad \text{or} \quad AC_v = TC_v/v = AC_1 \cdot v^{-\lambda} = AC_1/v^\lambda; \tag{2}
\]

where \(-\lambda = \frac{\Delta AC_v}{\Delta v} \cdot \frac{v}{AC_v}\) and \(0 \leq \lambda < 1\).

In Equation 3 \( AC_1 \) is divided by cumulative volume (or \( v \) has a negative exponent) because average costs decrease with experience and \( v \) is the proxy for experience. The term \( \lambda \) is essentially an index of the rate of decrease in costs as cumulative volume increases. It is known as the \textit{learning elasticity}, \textit{learning constant}, or \textit{index of learning}. The experience curve is a curvilinear function that has the restriction that the value of \( AC_v \) decreases at a constant rate every time \( v \) is doubled. That is, in the first period when \( t = 1 \), \( v \) and \( q_t \) coincide: \( v = q_t \); but as cumulative doublings progress: \( v = 2q_t, 4q_t, 8q_t, 16q_t, \ldots \); it is no longer the case that \( v = q_t \).

The experience effect is meant to capture relatively stable decreases in average costs as cumulative output doubles. This is what the value of \( \lambda \) ensures. For example, if \( \lambda = 0.322 \), then the average cost of a cumulative doubling is \( AC_{2v} = AC_1 \cdot (2v)^{-0.322} \approx 0.8 \times AC_v \). Here average costs after a cumulative doubling are 80\% of the previous level. Note that if \( \lambda = 0 \) no experience effect is present. From Equation 2 if \( \lambda = 1 \), then the total cost for any volume is constant and equal to \( AC_1 \), which is unrealistic. Consequently I assume \( 0 < \lambda < 1 \).

The reduction in average cost after a cumulative doubling in volume is known as the \textit{progress ratio}, denoted as \( \rho \), where

\[
\frac{\Delta AC_v}{\Delta v} \cdot \frac{v}{AC_v}\]

\[2\] Incremental cost is the change in total costs as output changes incrementally (equal to the sum of the marginal costs over the increment in output) divided by the change in output. Incremental costs are the “average” marginal costs over the range in output. Only at high levels of output will it be the case that incremental costs equal marginal costs. Hence, the Crawford model typically does not capture marginal costs. Johnson (1987) and Liao (1988) intuitively survey the difference between the Wright and Crawford models.

\[3\] For this reason it is assumed that technology is held constant and no “unlearning” occurs. Conversely, Rosen (1972) notes that failure to recognize the experience effect stemming from a given technology may lead to incorrect inferences of technical change.
\[ \rho = 1 - AC_{2n}/AC_{n} \Rightarrow \rho = 1 - \left[ AC_1 \cdot (2v)^{-\lambda} \right] / \left[ AC_1 \cdot v^{-\lambda} \right] = 1 - 2^{-\lambda}. \] (4)

The “2” corresponds to the cumulative doubling, making it evident that \( \lambda \) is the elasticity of learning that yields the progress ratio for every cumulative doubling. Returning to my example, when \( \lambda = 0.322 \) average costs will be reduced by \( \rho = 20\% \) after a cumulative doubling of output.

Finally, the experience curve is a significant departure from the neoclassical (duality-based) approach to cost and production. Cost-production duality stems from factor prices that correspond to current period output, \( q_t \). In the neoclassical approach output is specified on a per-period basis where output and inputs are contemporaneous flow variables. By contrast, the experience curve is expressed in terms of cumulative volume, \( v \), which is a stock. Moreover, the experience curve is regarded as a statistical relationship. Indeed, Berndt (1991)—who derives the cumulative average cost curve in Equation 3 from a Cobb-Douglas production function for \( q_t \)—purges factor prices from the resulting equation in order to properly estimate the experience curve as a function of \( v \) only.

Current period output, \( q_t \), and cumulative volume, \( v \), are dimensionally different; \( q_t \) is a flow whereas \( v \) is a stock. The two are related through the identity in Equation 1, which introduces a third variable, \( T \), corresponding to the number of periods. These intertemporal relationships can be illustrated with the following examples. First, suppose that output is \( q_t = 50 \) for all \( t \). Then for volume to reach \( v = 400 \) (i.e., four cumulative doublings: \( q_t = 50, 2q_t = 100, 4q_t = 200, 8q_t = 400 \) ) this requires \( T = 8 \). If output is instead increased by 30 for each period (\( q_t = 80 \)), then a volume of 400 requires \( \hat{T} = 5 \), and the last three output flows, \( q_6, q_7, \) and \( q_8 \), will now be produced in the first five periods.

3. Marginal Cost in the Statistical Learning Model

Increasing returns to scale occur when average cost is decreasing, corresponding to current output levels where marginal cost is less than average cost (illustrated in Figure 1). Given that the cost of the next additional unit of output (marginal cost) is less than current average cost, the production of that unit of output will cause average cost to decline. For \( TC = \text{total cost}, q = \text{current output}, AC = \text{average cost} = TC/q, MC = \text{marginal cost} = \Delta TC/\Delta q \), the elasticity of total cost, \( E_{TC} \), is

\[ E_{TC} = \frac{\% \Delta TC}{\% \Delta q} = \frac{\Delta TC/TC}{\Delta q/q} = \frac{\Delta TC}{\Delta q} \cdot \frac{q}{TC} = \frac{MC}{AC}. \] (5)

Under increasing returns to scale, \( 0 < E_{TC} < 1 \) as \( MC < AC \).

One benefit of focusing on this definition of increasing returns to scale is that it does not rely on the specification of an underlying production function. Such an exercise requires assumptions regarding the degree of substitutability among inputs and this in itself constrains the learning-scale relationship (Womer 1981; Berndt 1991). At the same time, even though returns to scale can be characterized in terms of the relationship between average and marginal costs within the production period, and the Wright version of the experience curve is expressed

---

4 Further discussion of this process is provided in section 4.
in terms of average cost, it is worthwhile recognizing that different measures of marginal cost stem from the Wright form of the experience curve. In particular, from Equation 2:

\[ \frac{\partial TC_v}{\partial v} = (1 - \lambda) \cdot AC_1 \cdot v^{-\lambda} = (1 - \lambda) \cdot AC_v. \] (6)

Equation 6 derives the slope of the total cost function with respect to cumulative volume, \( v \), rather than current output, \( q_t \). In this sense it is the marginal cost of cumulative volume, \( v \), denoted as \( MC_v \). From Equation 6, \( MC_v \) is a constant proportion of \( AC_v \). Moreover, it can readily be seen that \( \frac{\partial MC_v}{\partial v} = -\lambda \cdot (1 - \lambda) \cdot AC_1 \cdot v^{-\lambda} < 0 \); that is, the marginal cost of cumulative volume is decreasing. This is consistent with the experience effect being a shifter of the per-period average cost curve, \( AC_q \). In addition, \( MC_v \) can be thought of as the shadow value of learning in the sense of Rosen (1974).\(^5\) Without an experience effect the firm faces a (short-run) profitability constraint in that price must exceed minimum average (variable) cost. As a shadow value, \( MC_v \) captures the degree to which this constraint can be relaxed given experience effects. Rosen argues that the firm initially “overexpands” in comparison to myopic profit maximization so as to better capture experience effects that are, in a sense, an intermediate

\(^5\) This property was suggested by an anonymous referee.
product that is jointly produced with \( q_t \). The property that \(\frac{\partial MC}{\partial h_n} > 0\) implies that the firm’s ability to relax its profitability constraint is decreasing owing to the opportunity cost of overexpansion. Hence, short-term losses can only be sustained for low values of \( v \).

The marginal cost of cumulative volume, \( MC_v \), is not; however, the standard measure of a firm’s marginal cost found in economics, which is the slope of the firm’s total cost function with respect to current output, \( q \). The firm’s marginal cost is \(\frac{\partial TC}{\partial q} \), denoted as \( MC_q \). Although \(\frac{\partial MC}{\partial v} < 0\), it is generally assumed that \(\frac{\partial MC}{\partial q} \geq 0\). Indeed, increasing marginal costs (in current output) are not inconsistent with increasing returns to scale so long as \( MC_q < AC_q \) (as in Figure 1). This means that there are two competing notions of marginal cost that have different behaviors. The difference between the two marginal phenomena is further illustrated and discussed in Figure 2, which is an amalgamation of conjectures made in Alchian (1958), Hirshleifer (1962), and Oi (1967) that, to my knowledge, have never been checked against the Wright form of the experience curve given in Equation 3.

6 Womer (1981) analyzes how the relationship among inputs in an underlying homogenous production function affect the claims of Alchian (1958), Hirshleifer (1962), and Oi (1967).

4. Experience and Returns to Scale

From Equations 1 and 2 it is clear that total cost is a function of both \( v \) and \( q \): \( TC = TC(q, v) \). Hence, the possibility exists that cost reductions attributable to experience effects may instead be due to increasing returns to scale. As Abel and Hammond (1979) note, the two effects are easily confused because growth in experience coincides with growth in the size of an operation and scale effects involve taking advantage of the size of an operation. As discussed above, returns to scale can be characterized by the relationship between average and marginal costs for the production period in question.
Taking the logarithms of each side of Equation 2:

\[ \ln TC = \ln AC_1 + (1 - \lambda) \times \ln v. \]  

(7)

Taking the partial derivative of both sides with respect to \( q_t \):

\[ \frac{1}{TC} \cdot \frac{\partial TC}{\partial q_t} = (1 - \lambda) \frac{1}{v} \cdot \frac{\partial v}{\partial q_t}. \]  

(8)

Multiplying both sides by \( q_t \) and rearranging terms:

\[ \frac{\partial TC}{\partial q_t} \cdot \frac{q_t}{TC} = (1 - \lambda) \frac{\partial v}{\partial q_t} \cdot \frac{q_t}{v}. \]  

(9)

Given the expression for elasticity of total cost, \( E_{TC} \), in Equation 5 and recognizing that the last two terms in Equation 9 can be interpreted as the elasticity of volume with respect to current period output, which I denote as \( E_v = (\partial v/\partial q_t) \cdot (q_t/v) \), (Equation 9) becomes

\[ E_{TC} = (1 - \lambda) \times E_v. \]

(10)

Experience effects imply \( 0 < 1 - \lambda < 1 \). Moreover, \( E_v \) is the percentage change in \( v \), \( \% \Delta v = \Delta v/v \), divided by the percentage change in \( q_t \), \( \% \Delta q_t = \Delta q_t/q_t \). As \( v \geq q_t \) by definition, for a given \( \Delta q_t \), it is clearly the case that \( \% \Delta v \leq \% \Delta q_t \). It therefore follows that \( 0 < (1 - \lambda) \times E_v < 1 \), implying \( 0 < E_{TC} < 1 \), that is, increasing returns to scale.

**Result.** The experience effect occurs jointly with increasing returns to scale.

Alberts (1989) argues that “experience” should be understood as encompassing both the effects of learning and economies of scale. The above result confirms that this is, indeed, the case. The two phenomena are encompassed within the single equation, \( AC_v = AC_1 \times v^{-\lambda} \), used to measure decreasing costs. Notice that the result is a consequence of assuming that the functional form of the experience curve is that given in Equation 3, which is standard, and applying the identity in Equations 1–3. Hence the experience curve does not effectively distinguish between cost reductions due to learning/volume, which are permanent, and those stemming from increasing returns to scale, which have no long-run effect. Furthermore, the identity in Equation 1 does not force any relationship between current output and cumulative volume by requiring a specific relationship between \( q_t \) and \( q_{t+1} \). By contrast, Womer (1981) notes that the restriction \( q_{t+1} = q_t = v/T \) is implicit in Alchian (1958) and explicit in Hirshleifer (1962) and Oi (1967).

Berndt (1991, pp. 71–75) provides a related result where he integrates the learning curve with a Cobb-Douglas production function for \( q_t \). He shows that an integrated model is possible if both (i) the effects of input prices on costs of production can be accurately measured by using a GDP deflator rather than the individual input prices themselves and (ii) the production function exhibits constant returns to scale. Note that in Berndt returns to scale are defined in terms of the behavior of output when all inputs are changed by the same proportion. Increasing returns to scale occur when output increases by more than the equiproportional change in inputs, once again leading to decreasing average costs. In practice, equiproportional changes in inputs are rarely the way that managers increase production nor are equiproportional changes in inputs generally considered to be the means for operationalizing the strategic

\[ For the case of equiproportional changes in inputs, the elasticity of total cost is equal to the reciprocal of the elasticity of scale, \( (\Delta q/\Delta k)/(k/q) \), where \( k \) is the proportional change in inputs (Brown and Chachere 1980).
implications of economies of scale (Gold 1981). Instead, strategic aspects of economies of scale include spreading out overhead over a larger production run, the adoption of new technologies, and the boundaries of the firm (e.g., make versus buy versus make-and-buy decisions). This is why I have instead used the more empirically and strategically relevant definition of increasing returns to scale as decreasing current-period average costs. Nevertheless, Berndt’s result indicates that when returns to scale are not constant—as I have shown to be the case—then a regression based on Equation 3 without an accounting for scale suffers from omitted variable bias. In particular, it yields an overestimate of the value of $\lambda$.

I complete this section by returning to the issue of $MC_v$ versus $MC_q$ (refer to Figure 2). It has been shown above that $\frac{\partial^2 TC}{\partial v^2} < 0$. Equation 9 can be manipulated to show that $\frac{\partial^2 TC}{\partial q^2} \geq 0$. As $MC_q = \frac{\partial TC}{\partial q}$ this implies $\partial MC_q/\partial q \geq 0$. Hence, the marginal cost phenomena illustrated in Figure 2 are consistent with the experience curve in Equation 3.

5. Discussion and Conclusion: Strategic Implications

The experience curve captures the relationship between cumulative volume and average cost. It lies at the foundation for understanding the symbiotic relationship between a firm’s competitive environment and its costs. I have shown that the experience effect encompasses both reduced costs associated with learning as well as increasing returns to scale. In this way I provide a formal counterpart to Hall and Howell’s (1985) argument that, in the presence of economies of scale, transforming current output data by accumulation may produce a better empirical fit ex post but may not produce any improvement in predictive abilities over the use of current output as the independent variable.

Focusing on market share as a managerial objective is therefore problematic because the simple consideration of a single policy variable—cumulative volume—is not sufficient to strategically implement cost reductions. As the experience curve in Equation 3 is typically estimated via a log-log transformation, one must gauge the extent to which derived efficiencies reflect economies of scale. Decreasing average costs are consistent with experience effects but do not by themselves indicate the presence of learning efficiencies.

Experience encompasses both learning and increasing returns to scale because growth in experience comes by increasing current period output. Referring to Figure 1, increasing returns to scale are required in order for it to make economic sense to rely on increased volume to lower costs because otherwise the firm will be facing increasing average costs while expanding output. Indeed, the associated requirements on cash flow almost demand that the firm be exhibiting increasing returns. Under the classic experience curve strategy a firm initially prices below average costs in anticipation that the associated increased volume will reduce future costs via the learning effect. A firm is constrained in its ability to follow this strategy by its capacity to generate cash flow. Increasing returns to scale relax the constraint on cash flow because average

---

8 Gold (1981, p. 10) illustrates the shortcomings of the definition of economies of scale that relies on equiproportional changes in inputs by contrasting what a manager would expect from tripling production by adding two plants that are identical to the original versus adding one plant designed to yield twice the production of the original. Womer (1981) as well recognizes that increases in volume need not require equiproportional changes in labor and capital.

9 Technically, the log-log transformation is an estimate of $AC_v = AC_1v^{-\beta} \cdot e^\epsilon$ where $\epsilon$ is the error term.
costs are decreasing. By contrast, decreasing returns to scale place greater demands on cash flow as average costs are increasing.

Analytically, this can be seen by substituting the expression for $E_{TC}$ given in Equation 5 into the familiar inverse elasticity pricing rule:

$$
\frac{P - MC}{P} = \frac{1}{E_D} \Rightarrow \frac{P - (\frac{AC}{AC})MC}{P} = \frac{1}{E_D} \Rightarrow \frac{P - E_{TC} \cdot AC}{P} = \frac{1}{E_D}
$$

where $E_D$ is the firm’s elasticity of demand. Recall that $E_{TC} < 1$ under increasing returns to scale and $E_{TC} > 1$ for decreasing returns to scale. Hence, a firm can sustain a much lower price-cost margin, $P - AC$, under increasing returns to scale than it can under decreasing returns.

The aggregation of experience and scale effects raises strategic concerns as well. For example, consider the strategy of the initial use of below profit-maximizing prices (reduced margins) in order to gain cost efficiencies via learning effects. If a firm is simultaneously following a strategy of producing below capacity under increasing returns to scale in order to deter entry, then reduced pricing to capture learning effects may lead to increased sales/production that cause the firm to reach or exceed capacity, thereby nullifying its entry-deterrence strategy; that is, there exists an “experience curve trap” for the strategy of entry deterrence by maintaining excess capacity. If the purported efficiencies truly stem from the learning component of experience, the trap need not occur as the learning effect shifts the current period average cost curve down. At the same time, the impact of experience on minimum efficient scale (hereafter MES) is indeterminate because, by definition, $MC_q = AC_q$ at MES (refer to Figure 1), but under the standard approach the experience effect coincides with increasing returns to scale and under increasing returns to scale this equality never holds. One cannot assume that MES continually increases with experience. Certainly capacity cannot increase indefinitely for a given plant size. Yet my result shows that under the predominant functional form for the experience curve the possibility that the firm faces capacity constraints (i.e., $MC_q \neq AC_q$) is implicitly ruled out; consequently, the potential for the experience curve trap is not even recognized. Indeed, for MES the relevant experience is the experience gained from building plants with a higher capacity (e.g., the familiar rule that capital costs increase by six-tenths of the proportional increase in capacity). By contrast, the experience curve is meant to capture production economies only.

From an econometric perspective the experience equation is underspecified because it does not separate scale effects from experience effects. Consequently, average cost reductions need to be decomposed to reflect factors such as scale economies, learning-by-doing, investment in capital goods, and technological advances that primarily reduce production costs (rather than transforming the qualitative nature of the product). In the end it is likely the case that learning elasticity $\lambda$ itself is a dependent variable. Yet as Thompson (2012) observes, within economics there is remarkably little empirical understanding of the processes that contribute to the determination of $\lambda$. Such an analysis is therefore beyond the scope of the present article and is a topic for future research.

Finally, my result can seemingly be interpreted as an endorsement for the Crawford version of the learning curve, which expresses incremental costs as a function of cumulative volume and is widely used by industrial engineers. Yet in the same way there is very little understanding of the processes that contribute to the learning index in the Crawford curve. The Crawford curve also suffers from alternative sources of omitted variable bias because of the
coexistence of different sources of cost reduction (such as process R&D, expected future output, and the incentives underlying these two variables) within the firm (Sinclair, Klepper, and Cohen 2000). Moreover, an analysis such as mine on the relation between learning and scale is made more difficult with the Crawford curve, as only theoretical approximations of total cost can be made from an incremental cost function. Furthermore, the incremental cost-based Crawford curve is less strategically relevant that the Wright curve, which is expressed in terms of average cost.

References


