Envelopes for Economists:
Housing Hedonics and Other Applications

An e-Book Edited by John Yinger

With contributions by: Il Hwan Chung, William D. Duncombe, Yue Hu,
Phuong Nguyen-Hoang, and Pengju Zhang

Chapter 3: Long-Run Average Cost Curves

John Yinger
April 2020

3.0. Introduction

The best-known envelope among economists is undoubtedly the long-run average cost curve. The graphical derivation of this curve is a staple of intermediate microeconomics classes. Moreover, economists first started using the term "envelope" in connection with this topic. This chapter presents the intellectual history of cost curve envelopes and derives long-run average cost curves, that is, envelopes, for a wide range of cases.

3.1. Some History

3.1.1. Viner's Mistake and Harrod's Insight

The intellectual history of long-run average cost curves as envelopes centers on a drafting error. In a famous article about production costs (1931), Jacob Viner discussed and drew short-run and long-run average cost curves. In the case of constant costs, illustrated by the first case discussed in Section 3.2, Viner drew a horizontal long-run average cost curve, the envelope, with a series of U-shaped short-run average cost curves tangent to it (his Chart III). He did not use the term "envelope." In the case of a downward-sloping long-run average cost curve, however, Viner instructed his draftsman to make the short-run curves tangent to the long-run curves at their minimum points. His draftsman, Dr. Y. K. Wong, objected. This disagreement led to the following footnote:

> It may be noticed that at certain points the short-run *ac* curves are drawn so as to sink below the long-run *AC* curve…. [T]his is an error. My instructions to the draftsman were to draw the *AC* curve so as never to be above any portion of any *ac* curve. He is a mathematician, however, not an economist, and he saw some mathematical objection to this procedure which I could not succeed in understanding. I could not persuade him to disregard his scruples as a craftsman and to follow my instructions, absurd though they may be. (Viner 1931, p. 36, footnote 16)

I guess Viner became the draftsman himself, because the short-run average cost curves (his *ac* curves) in his Chart IV clearly all sink below the long-run average cost curve (his *AC* curve).

It took him a while, but Viner eventually figured out the problem.[1] As he said in a "Supplementary Note" when his article was reprinted almost 20 years later,

> I do not take advantage of the opportunity to revise my 1931 article. Even the error in Chart IV (page 215) is left uncorrected, so that future teachers and students may share the pleasure of many of their predecessors of pointing out that if I had known what an "envelope" was I would not have given my excellent draftsman the technically impossible and economically inappropriate assignment of drawing an *AC* which would past through the lowest points of all the *ac* curves and yet not rise above any *ac* curve at any point (Viner, 1950, p. 227)

This issue is also illustrated in Figure 3.1, which shows the envelope (derived below) and the minimum points for a series of parabolic short-run average cost (*SRAC*) curves.[2] In this symmetrical example, the minimum point for the middle *SRAC* curve is indeed on the envelope. However, the minimum points for all other *SRAC* curves, which are indicated by small black squares, are above the envelope. In other words, a *SRAC* curve's minimum point is above its tangency point with the associated long-run average cost (*LRAC*) curve, except at the minimum of the *LRAC* curve, if an interior minimum exists.

The left-most of the squares, which is the minimum point for the *SRAC* with the lowest *y*-intercept, is far above the *LRAC* and far to the right of the tangency point between this *SRAC* and the *LRAC*. It obviously makes no sense to try to force this minimum point down to the envelope, which is want Viner asked his draftsman to do. In recognition of the push-back from Viner's draftsman, the corrected envelope-*SRAC* relationship, such as the one in Figure 3.1, is often called the Viner-Wong diagram.

Despite all the attention that scholars have given to the Viner article, it was actually another article published in 1931 that introduced the term "envelope" into economists' lexicon. This article by Roy Harrod focused on the case of declining long-run average costs, which is the case that is incorrectly drawn by Viner. This case was particularly interesting to Harrod (and Viner) because it is associated with natural monopoly. In a book review published in 1959, Harrod reflects on his contribution:

> I had already drawn these curves myself (in an article which appeared in the *Economic Journal*, December, 1931), before Viner's article appeared not in the way that Viner wished his draughtsman to draw them, but in the way that his draughtsman insisted on doing. I remember vividly—it seems like yesterday—wandering across Tom Quad in Christ Church in quest of a scientific colleague. I showed him my curves. "Look at these curves," I said, "there must be some name in mathematics for this outside curve." "Yes," he replied, "it is called an envelope." As I used this word in my *Economic Journal* article (1931) and again in the *Quarterly Journal of Economics* (1934), I like to think that my walk across the quadrangle may have been responsible for initiating in economics the correct usage, for neglecting which Viner has been so much taunted! (Harrod 1959, p. 262)

All that is lacking from this story is the name of the colleague who handed Harrod the "envelope."[3]

Figure 2 in Harrod's 1931 article draws a series of parabolic short-run cost curves with the associated envelope.[4] His language is worth citing in detail because it is the first appearance of the term "envelope" in economic analysis:

> The cost of production may be represented by a family of parabolas, each of which shows the cost of any output from a plant of given size. The lowest point of the parabola shows the cost of the optimum output from its plant. The minimum point is supposed lower the larger the size of plant, and the locus of these points a curve falling smoothly for increasing values of x, the output. It is required to find the proper size of plant for any given prospective normal demand, x_l. This is the plant the parabola of which has of all the parabolas the lowest value for x_l; x_l units can be produced most cheaply from a plant

> of such a size. Plot a curve (see Fig. 2) the ordinate of which is equal to the lowest of the ordinates of all the parabolas for each value of x. Such a curve (the envelope) may be called the long-period productive cost curve, for it shows the cost of producing the normally required output x_l, if that is properly foreseen. (Harrod 1931, p. 575)

Because of this contribution, it seems to me that we should be talking about Viner-Wong-Harrod diagrams.[5]

Viner and Harrod had met in Oxford early in 1931, and they quickly became aware of each other's work on cost curves.[6] Sometime late in 1931 or early 1932, Harrod wrote a letter to Viner. This letter has been lost to time, but Viner's reply has not. Here are the key excerpts (Viner 1932):

> Department of Economics, University of Chicago
> 23 February 1932
>
> Dear Harrod:
>
> Deepest apologies for my delay in acknowledging both the reprint of your excellent article and your most generous note of comment on my own cost article. I was convalescing from an appendix operation when your article and letter arrived, and I postponed a reply until I had had time to give your argument careful consideration. There follow some comments, first, on your article, and next, on your letter.
>
> ….
>
> Finally, as to my article, the draftsman and yourself are right, and I was wrong. It was a natural but mistaken tendency to assume that optimum scale for a particular output and lowest cost output for that scale would be identical. But you saw what was wrong with my reasoning and my draftsman did not.
>
> Could you spare another reprint or two for use with my students?
>
> Most cordially yours,
> Jacob Viner

Curiously, however, both scholars appear to have forgotten about this exchange. In the "supplementary note" Viner published in 1950, he makes no mention of Harrod's article. A reprint of this supplementary note in 1958, adds a footnote, which is specifically dated 1951, that refers to an article by Harrod published in 1930, but not to his envelope paper of 1931.

Moreover, so far as I can tell, this footnote has nothing to do with the drawing of envelopes.[7] Harrod was then given the task of reviewing this book. The relevant passage from his review is

> Viner refers in a footnote to having seen my 1930 article between 1950 and 1951, and I take it that this applies to the 1931 article also, the latter being the one referring to the "envelope." It makes me a little wistful that Viner, despite his scholarly devotion to origins, did not see my article for twenty years after it appeared (Harrod 1959, p. 262).

Did Viner really forget about receiving a reprint of Harrod's 1931 article or is he leaving out the reference on purpose? Did Harrod really forget that he sent his article to Viner or is he just trying to be polite? I guess we will never know.

We are still not done with the story in 1931. In fact, there was even a third article. In his retrospective on Viner, Samuelson (1972) points out that despite the Viner mistake there was no reason for anyone to be confused about the relationship between short-run and long-run cost curves.

> Yet in the same 1931 volume of the *Zeitschrift für Nationalökonomie* there appears, bound with Viner's, a paper by Erich Schneider; here, in the guise of total rather than average-cost curves, appears a clear depiction of the proper envelope relations. And of course Roy Harrod later set the matter straight in terms of U-shaped curves. (1972, p. 9)

So far as I know, the Schneider article did not use the term "envelope," but, as we will see, his approach has been picked up by some other scholars.[8]

Although most economists link long-run average costs curves and the "envelope" terminology to the articles from 1931, these curves and this term actually appeared 42 years earlier in a book by Rudoph Auspitz and Richard Lieben (1889). As explained by Schmidt (2004), these authors were bankers[9] who had no formal training, were not academics,[10] wrote in German, and were known to only a few English-speaking economists.[11] Nevertheless, they drew a series of short-run average cost curves and indicated, as Schmidt (2004, p. 120-121) explains it, that[12]

> the optimizing decision maker who is able to switch modes would seek to trace out the lowest attainable cost over this collection of curves given by their lower boundary…. In their explanation of [this figure] they do not use the term *envelope*, notwithstanding their repeated use of the term in the appendixes (that is, much later in the book); nor do they refer back to [this figure] in the appendixes.

The punch line, it seems to me, is that Auspitz and Lieben invented the long-run average cost curve in 1889, and Harrod can only retain half credit for introducing the term "envelope" into the analysis of cost-curves.

3.1.2. The Envelope Theorem

Viner's article is well-known not only because it contained this drafting error, but also because this error opened the door to a very useful tool, known as the "Envelope Theorem." As Silberberg (1999, p. 75) put it, "The Envelope Theorem, now the fundamental tool in modem duality analysis, had its beginnings in Jacob Viner's classic 1931 article on short- and long-run cost curves." It seems reasonable to amend this statement to bring in Harrod and Schneider. After all, Harrod's 1931 diagram, unlike Viner's, was actually consistent with the envelope theorem. Moreover, Schneider (1931), like Silberberg, uses total-cost curve diagrams to illuminate the cost envelope. In short, the Envelope Theorem grew out of a series of articles on cost curves published in 1931. Samuelson, who introduced the Envelope Theorem to economists, recognizes these contributions when he refers to the "Wong-Viner-Harrod envelope theorem" (1947, p. 234). [13]

Before proceeding, it is worth noting that envelopes can be drawn and interpreted without reference to the Envelope Theorem and that the Envelope Theorem can be invoked without deriving or plotting an envelope. The Envelope Theorem is not required, for example, to understand the long-run cost curves derived in Section 3.2. Nevertheless, envelope graphs and the Envelope Theorem are inextricably connected. Every drawn envelope has an application of the Envelope Theorem imbedded in it, and every application of the Envelope Theorem can be illustrated with a graph. Of course, the theorem associated with a particular envelope graph may not be interesting and a graph may not add much insight to a particular application of the Envelope Theorem, but understanding this connection is helpful in many cases.

In the context of this book, a discussion of the Envelope Theorem is valuable for two reasons. First, it provides another example of the use of envelopes in economics. In fact, this book makes ample use of this tool in later chapters. Second, a discussion of the Envelope Theorem in general and of its link to the Viner-Wong-Harrod diagram in particular provides some insight into the interpretation of many envelopes that arise in economics. As a result, the rest of this section provides interested readers with a brief introduction to the Envelope Theorem.[14]

The standard version of the Envelope Theorem concerns a function, say F, to be maximized (or minimized) subject to a constraint, say $G = 0$. To keep the exposition simple, suppose that F and G both depend on one parameter, α, and two variables, X_1 and X_2. (The theorem can easily be generalized to any number of parameters and variables.) As a result, the initial problem is to select the values of X_1 and X_2 that maximize $F\{X_1, X_2, \alpha\}$ subject to $G\{X_1, X_2, \alpha\} = 0$. The associated Lagrangian expression is $F\{X_1, X_2, \alpha\} + \lambda(G\{X_1, X_2, \alpha\})$, where λ is the Lagrangian multiplier. The solution to this problem consists of optimal values for X_1 and X_2, say X_1^* and X_2^*, along with the optimal value for the Lagrangian multiplier, λ^*. These values obviously depend on α, that is, $X_1^* = X_1^*\{\alpha\}$, $X_2^* = X_2^*\{\alpha\}$, and $\lambda^* = \lambda^*\{\alpha\}$. Plugging these values into F yields the maximum value of F given α: $F\{X_1^*, X_2^*, \lambda^*\{\alpha\}, \alpha\} = F^*\{\alpha\}$.

Now suppose one wants to find out what happens to the optimal value of F, namely, $F^*\{\alpha\}$, changes when α changes. This is called an exercise in "comparative statics." How does the (static) solution with one value of α compare to the solution when α increases by a little bit? Because X_1^* and X_2^* are functions of α, it would seem that one would have to figure out how X_1^*, X_2^*, and $\lambda^*\{\alpha\}$ change when α changes in order to determine the change in $F^*\{\alpha\}$. The Envelope Theorem says that this is not the case—that a short-cut is available. More specifically, this theorem says that $dF^*\{\alpha\} /d\alpha = [dF/d\alpha - \lambda^*\{\alpha\}(dG/d\alpha)]$.

Now consider the case of the cost functions we are exploring in this chapter.[15] The relevant parameter is quantity, Q, and the variables are K and L. The long-run problem is to select the values of K and L that minimize average costs (or total costs) for a given value of Q as determined by the production function. In Figure 3.1, each short-run cost curve corresponds to a different, fixed value for K. So the way to minimize costs at a given Q is to select the lowest *SRAC* curve (and associated K) at that Q. Figure 3.1 makes it clear that the relevant *SRAC* curve is the one that is tangent to the *LRAC* curve. This figure also makes it clear that this tangency point is not the minimum point of that *SRAC* curve, except in the case of the *SRAC* curve in the middle, where the *LRAC* curve is flat.

More formally, the function to be minimized is $F = LRAC = (rK + wL)/Q$, where *LRAC* is long-run average costs, K is capital, L is labor, r is the rental rate on capital, and w is the wage rate. In addition, the constraint, G, is just the production function, say g, in implicit form, or $g\{L, K\} - Q = 0$, where Q is output.[16] The Lagrangian expression is $F + \lambda(g - Q)$. With Q treated as a parameter, the first-order conditions of this problem lead to $K^*\{Q\}$, $L^*\{Q\}$ and $\lambda^*\{Q\}$, and $F^*\{Q\} = [(rK^*\{Q\} + wL^*\{Q\})/Q]$. Now suppose we want to know how *LRAC* changes when Q changes. The Envelope Theorem tells us that $dF^*\{Q\}/dQ = (dF/dQ + \lambda dG/dQ)$. The trick in this case is that Q cannot change unless one of the inputs changes. So we hold K constant and allow L to change. This corresponds to moving along the *SRAC* curve that is tangent to the *LRAC* curve at the initial value of *Q*. Because the *SRAC* and *LRAC* curves are tangent, a small change in *Q* lead to the same change in average costs regardless of which curve we follow. To put it another way, for small changes in *Q* we can obtain a first-order approximation to movement along the *LRAC* curve by looking at movement along the *SRAC* curve that is tangent to the *LRAC* curve at our starting point. That is the essence of the Envelope Theorem.

Thus, Viner's mistake helped economists to recognize that it is helpful to focus on the tangency between a *SRAC* curve and the *LRAC* curve at a given *Q* instead of on the value of *Q* that leads to a minimum *SRAC* with a given *K*. This shift in focus opened the door to the Envelope Theorem. Would that all our mistakes, when corrected, could be so insightful![17] This insight also comes, of course, from the correct drawing in Harrod (1931). The Harrod article not only re-introduced economists to the "envelope" term, but should also be given credit, along with Viner (1931), for helping to inspire for the envelope theorem.

Schmidt (2004) makes a convincing case that the intellectual history of the Envelope Theorem, like that of the *LRAC* graph, actually began with Auspitz and Lieben in 1889. As Schmidt explains it (2004, p. 126):

> The envelope theorem asserts that the effect of a small parameter change on the optimum value of the decision maker's objective will be the same with, and without, optimizing adjustment of all decision variables. Auspitz and Lieben used the envelope principle on six occasions. They did not state it as a formal theorem on its own. Also, it is not self-evident that they recognized its wide-ranging applicability. But their analysis as presented leaves no doubt that they understood what they were doing…. Reliance on the fundamental principle of the envelope theorem emerges as one of Auspitz and Lieben's contributions to economic theory that went unrecognized.

This analysis leads Schmidt (2004, p. 126) to just the right conclusion: "If the naming of the envelope theorem should honor persons with some prior connection with some aspect of the general theorem, as has been the case, it would now seem appropriate to add the names of Auspitz and Lieben."

3.2. Deriving Long-Run Average Cost Curves

Long-run average cost curves can be derived in at least two different ways. The first way is to specify the algebraic form for a *SRAC* curve and then to use the mathematical concepts in Chapter 1 to derive the envelope, that is the *LRAC* curve. The second way is to specify a form for the production function, derive the form of the *SRAC* curves, and then use economic logic to derive the associated *LRAC* curve. Section 3.2.1 applies the first approach to parabolic *SRAC* curves; Section 3.2.2 applies the second approach to Cobb-Douglas production functions, and Section 3.3.3 applies to second approach to CES production functions.

3.2.1. Parabolic Average Cost Curves

Many textbooks illustrate short-run average cost curves using a parabolic form. This type of curve can be written as follows, where Q is output and C_0, β, and γ are parameters:

$$SRAC = c\{Q,\gamma\} = C_0 + \beta(Q-\gamma)^2 \tag{0.1}$$

In this formulation, γ is the parameter that varies across the members of a curve "family," as defined in Chapter 1, whereas C_0 and β are the same for all family members. In this case, the corresponding version of Equation (1.1) is just Equation (3.1) re-written in implicit form, that is,

$$F\{\gamma,Q,s\} = c - \left(C_0 + \beta(Q-\gamma)^2\right) = 0 \tag{0.2}$$

In addition, the corresponding version of Equation (1.2) is the derivative of equation (3.2) with respect to γ, or.

$$\frac{dF\{\gamma,Q,s\}}{d\gamma} = -2\beta(Q-\gamma) = 0 \tag{0.3}$$

Solving Equation (3.3) for γ and substituting the result into Equation (3.2) [or Equation (3.1)] yields the envelope

$$LRAC = C\{Q\} = C_0 \tag{0.4}$$

This long-run average cost curve is obviously characterized by constant costs. An example of this these short-run cost curves and their envelope is drawn in Figure 3.2.[18]

Other families of short-run cost curves lead to U-shaped long-run average cost curves. Consider the family

$$SRAC = c\{Q\} = C_0 + \beta\left(\gamma - \gamma^*\right)^2 + \beta(Q-\gamma)^2 \tag{0.5}$$

Writing this equation in implicit form and then differentiating with respect to the varying parameter, γ, leads to

$$F\{\gamma, Q, c\} = c - \left(C_0 + \beta\left(\gamma - \gamma^*\right)^2 + \beta(Q-\gamma)^2\right) = 0 \tag{0.6}$$

$$\frac{\partial F\{\gamma, Q, c\}}{\partial \gamma} = 2\beta\left(\gamma - \gamma^*\right) - 2\beta(Q-\gamma) = 0 \tag{0.7}$$

Solving Equation (3.7) implies that

$$\gamma = \frac{Q + \gamma^*}{2} \tag{0.8}$$

Substituting Equation (3.8) into Equation (3.6) yields the envelope:

$$LRAC = C\{Q\} = C_0 + \frac{\beta\left(Q - \gamma^*\right)^2}{2} \tag{0.9}$$

This long-run average cost is illustrated in Figure 3.3.

This technique can also be used to generate a figure that looks like Harrod's (1931) Figure 2, which was the first to show the *LRAC* curve as an envelope. Suppose that the short-run curve takes the following form

$$SRAC = c\{Q\} = C_0 - \alpha_1\gamma + \alpha_2\gamma^2 + \beta(Q-\gamma)^2 \tag{0.10}$$

where C_0, α_1, α_2, and β are all positive and fixed across firm sizes. Then following the same steps leads to the envelope

$$LRAC = C\{Q\} = C_0 - \alpha_1\left(\frac{2\beta Q + \alpha_1}{2(\beta + \alpha_2)}\right) + \alpha_2\left(\frac{2\beta Q + \alpha_1}{2(\beta + \alpha_2)}\right)^2 + \beta\left(\frac{2\alpha_2 Q - \alpha_1}{2(\beta + \alpha_2)}\right)^2 \qquad (0.11)$$

These short- and long-run cost curves are illustrated in Figure 3.4. The parameter values and formatting for this figure are selected to make the figure look as much as possible like Harrod's Figure 2.

A *LRAC* curve can also be derived using the technique developed in Chapter 1, Section 1.2.2. With this technique, the first step is to solve for the change in the intercept that is needed to generate the same *LRAC* at a given *Q* when the slope changes. In the case of Equation (3.1) with β constant across plant sizes but with variation in γ, this step leads to:

$$\frac{dC_0}{d\gamma}\Big|_{Q=Q\{\gamma\}} = -2\beta(Q-\gamma) \qquad (0.12)$$

The second step is to characterize the relationship between the slope parameter and the outcome, in this case *Q*, along the envelope. We know that the minimum *Q* for each member of the family of *SRAC* curves described Equation (3.1) equals the value of γ for that family member. But we also know, following Figure 3.1, that the points on the envelope do not equal these minimum points but are instead more spread out. Suppose we think they are twice as spread out, in the sense that a change in γ has twice the impact on the tangency point as it does on the minimum point. Finally, suppose we think this double impact occurs on either side of an overall minimum cost at $Q = \gamma^*$, where γ^* is a parameter that does not vary across firm sizes. These assumptions imply that the long-run relationship between Q and γ is

$$Q = 2\gamma - \gamma^* \qquad (0.13)$$

Equation (3.13) is equivalent to Equation (3.8); we just derived it through a different route. The two approaches do not necessarily lead to the same answer, however, because the method we are using now, unlike the method presented earlier, does not hold C_0 constant across members of the *SRAC* curve family.

Substituting Equation (3.13) into Equation (3.12) yields

$$\frac{dC_0}{d\gamma} = -2\beta\left(2\gamma - \gamma^* - \gamma\right) = -2\beta\left(\gamma - \gamma^*\right) \tag{0.14}$$

The solution to this simple differential equation is

$$C_0 = C^* + \beta\left(\gamma - \gamma^*\right)^2 \tag{0.15}$$

where C^* is a constant of integration.

Now the envelope can be found by substituting Equation (3.13) and Equation (3.8), which is the inverse of Equation (3.13), into Equation (3.1):

$$\begin{aligned} LRAC = C\{Q\} &= C^* + \beta\left(\gamma' - \gamma^*\right)^2 + \beta\left(Q - \gamma'\right)^2 \\ &= C^* + \beta\left(\frac{Q+\gamma^*}{2} - \gamma^*\right)^2 + \beta\left(Q - \frac{Q+\gamma^*}{2}\right)^2 \\ &= C^* + \frac{\beta\left(Q-\gamma^*\right)^2}{2} \end{aligned} \tag{0.16}$$

This is, of course, the same as Equation (3.9).

Assumptions about the long-run relationship between Q and γ other than Equation (3.13) would lead to different forms for the envelope.

3.2.2. Average Cost Curves for Cobb-Douglas Production Functions

Cobb-Douglas production functions are widely used for both theoretical and empirical work in economics. They provide a simple introduction to the theoretical derivation of a *LRAC* curve, that is, of an envelope.[19] The intellectual history of this production function is reviewed by Douglas (1976), who is one of the scholars after whom it is named.

A Cobb-Douglas production function can be written as follows, where Q is quantity, K is capital, L is labor, and A, α, and β are parameters:

$$Q = AK^{\alpha}L^{\beta} \tag{0.17}$$

The sum of α and β determines returns to scale. In the case of constant returns $[(\alpha + \beta) = 1]$ and competitive markets, the α and β parameters indicate the output shares of capital and labor, respectively. Increasing returns arise when $[(\alpha + \beta) > 1]$ and decreasing returns require $[(\alpha + \beta) < 1]$.[20]

In the short run, the amount of capital is fixed at $\bar{K}$, so

$$L = \left(\frac{Q}{A\bar{K}^{\alpha}}\right)^{1/\beta} \tag{0.18}$$

It follows that total costs in the short run are

$$SRTC = rK + wL = r\bar{K} + w\left(\frac{Q}{A\bar{K}^{\alpha}}\right)^{1/\beta} \tag{0.19}$$

Dividing Equation (3.19) by Q yields $SRAC$:

$$SRAC = s\{Q\} = \frac{r\bar{K}}{Q} + w\frac{Q^{(1/\beta)-1}}{\left(A\bar{K}^{\alpha}\right)^{1/\beta}} \tag{0.20}$$

In the long run, a firm will select the value of $\bar{K}$ that minimizes $SRAC$ at any given Q. The relevant first-order condition is found by differentiating Equation (3.20) and setting the result equal to zero:

$$\frac{ds}{d\bar{K}} = \frac{r}{Q} - \left(\frac{wQ^{(1/\beta)-1}}{A^{1/\beta}}\right)\left(\frac{\alpha}{\beta}\right)\bar{K}^{-(\alpha/\beta)-1} = 0 \tag{0.21}$$

Solving for $\bar{K}$ yields

$$\bar{K} = \left(\left(\frac{w\alpha}{r\beta} \right) \left(\frac{Q}{A} \right)^{1/\beta} \right)^{\beta/(\alpha+\beta)} \tag{0.22}$$

The final step is to substitute Equation (3.22) into (3.20), which gives the envelope of the *SRAC* curves or

$$LRAC = Q^{\frac{1-(\alpha+\beta)}{\alpha+\beta}} \left(\frac{w^{\beta} r^{\alpha}}{A} \right)^{1/(\alpha+\beta)} \left(\left(\frac{\beta}{\alpha} \right)^{\alpha/(\alpha+\beta)} + \left(\frac{\alpha}{\beta} \right)^{\beta/(\alpha+\beta)} \right) \tag{0.23}$$

Figures 3.4, 3.5, and 3.6 provide examples of this envelope. Figure 3.4 considers the case of constant returns ($\alpha + \beta = 1$); Figure 3.5 shows the case of decreasing returns ($\alpha + \beta < 1$); and Figure 3.6 applies to increasing returns ($\alpha + \beta > 1$). In standard terminology, a production function with decreasing returns to scale leads to cost curves with "diseconomies of scale." Similarly, increasing returns to scale in production lead to "economies of scale" in the associated cost curves.

3.2.3. Average Cost Curves for CES Production Functions

A CES production function was first derived in a famous article by Arrow, Chenery, Minhas, and Solow (1961). The elasticity of substitution is defined as the elasticity of the ratio of two inputs to the ratio of their marginal products. It is a measure of the ease with which one input can be substituted for another. The acronym "CES" stands for a constant elasticity of substitution.

A CES production function can be written as follows, where A, α, σ, and ρ are fixed parameters.

$$Q = A\left(\alpha K^{\rho} + (1-\alpha) L^{\rho} \right)^{\sigma/\rho} \tag{0.24}$$

In this formulation, α is often called the share parameter, ρ indicates the degree of substitutability between the inputs, and σ indicates returns to scale. More specifically, $\rho <= 1$ and $1/(1 - \rho)$ is the elasticity of substitution. With constant returns to scale ($\sigma = 1$) and a unitary elasticity of substitution ($\rho = 0$), Equation (3.24) simplifies to the constant return version of Equation (3.17), which requires $\beta = 1 - \alpha$. The CES form also covers the case of perfect substitutes with isoquants that are straight lines ($\rho = 1$) and of no substitution with right-angle isoquants ($\rho = -\infty$).

Now fixing K at a given value, $\bar{K}$, the labor input, L, required for a given output Q, is

$$L = \left(\frac{\left(\frac{Q}{A}\right)^{\rho/\sigma} - \alpha \bar{K}^{\rho}}{1-\alpha} \right)^{1/\rho} \tag{0.25}$$

With capital rental rate r and wage w, the short-run average cost curves can therefore be written

$$SRAC = s\{Q\} = \frac{r\bar{K}}{Q} + \left(\frac{w}{Q}\right)\left(\frac{\left(\frac{Q}{A}\right)^{\rho/\sigma} - \alpha \bar{K}^{\rho}}{1-\alpha} \right)^{1/\rho} \tag{0.26}$$

Long-run average costs can be found by selecting K and L to minimize total costs subject to the production function. In symbols, the problem is to

$$\begin{aligned} &\text{Minimize:} && rK + wL \\ &\text{Subject to:} && A\left(\alpha K^{\rho} + (1-\alpha)L^{\rho}\right)^{\sigma/\rho} = Q \end{aligned} \tag{0.27}$$

Dividing the first-order condition for K by the first-order condition for L and re-arranging terms, we find that

$$K = L\left(\frac{(1-\alpha)r}{\alpha w}\right)^{1/(\rho-1)} \tag{0.28}$$

Substituting this expression into the production function (that is, into the constraint) leads to a solution for L and hence, using Equation (3.28), to a solution for K:

$$K = \frac{\left(\frac{Q}{A}\right)^{1/\sigma}}{\theta} \quad \text{and} \quad L = \left(\frac{\alpha w}{(1-\alpha)r}\right)^{1/(\rho-1)} \frac{\left(\frac{Q}{A}\right)^{1/\sigma}}{\theta} \tag{0.29}$$

where

$$\theta = \left(\alpha + (1-\alpha)\left(\frac{\alpha w}{(1-\alpha)r}\right)^{\rho/(\rho-1)}\right)^{1/\rho} \tag{0.30}$$

Long-run average costs equal $(rK + wL)/Q$ or

$$LRAC = l\{Q\} = \left(r + w\left(\frac{\alpha w}{(1-\alpha)r}\right)^{1/(\rho-1)}\right)\left(\frac{Q^{(1/\sigma)-1}}{\theta A^{1/\sigma}}\right) \tag{0.31}$$

Figures 3.7 and 3.8 illustrate this result with constant ($\sigma = 1$) and increasing ($\sigma > 1$) returns, respectively.

3.2.4. A Brief Return to the Envelope Theorem

The link between an envelope and the envelope theorem was derived in Section 3.1.3, and it is clear in the figures presented in this chapter. It can also be shown algebraically. All the preceding examples involve a parameter that varies across *SRAC* curves, such as $\bar{K}$. For any given quantity, Q, the derivative of the *SRAC* curve with respect to Q, evaluated at the value of this parameter that sets *SRAC* = *LRAC*, equals the derivative of the *LRAC* curve with respect to Q. A first-order approximation to the impact of a small change in Q on long-run average costs can be found, therefore, by finding the derivative of an *SRAC* curve where it is tangent to the envelope (= *LRAC* curve).

Consider the case of the *SRAC* curve in Equation (3.5) and the associated *LRAC* curve in Equation (3.9). Differentiating each of these equations with respect to Q yields

$$\frac{\partial SRAC}{\partial Q} = 2\beta(Q-\gamma) \tag{0.32}$$

and

$$\frac{\partial\, LRAC}{\partial Q} = \beta\left(Q - \gamma^*\right) \tag{0.33}$$

Now setting *SRAC* = *LRAC* and the quadratic formula to solve for the parameter that defines a *SRAC* curve, *γ*, yields Equation (3.7). Substituting this result into Equation (3.32) leads to Equation (3.33). When evaluated at their point of tangency, the *LRAC* and *SRAC* curves have the same slope with respect to *Q*. We already knew this, of course; this algebra just provides another way to see how envelopes and the Envelope Theorem are connected.

3.3. Conclusions

Mathematical envelopes play an important role in the development of modern microeconomics. As far back as Auspitz and Lieben in 1889, economists recognized that long-run average or cost curves are the envelope of the family of short-run average cost curves at different scales. The details of this analysis are widely known, thanks in part to the famous error made by Viner (1931).

The development of his cost-curve analysis also contributed to the development of the envelope theorem, which is an important tool for many microeconomic problems. Several of the derivations in this book rely on this theorem. As pointed out earlier, mathematical envelopes and the envelope theorem each stand on their own, but the link between them provides powerful intuition for each one.

Because a long-run cost function is the envelope of a family of short-run cost functions with different plant sizes or scales, a long-run cost function can be derived for a wide range of assumptions about the functional form of short-run costs. This chapter shows how to derive long-run average cost functions, i.e. envelopes, for parabolic short-run average cost curves, Cobb-

Douglas production functions, and CES production functions. These forms, which are widely known, have proven to be useful in many cost-function studies.[21] The key analytical lesson here, namely, that different assumptions about the form of a family of curves to describe some type of economic behavior can lead to very different forms for the associated hedonic function, plays an important role in many of this e-book's remaining chapters.

References

Arrow, K. J., H. B. Chenery, B. S. Minhas and R. M. Solow. 1961. "Capital-Labor Substitution and Economic Efficiency." *The Review of Economics and Statistics* 43 (3) (August): 225-250.

Auspitz, Rudolph and Richard Lieben. 1889. *Untersuchungen über die Theorie des Preises* [Investigations in the Theory of Price]. Leipsig: Verlag von Duncker & Humblot.

Besomi, Daniele. Undated. "Jacob Viner." In Besomi, Daniele, *Collected Interwar Papers and Correspondence of Roy Harrod*, Electronic Version. Available at: http://economia.unipv.it/harrod/edition/welcome.htm.

Carlaw, Kenneth I. 2004. "Returns to Scale Generated from Uncertainty and Complementarity." *Journal of Economic Behavior and Organization* 53 (2) (February): 261–282.

Douglas, Paul H. 1976. "The Cobb-Douglas Production Function Once Again: Its History, Its Testing, and Some New Empirical Values." *Journal of Political Economy* 84 (5) (October): 903-916.

Duncombe, William D., and John Yinger. 1993. "An Analysis of Returns to Scale in Public Production, With an Application to Fire Protection." *Journal of Public Economics* 52 (1) (August): 49-72.

Duncombe, William D., and John Yinger. 2007. "Does School District Consolidation Cut Costs?" *Education Finance and Policy* 2 (4) (Fall): 341-375.

Duncombe, William D., and John Yinger. 2011. "Are Education Cost Functions Ready for Prime Time? An Examination of their Validity and Reliability." *Peabody Journal of Education* 86 (1): 28-57.

Harrod, Roy F. 1930. "Notes on Supply," *The Economic Journal* 40 (158) (June): 232-241.

Harrod, Roy F. 1931. "The Law of Decreasing Costs." *The Economic Journal* 41 (164) (December): 566-576.

Harrod, Roy F. 1934. "Doctrines of Imperfect Competition." *The Quarterly Journal of Economics* 48 (3) (May): 442-470.

Harrod, Roy F. 1959. "Review of 'The Long and the Short' by Jacob Viner." *Economica*, New Series 26 (103) (August): 260-262.

Harrod, Roy. 1972. "Imperfect Competition, Aggregate Demand and Inflation." *The Economic Journal* 82 (325), March): 392-401.

Samuelson, Paul Anthony. 1947. *Foundations of Economic Analysis*. Cambridge, MA: Harvard University Press.

Samuelson, Paul Anthony. 1972. "Jacob Viner, 1892-1970. " *Journal of Political Economy* 80 (1) (Jan. - Feb.): 5-11.

Samuelson, Paul A. 1988. "The Passing of the Guard in Economics." *Eastern Economic Journal* 14 (4) (October-December): 319-329.

Samuelson, Paul A. 1998. "How *Foundations* Came to Be." *Journal of Economic Literature* 36 (3) (September): 1375-1386.

Schmidt, Torsten. 2004. "Really Pushing the Envelope: Early Use of the Envelope Theorem by Auspitz and Lieben." *History of Political Economy* 36 (1) (Spring): 103-129.

Schneider, Erich. 1931. "Kostentheoretisches zum Monopolproblem." *Zeitschrift für Nationalökonomie* 3 (2):185–211.

Silberberg, Eugene. 1978. *The Structure of Economics: A Mathematical Analysis*. New York, McGraw-Hill)

Silberberg, Eugene. 1999. "The Viner-Wong Envelope Theorem." *Journal of Economic Education* 30 (1) (Winter): 75-79.

Viner, Jacob. 1927. “141R. J. Viner to Harrod, 21 October 1927.” In Besomi, Daniele, *Collected Interwar Papers and Correspondence of Roy Harrod*, Electronic Version. Available at: http://economia.unipv.it/harrod/edition/welcome.htm.

Viner, Jacob. 1931. “Cost Curves and Supply Curves.” *Zeitschrift für Nationalökomomie* 3 (2): 23-46.

Viner, Jacob. 1932. “227. Viner to Harrod , 23 February 1932.” In Besomi, Daniele, *Collected Interwar Papers and Correspondence of Roy Harrod*, Electronic Version. Available at: http://economia.unipv.it/harrod/edition/welcome.htm.

Viner, Jacob. 1950. “Supplementary Note,” In *Readings in Economic Analysis*, Vol. II, edited by R. V. Clemence (Cambridge, MA: Addison-Wesley Press), pp, 31-35.

Viner, Jacob. 1958. *The Long View and the Short*. Glencoe. IL: The Free Press.

Yinger, John, Forthcoming. “Still Unknown: The Impact of School Capital on Student Performance.” In J. Yinger, Editor. *Poverty and Proficiency, the Cost of and Demand for Local Public Education*, Chapter 21. World Scientific Publishing.

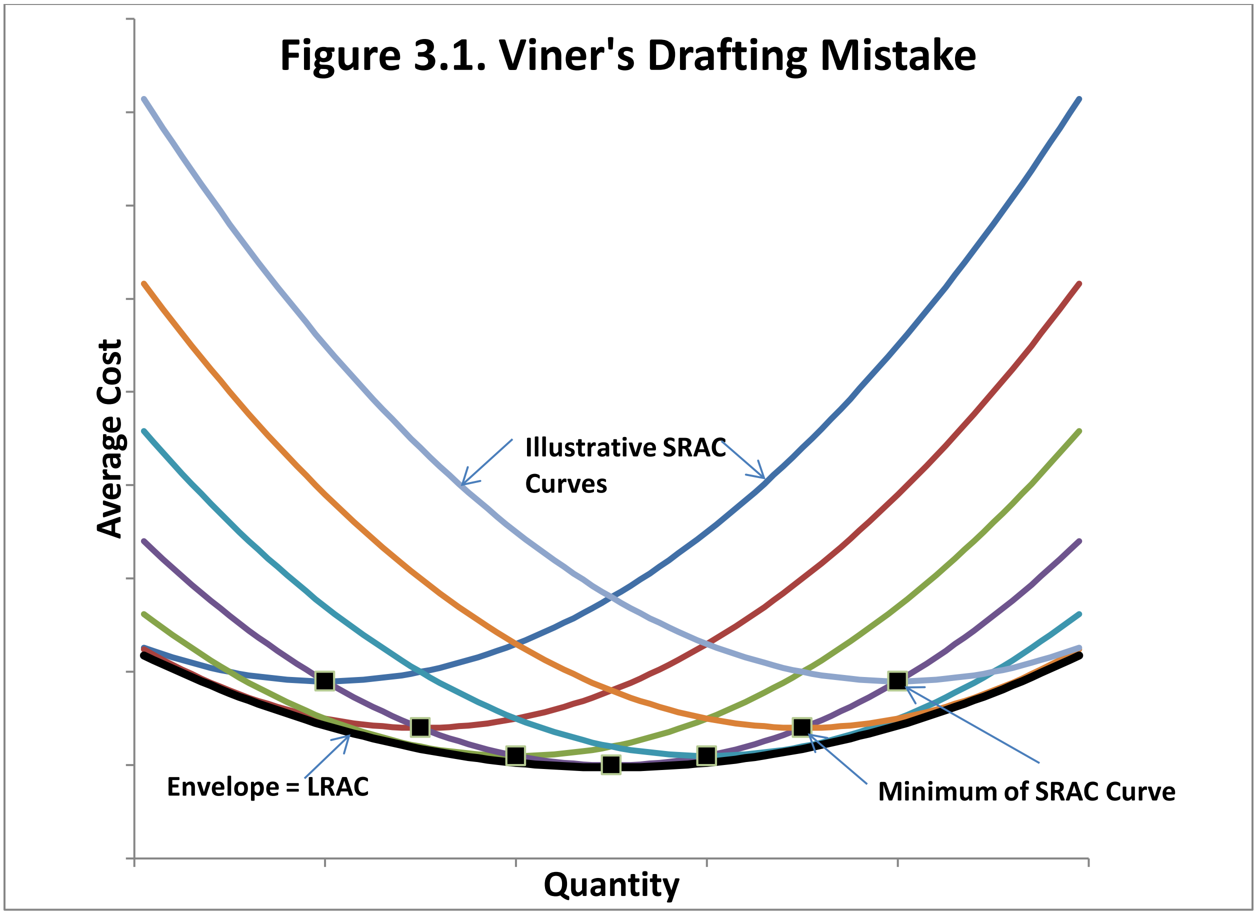


Figure 3.1. Viner's Drafting Mistake
Average Cost
Illustrative SRAC Curves
Envelope = LRAC
Minimum of SRAC Curve
Quantity

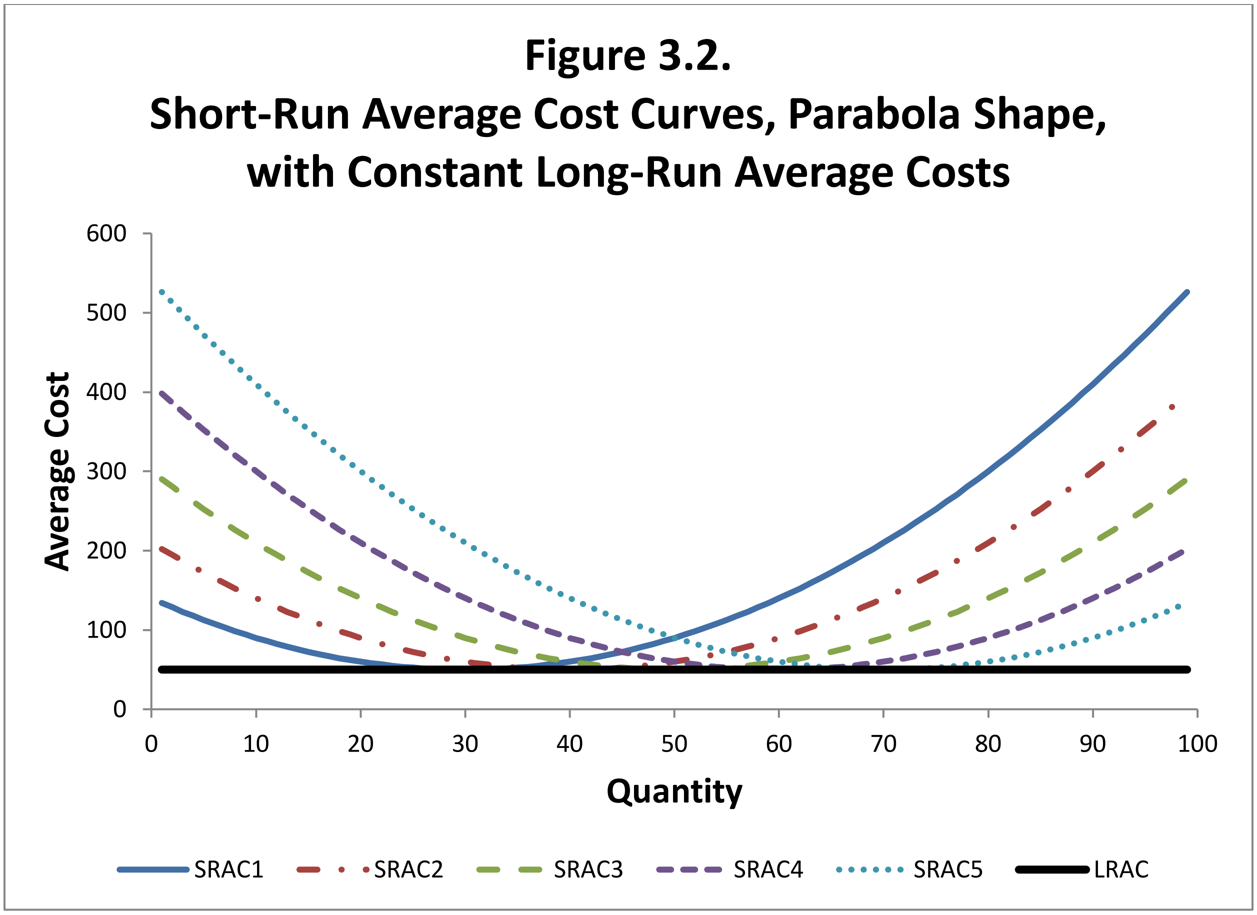


Figure 3.2.
Short-Run Average Cost Curves, Parabola Shape, with Constant Long-Run Average Costs
600
500
400
300
200
100
0
Average Cost
0
10
20
30
40
50
60
70
80
90
100
Quantity
SRAC1
SRAC2
SRAC3
SRAC4
SRAC5
LRAC

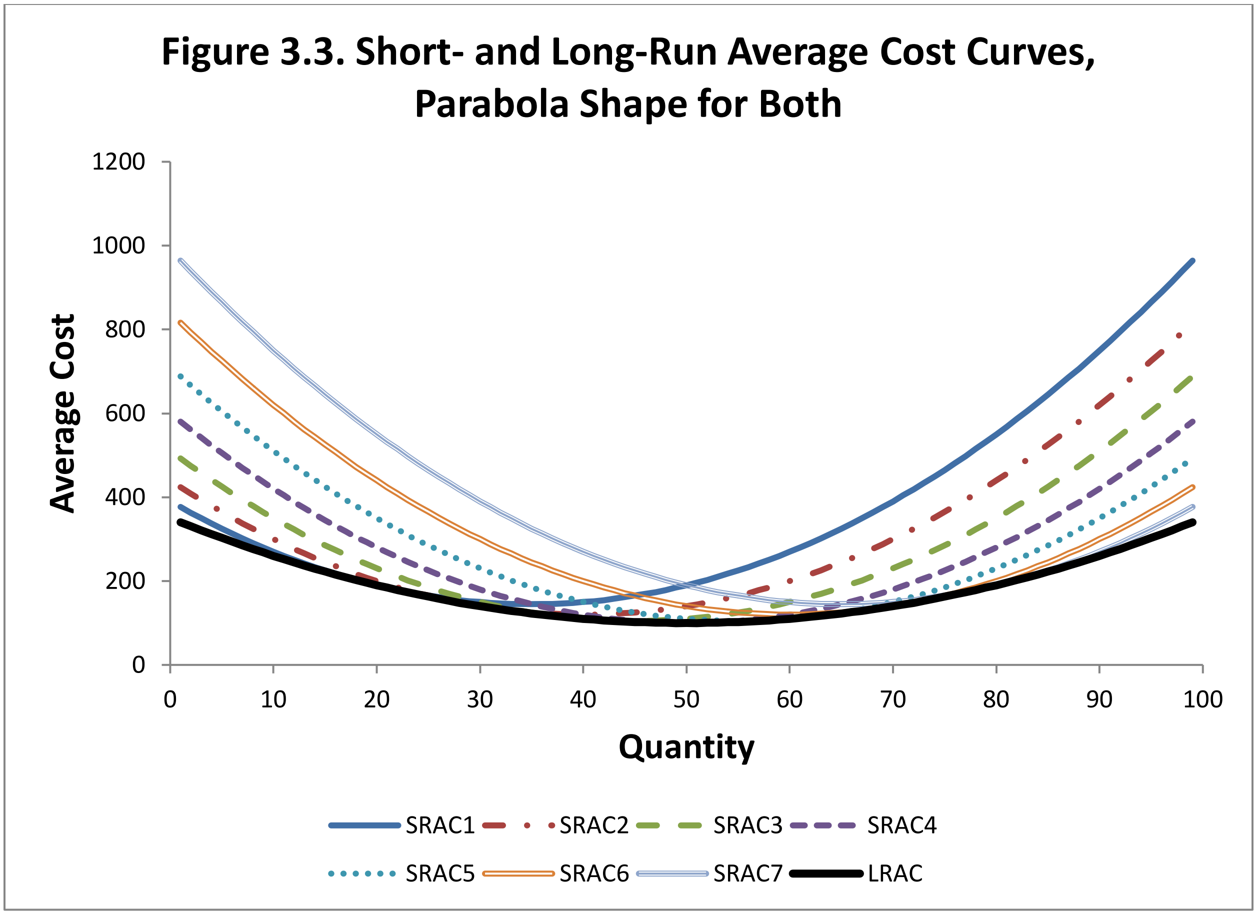


Figure 3.3. Short- and Long-Run Average Cost Curves,
Parabola Shape for Both
Average Cost
1200
1000
800
600
400
200
0
0
10
20
30
40
50
60
70
80
90
100
Quantity
SRAC1
SRAC2
SRAC3
SRAC4
SRAC5
SRAC6
SRAC7
LRAC

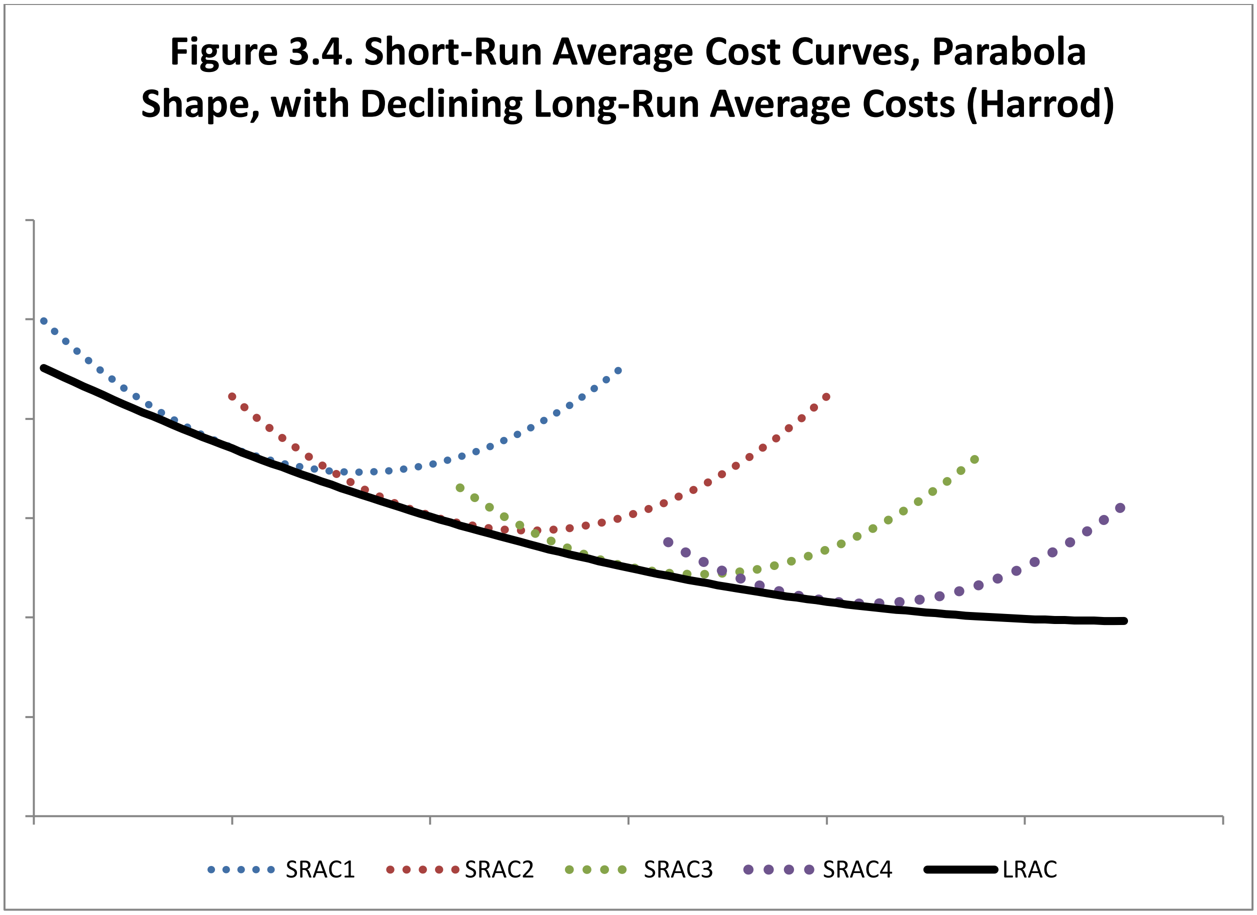


Figure 3.4. Short-Run Average Cost Curves, Parabola Shape, with Declining Long-Run Average Costs (Harrod)
SRAC1
SRAC2
SRAC3
SRAC4
LRAC

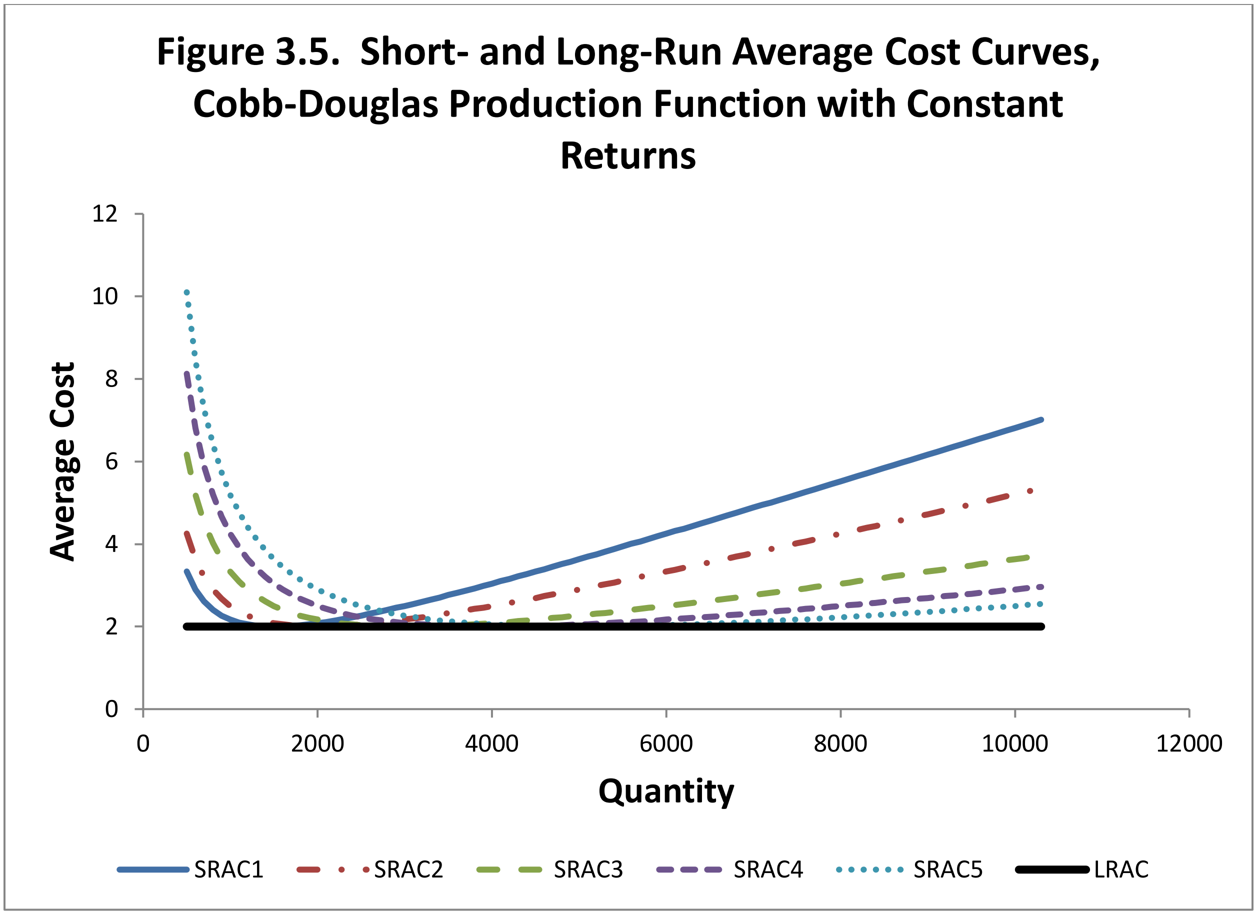


Figure 3.5. Short- and Long-Run Average Cost Curves, Cobb-Douglas Production Function with Constant Returns
Average Cost
12
10
8
6
4
2
0
0
2000
4000
6000
8000
10000
12000
Quantity
SRAC1
SRAC2
SRAC3
SRAC4
SRAC5
LRAC

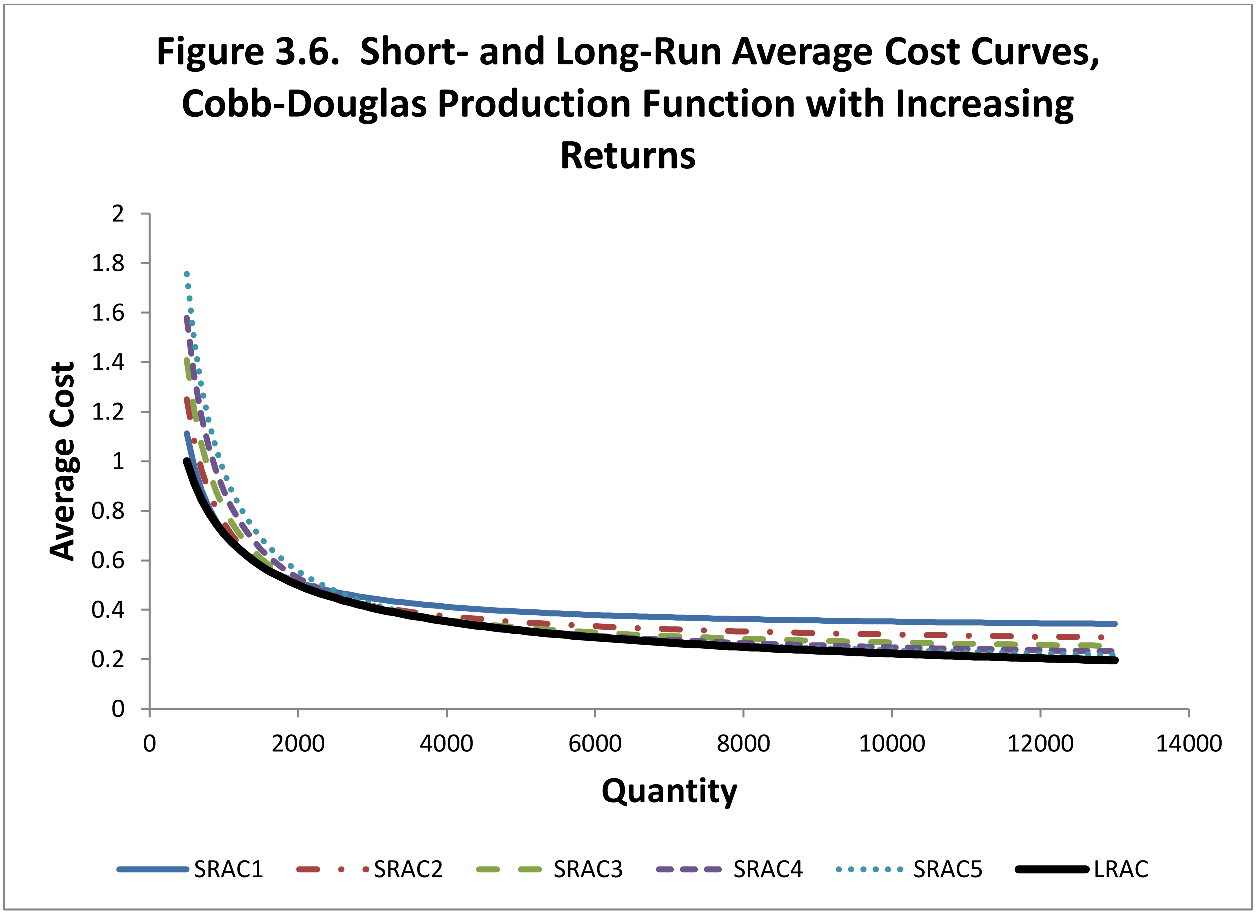
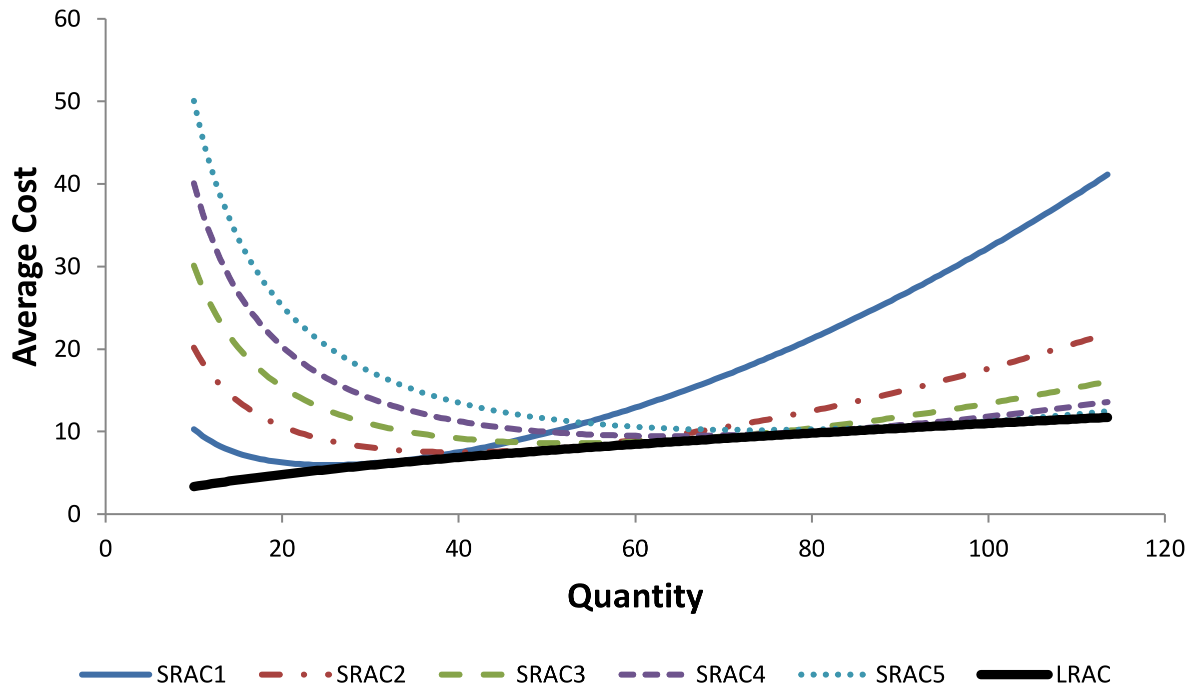


Figure 3.6. Short- and Long-Run Average Cost Curves, Cobb-Douglas Production Function with Increasing Returns
Average Cost
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
2000
4000
6000
8000
10000
12000
14000
Quantity
SRAC1
SRAC2
SRAC3
SRAC4
SRAC5
LRAC

Figure 3.7. Short- and Long-Run Average Cost Curves, Cobb-Douglas Production Function with Decreasing Returns



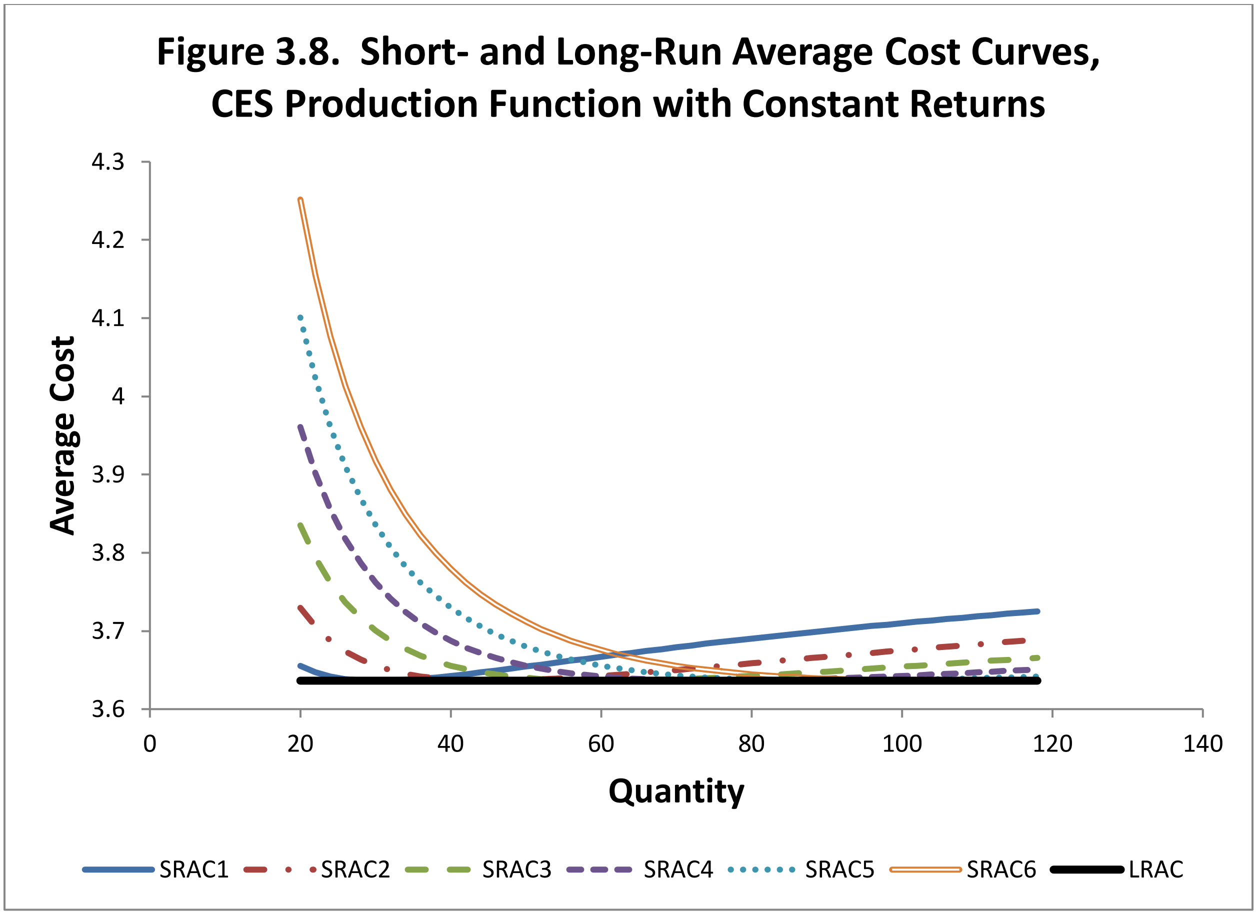


Figure 3.8. Short- and Long-Run Average Cost Curves,
CES Production Function with Constant Returns
Average Cost
4.3
4.2
4.1
4
3.9
3.8
3.7
3.6
0
20
40
60
80
100
120
140
Quantity
SRAC1
SRAC2
SRAC3
SRAC4
SRAC5
SRAC6
LRAC

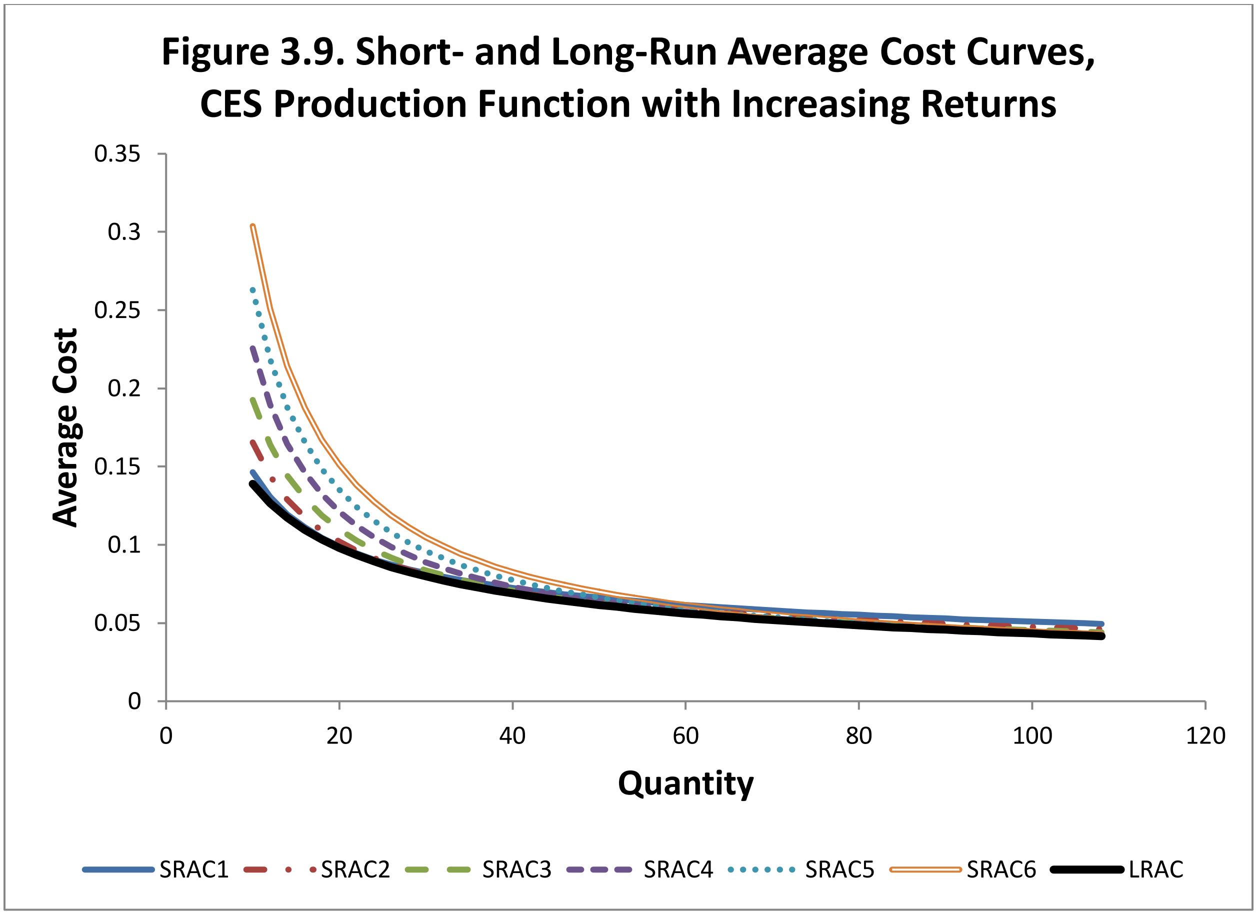


Figure 3.9. Short- and Long-Run Average Cost Curves, CES Production Function with Increasing Returns
Average Cost
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
20
40
60
80
100
120
Quantity
SRAC1
SRAC2
SRAC3
SRAC4
SRAC5
SRAC6
LRAC

Endnotes

[1] Viner acknowledged his error in a 1932 letter, which is discussed later in the text. Apparently, however, it took a little longer for him to be fully convinced. According to Samuelson (1972, p. 9), who was Viner's student,

> By 1935 Viner reported to the class that Wong had been right in 1931 and he, Viner, had been wrong, mathematically and economically. "But" he said to me privately just as the class bell had rung, "although there seems to be some esoteric mathematical reason why the envelope cannot be drawn so that it passes smoothly through the declining bottoms of the U -shaped cost curves, nevertheless I can do it!" "Yes," I replied impishly, "with a good thick pencil, you can do it."

Samuelson (1988, p. 322) repeats this story and adds "Viner's boner in trying to get Wong, the mathematical draftsman, to make the envelope to the family of descending U-shaped cost curves pass through their bottoms amused his admirers and mortified him."

[2] Viner's (1931) Chart IV depicts declining long-run average costs and therefore looks like the left half of Figure 3.1—but with distorted *SRAC* curves.

[3] In a later reflection, Harrod seems to forget that the long-run average cost curve published by Viner in 1931 was not correct—despite his "draughtsman's" attempts to correct it.

> The relation between these short-period and long-period phenomena may be represented by an envelope curve (cf. my article "The Law of Decreasing Costs," *Economic Journal*, December 1931). This curve was also published by Jacob Viner, not in consequence of his own thinking, but because his "draughtsman" insisted that the long- and short-period curves must be thus related. Being at Oxford, I had no draughtsman! (Harrod 1972, p. 396).

[4] Algebraic and graphical versions of Harrod's (1931) parabolic short-run cost curves and their declining envelope are provided in Section 2.2.1.

[5] Harrod (1934, pp. 450-451) provides another version of this explanation in his 1934 article, with a footnote referencing Viner:

> It might be thought that the long-period average total cost per unit curve was the locus of the lowest points of these parabolas. But that is not so. The long-period average total cost

of producing x units is the cost at which x units can be produced most cheaply. The lowest cost of producing x units is shown by the lowest point on any member of the family of parabolas for that value of x. The locus of these points is not the line joining the lowest points of the parabolas, but the envelope of the family[8] (see Fig. 2).

[8.] Cf. Professor J. Viner, *Zeitschrift fur Nationaldkonomie*, Bd. III, September 1931, p. 36. The draughtsman, to his argument with whom Professor Viner refers in a footnote, was right in economics as well as in mathematics. Cf. also R. F. Harrod, *Economic Journal*, 1931, December, p. 575.

Harrod's Figure 2 in this article is very similar to the Figure 2 in his 1931 article.

[6] The 1931 visit is documented in Besomi (Undated). Viner also visited Oxford in 1927 and may have met Harrod then. See Viner (1927).

[7] Here is the footnote and associated text from Viner (1958, p. 227):

The partial equilibrium nature of the Marshallian assumptions leaves a wider range of possibilities to the long-run tendencies of costs for an expanding industry than is consistent with general-equilibrium analysis. I first saw this in 1938, and thereafter pointed in out to my students at the University of Chicago. But the first, and to my knowledge, still the only analysis in print similar to what I have in mind[31] is in Joan Robinson's excellent article, "Rising Supply Price," Economica, VIII, February, 1941.

[31] I have since found the same doctrine expounded in an earlier article by R. F. Harrod, "Notes on Supply," The Economic Journal, Vol XL (1930), pp. 232-241, especially pp. 240-241. [Note added in 1951.]

[8] Schneider (1931) is discussed, and some of its figures are reproduced, in Schmidt (2004). My comment are based on this material.

[9] A more complete description from Schmidt (2004, p. 127):

Their theorizing efforts may have benefited from real-world experience: Auspitz was a sugar producer and a longtime member of the Austrian parliament; his cousin Lieben was a banker who held offices in several commercial and educational institutions; both were partners in the banking company founded by their fathers.

[10] Schmidt (2004, p. 127) reports that "In fact, the academic establishment in their native Austria and in Germany shunned them and their book."

[11] One important exception was F. Y. Edgeworth. According to Schmidt, Edgeworth reviewed the Auspitz-Lieben book, *Untersuchungen*, in 1889. In this review Edgeworth "noted the

presence of the envelope cost curve in *Untersuchungen* and even highlighted it as a particularly interesting feature of the work of Auspitz and Lieben" (2004, p. 116). Edgeworth also cited this book in another publication, but without explanation (Schmidt, 2004, p. 126n). Another exception was Irving Fischer who, in a preface to one of his books, "had declared that Auspitz and Lieben, along with Jevons, had 'influenced [him] the most'" (Schmidt, p. 126n).

[12] As reproduced by Schmidt (2004), the figure drawn by Auspitz and Lieben (1889) is similar to the right half of Figures 3.1 and 3.3. The main difference is that their (unlabeled) envelope follows the lower points of the drawn short-run cost curves instead of using a mathematical formula (or a smoothed drawing) to indicate the envelope at other plant sizes.

[13] As far as I know, Samuelson (1947), the Viner student, was the first to point out the relationship between the long-run cost curve diagram and the Envelope Theorem. His initial discussion of this topic (pp. 34-5) mentions only Wong and Viner. It cites the Viner mistake, the Wong correction, and Viner's glimmer of understanding about the intuition of the envelope theorem in his work on international trade. "As Professor Viner has pointed out with great insight, at the margin … all factors are perfectly indifferent substitutes," Samuelson says (1947, p. 35). At another point, however, Samuelson (1947, p. 243) specifically acknowledges Harrod's contribution by referring to the "Wong-Viner-Harrod envelope theorem," although he does not cite any of Harrod's articles on the subject. Some scholars may not be aware of this acknowledgement, because Samuelson's own usage has not been consistent. His book (1947, p. 66) also refers to "Mr. Wong's famous envelope theorem." Moreover, his reflections on the origins of his 1947 book refer to what he "waggishly called in *Foundations* the Wong-Viner *Envelope* Theorem" (1998, p. 1377). This name does not actually appear in his book. In addition, Samuelson gave Harrod a back-handed compliment on this issue in his survey of contributions

by economists in the Harrod-Viner cohort (1988). In writing about Harrod, Samuelson says "His microeconomic contributions were first class: the correct form of Viner's envelope is given in Harrod's 1934 paper on imperfect competition" (p. 324). Harrod's 1931 paper is not mentioned. What are we to make of this inconsistency? My own guess is that Samuelson recognized Harrod's contribution to the drawing of a long-run cost curve, but also believed that it was Viner's intuition about marginal changes from an optimum that really opened the door to the Envelope Theorem. Samuelson apparently did not know about Auspitz and Lieben in 1947.

[14] A more thorough discussion and proof of the Envelope Theorem can be found in most graduate microeconomics textbooks, including Silberberg (1978).

[15] For a more extensive discussion, see Silberberg (1999). By the way, Silberberg reports (in his note 1, p. 78) that "Samuelson's proof of the envelope theorem was so opaque to me when I first encountered it in graduate school in a reading class with Jim Quirk that I promptly dropped the course!" I guess he figured it out eventually!

[16] A version of this problem with a CES production function is presented in Section 3.2.3.

[17] As usual, Samuelson (1972, p. 9) says it better:

> Yet I would argue that it is the occasional errors of geniuses like Viner which make the reputations of mere mortals, and which also seminally advance the body of science. Who in economics would remember Dr. Wong if his memory had not been perpetuated by his correcting of Viner's long-run cost-curve envelope? Precisely because Viner was so Jovianly impervious to error, the economics profession got a modicum of *Schadenfreude* at his expense over the envelope incident.

[18] This figure is equivalent to Viner's (1931) Chart III.

[19] The cost curve derivations in this section and the next can be found in many microeconomics textbooks. Production functions and cost curves are relevant for both the private and the public sectors. The relationship between public education production functions and cost curves is derived in Duncombe and Yinger (2011).

[20] An important topic that goes beyond the scope of this chapter is the determinants of returns to scale or, equivalently, of economies and diseconomies of scale. For a recent contribution to this literature, see Carlaw (2004). Note that the issue here is of "internal," that is, within firm, economies of scale, not "external" economies of scale, which involve the size of an industry. The causes of scale economies and diseconomies in public production are discussed in Duncombe and Yinger (1993, 2007).

[21] Despite the extensive knowledge about short- and long-run cost functions, some scholars make incorrect statements about the relationship between these two concepts. See Yinger (Forthcoming).