Scissors or Horizon: Neoclassical Debates about Returns to Scale, Costs, and Long-Run Supply, 1926–1942

Nahid Aslanbeigui* and Michele I. Naples†

Modern treatment of long-run (U-shaped) cost curves developed from reactions to Sraffa’s criticisms of Marshall. He argued that internal (dis)economies were incompatible with partial-equilibrium analysis under perfect competition. Pigou concurred and drew L-shaped cost curves; Viner realized that this made firm size indeterminate and industry output volatile. Using Austin and Joan Robinson’s analyses, Stigler justified rising costs/supply, determinacy, and stability by irrational entrepreneurs enduring coordination failure and by factor price changes. We conclude that consistency requires constant costs but firm employment, output, and factor incomes remain theoretically indeterminate. It becomes likely that large firms will undermine perfect competition.

1. Introduction

The short-run U-shaped cost curves of economics textbooks have been the subject of a growing literature, suggesting that firm costs are constant over a wide range before they rise (Larson 1991; Maxwell 1965; Yordon 1992). Using Machlup’s (1952) distinction and building on Beattie’s framework (1988), Larson (1991) has shown that, if factors are indivisible in acquisition but divisible in use, firms will choose fixed input proportions at output levels below full-capacity utilization. This implies constant and equal marginal and average factor productivity as well as constant and equal marginal and average variable costs. With the majority of the real-world cases falling in this category, Larson has suggested that short-run constant costs require “reworking of many aspects of traditional economic theory” (p. 473).

In this paper, we focus on a related aspect of textbook analysis, the prevailing view that long-run cost curves are U-shaped as well. We document the challenge posed by Sraffa’s (1926) seminal article, which showed the incompatibility of downward- or upward-sloping cost curves with partial-equilibrium analysis under perfect competition. We discuss and evaluate the ensuing debate, which lasted for almost two decades and led to the emergence of long-run U-shaped costs/rising supply curves as dominant. We show that the interwar debate was resolved at the expense of consistency, since partial-equilibrium analysis is logically coherent only for the

---

* Department of Economics and Finance, Monmouth University, West Long Branch, NJ 07764, USA; E-mail naslanbe@mondec.monmouth.edu.

† Economics, The College of New Jersey, CN4700, Ewing, NJ 08650-4700, USA; E-mail naples@tcnj.edu.

The authors would like to thank Jim Devine, Frank Thompson, and the anonymous referee for insightful comments and Young Back Choi and Cristina Marcuzzo for helpful conversations. The usual disclaimer applies.

Received August 1996; accepted April 1997.

1 On short-run cost curves, see also Johnston (1958) and Dean (1976).

2 The empirical evidence for L-shaped/U-shaped cost curves in the long run is beyond the purview of this paper. For studies on this issue, see Lee (1984), Dean (1976), and Johnston (1958, 1960).
constant-cost case. We suggest that consistency requires a reworking of traditional price theory under constant costs, as well as the theories of the functional distribution of income, factor employment, and firm output in the long run.

The paper begins with background sections on Marshall’s (1920) and Pigou’s (1912) analyses of the laws of returns and Sraffa’s challenge to long-run models of perfect competition. The next two sections show that Sraffa’s essay triggered two different reactions. Some (Pigou 1927, 1928) argued that long-run firm cost and industry supply curves were L-shaped: Costs declined until they reached a horizontal range. This suggested a return to the classical view that price was determined on the supply side of the economy. It also implied indeterminate firm size. Others (Viner 1931; E. A. G. Robinson 1931; J. V. Robinson 1941) justified U-shaped firm costs and/or upward-sloping industry supply curves. Stigler’s (1942) textbook synthesis secured the need for “both blades of the scissors,” demand as well as supply, to determine long-run price and firm size. Sraffa’s arguments notwithstanding. We then evaluate the remaining inconsistencies in Stigler’s synthesis. We conclude that only constant costs are theoretically consistent with perfect competition. However, constant costs make it impossible to determine firm output, employment, and factor incomes a priori and make perfect competition unlikely.

2. Costs and Industry Supply Before Sraffa

Marshall

According to Marshall (1920), increased production in agriculture yielded diminishing returns since the supply of land was limited; this led to an upward-sloping supply curve for the industry. In manufacturing, however, increased scale of production “improved organization, which increase[d] the efficiency of the work of labour and capital” (p. 265), implying increasing returns. Economies of scale were internal if they depended on “the resources of the individual houses of business engaged in it and the efficiency of their management” (pp. 262–263). These were due to such factors as division of labor and improved skill–task fit, improved recognition for the product of one firm, and diminishing marketing costs. Economies were external if they depended on “the general development of the industry” (p. 262). Examples of such economies included “the growth of correlated branches of industry which mutually assist one another, perhaps being concentrated in the same localities, but anyhow availing themselves of the modern facilities for communication offered by steam transport, by the telegraph and by the printing-press” (p. 264). Regardless of their origins, economies of scale resulted in downward sloping supply curves (p. 664).

Most of Marshall’s discussion was couched in terms of cost and supply at the industry level. The firms within any industry were dynamic in nature: Like the trees of a forest, they were in different stages of their life-cycle. Any discussion about individual production units was therefore in terms of the representative firm, which linked the static notion of industry to the dynamic and historical firm (Moss 1984, p. 309). The representative firm was a production

---

3 In some quarters, Sraffa’s challenge was met by the development of models of imperfect/monopolistic competition (Chamberlin 1933; Robinson 1933). This has been well documented elsewhere and is beyond the purview of our paper (see Blaug 1978; Marcuzzo 1994; Sardoni 1994; Shackle 1983).

4 Most Cambridge economists who were trained by Marshall or Pigou used the term diminishing returns to denote both short-run and long-run phenomena.
unit that “had a fairly long life, and fair success, which [was] managed with normal ability, and which [had] normal access to the economies, external and internal” (Marshall 1920, p. 265).

Marshall’s analysis has been touted for its richness (see, e.g., Abouchar 1990). Perhaps because of it, his treatment of the relationships among cost, output, and price lacked precision (Viner 1931, p. 199). Diminishing returns described both short-run and long-run phenomena. In addition, empirical, historical, and analytical costs were treated at the same level, and shifts in and movements along cost curves were conflated.

**Pigou**

As Marshall’s successor, Pigou retained the analysis of laws of returns and external and internal economies, but added a few ideas showing how to theoretically incorporate external factors into firm supply decisions. Firm and industry costs were interdependent, though the ... supply ... schedule of the market can be represented by a plane curve, the ... supply ... schedules of the separate sources that make up the market cannot be so represented, and cannot be simply added together to constitute the aggregate ... supply ... schedule. (Pigou 1913, p. 21)

The interdependence was graphically analyzed through two curves, the supply price and the marginal supply price. Pigou defined the supply price of a given quantity of output as the “price which tends to call out the production of that quantity annually”; at this price, total revenue and total costs were equal (Pigou 1912, p. 174). Marginal supply price was defined as “the difference between the aggregate expenses of the annual production of x units and of (x + Δx) units respectively” (p. 176). The two curves coincided only when constant returns were prevalent. Under increasing returns, supply price was higher than marginal supply price; an increase in the output of the individual firms could lead to external economies of the type specified by Marshall (1920) that reduced the cost of production for the industry as a whole. Under diminishing returns, supply price was less than marginal supply price, implying that the increase in the output of the individual firm would lead to increased costs for the industry as a whole (Pigou 1912, p. 176).

What did the above signify for the industry as a whole? The supply curve for the industry followed the laws of returns: Increased scale of production under diminishing returns increased the costs of production through external diseconomies, leading to an upward-sloping supply curve; the inverse applied to increasing returns.

In discussing the supply curve for the firm, Pigou did not mention Marshall’s representative firm. But he did realize that the individual firm could not have internal economies at equilibrium, otherwise it would drive the competition out of the market and lead to monopoly. Stability required that external economies be “common to all the suppliers jointly” and that the firm

---

5 A plane curve lies on a single plane. If the supply price of firm output were independent of other firms’ prices, its supply curve could be drawn on a flat plane. But since a firm’s supply price depends on other firms’, the firm’s supply curve can only be accurately represented in multiple dimensions and would be a curvilinear surface rather than a simple two-dimensional curve. Industry supply differs because, by assumption, in partial equilibrium, the prices of other industries’ products are parameters.

6 Pigou (1912, pp. 173–179) argued that under both decreasing- and increasing-cost conditions, marginal private and social net products diverged, implying a prima facie case for government action. Subsidies or taxes in these circumstances could help equate social and private interests, allocate resources more efficiently, and increase the size of the national dividend. Young (1913) attacked Pigou’s welfare proposals for increasing-cost industries and Pigou conceded the point in 1924 (Pigou, 1924, pp. 194–196).
itself be at a point beyond which it would experience diminishing returns (Pigou 1912, p. 177). This analysis did not, however, specify the shape of the supply curve for the individual firm.

In the 1920s, the Marshallian treatment of the laws of returns was criticized heavily on various grounds. Clapham (1922) declared that the laws of returns were empty economic boxes, representing a logical/theoretical apparatus devoid of empirical content, precision, and clarity. Others repeated Young’s (1913) concern with the accuracy of Pigou’s policy conclusions with respect to decreasing and increasing cost industries (Knight 1924; Robertson 1924). However, everyone seemed to agree, explicitly or implicitly, that cost curves were U-shaped and supply curves upward sloping.

3. Sraffa’s Criticisms

Sraffa’s (1926) seminal article showed that downward-sloping or upward-sloping industry supply curves violated the conditions necessary for a competitive industry studied in a partial-equilibrium context. Perfect competition required that the demand for and the supply of a commodity be independent from each other and from the demand for and supply of other commodities. When conditions of decreasing or increasing costs were present, a change in the output of one industry introduced precisely such interdependence.

First, the Marshallian treatment assumed that increasing costs were due to the existence of a factor in limited supply (e.g., land). Increased industry output implied an increase in the price of this factor; this then implied a rise in the cost of production in other industries that used this factor. If their products were potential substitutes, the change in relative prices would affect the demand for the product of the original industry (Sraffa 1926, p. 539).

Second, increasing returns faced similar problems. Internal economies were incompatible with perfect competition because they led to monopoly. A static framework required assuming away economies due to general economic progress. As a result, one had to focus on economies that were external to the firm but internal to the industry (hereafter referred to as external economies). Such economies were rare in reality, argued Sraffa, and thus so were “supply curves showing decreasing costs” (Sraffa 1926, p. 540). The “cost of production of commodities produced competitively . . . must be regarded as constant in respect of small variations in the quantity produced” (p. 541).

Sraffa’s article sparked a debate about the shape of cost and supply curves. Although by the early 1930s, perfect competition was a familiar term and its prerequisites were well specified, no such consensus was yet reached over the shape of the cost and supply curves. The following two sections will discuss these debates about L-shaped and U-shaped cost curves and their implication for the slope of the industry supply curve.

4. Long-Run Cost and Supply Curves

L-Shaped Cost Curves

Pigou’s Reaction. Six months after the publication of Sraffa’s article, Pigou accepted his argument that diminishing returns required general-equilibrium analysis and were incompatible

---

7 For a comprehensive survey of this controversy, see Aslanbeigui (1996).
with perfect competition under partial equilibrium. In two articles, Pigou (1927, 1928) spelled out the conditions for the existence of equilibrium in that framework. Events exogenous to the operation of an industry, for example, general technological changes, inventions, or changes in tastes, were assumed away. In addition, the scope of the analysis was necessarily and logically confined to "commodities which individually employ so small a proportion of each of the several factors of production that no practicable changes in the scale of their output could sensibly affect the relative values of these factors" (Pigou 1927, p. 192). With no limitation in the supply of factors of production, including managerial ability, Pigou concluded that

...it is impossible for production anywhere to take place under conditions of increasing costs. In this matter my conclusion agrees with that reached by Professor Sraffa in his recent article. (p. 193; emphasis added)

However, Pigou did not rule out the existence of external economies. A change in the industry's scale of production could lead to the use of more productive techniques and "increased specialization among the makers of the machines used in the industry" (Pigou 1927, p. 195). Therefore, the industry supply curve could slope downward and/or be horizontal. Where average cost for the industry declined (i.e., marginal cost was below the average cost), the supply curve would be the same as the average-cost curve since no industry could survive at prices that did not cover average costs. Where marginal costs were constant, the average- and marginal-cost curves would be horizontal and coincide. In other words, the general form for the industry supply curve was a rectangular hyperbola (p. 197).

In 1928, Pigou turned his focus to the firm's cost curves. He concurred with Marshall that the firms in an industry were at different stages of their life-cycle. He, too, accepted the need to study the cost and supply conditions for a firm through the medium of a typical unit, which he chose to call the equilibrium firm, implying

...that there can exist some one firm, which, whenever the industry as a whole is in equilibrium, in the sense that it is producing a regular output ... in response to a normal supply price ... will itself also individually be in equilibrium ... The conditions of the industry are compatible with the existence of such a firm; and the implications about these conditions, which, whether it in fact exists or not, would hold good if it did exist, must be valid. (Pigou 1928, pp. 239–240)

Unlike Marshall's representative firm, however, Pigou's equilibrium firm did not enjoy internal economies; it reached an optimum size, "trespass beyond which yields no further internal economies" (Pigou 1927, p. 195). Its costs therefore declined for the initial ranges of output and became constant later. Pigou was one of the first economists to publish U-shaped cost curves for the firm; however, it was implicit in his diagrammatic analysis and discussion that the firm would not move beyond the lowest point of its average cost curve. Stable equilibrium for this

---

8 Pigou's acquiescence had been in the making for many years. In 1913, he had recognized the importance of interdependence between individual and market supply (demand) curves (Pigou 1913, p. 21). In 1920, he had made explicit that, under partial-equilibrium conditions, each industry must be small enough so that an increase in its output did not appreciably affect the price of factors of production (Pigou 1920, p. 935). In 1924, in a response to Allyn Young, he had agreed that the increase in long-run costs due to diminishing returns could not be caused by technical diseconomies. Such diseconomies as rising factor prices were only pecuniary, involving income redistribution but not increased input requirements (Pigou 1924, p. 194). Once Sraffa's article pulled all the above arguments together, Pigou agreed that increasing costs and, by implication, upward-sloping supply curves were incompatible with the Marshallian partial-equilibrium, perfect-competition model.

9 Sraffa had objected that such economies were not very common empirically. For Pigou, an analytical study merely exposed the possibility of such economies (Pigou 1927, p. 196).

10 Pigou (1912, p. 174) defined total cost to include interest or the opportunity cost of capital.
firm dictated that the industry price be equal to its marginal and average costs (Pigou 1928, p. 243).\footnote{For other types of equilibria (neutral and unstable), see Pigou (1928, pp. 245–249).}

In the absence of external economies, the equilibrium firm’s output and size would be fixed, the firm always operating at the lowest point of its U-shaped average cost curve. Industry’s output would have to be met by the entry of new firms or by the expansion of output of nonequilibrium firms. For firms with L-shaped cost curves, output was indeterminate. Because Pigou’s analytical framework, following Marshall, allowed for heterogeneous, ever-changing firms, he was not concerned with this result. Instead, he focused on the concomitant stability of industry price.

For Pigou, in the presence of external economies,\footnote{Pigou also analyzed the logic of external diseconomies, but as observed above, he had already rejected its empirical relevance.} the industry supply curve was downward sloping. As the industry expanded, the equilibrium firm’s cost curves shifted, the firm would produce at the lowest point of its new average-cost curve, although its size might change (Pigou 1928, p. 250). “[A]s the output of the industry as a whole varies, both the output of the equilibrium firm which will make these two things [marginal and average costs] equal and also their magnitude when they are equal may vary indefinitely in either direction” (p. 250), given the benefits of specialization within the expanding industry. That is, external economies introduced theoretical indeterminacy into the analysis of the supply curve and equilibrium price.

\textit{Viner’s Model.} In his celebrated article, Viner (1931) argued that, in the short run, firm costs were U-shaped (average) or upward sloping (marginal). In the long run, if there were some “absolutely limited factor” (i.e., land) (Viner 1931, p. 207), costs would also be U-shaped. Absent such a factor, the envelope of short-run U-shaped curves would be a horizontal long-run average cost-curve (p. 211) for atomistic (competitive) firms.

Although Viner addressed net internal economies and diseconomies, he argued that, in the long run, neither could exist at the firm level. The first was inconsistent with perfect competition since it led to monopoly (Viner 1931, p. 215); the second was hard to imagine since the firm could avoid cost increases by multiplying the number of plants, a conclusion similar to Pigou’s (p. 217). By implication, this left the constant-cost case as most common in industries that do not heavily use land or other absolutely fixed factors. Here Viner implicitly recognized the fallacy of composition in going from upward-sloping marginal-cost curve for a plant to upward-sloping marginal-cost curve for the industry, as today’s textbooks commonly do.

Like Pigou, Viner had realized that constant returns introduced indeterminacy for the firm. Consequently, Viner found the coexistence of constant returns and perfect competition uneasy, to say the least. If each individual firm experienced constant returns, marginal and average costs would be constant and the firm’s supply curve horizontal (Viner 1931, p. 211). With heterogeneous costs across firms, the firm with the lowest costs would monopolize the industry. With identical costs, each producer could produce any amount at the equilibrium price (the minimum of the long-run average-cost curve), nothing below it, and unlimited quantities above it. In this case, the industry supply curve would not be definable; “it is impossible to indicate graphically the relationship between the long-run supply curves of individual concerns and the industry as a whole” (p. 211).

\[\text{In the long-run there would be a constant tendency toward overproduction, with consequent losses and a} \]
reaction toward underproduction. Actual long-run price and output would be unstable, but would oscillate above and below stable points of equilibrium price and equilibrium output. (p. 211)

Viner’s discussion is vague. He appears to say that the industry supply curve cannot be defined, in light of firm behavior. However, the quote above addresses the stability of market equilibrium, not the shape of the supply curve; industry supply would be horizontal and therefore infinitely elastic, given uniform and horizontal long-run average cost curves for all firms. The problem that produces the instability is that perfectly elastic industry and firm supply cause an indeterminate division of industry output among firms. Therefore, firm output will be subject to swings that economists cannot accurately predict and market forces will not attenuate. The precision of the price–quantity solution implied by a horizontal industry supply curve and downward-sloping demand curve belies the market volatility and indeterminacy at the firm level, of which Viner was acutely conscious.13

U-Shaped Costs and Upward-Sloping Industry Supply

Viner’s Pecuniary Diseconomies. Viner sought to remedy this indeterminacy by incorporating externalities, appealing to external economies and diseconomies of production as the only possible sources of increasing or decreasing costs for the perfectly competitive industry. External economies and diseconomies were mainly pecuniary14—due to changes in “prices of services and materials” (Viner 1931, p. 218), although initially Viner had identified “technological conditions” as determining the shape of the long-run supply curve (p. 206). In fact, he found pecuniary external diseconomies of scale to be of “indisputable practical importance” (p. 220).15

For Viner, external economies and diseconomies of scale shifted a plant’s (short-run) average- and marginal-cost curves in a manner that kept its output constant.16 The increase in the output of the industry would have to take place through the entry of new plants (Viner 1931, p. 217) or producers (p. 221). Viner did not explicitly identify firm supply. Industry supply, however, could be upward- or downward-sloping, coinciding with the industry’s average-cost curve. Unlike the modern portrayal of Viner’s cost curves, the firm would not have a U-shaped long-run average-cost curve, but the industry could.

Viner presumed that the incorporation of external diseconomies had remedied the indeterminacy of firm supply. This, however, is not necessarily the case. While plants had U-shaped cost curves, firms with reproducible plants had constant internal returns, suggesting indeterminate size. As Viner himself observed, “the individual producers will not concern themselves

13 Viner’s concern with the problem of instability seems to be derived from his static notion of the industry and the associated assumption of identical firms.
14 Viner had difficulty locating examples for technological external economies and even more so for diseconomies (Viner 1931, pp. 217–218). One possible example of technological external diseconomies “might be higher unit highway transportation costs when an industry which provides its own transportation for materials and products expands its output and thereby brings about traffic congestion on the roads” (p. 221).
15 Viner (1931, p. 220) suggested that industry expansion caused “increased purchases of primary factors and materials which . . . must tend to raise their unit prices.” This reference to nonmanufactured inputs implicitly suggested diminishing returns to land as the source of diseconomies. However, he provided no concrete examples or supporting evidence.
16 Viner did not explain why the output of the plant or the individual concern (he used the two terms interchangeably (Viner 1931, pp. 218–219)) would remain constant. If the price of an input changed relatively, factor ratios in each plant would be adjusted. This would modify the firm’s blueprints and could affect optimal plant output. For Pigou, under the same circumstances, the firm’s size could easily change.
with the effect on the costs of other producers of their own withdrawal from or entrance into the industry” (Viner 1931, p. 222). Therefore, firms would decide whether to expand or contract based on internal costs and returns, not full information about general-equilibrium input-price changes. Firms would face a prisoner’s dilemma: A good output choice for one firm would turn out to be unprofitable if all firms simultaneously made the same output choice. Volatility would result. Not only would expansion mean excess supply for the industry, but input prices would also have risen due to external diseconomies. Both phenomena provide incentives to contract output. Note that, since firms in the aggregate would probably overcontract and the industry experience excess demand, this would raise output prices, while input price was falling. Volatility would be a common state of affairs, despite external diseconomies and upward-sloping industry supply.

*Austin Robinson’s Coordination Failure.* In the same year that Viner published his article, Austin Robinson had discovered a way to reduce the theoretical possibility of unstable equilibrium through internal diseconomies for the firm (see Robinson 1931). Robinson discussed the supply curve for the optimum (representative) firm, the optimum size being determined by the interplay of five different forces: technical, managerial, financial, marketing, and risk and fluctuation (p. 17).

The optimum firm could experience technical economies due to increased division of labor and dexterity as well as to purchases of larger and more efficient machines. However, Robinson agreed with Pigou and Viner that, beyond some point, the firm would exhaust technical internal economies and would never experience technical internal diseconomies. “If other considerations require a scale larger than the technical optimum, the technical scale of production can be increased by mere multiplication” (Robinson 1931, p. 35).

Robinson believed, however, that other diseconomies could emerge. Economies of scale in administration would be exhausted at some level of output (Robinson 1931, p. 43). This coordination failure would increase costs for the firm; there is such a thing as too big a firm (p. 44).

How big a firm can successfully grow will depend upon how it solves this problem of the co-ordinating of separated departments and separated specialists. (p. 45)

In this the managerial optimum differs from the technical optimum. The latter affords only a minimum scale below which the greater efficiency cannot be achieved. Additional output may be produced under conditions of approximately constant cost. But if the managerial optimum is exceeded, costs, through declining efficiency and the need for additional co-ordination, begin to rise. The managerial optimum sets, therefore, not only a lower but also an upper limit to the scale of operations. (p. 48)

Financial factors could equally lead to economies and diseconomies of scale. Large firms could borrow at more advantageous interest rates from banks and by issuing stock to the general public. These economies would likely, though not necessarily, come to a halt at some point, however, since the leaders of larger firms would have to face the “cramping influence upon their enterprise” from the shareholders and others; “the autocracy which is the foundation of their efficiency” would be “confined and limited” (Robinson 1931, pp. 59–60).

Marketing factors related to scale economies included the cost of buying (raw material, etc.) and the cost of selling the product. Larger firms could buy at discounted prices, hire expert

---

17 The optimum firm, according to Robinson (1931, pp. 14–15), represented “the scale of production which, having regard to the circumstances of the industry, was looked upon as the best scale of production sometime in the recent past”; it was the firm with the lowest long-run average cost.
buyers (Robinson 1931, p. 65), and/or maintain a separate marketing department (p. 67). Without providing any evidence or discussion, Robinson maintained that beyond some level these expenditures would increase "the more rapidly, the more rapidly the firm attempts to grow" (p. 79).

All firms would be subject to four types of fluctuations in the demand for their products: permanent changes due to changes in tastes, cyclical variations due to general reductions in income, seasonal variations, and random changes "due to the failure of individual orders to combine in such a way that they form a single continuous stream" (Robinson 1931, p. 83). Robinson (p. 102) concluded that smaller firms were generally more adept at adapting to changes in demand than larger firms.

Robinson's scale diseconomies resolved the problem implicit in constant costs that Viner had addressed. By identifying sources of internal diminishing returns even in the long run, he secured an upward-sloping firm cost curve and therefore determinate output choice. However, except for coordination failure, the other factors Robinson identified would not be applicable to perfectly competitive firms: They need not market their homogeneous outputs, cannot seek competitive advantage in perfect capital markets, and cannot grow so large as to be unresponsive to market changes. Moreover, the argument that small firms are better able to survive demand shocks relates to disequilibrium adjustment as distinct from the analysis of the long run. Finally, Robinson himself identified ways in which the entrepreneur could expand firm size while avoiding diseconomies. These included limiting the number of products provided by the firm, vertical disintegration, or dividing the firm into smaller, more manageable units (Robinson 1931, pp. 110–111).

Joan Robinson's Factor Proportions. In 1941, Joan Robinson adopted an entirely different approach to justify increasing costs. In a fully employed, perfectly competitive economy, an increase in the demand for a commodity could only be met if resources were released from other sectors. If factor proportions were identical across all industries, relative factor prices would stay constant. In reality, argued Robinson, factor proportions were different. The factors used more intensely for the production of the commodity at hand would be demanded more and therefore their price would go up. The factors used less intensely compared to the other sector would experience a decrease in their relative prices. The price of the commodity for which the demand has increased would go up; the price of other commodities would drop (Robinson 1941, p. 237). The supply price, according to Joan Robinson, was almost always increasing.

However, Robinson also explored the possibility that elastic supply of the inputs intensely used by the industry and relatively inelastic supply of its other inputs would mean that industry expansion would lead to net pecuniary economies (decline in unit costs). She then applied this possibility to Marshall's intuition that agriculture and mining would face pecuniary diseconomies because their required inputs of particular kinds of land were highly inelastic and had few close substitutes (idiosyncratic factor proportions). That manufacturing faced virtually constant costs, she argued, reflected the more elastic supply of reproducible goods and the fact that

18 Robinson's earlier (1933) analysis of costs combined Viner's and Austin Robinson's. Cost curves were U-shaped for both the firm and the industry. Internal economies were exhausted at the firm level and costs would eventually go up due to increased prices of scarce factors (e.g., agricultural goods) as well as coordination failure (Robinson 1933, pp. 24–25, 48–49).

19 Joan Robinson discussed industry costs, not whether firm cost curves shifted. Viner later agreed with Joan Robinson's factor-proportions analysis (Viner 1950, p. 228).
manufacturing processes were typically not cranky or idiosyncratic in finding substitutes for inputs in short supply.

Joan Robinson’s analysis of the elasticity of substitution as well as supply can be interpreted as supporting Sraffa’s views on constant costs. Sraffa had said that if supply elasticity for the input were infinite for all factors, which is true by assumption in partial equilibrium, constant returns would be implied. Yet Viner and Stigler focused on the first part of her paper and rising supply price at full employment, given imperfect supply elasticities.

Stigler’s Textbook Treatment

By the early 1940s, U-shaped cost and upward-sloping supply curves were part of the orthodox economist’s tool box. Stigler’s (1942) textbook, The Theory of Competitive Price, which came to be the basis of microeconomics education for the next several decades (Yordon 1992), synthesized the theories of perfect competition emerging from this literature.20

Stigler introduced the firm first and foremost. He separated the short-run U-shaped cost curves from cost curves in the long run, the latter being an envelope of the many possible short-run cost curves. Unlike the short run, the long run was defined as a period in which the size of the plant could vary.

Like many before him, Stigler recognized that, under perfect competition, internal economies would be exhausted21 and the firm would reach a point of constant returns to scale. However, beyond some size, the firm would face internal diseconomies of scale:

... the growth of a firm puts heavier and heavier burdens on the management. Quite aside from the difficult problem of expansion itself, large groups are much harder to coordinate than smaller units. For management, and control in general, inherently face a problem: the final authority to make decisions cannot be subdivided or delegated. Large units are, in fact, confronted by a dilemma. At one extreme all authority may be delegated. Then there will be no unity of policy or uniformity of performance. At the other extreme, all decisions may be made by a final center. This system involves bureaucracy in its worst form: “red tape,” hopeless delay, decisions based on diluted memoranda. Between these two extremes the large firm attempts to steer a middle course, but it never achieves that compactness, flexibility, and singleness of purpose which are possessed by every well-managed medium-sized firm. The growing difficulty of coordination and decision-making eventually stops the growth of every firm. (Stigler 1942, p. 138)

As is evident from the above quotation, Stigler replicated Austin Robinson’s arguments—without attribution—with a minor twist. In his hands, coordination failure translated into a fixed factor of production, the final authority or entrepreneur. Stigler did not address the possibilities that Robinson had identified whereby competitive firms could avoid these sources of rising costs (e.g., decentralization). Stigler assumed that “diseconomies of large-scale production set in soon enough to insure numerous firms and therefore competition” (Stigler 1942, p. 160), implying that the long-run envelope of short-run average costs is also U-shaped, albeit flatter.

The firm would operate at the minimum point on this curve or be competed out of business. Stigler concluded that all firms must have identical minimum costs in the long run, although their sizes and technologies might differ; firms with some advantage due to a superior resource

---

20 Samuelson’s text, Economics: An Introductory Analysis, first published in 1948, suggested that perfect competition was rare, existing only in a few lines of agriculture (Samuelson 1948, pp. 491, 509). Since he focused on monopolistic or imperfect competition, we do not include his text in our analysis.

21 Increasing returns were possible due to indivisible machinery. Other forms of indivisibility included marketing, financial, research, and management indivisibilities (Stigler 1942, pp. 136–138); Stigler did not mention that three of these four would have no place under perfect competition. If increasing returns were ongoing, the long-run average cost curve for the firm would be downward sloping, and monopoly or oligopoly would replace competition.
would be forced by competition to pay that resource a rent that would bring unit costs in line with the industry’s. However, upward-sloping firm supply need not mean upward-sloping industry supply. Stigler realized what Viner had realized, that rising costs due to coordination failure need not mean rising industry costs. Industry expansion could “take place primarily through an increase in the number of firms” (Stigler 1942, p. 162).

Hence, external pecuniary factors proved key to generating a rising-cost industry. Stigler (1942, p. 164) first argued that inputs were heterogeneous (although under perfect competition output is homogeneous [p. 161]), and therefore the expanding industry attracts productive services less suitable for the product. He then reproduced J. Robinson’s full-employment argument that factor prices would tend to rise as any industry expands (p. 164). Stigler’s synthesis of the debates on perfect competition provided a basis for partial-equilibrium analysis that produced determinate results for the firm and the industry. It rested on two crucial assumptions: coordination failure leading to internal diseconomies and external diseconomies reflecting heterogeneous inputs and rising input prices. Like any synthesis, Stigler’s leaves out certain insights from the historical literature. In this case, these challenged the viability of his assumptions.

5. Stigler’s Synthesis: An Assessment

The Irrational Entrepreneur

Internal diseconomies for the firm are problematic. Once a firm has achieved minimum-cost conditions, it is illogical to assume it does not have the know-how to duplicate them. Austin Robinson’s explanation, which was later seized upon by Stigler and others, begs the question: Why should coordination fail? Given perfect foresight, the entrepreneur can find other ways to expand besides shooting him/herself in the foot, as Austin Robinson himself emphasized. If managing a larger firm is dysfunctional, the entrepreneur has the alternative of setting up smaller subsidiaries under a holding company or of decentralizing (consider the 1995 breakup of AT&T into three autonomous entities). When Stigler assumed the entrepreneur puts his/her desire for control ahead of the drive for profits, he created an inconsistency in a model premised on profit-maximization—determinate firm size rests on irrational behavior by the entrepreneur.

22 The heterogeneity of inputs is a corollary of the rising supply price. In perfect competition, an industry’s output is homogeneous. To say, as Stigler did, that expansion requires using inferior inputs means substituting lower quality inputs for the preferred input, which is only rational if the price of the preferred input has risen.

23 Stigler acknowledged that there might exist constant-cost industries where the industry was small relative to its supplying industries and changes in its size did not affect the price of inputs. In that case, “[the long-run supply curve of the industry . . . is of course a horizontal line” (Stigler 1942, p. 163). He allowed for the possibility that factor prices might fall as the competitive industry expanded, because it would buy inputs from a monopoly or from another decreasing-cost industry, but deemed this case rare. Stigler also addressed what he called “Sraffa’s Case,” where industry demand would not be well defined since changes in prices of productive services potentially affect output prices in industries whose products are complements or substitutes for this industry’s output. He claimed that, nevertheless, “it is always possible to draw an unambiguous supply curve for a competitive industry” (p. 165).

24 As Sargent Florence observed, “there is little to prove the universality and inevitability of any such law of increasing costs in the long run when manufacturers have time to get new equipment in order to meet enlarged orders or anticipated orders and have time to reorganize and delegate responsibilities” (quoted in Johnston 1960, pp. 23–24).
Partial General Equilibrium

After the 1920s, a consensus emerged that external economies and diseconomies could affect the costs of production of individual firms and industry. Stigler's synthesis of Viner's and Joan Robinson's views found that increased output of the industry increased factor prices and hence led to upward-sloping industry supply curves.

This argument, which is still being used in introductory and intermediate economics textbooks, fails to respond to Sraffa's criticism: It makes the demand for and the supply of industry output interdependent. Joan Robinson's argument rested on the existence of a fully employed economy where an increase in the demand for the output of one industry would have to be met with decreased supply of the products of another (and associated demand adjustments). This led to increased prices of factors of production more heavily used by the expanding industry.

To invoke the factor scarcity of macroeconomic full employment requires a general-equilibrium analysis: Factors available to the industry are scarce because they are in demand in all other sectors. What is the logical basis for taking some consequences of that general equilibrium into account but neglecting others? If long-run industry supply is upward sloping because of input-price increases, all prices—including prices of factors used in the production of substitutes and complements as well as output prices of goods using any of these inputs—will change. Since the demand for any product depends on the price of all others, any movement along that one industry supply curve is associated with widespread price changes that cause that industry's demand curve to shift and change slope in ways that are not obvious. Stigler suggested that demand would not be well defined. Without a definite demand curve, partial-equilibrium analysis becomes impossible if factor proportions and prices are the source of upward-sloping industry supply.

Discussion

We conclude with Sraffa that perfect competition under partial equilibrium is only logically consistent if the firm and industry supply curves are horizontal rather than upward sloping. However, this means that demand plays no role in price determination. For the long run, Marshall's scissors are replaced by the classical supply-side horizon for price determination.

This result makes the determination of factor incomes open-ended. Absent coordination failure and given reproducible plants, the counterpart of constant costs is a horizontal marginal product of labor equal to its average product in the long run. If competition forces the wage to equal that marginal product, there will be nothing left for capital and normal profits will be zero. However, since in that framework the marginal product of capital would also equal its average product, if competition forced the profit rate to the same level, there would be nothing left for wages. The functional distribution of income cannot be established a priori and cannot be determined by factors' marginal productivity.

Such a range of indeterminacy for wages in the short run has long been accepted in the industrial-relations literature as the basis for wage setting by union negotiations (Kaufman 1993). Our results suggest that the extension of this type of wage analysis to the long run for constant-cost firms may prove fruitful.

Moreover, horizontal cost curves make firm size indeterminate and therefore perfect competition precarious. This indeterminacy implies that, instead of rationally allocating resources among firms, the market cannot prevent ongoing interfirm conflict over market share and tendencies to under- and over-produce. By implication, factor employment will be equally unstable.
Those seeking to construct the microfoundations of macroeconomic fluctuations might do well to consider such a constant-cost model.

Finally, since factor incomes and employment and firm output are indeterminate in theory, economists will have to use the inductive method and infer generalizations from empirical research, a method at odds with the positivist tradition. Notably, such induction dominates case research in the business disciplines (accounting, finance, production management, marketing), which typically treat constant production costs as the referent. Further research on the theory of the firm under constant costs might fruitfully cross these disciplinary boundaries.

6. Conclusion

The cost controversy (1926–1942) led to the articulation of the relationship between industry supply and firm costs under perfect competition. It shifted the analytical focus from the industry to the firm and led to the development of U-shaped cost curves. The Marshallian attention to the dynamic nature of the firm and to historical costs was supplanted by a conceptualization of the long run as composed of static, identical firms. The mathematical and diagrammatic analyses of costs became much easier and the theory more precise and standardized. However, we have shown that the standard model is still logically inconsistent.

To be consistent, partial-equilibrium analysis under perfect competition requires constant costs. Ironically, if firm size is indeterminate but large size is known to facilitate market power, constant-cost firms will arguably tend to grow beyond the size consistent with perfect competition. Then the only cost curves consistent with the partial-equilibrium analysis of perfect competition are likely to lead to its opposite: monopolistic competition or oligopoly.

References


