The Division of Labor is Limited by the Extent of the Market

George J. Stigler

The Journal of Political Economy, Vol. 59, No. 3. (Jun., 1951), pp. 185-193.

Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28195106%2959%3A3%3C185%3ATDOLIL%3E2.0.CO%3B2-J

The Journal of Political Economy is currently published by The University of Chicago Press.

THE JOURNAL OF
POLITICAL ECONOMY

Volume LIX JUNE 1951 *Number 3*

THE DIVISION OF LABOR IS LIMITED BY THE EXTENT OF THE MARKET

GEORGE J. STIGLER
Columbia University

ECONOMISTS have long labored with the rate of operation of firm and industry, but they have generally treated as a (technological?) datum the problem of what the firm does—what governs its range of activities or functions. It is the central thesis of this paper that the theorem of Adam Smith which has been appropriated as a title is the core of a theory of the functions of firm and industry, and a good deal more besides. I shall (1) make some brief historical remarks on the theorem, (2) sketch a theory of the functions of a firm, (3) apply this theory to vertical integration, and (4) suggest broader applications of the theorem.

I. HISTORICAL INTRODUCTION

When Adam Smith advanced his famous theorem that the division of labor is limited by the extent of the market, he created at least a superficial dilemma. If this proposition is generally applicable, should there not be monopolies in most industries? So long as the further division of labor (by which we may understand the further specialization of labor and machines) offers lower costs for larger outputs, entrepreneurs will gain by combining or expanding and driving out rivals. And here was the dilemma: Either the division of labor is limited by the extent of the market, and, characteristically, industries are monopolized; or industries are characteristically competitive, and the theorem is false or of little significance. Neither alternative is inviting. There were and are plenty of important competitive industries; yet Smith's argument that Highlanders would be more efficient if each did not have to do his own baking and brewing also seems convincing and capable of wide generalization.

In the pleasant century that followed on the *Wealth of Nations*, this conflict was temporarily resolved in favor of Smith's theorem by the simple expedient of ignoring the conditions for stable competitive equilibrium. Ricardo, Senior, and J. S. Mill—and their less famous confreres—announced the principle of increasing returns in manufacturing—for

Senior it was even an axiom. The exclusion of agriculture was based on the empirical judgment, not that further division of labor was impossible, but that it was a weaker tendency than that of diminishing returns from more intensive cultivation of a relatively fixed supply of land.

This was hardly a satisfactory solution, and, when Marshall came to reformulate classical economics into a comprehensive and internally consistent system, the dilemma could no longer be ignored. He refused to give up either increasing returns or competition, and he created three theories (of course, not only for this purpose) which insured their compatibility. First, and perhaps most important, he developed the concept of external economies—economies outside the reach of the firm and dependent upon the size of the industry, the region, the economy, or even the whole economic world. Second, he emphasized the mortality of able entrepreneurs and the improbability that a single business would be managed superlatively for any length of time. Third, he argued that each firm might have a partial monopoly —a separate, elastic demand curve for its product—so that, with expansion of its output, the price would usually fall faster than average costs would.

For a time this reconciliation of competition and increasing returns served its purpose, but, as the center of price theory moved toward the firm, Smith's theorem fell into the background. External economies were a rather nebulous category relative to anything so concrete and definite as economists for a time believed the costs of a firm to be. It was pointed out by Professor Knight, moreover, that economies external to one industry may (and perhaps must) be internal to another. The industries in which the economies are internal will tend to monopoly; and, incidentally, it is no longer a foregone conclusion that such economies will be shared with the buyers. Since external economies seemed a refractory material for the popular analytical techniques, they were increasingly neglected.

Marshall's theory of business mortality was also increasingly neglected, with even less explicit consideration. It was not an approach that harmonized well with the economics of a stationary economy, and again the theory was very inconvenient to incorporate into cost and demand curves (especially if one will not use the concept of a representative firm). If the economies of scale within the firm were as strong as Marshall pictured them, moreover, it was not clear that continuously high-quality entrepreneurship was necessary to achieve monopoly. And could the giant firm not grow quickly by merger as well as hesitantly by internal expansion?

Marshall's third theory, of the falling demand curve for the individual firm, lost popularity for a generation because it was incompatible with perfect competition rigorously defined, and this became increasingly the standard model of analysis. And, paradoxically, when the falling demand curve was rediscovered and popularized in the 1930's by the proponents of imperfect and monopolistic competition, they used it not to examine the broad movements of industries and of economies but to focus price theory on the physiology and pathology of the firm.

In 1928, to retrace a step, the neglect of increasing returns had gone so far that Allyn Young felt the need to restore perspective by an emphatic indorsement of the fundamental importance of Smith's theorem: "That theorem, I have always thought, is one of the most illuminating

and fruitful generalizations which can be found anywhere in the whole literature of economics."[1] His position seemed persuasive, but he did not resolve the technical difficulties of incorporating the extent of the market into competitive price theory. Indeed, he openly avoided this problem, asserting that the firm and perhaps also the industry were too small to serve as units of analysis in this area. And so, although Young's and Marshall's and Smith's position is often given lip service to this day, the tributes are tokens of veneration, not evidences of active partnership with the theory of the firm and the competitive industry.

II. THE FUNCTIONS OF A FIRM

The firm is usually viewed as purchasing a series of inputs, from which it obtains one or more salable products, the quantities of which are related to the quantities of the inputs by a production function. For our purpose it is better to view the firm as engaging in a series of distinct operations: purchasing and storing materials; transforming materials into semifinished products and semifinished products into finished products; storing and selling the outputs; extending credit to buyers; etc. That is, we partition the firm not among the markets in which it buys inputs but among the functions or processes which constitute the scope of its activity.

The costs of these individual functions will be related by technology. The cost of one function may depend upon whether the preceding function took place immediately before or in the immediate vicinity, as when hot ingots are processed with a saving of heat. Or the interrelationships among processes may be remote, as when the entrepreneur must neglect production in order to supervise marketing.

Let us ignore for a moment these interrelationships of costs of various functions, in order to achieve a simple geometrical picture of the firm's costs of production. If the cost of each function depends only on the rate of output of that function, we may draw a unique cost curve for it. Furthermore, if there is a constant proportion between the rate of output of each function and the rate of output of the final product (as when every 100 pounds of cement is bagged), we may draw the cost curves of all functions on one graph, and the (vertical) sum of these costs of various functions will be the conventional average-cost curve of the firm.

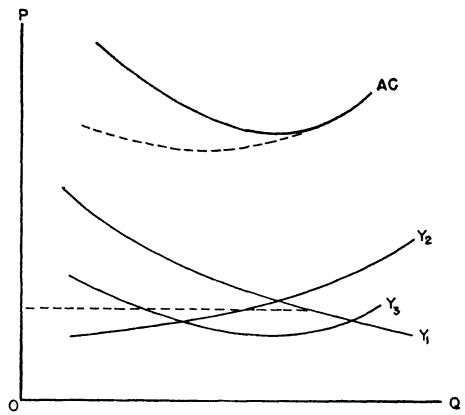

FIG. 1

We should expect to find many different patterns of average costs of functions: some falling continuously (Y_1); some rising continuously (Y_2); some conventionally U-shaped (Y_3) (see Fig. 1). It is not impossible, of course, that the average cost of some operations first rises and then falls.

Now consider Smith's theorem. Certain processes are subject to increasing returns; why does the firm not exploit

[1] "Increasing Returns and Economic Progress," *Economic Journal*, XXXVIII (1928), 529.

them further and in the process become a monopoly? Because there are other functions subject to diminishing returns, and these are, on balance, at least so costly that average cost of the final product does not diminish with output. Then why does the firm not abandon the functions subject to increasing returns, allowing another firm (and industry) to specialize in them to take full advantage of increasing returns? At a given time these functions may be too small to support a specialized firm or firms. The sales of the product may be too small to support a specialized merchant; the output of a by-product may be too small to support a specialized fabricator; the demand for market information may be too small to support a trade journal. The firm must then perform these functions for itself.

But, with the expansion of the industry, the magnitude of the function subject to increasing returns may become sufficient to permit a firm to specialize in performing it. The firms will then abandon the process (Y_1), and a new firm will take it over. This new firm will be a monopoly, but it will be confronted by elastic demands: it cannot charge a price for the process higher than the average cost of the process to the firms which are abandoning it. With the continued expansion of the industry, the number of firms supplying process Y_1 will increase, so that the new industry becomes competitive and the new industry may, in turn, abandon parts of process Y_1 to a new set of specialists.

The abandonment of function Y_1 by the original industry will alter each firm's cost curves: the curve Y_1 will be replaced by a horizontal line (ignoring quantity discounts) at a level lower than Y_1 in the effective region. The cost curve of the product (drawn with broken lines in Fig. 1) will be lower, and, on present assumptions, the output at which average costs are a minimum (if only one such output exists) becomes smaller.

Certain functions are also subject to increasing cost; why not abandon or at least restrict the scale of operation of these functions? The foregoing discussion is also applicable here, with one change. When the industry grows, the original firms need not wholly abandon the increasing-cost processes. Part of the required amount of the process (say, engine castings for automobiles) may be made within the firm without high average (or marginal) costs, and the remainder purchased from subsidiary industries.

In order to give a simple geometrical illustration, we have made two assumptions. The first is that the rate of output of the process and the rate of output of the final product are strictly proportional. This will be approximately true of some functions (such as making parts of a single final product), but it will also be untrue of other functions (such as advertising the product). If we drop the assumption, the substance of our argument will not be affected, but our geometrical picture becomes more complicated.[2]

Our second assumption, that the costs of the functions are independent, is more important. Actually, many processes will be rival: the greater the rate of output of one process, the higher the cost of a given rate of output of the other process or processes. Sometimes the rivalry will be technological (as in many multiple-product firms), but almost always it will also be managerial: the wider the range of functions the firm undertakes, the

[2] We can either draw separate cost curves for the various functions or combine them on one chart, with the scales of the functions chosen so that the optimum amount of each function is shown for the given rate of final output.

greater the tasks of co-ordination. Other processes will be complementary: the greater the rate of output of one process, the lower the cost of a given rate of output of the other processes. A most curious example of complementarity is the circular flow of materials within a plant; thus, in the course of making steel, steel plants supply a large part of their requirements for scrap.

If, on balance, the functions are rival, then usually the firm will increase its rate of output of the final product when it abandons a function; and I think that this is generally the case. For example, in the famous study of the Lancashire textile industry by Chapman and Ashton, it was found that firms engaged in both spinning and weaving in 1911 had, on average, 47,634 spindles, while those engaged only in spinning had, on average, 68,055 spindles.[3] But this is not necessary—indeed, they found the converse relationship in number of looms—and the effect of the range of functions on the size of the firm requires much study before we can reach safe generalizations.

III. VERTICAL INTEGRATION

Many economists believe that, with the growth of firms (and industries?), functions are usually taken over from previous independent industries. For example, United States Steel Corporation now mines its ores, operates its own ore-hauling railroads and ships, and, at the other end, fabricates barrels, oil-field equipment, and houses. (The number of economic views based chiefly on half-a-dozen giant corporations would repay morbid study.)

Broadly viewed, Smith's theorem suggests that vertical disintegration is the typical development in growing industries, vertical integration in declining industries.[4] The significance of the theorem can therefore be tested by an appeal to the facts on vertical integration.

Unfortunately, there are no wholly conclusive data on the trend of vertical integration. The only large-scale quantitative information at hand comes from a comparison of the 1919 study by Willard Thorp with the 1937 study by Walter Crowder of central offices (companies with two or more manufacturing establishments). In 1919, 602 manufacturing companies, or 13.0 per cent of a moderately complete list of 4,635 companies, had two or more establishments making successive products, that is, the product of one establishment was the raw material of another establishment.[5] In 1937, successive functions were found in 565 companies (or 10.0 per cent of a more complete list of 5,625 companies).[6] In 1919, successive functions were found in 34.4 per cent of all complex central offices (companies with establishments in two or more industries); in 1937, in only 27.5 per cent. Multiplant companies probably grew in relative importance during this period, so it is possible that a larger share of manufacturing output came from vertically integrated firms. But, so far as these multiplant com-

[3] S. J. Chapman and T. S. Ashton, "The Sizes of Businesses, Mainly in the Textile Industries," *Journal of the Royal Statistical Society*, LXXVII (1914), 538.

[4] This is not a wholly rigorous implication, however. With the growth of industries, specialism of firms may take the form of dealing with a narrower range of products as well as performing fewer functions on the same range of products.

[5] W. Thorp, *The Integration of Industrial Operation* (Washington, 1924), p. 238. I have omitted railroad repair shops and also the 301 companies having establishments which made successive products, because mining establishments were included.

[6] W. F. Crowder, *The Integration of Manufacturing Operations* ("T.N.E.C. Monographs," No. 27 [Washington, 1941]), p. 197.

panies are concerned, there seems to have been a tendency away from vertical integration.[7]

If one considers the full life of industries, the dominance of vertical disintegration is surely to be expected. Young industries are often strangers to the established economic system. They require new kinds or qualities of materials and hence make their own; they must overcome technical problems in the use of their products and cannot wait for potential users to overcome them; they must persuade customers to abandon other commodities and find no specialized merchants to undertake this task. These young industries must design their specialized equipment and often manufacture it, and they must undertake to recruit (historically, often to import) skilled labor. When the industry has attained a certain size and prospects, many of these tasks are sufficiently important to be turned over to specialists. It becomes profitable for other firms to supply equipment and raw materials, to undertake the marketing of the product and the utilization of by-products, and even to train skilled labor. And, finally, when the industry begins to decline, these subsidiary, auxiliary, and complementary industries begin also to decline, and eventually the surviving firms must begin to reappropriate functions which are no longer carried on at a sufficient rate to support independent firms.

[7] The ratio of "value-added" to value of product is a crude index of the extent of vertical integration *within* establishments. It is interesting to note that in the 17 industries in which this ratio was highest in 1939 in manufacturing, the average number of wage-earners was 16,540. In the 17 industries in which the ratio was lowest, the average number of wage-earners was 44,449. Thus the vertically integrated establishments were in smaller industries than the vertically disintegrated establishments (see National Resources Planning Board, *Industrial Location and National Resources* [Washington, 1943], p. 270).

We may illustrate this general development from the cotton textile machinery industry, much of whose history has recently become available.[8] This industry began as a part of the textile industry: each mill built a machine shop to construct and repair its machines. The subsequent history is one of progressive specialism, horizontal as well as vertical: at various times locomotives, machine tools, the designing of cotton mills, and direct selling were abandoned. When the cotton textile market declined in the 1920's, the machinery firms added new products, such as paper machinery, textile machinery for other fabrics, and wholly novel products, such as oil burners and refrigerators. Indeed, one is impressed that even the longer cyclical fluctuations seem to have affected the extent of specialism in much the same way as have the secular trends.

Of course, this is not the whole story of vertical integration, and it may be useful to sketch some of the other forces at work. The most important of these other forces, I believe, is the failure of the price system (because of monopoly or public regulation) to clear markets at prices within the limits of the marginal cost of the product (to the buyer if he makes it) and its marginal-value product (to the seller if he further fabricates it). This phenomenon was strikingly illustrated by the spate of vertical mergers in the United States during and immediately after World War II, to circumvent public and private price control and allocations. A regulated price of OA was set (Fig. 2), at which an output of OM was produced. This quantity had a marginal value of OB to buyers, who were rationed on a nonprice basis. The gain to buyers

[8] G. S. Gibb, *The Saco-Lowell Shops* (Cambridge: Harvard University Press, 1950); T. R. Navin, *The Whitin Machine Works since 1831* (Cambridge: Harvard University Press, 1950).

and sellers combined from a free price of *NS* was the shaded area, *RST*, and vertical integration was the simple way of obtaining this gain. This was the rationale of the integration of radio manufacturers into cabinet manufacture, of steel firms into fabricated products, etc.

Although nonprice rationing provides the most striking examples of this force toward vertical integration, private monopolies normally supply the same incentive. Almost every raw-material cartel has had trouble with customers who

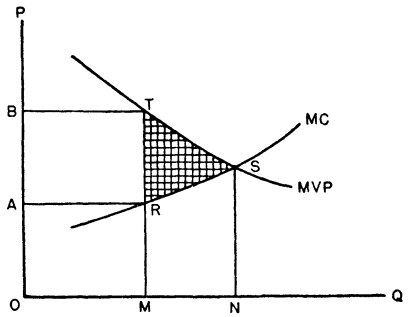

FIG. 2

wish to integrate backward, in order to negate the cartel prices. Since the cartel members are sharply limited in their output quotas, the discounted future profits of a cartel member need not be high, even with very high prices; so it is profitable for buyers to integrate backward by purchase (as well as by seeking noncartelized supply sources). The Rhenish-Westphalian Coal Cartel, for example, was constantly plagued by this problem:

> While a few of the members of the original syndicate agreement of 1893 had been steel companies which produced a part of their own coal and coke requirements, the steel industry, for the most part, had relied upon fuel purchased in the market. The stiffening of prices, coupled with the inelastic terms of sale resulting from the operation of the coal syndicate, now caused the steel companies to seek to free themselves from dependence upon the syndicate.
>
> . . . defensive measures were adopted by all classes of consumers. Some of the large industrial consumers . . . acquired their own mines individually or in groups. Among these were such important companies as the Vereinigte Stahlwerke, Rhenische Stahlwerk-Admiral, Badische Anilin- und Sodafabrik, Norddeutsche Lloyd, Friedrich Krupp, and a number of others representing the electric, gas, railway equipment, rubber and other industries. Also some cities such as Cologne and Frankfurt were among them.[9]

Monopoly is a devious thing, and it leads to vertical integration for other reasons also. A firm cannot practice price discrimination in the stages in which it does not operate; only by fabricating cable could the Aluminum Company of America sell cable at less than the ingot price in competition with copper, while maintaining a higher price on less competitive products.[10] Again, it is possible that vertical integration increases the difficulty of entry by new firms, by increasing the capital and knowledge necessary to conduct several types of operation rather than depend on rivals for supplies or markets.

These remarks are not intended to constitute a theory of vertical integration. There is doubt, indeed, that we want a theory of vertical integration except as part of a theory of the functions of a firm. As soon as one tries to classify the variegated details of production, one finds how artificial and arbitrary "vertical" relationships are. Whether one wishes to treat vertical relationships separately or as part of a general theory, however, Smith's theorem promises to be a central part of the explanation.

IV. WIDER IMPLICATIONS

If Smith's theorem is less than a complete theory of the division of functions

[9] A. H. Stockder, *Regulating an Industry* (New York, 1932), pp. 8, 11, and 36.

[10] D. H. Wallace, *Market Control in the Aluminum Industry* (Cambridge: Harvard University Press, 1937), pp. 218–19, 380.

among industries, it is also something more than this: it sheds light on several aspects of the structure and workings of economies. A few of the implications of the principle of increasing specialization will be discussed very tentatively.

One expects to find some relationship between the functional structure of an industry and its geographical structure —after all, reductions of transportation costs are a major way of increasing the extent of the market. (A reminder is hardly necessary that we are dealing with highly interdependent forces and that unilateral causation is implicitly assumed for simplicity and emphasis.) Localization is one method of increasing the economic size of an industry and achieving the gains of specialization. The auxiliary and complementary industries that must operate in intimate co-operation can seldom do so efficiently at a distance. I venture that, within a market area, geographical dispersion is a luxury that can be afforded by industries only after they have grown large (so that even the smaller production centers can reap the major gains of specialization) and that it must be sacrificed for geographical concentration, once the industry begins to shrink in size.

Closely related to this is the influence of localization upon the size of plant. The individual plants can specialize in smaller ranges of products and functions in highly localized industries (the size of the industry in some sense being held constant). In the United States geographically concentrated industries usually have fairly small plants.[11] There is also some evidence that the plants of an industry are smaller in the larger production centers. For example, in 1937 the average shoe factory in industrial areas had 137 employees, in other areas, 314 employees.[12] The dominance of medium-sized plants in highly localized industries has also been found in England.[13]

During the nineteenth century it was often said that England had the advantage of an "early start"; and this ambiguous statement had an element of truth which Smith's theorem more clearly expresses. As the largest economy in the world, it could carry specialism further than any other country, especially those "general" specialties (like railroads, shipping, banking, etc.) which are not closely attached to any one industry. England's advantage was a big start, as well as an early one.

Those too numerous people who believe that transactions between firms are expensive and those within firms are free will do well to study the organization of England during this period of eminence. In Birmingham, the center of the metal trades, specialism was carried out to an almost unbelievable extent. Consider the small-arms industry in 1860, when Birmingham was still the leading production center of the world:

> Of the 5800 people engaged in this manufacture within the borough's boundaries in 1861 the majority worked within a small district round St Mary's Church. . . . The reason for the high degree of localization is not difficult to discover. The manufacture of guns, as of jewellery, was carried on by a large number of makers who specialized on particular processes, and this method of organization involved the frequent transport of parts from one workshop to another.
>
> The master gun-maker—the entrepreneur—seldom possessed a factory or workshop. . . . Usually he owned merely a warehouse in the gun quarter, and his function was to acquire semi-finished parts and to give these out to specialized craftsmen, who undertook the assembly and finishing of the gun. He purchased

[11] National Resources Planning Board, *op. cit.*, pp. 250 ff.

[12] *Ibid.*, p. 257.

[13] P. S. Florence, *Investment, Location, and Size of Plant* (Cambridge, 1948).

materials from the barrel-makers, lock-makers, sight-stampers, trigger-makers, ramrod-forgers, gun-furniture makers, and, if he were engaged in the military branch, from bayonet-forgers. All of these were independent manufacturers executing the orders of several master gun-makers. . . . Once the parts had been purchased from the "material-makers," as they were called, the next task was to hand them out to a long succession of "setters-up," each of whom performed a specific operation in connection with the assembly and finishing of the gun. To name only a few, there were those who prepared the front sight and lump end of the barrels; the jiggers, who attended to the breech end; the stockers, who let in the barrel and lock and shaped the stock; the barrel-strippers, who prepared the gun for rifling and proof; the hardeners, polishers, borers and riflers, engravers, browners, and finally the lock-freers, who adjusted the working parts.[14]

At present there is widespread imitation of American production methods abroad, and "backward" countries are presumably being supplied with our latest machines and methods. By a now overly familiar argument, we shall often be a seriously inappropriate model for industrialization on a small scale. Our processes will be too specialized to be economical on this basis. The vast network of auxiliary industries which we can take for granted here will not be available in small economies. Their educational institutions will be unable to supply narrowly specialized personnel; they will lack the specialists who can improve raw materials and products. At best, the small economies that imitate us can follow our methods of doing things this year, not our methods of changing things next year; therefore, they will be very rigid. This position has been stated well by one observant citizen of a backward economy, Benjamin Franklin:

Manufactures, where they are in perfection, are carried on by a multiplicity of hands, each of which is expert only in his own part, no one of them a master of the whole; and if by any means spirited away to a foreign country, he is lost without his fellows. Then it is a matter of extremest difficulty to persuade a complete set of workmen, skilled in all parts of a manufactory, to leave their country together and settle in a foreign land. Some of the idle and drunken may be enticed away, but these only disappoint their employers, and serve to discourage the undertaking. If by royal munificence, and an expense that the profits of the trade alone would not bear, a complete set of good and skilful hands are collected and carried over, they find so much of the system imperfect, so many things wanting to carry on the trade to advantage, so many difficulties to overcome, and the knot of hands so easily broken by death, dissatisfaction, and desertion, that they and their employers are discouraged altogether, and the project vanishes into smoke.[15]

The division of labor is not a quaint practice of eighteenth-century pin factories; it is a fundamental principle of economic organization.

[14] G. C. Allen, *The Industrial Development of Birmingham and the Black Country, 1860–1927* (London, 1929), pp. 56–57, 116–17. Commenting on a later period, Allen says: "On the whole, it can be said that specialization was most apparent in the engineering industries in which output was rapidly expanding; while the policy of broadening the basis [product line] was found, mainly, either in the very large concerns, or in industries in which the decline of the older markets had forced manufacturers to turn part of their productive capacity to serve new demands" (*ibid.*, pp. 335–36). The later history of the gun trade, in which American innovations in production techniques were revolutionary, suggest that the organization in Birmingham was defective in its provision for technical experimentation.

[15] "The Interest of Great Britain in America," cited by V. S. Clark, *History of Manufactures in the United States* (New York, 1949), I, 152. Clark adds: "In these words Franklin was but reciting the history of the more important colonial attempts to establish a new industry or to enlarge an old one with which he was personally familiar."

[Footnotes]

[1] **Increasing Returns and Economic Progress**
Allyn A. Young
The Economic Journal, Vol. 38, No. 152. (Dec., 1928), pp. 527-542.
Stable URL:
http://links.jstor.org/sici?sici=0013-0133%28192812%2938%3A152%3C527%3AIRAEP%3E2.0.CO%3B2-R

[3] **The Sizes of Businesses, Mainly in the Textile Industries**
S. J. Chapman; T. S. Ashton
Journal of the Royal Statistical Society, Vol. 77, No. 5. (Apr., 1914), pp. 469-555.
Stable URL:
http://links.jstor.org/sici?sici=0952-8385%28191404%2977%3A5%3C469%3ATSOBMI%3E2.0.CO%3B2-G