Horizontal Boundaries of the Firm
Revised: August 31, 2001

Today we'll look at a firm's horizontal boundaries: How large is it? How many products does it sell? Are there natural limits to either one?

Consider size. Although there is no obvious tendency for firms to get larger over time (some firms grow, but small firms are born all the time), the modern business landscape obviously includes some very large firms. Citigroup employs 233k, Daimler-Chrysler 417k, Sony 190k, Bertelsmann 65k, Novartis 82k, and Telefonos de Mexico 72k. Why are these firms so large? Is there a cost advantage to size? Are there disadvantages?

The size of firms reflects, in large part, the cost functions in their industries, and differences across industries reflect differences in industry cost structures. Boeing and Airbus are the two dominant players in an industry that favors size. Smaller firms would have a substantial cost disadvantage in the manufacture of large airframes (although there are several firms in the market for smaller planes). At the other end of the spectrum, our local plumber, barber or hairdresser, and accountant are typically part of very small firms in industries where size is less important. In exploring these differences, we will look at economies and diseconomies of scale, economies of scope, and learning curves as properties of cost functions that affect the sizes of firms.

Economies of Scale

We say there are “economies of scale” when average cost declines with output. If you double output, costs increase by less than a factor of two. Or if you double costs, output more than doubles. If there are economies of scale, large firms are more efficient than small ones. If the average cost curve is U-shaped, as we've drawn it earlier, then we say there are economies of scale over the part in which average costs are decreasing. We can all see that a telecomm firm with one employee is likely to be less efficient than one with thousands. But why? What are the sources of this increasing efficiency with scale? Should we expect them to continue without limit?

Fixed costs are probably the most important source of scale economies. If there is a fixed cost to production and (say) a constant marginal cost, then if we spread the fixed cost over more units, average cost declines. This is especially important when the fixed costs are high relative to the variable costs. Consider the manufacture of a classical music CD. There are costs of producing the music (recording it, mixing it, etc) that must be paid regardless of the number of CDs that are produced. For a string quartet, this is at least $100,000. The marginal cost is in the neighborhood of 25¢ per CD. For a production run of 50,000 (large for this market), the average fixed cost is eight times the variable cost.
Ditto computer software. Similarly, the fixed cost of developing a new version of (say) Adobe Photoshop is far greater than the manufacturing and distribution costs of an additional unit. (Internet distribution may reduce the latter costs further, but technical support services are not fixed in this sense.) Ditto pharmaceuticals (fixed costs of research and development, regulatory approval), movies and broadcast television (content production), and wireless telephony (wireless infrastructure). Many people think advertising is partly a fixed cost, making it an important source of scale economies in many industries.

A second source of scale economies is specialization. This idea was laid out more than two centuries ago by Adam Smith in his famous example of a pin factory. In this example, we consider the manufacture of pins in a factory with small output and large output. In a factory with small output, it makes sense for the one or few workers there to make a pin from start to finish. Thus one person will draw the wire, cut the wire, sharpen the point, add the top, etc. In a larger plant, however, it makes sense to specialize, with one person drawing wire, another cutting, etc. This generally reduces costs, since individuals become very good at their tasks. Another example is a business school. In a small school, there would be fewer faculty, and faculty members would teach more courses. Without the benefits of faculty specialization, you might expect to see lower quality of teaching: Imagine an economist teaching Managing Organizations! You'd also see fewer choices of courses.

Another source of scale economies is the “two-thirds rule.” This is a mathematical result: as you increase the size of a container, the surface area increases with the square of the radius, but the volume increases with the cube. If cost is proportional to surface area (materials used to make to make it) and output is proportional to volume, then the cost per unit of output declines with scale. Formally, cost increases with output to the “two-thirds power,” so average cost decreases with scale. Enough math. What this means is that things like brewing, oil refining, and shipping by ocean tankers exhibit scale economies. Mathematically similar is inventory management. For reasons your Data Analysis and Operations professors can best explain, the quantity of inventories needed to maintain a given (presumably low) probability of a “stock-out” increases less quickly than output. That’s one of the reasons Amazon's Jeff Bezos went into book retailing: inventories are a substantial cost in that business, so it was a natural place to look for economies of scale from centralizing distribution.

**Diseconomies of Scale**

Most scale economies have limits. At high enough levels of production, fixed costs are essentially irrelevant. At some point, specialization, too, becomes counterproductive, as the cost of organizing a highly specialized production process overcomes its direct economies. Even the two-thirds rule has limits, since the cost of reinforcements to keep a very large container from collapsing reverses some of the benefits of scale. In other cases, the physical scale economies are limited by the market: large tankers and airplanes aren’t efficient on some routes, because the market isn’t large enough to support them.
When average cost increases with output, we say there are “diseconomies of scale.” In addition to the reasons just cited, an important source of diseconomies of scale is the difficulty of managing a very large organization effectively. In large organizations, the bureaucracy of running it may reduce some of the other benefits of scale. To a great extent, what distinguishes effective organizations is their ability to manage a large number of employees efficiently. For any given firm, it's easy to imagine (easier with experience!) that size increases the challenges of management. Consider a stylized by illustrative organization structure, in which each manager has two workers on the previous level reporting to her and we keep adding layers of management until we reach a single person at the top. [Picture a pyramid.] Think of output as proportional to the number of workers on the lowest level (n, say). Think of cost as proportional to the number of workers \( [n(n+1)/2] \). In this setting, the extra layers of management result in costs that increase with size. The question in practice is how large these “span of control” costs are relative to the various sources of scale economies.

To a large extent, the accelerating globalization we have seen over the last two decades has been supported by economies of scale. Lower costs of global communications and transportation has likely expanded the effective size of many markets and allowed firms to exploit economies of scale on a global basis.

Putting economies and diseconomies together, we might say that economies of scale explain why we see large firms in many industries, and diseconomies suggest that firms will not grow without limit. With a U-shaped average cost curve, we would say that the upward-sloping right side of the curve exhibits diseconomies of scale. The U-shape would indicate that scale economies are relevant at very small scales, and diseconomies at very large scales. In the middle, there may be a range over which average cost is roughly constant. We refer to the (smallest) level of output at which average cost is lowest as the “minimum efficient scale.” Firms producing lower output would be expected to have higher costs as a result.

**Scale of What?**

You might think that economies of scale, when they exist, mean that large firms are more efficient than small ones. In fact, economies of scale often apply to particular activities, not entire firms, and depending on how a firm is organized we might find them even in relatively small firms – or not find them in large firms.

Southwest Airlines is a good example. Despite being relatively small, it has the lowest costs in the business. One reason is specialization. Using only Boeing 737s, their costs of training pilots and maintaining equipment are substantially lower than those of competitors using mixed fleets. (To be fair, scale economies play a minor role in Southwest’s success. Observers generally cite their organizational structure, which delivers high-quality service at low cost.)
A more complex example is Hewlett-Packard. Since its start, HP had followed a strategy of avoiding the diseconomies of managing a large organization by creating new units when existing ones got too big. However, this strategy failed to exploit potential economies of scale in routine support services that had to be replicated in each unit. It also presented a complex front to large customers, who often had to open separate accounts to purchase different products.

When Carly Fiorina took over as CEO in 1999, HP had 83 autonomous business units. While the bureaucratic burdens of HP’s size were arguably low, the benefits were nonexistent. An HP customer, for example, could expect to get separate bills from the printer, desktop, and consulting divisions, even if all three were involved in the same sale. Fiorina reduced the number of units to 12 and added a unified sales team to deal with leading customers. Her integrated approach had a highly visible early success, when Amazon bought a combined package of servers, printers, and PCs. And this despite HP’s relatively late start in the server market. The challenge for HP going forward is to get the benefits of coordination across divisions without excessive costs of coordination.

Generally there is a tradeoff between scale economies and product differentiation. Henry Ford, for example, drove down costs by going after the mass market with the Model T and offering no options: “You can have any color you like, as long as it’s black.” In today’s world, customers want choice, and the question is how to give it to them without driving costs through the roof. The Holy Grail is manufacturing is “mass customization,” in which manufacturing processes generate diverse goods at low cost. An easy example to picture is Dell. You can order a PC on their web site and choose any of dozens of options: choice of processor, hard drive, monitor, etc. In many cases, they can build it to order and have it delivered within 2 days. All this is possible because components can be plugged into the PC with relatively ease. (Not to deny Dell’s innovation: they do it extremely well, and have a terrific reputation for service, too.) Automobiles are more challenging, since the parts are larger (hence harder to keep in inventory) and assembly more complex. Take color. It’s much more efficient to paint panels in bunches, rather than one at a time, which makes it difficult to meet orders for (say) blue, white, and green in succession. Nevertheless, Toyota claims to be able to build-to-order and deliver in 3 weeks. More on this in your Operations course.

**Economies of Scope**

We say that two products exhibit “economies of scope” if the cost of producing them together is less than producing them separately. Or, equivalently, if for the same cost the quality of the products is higher. Thus when you hear someone refer to a “1+1=3” synergy, they’re talking about economies of scope.

Synergy has become one of those infamous lies, like "the check is in the mail" and "my dog ate my homework," but there are examples that make a lot of sense. The various mergers that led to Citigroup were largely justified by economies of scope: between investment banking and retail brokerage, between insurance and banking, and so on.
Ditto the 1997 merger of Morgan Stanley and Dean Witter. Other examples are less obvious. What are the benefits of Anheuser-Busch producing snacks as well as crackers? Or Virgin’s Richard Branson selling music CDs, airplane flights, and mutual funds? (We're not saying he won't find them, just pointing out they're not obvious.)

One source of economies of scope is natural by-products: one product is a natural byproduct of another. For example, if you are refining oil, you would tend to produce both gasoline and kerosene as part of the refining process. It would be cheaper for your firm to produce and sell both than for two firms to produce them separately. In the latter case, each firm would be throwing away the other good as a by-product of the process and therefore would not be producing as cheaply as one firm producing both. An intriguing example was suggested by MBA02 Erin Yau: Gas pipelines and telecommunications are natural joint products, since the right-of-way for the pipeline provides a useful path for fiber optic cable. Another example (trust us, we read this in the *Times*) is chicken and alligator farming. Apparently a large number of chickens die from such things as heat exhaustion on chicken farms. These chickens are not fit for human consumption, but alligators seem to like them just fine. Running an alligator farm at the same time as a chicken farm, then, saves one the expense of disposing of the chickens and buying food separately for the alligators. Economies of scope indeed.

Another source of economies of scope are inputs that are fixed costs at the firm level, making a multi-product firm more efficient. GE’s manufacture of airplane engines and power turbines has this flavor, since the technologies are similar for the two products. Merging partners often cite overlap of support services: HR, legal, research, marketing, technology, and so on. Some of these really are fixed costs, but others increase with the scale of the firm. Maintaining a brand name might also be considered in this category. For example, Bic made a name for itself selling disposable ballpoint pens. They successfully extended their brand into disposable razors and lighters. This saved them the money of having to establish an entirely new brand for these lines. Canon is another example of a successful brand expansion, in this case from cameras to copiers and printers.

There are diseconomies of scope, too. One reason is – again! – the difficulty of managing a large organization. The effect is conceivably greater across different product lines, since the complexity of an organization depends on both scale and the range of products. A second reason is brand overstretching. A brand cannot be stretched beyond a certain limit without either hurting the overall brand or not being useful for the particular product. For example, Levis came up with a line of business suits. Since people associated the company with casual wear, this proved to be a disaster. Other examples (yes, these really happened) are McDonald’s photo processing and Heineken popcorn.

**Learning Economies**
Learning economies occur when the cost of producing a given level of output declines with past experience. The difference between learning and scale is that learning take place over time and its effects are cumulative. The usual method people measure learning economies is the decrease in costs associated with cumulative output. That is, the cost per unit decreases with the total number of units produced since the firm began production. This can be illustrated in industries such as semiconductors. In the semiconductor industry chips are produced on wafers with some of the chips being unusable. The yield of good chips is the key to profitability in the industry. Improving to a 90% yield from an 80% yield adds a huge amount to the company’s bottom line. A lot of how this is done is trial and error, with the yield increasing as you go along and thus the cost per unit decreasing. Another good example of this process in action is aircraft manufacturing. Aircraft have so many parts that manufacturing can be improved significantly by assembling them in the right way. Boeing usually disassembles and reassembles prototype aircraft several times in order to gain process improvements and decrease assembly costs. They have gone from an 18-month assembly period to 14 months and have a goal of 10-12 months for the future.

Many times people point to learning curves when in fact there are scale economies, or vice versa. Or in some cases, both may be relevant. A good way to think about the difference is that learning economies cannot be reversed simply by lowering production. If you have learned a process to assemble an aircraft in 14 months you can do it in that time whether you are producing 1 or 10 aircraft at once.

**External Economies**

“External economies” occur when advances in one company reduce costs in another. The mechanism is mysterious, but there's a great deal of evidence for it. External economies can come from such things as a skilled labor pool, customers knowing where to look for particular goods, etc. Examples of this are Mr. Woo and Toytown, where a man setting up a toy distributor in LA encouraged competitors to move to the same neighborhood. This allowed the creation of external economies for the area and created LA’s Toytown. Retail stores in New York follow this pattern also with such areas as the Garment District, the Flower District, the Diamond District, the Meatpacking District, etc. Perhaps the most studied and emulated example is Silicon Valley and the microcomputer industry there. Real estate in the area is extremely expensive, yet firms attempt to move and start there all the time. There is no little irony here that geography remains so important in the center of internet innovation. The importance of a skilled labor market that brings ideas from other firms is important, as well as the generation of ideas by people who happen to be in the same office or area. Thus the idea that we can work from anywhere seems to be limited by the loss of external economies.

Written by Kevin Cassidy under the supervision of Luís Cabral and David Backus. © 2001 Luís Cabral and David Backus.