MANAGERIAL DISECONOMIES OF SCALE

Literature Survey and Hypotheses Anchored in Transaction Cost Economics

Working paper

by

Staffan Canbäck

Henley Management College

April 1997
[Updated February 2002]

CONTENTS

FIGURES

TABLES

1. SUMMARY

This paper tests whether diseconomies of scale influence corporate performance. It uses Coasian transaction cost economics (Coase 1937) and Williamson's thinking on the nature of diseconomies of scale and the limits of firm size (Williamson 1975, 1985; Riordan and Williamson 1985) to develop a theoretical framework for describing diseconomies of scale, economies of scale, and moderating factors. It validates the framework against the relevant literature and translates it into five hypotheses. The findings are consistent with Williamson's limits-of-firm-size framework.

Diseconomies of scale are a neglected area of study (see also Chapter 2). Observers from Knight ([1921] 1964) to Holmström and Tirole (1989) have pointed out that our understanding of bureaucratic failure is low. The neglect is to some extent due to a disbelief in the existence of diseconomies of scale (e.g., Florence 1933, 12; Bain 1968, 176). It is also due to a dearth of theoretical frameworks that can help inform our understanding of the nature of diseconomies of scale. However, if diseconomies of scale did not exist, then we would presumably see much larger firms than we do today (Panzar 1989, 38). No business organisation in the United States has more

---

This working paper is the foundation for the doctoral thesis "Bureaucratic Limits of Firm Size: Empirical Analysis Using Transaction Cost Economics" (Canbäck 2002) available at http://canback.com/henley.htm. The thesis contains a full statistical analysis of the hypotheses described in this paper, based on a sample of 784 US manufacturing firms.

than one million employees[1] or more than ten hierarchical levels. No firm has ever been able successfully to compete in multiple markets with a diverse product range for an extended period of time. Common sense tells us that there are limits to firm size. Common sense does not, however, prove the point. Unfortunately, scientific inquiry has not yet focused on finding such proof.

The US manufacturing sector has, as a whole, been remarkably stable over the last century. Contrary to popular opinion, markets have on average not become more concentrated (e.g., Nutter 1951; Scherer and Ross 1990). Large firms are not increasingly dominant. Large manufacturing firms in the United States employed 16 million people in 1979 versus 11 million in 1994, while private sector employment grew from 99 to 123 million people (Council of Economic Advisers 1998; Fortune 1995a).

Williamson (1975, 117-131) found that the limits of firm size are bureaucratic in origin and can be explained by transaction cost economics (see also Chapter 3). He identified four main categories of diseconomies of scale: *atmospheric consequences* due to specialisation, *bureaucratic insularity*, *incentive limits* of the employment relation and *communication distortion* due to bounded rationality.

---

[1] The largest company, Wal-Mart Stores, Inc., had 910,000 employees in 1998. The largest manufacturing company, General Motors Corporation, had 594,000 employees.

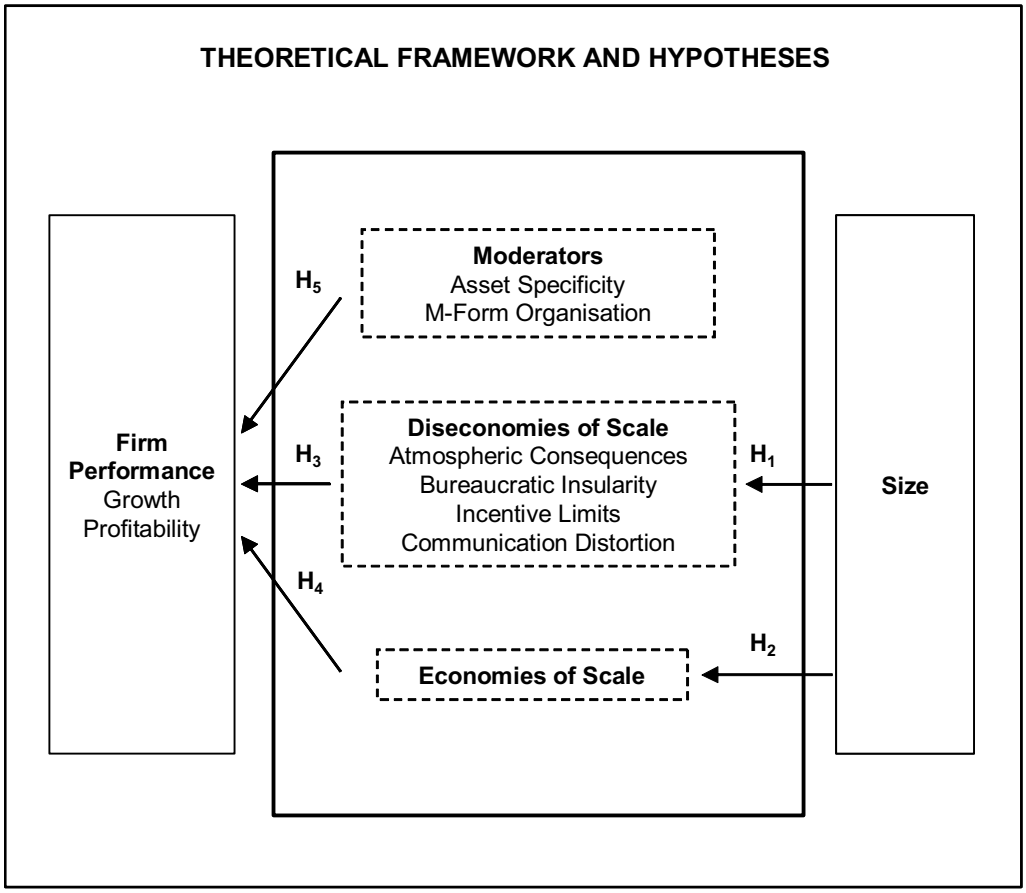
Economies of scale[2] in production costs and transaction costs tend to offset these diseconomies of scale (Riordan and Williamson 1985). Moreover, the disadvantages of bureaucracy can be moderated by using the *multidivisional organisation form* (M-form) and by a judicious optimisation of the degree of integration through high internal *asset specificity* (Williamson 1975, 1985). Together, these influences on firm performance form the theoretical framework used in this research.

The literature review supported the framework. There are, as far as this researcher could determine, around 60 pieces of work that deal with diseconomies of scale in a substantial manner (see Appendix). Based on these and other more fragmentary sources, it was possible to validate Williamson's framework and his categorisation of the factors driving diseconomies of scale, economies of scale and the moderating factors, except that the literature review was inconclusive regarding economies of scale. The framework was translated into five testable hypotheses, summarised in Figure 1 (see also Chapter 4).

---

[2] A standard definition of economies of scale, taken from *The New Palgrave: A Dictionary of Economics*, is that they exist if the unit cost of producing one additional unit of output decreases. They are driven by (a) the existence of indivisible inputs, (b) set-up costs and (c) the benefits of division of labour (Eatwell, Milgate and Newman 1987, 80-81). In the case of the multi-product firm, economies of scale exist if the ray average cost decreases as output increases.

Figure 1. Theoretical Framework and Hypotheses



The first two hypotheses test the tautological statement that diseconomies of scale and economies of scale increase with firm size. The last three hypotheses test how a firm's performance is affected by the diseconomies of scale, economies of scale and moderating influences.

H_1: Bureaucratic failure, in the form of atmospheric consequences, bureaucratic insularity, incentive limits and communication distortion, increases with firm size.

H_2: Large firms exhibit economies of scale

H_3: Diseconomies of scale from bureaucratic failure have a negative impact on firm performance

H_4: Economies of scale increase the relative profitability of large firms over smaller firms

H_5: Diseconomies of scale are moderated by two transaction cost-related factors: organisation form and asset specificity

The third hypothesis has four sub-hypotheses, which test each of the diseconomies of scale factors.

H_{3a}: Atmospheric consequences have a negative impact on the performance of large firms

H_{3b}: Bureaucratic insularity has a negative impact on the performance of large firms

H_{3c}: Incentive limits have a negative impact on the performance of large firms

H_{3d}: Communication distortion has a negative impact on the performance of large firms

The fifth hypothesis has two sub-hypotheses for organisation form and asset specificity, respectively.

H_{5a}: Large M-form firms perform better than large U-form firms

H_{5b}: High internal asset specificity affects a firm's performance positively

Table 1 summarises the findings for each hypothesis (see also Chapter 5). The strongest negative influence from diseconomies of scale on a large firm's performance appears to be on its ability to grow, while there is less negative influence on profitability. Thus, Penrose's claim ([1959] 1995, 261-263) that diseconomies of scale reduce the growth capability of large firms, appears to be validated.

Table 1. Summary of Findings

SUMMARY OF FINDINGS[a]	
Hypothesis	**Literature Finding**
H_1: Bureaucratic failure, in the form of atmospheric consequences, bureaucratic insularity, incentive limits and communication distortion, increases with firm size	Confirmed
H_2: Large firms exhibit economies of scale	Confirmed
H_3: : Diseconomies of scale from bureaucratic failure have a negative impact on firm performance	Confirmed
H_{3a}: Atmospheric consequences have a negative impact on the performance of large firms	Confirmed
H_{3b}: Bureaucratic insularity has a negative impact on the performance of large firms	Confirmed
H_{3c}: Incentive limits have a negative impact on the performance of large firms	Confirmed
H_{3d}: Communication distortion has a negative impact on the performance of large firms	Confirmed
H_4: Economies of scale increase the relative profitability of large firms over smaller firms	Inconclusive
H_5: Diseconomies of scale are moderated by two transaction cost-related factors: organisation form and asset specificity	Confirmed
H_{5a}: Large M-form firms perform better than large U-form firms	Confirmed
H_{5b}: High internal asset specificity affects a firm's performance positively	Confirmed
[a] For simplicity, the word "confirmed" is used, although "not rejected" is more accurate.	

The implications may be that diseconomies of scale are real and important contributors to a firm's performance, in a negative way. However, economies of scale can offset some of these negative consequences. Finally, the use of M-form organisation and pursuit of high internal asset specificity can moderate the negative impact of diseconomies of scale.

These findings make it possible to create conceptual cost curves and growth curves that extend neoclassical theory. The curves are found in Chapter 5.

There are several practical implications (see also Chapter 6). Among them are that corporate strategies are interconnected with the organisational choices made. That is, structure does not necessarily follow strategy. In light of this, it is understandable that mergers or acquisitions often fail, especially when the rationale for the merger-and-acquisition activity is to capture revenue growth opportunities. It is also evident that the focus on corporate governance over the last decade has its benefits. Other things equal, good governance allows large corporations to expand their limits-of-firm-size horizon. Moreover, as initiatives in large corporations are increasingly team-oriented, it is not surprising that senior executives pay more attention to motivation and how to structure incentives to extract optimal effort from the employees.

In the next chapter, the research objectives are defined and the importance of the research is discussed, linking it back to perspectives on economies of scale and diseconomies of scale in neoclassical theory and transaction cost economics. The chapter then explores the definition of the firm and metrics for measuring firm size. Finally, trends in firm size and concentration in the US manufacturing sector are discussed.

2. INTRODUCTION TO THE RESEARCH

Why are large firms so small? What stops firms from effortlessly expanding into new businesses? Only fragmentary research exists today as to why the largest business organisations do not have ten, twenty or a hundred million employees rather than a few hundred thousand.

According to Arrow (1974, 55) a "tendency to increasing costs with scale of operation" due to the cost of handling information and the irreversible cost of building organisational knowledge leads to limits of firm size. Coase (1937, 397) found that these costs—labelled "diseconomies of scale" in this paper to contrast them with "economies of scale"—are associated with the resources required to manage the firm's internal planning processes, as well as the cost of mistakes and the resulting misallocation of resources, especially under conditions of uncertainty.

This paper builds on original research carried out in the subject area. Specifically, it tests whether Williamson's "limits of firm size" discussion in *Markets and Hierarchies: Analysis and Antitrust Implications* (1975, 117-131) and in *The Economic Institutions of Capitalism* (1985, 131-162), which extend Arrow's and Coase's arguments, are valid. The findings include a look at the nature of diseconomies of scale and factors which moderate their

impact, as well as estimates of the impact of diseconomies of scale on firm performance.

Transaction cost economics (TCE) provides the theoretical foundation for this research. There are other partial explanations of diseconomies of scale, such as those found in neoclassical economics (e.g., Mas-Colell, Whinston and Green 1995; Scherer and Ross 1990); agency theory (e.g., Pratt and Zeckhauser 1985; Jensen and Meckling 1976); growth theory (e.g., Penrose [1959] 1995); evolutionary theory (e.g., Nelson and Winter 1982); sociology (e.g., Blau and Meyer 1987); and Marxist theory (e.g., Marglin 1974). These explanations are not the focus here, although they will be used to illuminate and test particular aspects of the TCE argument described in Chapter 3.

The purpose of the research is to create a theoretically robust framework that can be used by executives and others to inform strategic and organisational choices for large corporations. These choices may help decision-makers achieve higher growth and profitability by minimising diseconomies of scale due to atmospheric consequences, bureaucratic insularity, incentive limits and communication distortion (as defined in Section 3.1.2); to capture economies of scale; to optimise organisational structures; and to maximise asset specificity within the corporation.

The remainder of this chapter describes the research objectives and their importance in more detail, defines firm size, and documents trends in firm size over the last century.

2.1 RESEARCH OBJECTIVES

This section gives an initial problem definition and discusses the importance of the research. It spells out why diseconomies of scale are real and pervasive, yet poorly understood. In fact, while the economics literature often includes cost curves that bend upward at large firm sizes, there are only around 60 pieces of work that explicitly discuss the nature of the diseconomies,[3] and only a few of these have attempted to quantify the diseconomies of scale.

2.1.1 Problem Definition

In the early 1920s, Knight ([1921] 1964, 286-287) observed that "the diminishing returns to management is a subject often referred to in economic literature, but in regard to which there is a dearth of scientific discussion". Since then, many authorities have referred to the existence of diseconomies of scale, but no systematic studies of the general issue exist.

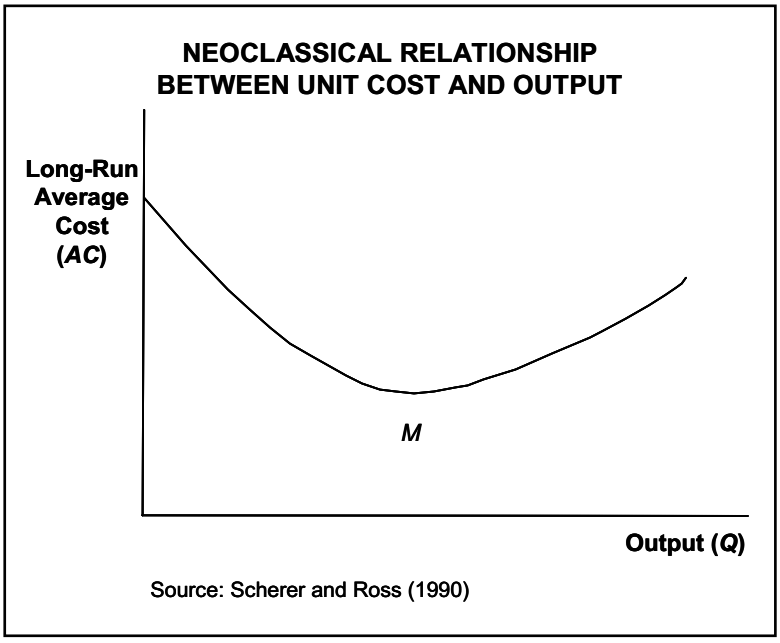
---

[3] There is also a vast literature on the size-distribution of firms, but it generally does not discuss the specific mechanisms underlying bureaucratic failure.

The basic dilemma is illustrated by the mismatch between theoretical expectations and real-world observations. On the one hand, if diseconomies of scale do not exist, then there should be no limits to firm growth and size. We would observe an inexorable concentration of industries and economies until only one global firm was left. The answer to Coase's question (1937, 394): "Why is not all production carried on by one big firm?" would be: it will. Similarly, Stigler (1974, 8) wrote that "if size were a great advantage, the smaller companies would soon lose the unequal race and disappear". This is not happening. On the other hand, if a given industry has an optimum firm size, then we would expect increased fragmentation as the overall economy grows. This would be in line with Stigler's survivor-principle argument which holds that "the competition between different sizes of firms sifts out the more efficient enterprises" (1958, 55). Again, this is not happening. Lucas (1978, 509) observed that "most changes in product demand are met by changes in firm size, not by entry or exit of firms". The size distribution of firms has been remarkably stable over time for most for the last century, when measured by number of employees or as a share of the total economy (as discussed in Section 2.3).

Cost curves (Figure 2) are used in neoclassical theory to illustrate economies and diseconomies of scale (e.g., Marshall [1920] 1997, 278-292; Scherer and Ross 1990, 101).

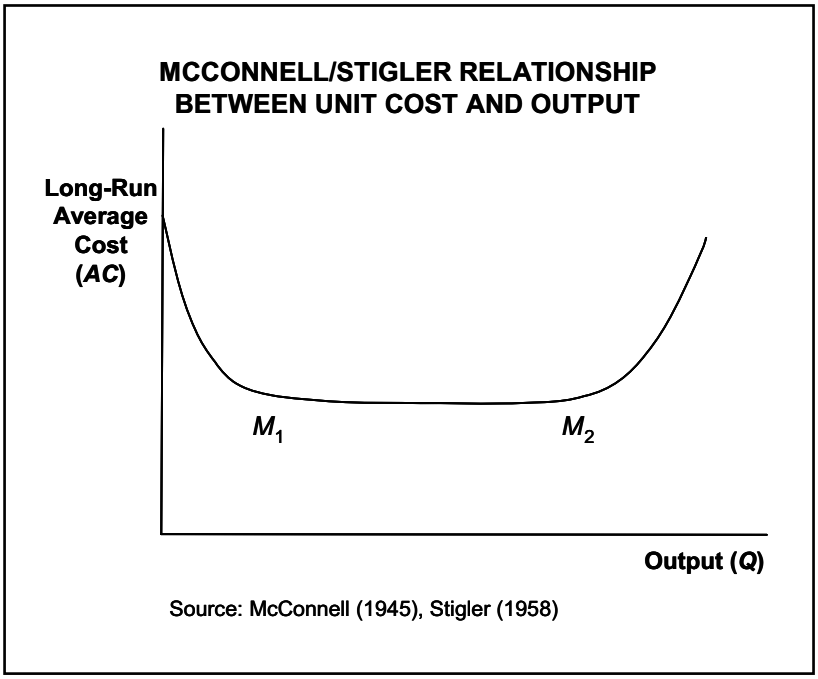
Figure 2. Neoclassical Relationship between Unit Cost and Output



As the output *Q* increases, the average cost decreases due to economies of scale. At a certain point (*M*) the economies of scale are exhausted, while diseconomies of scale, presumably driven by diminishing returns to management (e.g., Coase 1937, 395), start to influence the unit cost. As output increases, the unit cost increases. In a competitive market, this implies an equilibrium output *M* where marginal cost not only equals marginal revenue, but also intersects long-run average cost at its minimum (e.g., Mankiw 1998, 296).

In reality, however, this is not what is observed. Rather, the cost-minimising part of the curve covers a wide range of outputs, and only at high output levels do diseconomies set in, if ever (Panzar 1989, 37-38). McConnell's quantification (1945, 6) and Stigler's illustration (1958, 59), reproduced in Figure 3, are typical.

Figure 3. McConnell/Stigler Relationship between Unit Cost and Output



This shape of the cost curve reconciles several real-world observations. (1) It explains why large and small firms can coexist in the same industry. There is a wide range of outputs, between the points M_1 and M_2, for which the unit cost is more or less constant. (2) It is consistent with Lucas's observation (1978, 509) that, as the economy grows, existing firms tend to expand supply to meet additional demand, because most firms operate

with outputs Q below the M_2 inflexion point. (3) It eliminates the supposition that economies of scale are exhausted at approximately the same point that diseconomies of scale start increasing unit cost, which is indicated with M_1 being much to the left of M_2. (4) It demonstrates that there are indeed limits to firm size due to diseconomies of scale, as shown by the increasing unit cost beyond M_2 — large firms have not expanded indefinitely.

However, if the reasoning above is correct, it is still unclear why the cost curve bends upwards at M_2. Neoclassical theory does not provide a satisfactory answer. As Simon ([1947] 1976, 292) said: "the central problem is not how to organize to produce efficiently (although this will always remain an important consideration), but how to organize to make decisions".[4] The first part of this statement refers to the negative derivative of the cost curve at outputs smaller than M_1, where economies of scale in production have not yet been exhausted, while the second part applies to the upward slope, where diseconomies of scale due to diminishing returns to management set in beyond M_2.

[4] Simon echoed the writing of Robertson (1923, 25): "It is the economies of large-scale government rather than of large-scale technique which dictate the size of the modern business unit". (Note: *government* here refers to *corporate organisation and governance*, not national government.)

Clarifying "how to organise to make decisions"—and thus the upward bend of the cost curve—will help executives optimise corporate performance. The current research investigates whether transaction cost economics can more thoroughly explain diseconomies of scale and what drives these diseconomies. It picks up on a debate that harks back to the early 1930s when Florence (1933) and Robinson (1934), respectively, argued the case against and for limits of firm size. Florence believed that optimum firm size meant maximum firm size: "the more the amount of any commodity provided the greater the efficiency" and "there is in my view no theoretical limit to the increase in the physical return obtainable by larger-scale operations" (p. 12). He argued that no organisation would be too large for a single leader to control and thought that the only reason this had not happened yet was a certain lag between what managers at the time assumed they could do and the inevitable outcome (p. 47).

In contrast, Robinson did not subscribe to this reasoning and he believed strongly in "the increasing costs of coordination required for the management of larger units" (p. 242). He argued that the existing facts—the then newly released first report on the size distribution of British firms—supported the notion that optimum firm size was less than maximum firm size (p. 256).

2.1.2 Importance of the Research

Diseconomies of scale have not been extensively studied and thus there may be a genuine gap in our understanding of the firm. Transaction cost economics may help fill this gap because the theory embeds a number of concepts relating to the limits of the firm. Filling the gap may not only affect the way we think about strategy and structure, but also help executives make more effective decisions.

Limits-of-firm-size is not a major field of study (Coase 1993a, 228; Holmström and Tirole 1989, 126). There are around 60 articles or books that deal with the topic in a meaningful way (see Chapter 3 for a review and Appendix for a list of references). Williamson (1985, 153), for example, stated that our understanding of bureaucratic failure is low compared with what we know of market failure. Given the relative slowdown in the growth of large firms over the last 30 years (see Section 2.3), understanding why market-based transactions are slowly winning over internally-based transactions matters more than ever.

The second reason why this research is academically important is that it uses transaction cost economics in a somewhat new fashion. The 1970s were the defining years of TCE. At that time, large firms still appeared set to become ever more dominant, and TCE reflects this *zeitgeist*. Thus, many

of the theory's applications have been in antitrust cases, rather than in studies of internal organisation. Further, TCE has arguably evolved over time from a general theory for understanding industrial organisation to a tool for primarily analysing vertical integration. For example, Shelanski and Klein (1995) surveyed the empirical transaction-cost-economics literature; out of 118 journal articles published between 1976 and 1994, 87 (74 per cent) related to vertical integration, make/buy decisions, or hybrid forms of vertical integration.[5] Williamson's introductory overview of TCE in the *Handbook of Industrial Organization* (1989, 150) called vertical integration the paradigm problem of TCE. This research breaks with that tradition by looking at the firm as a whole, rather than its vertical integration characteristics.

Limits of firm size are also a real and difficult problem for business executives. The cost of suboptimal size—that is, a firm that is too large—is probably significant. For example, up to 25 per cent (Riahi-Belkaoui 1994, 35-64) of the cost of goods sold of a large manufacturing firm can be attributed to organisational slack, often embedded in communication problems, bureaucratic inefficiencies and other diseconomies of scale discussed in detail in Chapter 3. Moreover, large firms have a tendency

[5] Shelanski and Klein claimed that vertical integration research has declined as a share of the total over time, but a categorisation by year shows that the share is stable or may in fact have increased. 1976-1979: 5 articles, 40 per cent vertical integration; 1980-1984: 26 articles; 73 per cent vertical integration; 1985-1989: 53 articles, 72 per cent vertical integration; 1990-1994: 34 articles, 82 per cent vertical integration.

slowly to decline and disappear (Hannah 1996, 1). Shedding light on why this is the case may be socially and privately beneficial, Hannah pointed out, because "we have made great strides in storytelling, but a clearer, surer recipe for sustained success for large corporations has remained elusive" (p. 24).

2.2 DIMENSIONS OF FIRM SIZE

This section defines size and shows the trends in firm size in the US manufacturing sector. Large manufacturing firms in the US have shrunk relative to the total manufacturing sector and the economy as whole over the last 20 to 25 years, while overall industry concentration has been rather stable over the last 100 years. Applying the survivor principle (see p. 15, above), this implies that there are indeed limits to firm size.

2.2.1 Definition of the Firm

To begin with, there are a number of definitions of what a firm is. The *first*, based on Coase (1937, 389), Penrose ([1959] 1995, 15), and Arrow (1964, 403; 1974, 33) holds that the boundary of the firm is where the internal planning mechanism is superseded by the price mechanism. That is, the firm's border is at the point where transactions are regulated by the market rather than by administration. In most cases this means that the

operating firm is equivalent to the legal corporation. An important, if rare, exception is a corporation in which divisions are totally self-contained profit centres. In this case the parent company is not a firm, because the company's divisions by definition trade between themselves through market-based transfer prices.

The *second* definition is that ownership sets a firm's boundaries (e.g., Hart 1995, 5-8). With this definition, a firm is the combination of activities for which the bearers of residual risk are one and the same. One problem with this definition is that employees are not "owned", so they therefore would not be considered part of the firm. Another issue is how units such as a partly-owned subsidiary should be treated. For example, General Motors Corporation owned 82 per cent of Delphi Automotive Systems in early 1999, but Delphi would not be viewed as part of General Motors under the above definition. Still, this definition is quite similar to Coase's because employment contracts can be viewed as temporary ownership claims, and partial ownership is still uncommon even though alliances and carve-outs have grown in popularity.

A *third* definition sees the firm as a network (Richardson 1972, 884-887). McDonald's Corporation, for example, extends far beyond its corporate ownership, because it also consists of a network of thousands of

franchisees over whom McDonald's have a high degree of contractual control (Rubin 1990, 134-144).[6]

The *fourth* definition is based on the firm's sphere of influence. This includes distributors, alliance partners, first- and second-tier suppliers, and so on (Williamson 1985, 120-122). Toyota Motor Corporation, for example, directly employed 215,000 people in 2000, but its sphere of influence probably extended over more than one million people.

In all four cases, it is theoretically somewhat difficult to draw the boundaries of the firm and to distinguish the firm from the whole economy. Nevertheless, it is, to use the words of Kumar, Rajan and Zingales (1999, 10), possible to create an "empirical definition". For the purposes of this paper, the firm is defined as having commonly owned assets—the ownership definition—but employees are also treated as part of the firm. This definition relates closely to Hart's definition (1995, 7), and publicly available data builds on it. It is also commonly used in research (Kumar, Rajan and Zingales 1999, 11). Thus, a firm is an incorporated company (the legal entity) henceforth.

---

[6] 18,265 at the end of 1999.

2.2.2 Definition of Size

There are various ways to measure the size of a firm. Size is most often defined as annual revenue, especially by the business press. However, this measure is basically meaningless because it tells nothing about the depth of the underlying activity. Based on this measure, the world's four largest companies were Japanese trading houses in 1994 (Fortune 1995b). They had between 7,000 and 80,000 employees, but almost no vertical integration.

A better measure of size is value added, which is more or less equivalent to revenue less externally purchased products and services. This metric gives a precise measure of activity, but it is usually not publicly available for individual firms.

Number of employees is the most widely used measure of size. A review by Kimberley claims that more than 80 per cent of academic studies use this measure (1976, 587). In line with Child's observation (1973, 170) that "it is people who are organized", it is not surprising that the number of employees is the most used metric for measuring firm size.

Finally, assets can define size (e.g., as described by Grossman and Hart 1986, 693-694). As with revenue, this measure may not reflect underlying activity; but for manufacturing firms, asset-to-value-added ratios are fairly

homogeneous. Asset data for individual firms are usually available back to the 1890s and are therefore a practical measure in longitudinal studies.

In sum, the best measures of size are value added and number of employees, although assets can be used in certain types of studies. This research uses number of employees as the size metric because the data is available and diseconomies of scale should be associated with human frailties. Moreover, this research deals with bureaucratic failure, which in the end is the result of coordination costs. Such costs are best measured in relation to number of employees (Kumar, Rajan and Zingales 1999, 12).

The definitions are summarised in Table 2 with the suitability for the research at hand indicated by the shadings, ranging from high (black) to low (white).

Table 2. Definition of the Firm and Firm Size

DEFINITION OF THE FIRM AND FIRM SIZE				
	Firm Definition			
Size Metric	Internal Planning (Coase)	Ownership	Network	Sphere of Influence
Revenue				
Value Added				
Employees				
Assets				

2.3 TRENDS IN FIRM SIZE

The US economy is the basis for the analysis in the current research because it is large, fairly homogenous and transparent, and it has a high level of competition between firms. Within this economy, the research focuses on the manufacturing sector.[7]

Large manufacturing firms play a major role in the US economy. The Fortune industrial 500 companies controlled more than 50 per cent of corporate manufacturing assets and employed more than eleven million people in 1994, the last year for which the Fortune industrial ranking was compiled (Fortune 1995a). Their sphere of influence was approximately 40 million employees out of a total private sector workforce of 123 million. Contrary to popular belief, however, the importance of large firms is not increasing and has not done so for many years. Studies show that large manufacturing firms are holding steady as a share of value added since circa 1965 (Scherer and Ross 1990, 62). Their share of employment in the manufacturing sector has declined from around 60 per cent (1979) to around 50 per cent (1994). Moreover, as a share of the total US economy, they are in sharp decline. Large manufacturing firms employed 16 million people in 1979 versus 11 million in 1994 (Fortune 1995a, 185), while private

[7] Alternative approaches would be to study the global manufacturing sector, the total US private sector, or both. However, statistics on the global manufacturing sector are not reliable, and the non-manufacturing sectors are often highly regulated.

sector employment grew from 99 to 123 million people (Council of Economic Advisers 1998, 322) over the same time period.

Further evidence that large firms do not increasingly dominate the economy is available from a number of historical studies. Aggregate industry concentration has changed little since the early part of the last century.[8] Nutter (1951) studied the concentration trend between 1899 and 1939 and found no signs of increased aggregate concentration during this period, mainly because new, fragmented industries emerged, while older ones consolidated (pp. 21, 33). Bain (1968) found the same trend between 1931 and 1963, but with less variability between industries. Scherer and Ross (1990, 84) used Nutter's methodology and showed that aggregate concentration increased slightly, from 35 per cent in 1947 to 37 per cent in 1982. Similarly, Mueller and Hamm (1974, 512) found an increase in four-firm concentration from 40.5 per cent to 42.6 per cent between 1947 and 1970, with most (70 per cent) of the increase between 1947 and 1963.

Bain (1968, 87) calculated that the assets controlled by the largest 200 nonfinancial firms amounted to about 57 per cent of total nonfinancial assets in 1933.[9] He also estimated that the 300 largest nonfinancial firms

---

[8] Note that there have been significant changes within individual industries.

[9] A similar study by Berle and Means ([1932] 1991) has been partly discredited. For example, Scherer and Ross (1990, 60) found that Berle and Means, based on the "meager data then available,...overestimated the relative growth of the largest enterprises".

accounted for 55 per cent of nonfinancial assets in 1962. The largest 200 firms therefore accounted for approximately 50 per cent of nonfinancial assets in 1962 (using the current researcher's estimate of the assets controlled by the 100 smallest firms in the sample). This researcher's data showed that the top 200 nonfinancial firms controlled less than 50 per cent of the total nonfinancial assets in 1994. Adelman (1978) observed a similar pattern when he studied the 117 largest manufacturing firms between 1931 and 1960. He found that concentration was the same at the beginning and at the end of the period (45 per cent). He concluded that "overall concentration in the largest manufacturing firms has remained quite stable over a period of 30 years, from 1931 to 1960". Allen (1976) updated Adelman's number to 1972 and reached the same conclusion. The current research replicated the analysis for 1994 and found the same concentration number to be 45 per cent. Both sets of longitudinal data indicate that large firms represent a stable or declining fraction of the manufacturing sector.

Finally, Bock (1978, 83) studied the share of value added contributed by the largest manufacturing firms between 1947 and 1972. There was a large increase between 1947 and 1954, and a further slight increase until 1963. Between 1963 and 1972, there was no increase. Scherer and Ross (1990, 62) confirmed the lack of increase through the end of the 1980s. Sutton

(1997, 54-55) reached a similar conclusion in a comparison of concentration in the US manufacturing sector between 1967 and 1987.

As for the future, the stock market does not expect the largest firms to outperform smaller firms. The stock market valuation of the largest firms, relative to smaller firms, has declined sharply between 1964 and 1998 (Farrell 1998). In 1964 the largest 20 firms comprised 44 per cent of total stock market capitalisation in the United States; in 1998 they accounted for 19.5 per cent. Market value primarily reflects future growth and profit expectations, and thus the market is increasingly sceptical of large firms' ability to compete with smaller firms. This could be due to industrial evolution, but if it is assumed that diseconomies of scale do not exist, then the largest 20 firms should presumably be able to compensate for a relative decline in their mature businesses by effortlessly growing new businesses.

A study of firms on the New York stock exchange (Ibbotson Associates 1999, 127-143) similarly showed that small firms outperformed large firms between 1926 and 1998. The total annual shareholder return over the period was 12.1 per cent for the largest size decile and 13.7 per cent for the second largest size decile. It increased steadily to 21.0 per cent for the smallest size decile (p. 129). The real return to shareholders after adjustment for risk (using the capital asset pricing model) was -0.28 per cent for decile 1, +0.18 per cent for decile 2 and rising steadily to +4.35 per

cent for decile 10 (p. 140). Note, however, that market capitalisation was used as the definition of size in this study.

The above evidence shows that concentration in the manufacturing sector—defined as the share of value added, employment, assets or market capitalisation held by large firms—has changed little or has declined over much of the last century. The size of large manufacturing firms has kept pace with the overall growth of the manufacturing part of the economy since the 1960s in value-added terms, but has declined in employment terms since 1979 (and has declined relative to the total US corporate sector and the global corporate sector). This indicates that there is a limit to firm size and that this limit may be decreasing in absolute terms, all of which supports the research findings of this paper.

The next chapter explores these limits of firm size through a review of the relevant literature. A theoretical framework is constructed based on transaction cost economics, and the literature is surveyed to validate the framework.

3. LITERATURE REVIEW

The literature review is divided into two parts. The first part defines the theoretical framework and discusses the transaction-cost-economics literature relating to the framework. The second part examines the evidence in transaction cost economics and other fields which supports (and occasionally contradicts) the theoretical framework. The chapter shows that a robust theoretical framework can be constructed based on transaction cost economics, and that the theoretical and empirical literature is congruent with this framework.

3.1 THEORETICAL FRAMEWORK

Transaction cost economics focuses on the boundary of the firm (Holmström and Roberts 1998, 73; Williamson 1981, 548)—that is, the distinction between what is made internally in the firm and what is bought and sold in the marketplace. The boundary can shift over time and for a number of reasons, and the current research looks at one aspect of these shifts. As firms internalise transactions, growing larger, bureaucratic diseconomies of scale appear. Thus, a firm will reach a size at which the benefit from the last internalised transaction is offset by bureaucratic failure. Two factors moderate these diseconomies of scale. First, firms can lessen the negative impact of diseconomies of scale by organising activities

appropriately and by adopting good governance practices. Second, the optimal degree of integration depends on the level of asset specificity, uncertainty and transaction frequency.

Coase's article "The Nature of the Firm" (1937) establishes the basic framework. "Limits of Vertical Integration and Firm Size" in Williamson's book *Markets and Hierarchies* (1975) suggests the nature of size limits. "The Limits of Firms: Incentive and Bureaucratic Features" in Williamson's book *The Economic Institutions of Capitalism* (1985) expands on this theme and explains why the limits exist.[10] Riordan and Williamson's article "Asset Specificity and Economic Organization" (1985) augments the theoretical framework presented here by combining transaction costs with neoclassical production costs. The remainder of the section discusses the details of the argument.

3.1.1 Reasons for Limits

Coase's paper on transaction costs (1937) is the foundation of the New Institutional Economics branch of industrial organisation. Coase asked two fundamental questions "Why is there any organisation?" (p. 388) and "Why is not all production carried on by one big firm?" (p. 394). He

[10] Published earlier by Williamson in a less-developed form (1984).

answered these questions by emphasising transaction costs, which determine what is done in the market—where price is the regulating mechanism, and what is done inside the firm—where bureaucracy is the regulator. Coase pointed out that "the distinguishing mark of the firm is the supersession of the price mechanism" (p. 389). To Coase, all transactions carry a cost, whether it is an external market transaction cost or one that accrues from an internal bureaucratic transaction. "The limit to the size of the firm would be set when the scope of its operations had expanded to a point at which the costs of organizing additional transactions within the firm exceeded the costs of carrying out the same transactions through the market or within another firm" (Coase 1993b, 48).

According to Coase, the most important market transaction costs are the cost of determining the price of a product or service; the cost of negotiating and creating the contract; and the cost of information failure. The most important internal transaction costs are associated with the administrative cost of determining what, when and how to produce; the cost of resource misallocation, because planning will never be perfect; and the cost of lack of motivation on employees' parts, given that motivation is lower in large organisations. In any given industry, the relative magnitude of market and internal transaction costs will determine what is done where.

Coase thus created a theoretical framework which potentially explains why firms have size limits. However, this is only true if there are diminishing returns to management within the firm (Penrose [1959] 1995, 19). Williamson (1975, 130) later argued that this is the case, asking his own rhetorical question: "Why can't a large firm do everything that a collection of small firms can do and more?" (Williamson 1984, 736). Williamson pointed out that the incentive structure within a firm has to differ from market incentives. Even if a firm tries to emulate the high-powered incentives of the market, there are unavoidable side effects, and the cost for setting up incentives can be high. In other words, combining small firms into a large firm will never result in an entity that operates in the same way as when independent small firms respond directly to the market.

3.1.2 Nature of Limits

Williamson (1975, 126-130) found that the limits of firm size are bureaucratic in origin and can be explained by transaction cost economics. He identified four main categories of diseconomies of scale: atmospheric consequences due to specialisation, bureaucratic insularity, incentive limits of the employment relation and communication distortion due to bounded rationality.

Williamson's categories are similar to those Coase described in 1937. Coase talked about the determination (or planning) cost, the resource misallocation cost and the cost of lack of motivation. Williamson's first and second categories correspond broadly to the determination cost; the third category to the demotivation cost, and the fourth category to the resource misallocation cost. Williamson's categories are, however, more specific and allow for easier operationalisation. The four categories are detailed below:

Atmospheric consequences. According to Williamson (1975, 128-129), as firms expand there will be increased specialisation, but also less commitment on the part of employees. In such firms, the employees often have a hard time understanding the purpose of corporate activities, as well as the small contribution each of them makes to the whole. Thus, alienation is more likely to occur in large firms.

Bureaucratic insularity. Williamson (1975) argued that as firms increase in size, senior managers are less accountable to the lower ranks of the organisation (p. 127) and to shareholders (p. 142). They thus become insulated from reality and will, given opportunism, strive to maximise their personal benefits rather than overall corporate performance. According to Williamson, this problem is most acute in organisations with well-established procedures and rules and in which management is well-

entrenched. The argument resembles that of agency theory (Jensen and Meckling 1976; Jensen 1989), which holds that corporate managers tend to emphasise size over profitability, maintaining excess cash flow within the firm rather than distributing it to a more efficient capital market (a lengthier comparison of agency theory and transaction cost economics appears in Section 3.2.1.3). As a consequence, large firms tend towards organisational slack, and resources are misallocated. If this is correct we would expect, for example, to see wider diversification of large firms and lower profits.

Incentive limits of the employment relation. Williamson (1975, 129-130) argued that the structure of incentives large firms offer employees is limited by a number of factors. First, large bonus payments may threaten senior managers. Second, performance-related bonuses may encourage less-than-optimal employee behaviour in large firms. Therefore, large firms tend to base incentives on tenure and position rather than on merit. Such limitations may especially affect executive positions and product development functions, putting large firms at a disadvantage when compared with smaller enterprises in which employees are often given a direct stake in the success of the firm through bonuses, share participation, and stock options.

Communication distortion due to bounded rationality. Because a single manager has cognitive limits and cannot understand every aspect of a complex organisation, it is impossible to expand a firm without adding hierarchical layers. Information passed between layers inevitably becomes distorted. This reduces the ability of high-level executives to make decisions based on facts and negatively impacts their ability to strategise and respond directly to the market. In an earlier article (1967), Williamson found that even under static conditions (no uncertainty) there is a loss of control. He developed a mathematical model to demonstrate that loss of control is a critical factor in limiting firm size, and that there is no need to assume rising factor costs in order to explain such limits (pp. 127-130). His model showed that the number of employees can not expand indefinitely unless span of control can be expanded indefinitely. Moreover, he applied data from 500 of the largest US firms to the model, showing that the optimal number of hierarchical levels was between four and seven. Beyond this, control loss leads to "a static limit on firm size" (p. 135).

Williamson pointed out a number of consequences for these four diseconomies of scale.[11]

[11] Williamson's descriptions are confusing. They are scattered throughout the chapters referenced, inserted between theory and examples. The consequences discussed here are this researcher's attempt to clarify Williamson's descriptions.

- Large firms tend to procure internally when facing a make-or-buy decision (1975, 119-120).

- They have excessive compliance procedures and compliance-related jobs tend to proliferate. Thus, policing costs, such as the cost of audits, can be disproportionately high (1975, 120).

- Projects tend to persist, even though they clearly are failures (1975, 121-122).

- Information is often consciously manipulated to further individual or sub-unit goals (1975, 122-124).

- Asset utilisation is lower because high-powered market incentives do not exist (1985, 137-138).

- Transfer prices do not reflect reality, and cost determination suffers (1985, 138-140).

- Research and development productivity is lower (1985, 141-144).

- Large firms often operate at a suboptimal level by trying to manage the unmanageable, forgiving mistakes, and politicising decisions (1985, 148-152).

Table 3 outlines the links between limiting factors and the consequences listed above.

Table 3. Links between Limiting Factors and Consequences

LINKS BETWEEN LIMITING FACTORS AND CONSEQUENCES				
	Factor			
Consequence	Communication Distortion	Bureaucratic Insularity	Atmospheric Consequences	Incentive Limits
Internal procurement		Strong	Moderate	Strong
Excessive compliance procedures	Strong	Strong	Strong	Strong
Project persistence		Strong	Strong	Moderate
Conscious manipulation of information	Strong	Strong		
Low asset utilisation	Strong		Strong	
Poor internal costing	Strong			Strong
Low R&D productivity	Strong	Moderate	Strong	Strong
Dysfunctional management decisions	Moderate	Strong	Strong	

Each of the factors which limit size appears to have several negative consequences for firm performance. Given the strength of many of these links, it is plausible to assume that a large firm will exhibit lower relative

growth and profitability than a smaller firm with the same product and market mix.

3.1.3 Economies of Scale

Transaction cost economics does not usually deal with economies of scale, which are more often associated with neoclassical production costs. However, Riordan and Williamson (1985) made an explicit attempt to reconcile neoclassical theory and transaction cost economics and showed, among other things (see also pp. 44-45, below), that economies of scale are evident in both production costs (p. 371) and transaction costs (p. 373), and that both can be kept internal to a firm if the asset specificity is positive. That is, the economies of scale can be reaped by the individual firm and are not necessarily available to all participants in a market (pp. 367-369).

3.1.4 Moderating Influences on Firm-Size Limits

While the four categories relating to diseconomies of scale theoretically impose size limits on firms, two moderating factors tend to offset diseconomies of scale: organisation form and degree of integration. Both are central to transaction cost economics, and in order to test the validity of the diseconomies-of-scale argument, it is necessary to account for these factors.

Organisation form. Williamson (1975, 117) recognised that diseconomies of scale can be reduced by organising appropriately. Based on Chandler's pioneering work (e.g., 1962) on the evolution of the American corporation, Williamson argued that the M-form organisation lowers internal transaction costs compared to the U-form organisation.[12] It does so for a key reason: The M-form allows most senior executives to focus on high-level issues rather than day-to-day operational details, making the whole greater than the sum of its parts (p. 137). Thus, large firms organised according to the M-form should perform better than similar U-form firms.

Degree of integration. Williamson showed that three factors play a fundamental role in determining the degree of integration: *uncertainty*, *frequency of transactions* and *asset specificity,* under conditions of bounded rationality (Simon [1947] 1976, xxvi-xxxi) and opportunism (Williamson 1993).

High uncertainty, such as business-cycle volatility or rapid technological shifts, often leads to more internal transactions; it is difficult and prohibitively expensive to create contracts which cover all possible outcomes. Thus, with higher uncertainty, firms tend to internalise activities. In addition, if the transactions are frequent they tend to be

[12] Often referred to as "functional organisation" by other authorities, including Chandler.

managed internally because the repeated market contracting cost usually is higher than the internal bureaucratic cost.

While uncertainty and frequency play some role in creating transaction costs, Williamson considered asset specificity the most important driver of integration (e.g., Riordan and Williamson 1985, 366). Asset specificity is relatively independent of the other factors that affect firm-size limits (p. 368), and therefore the current research focuses on it.

With high asset specificity, market transactions become expensive. Asset specificity refers to physical, human, site, or dedicated assets (Williamson 1985, 55), which have a specific use and cannot easily be transferred.[13] Opportunistic behaviour can be expected if the asset is part of a market transaction. For example, a supplier invests in specific tooling equipment dedicated to one customer. Over time, the customer will be able to put pressure on the supplier because the supplier has no alternative use for the investment. The supplier ultimately lowers its price to the variable cost of production in order to cover fixed costs. But by owning the asset, a firm's incentive to cheat disappears, and the cost of creating contractual safeguards is reduced (Williamson 1985, 32-35).

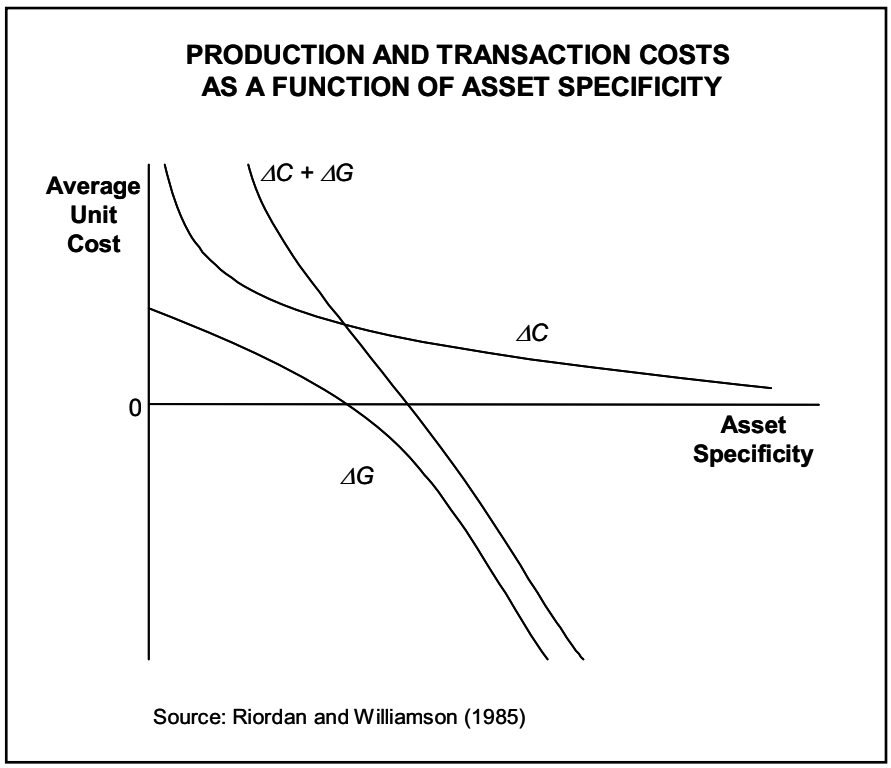
[13] Williamson (1996, 59-60) added brand name capital and temporal specificity.

Neoclassical production costs also exhibit diseconomies as a function of asset specificity (Riordan and Williamson 1985, 369):

> The diseconomies are arguably great where asset specificity is slight, since the outside supplier here can produce to the needs of a wide variety of buyers using the same (large scale) production technology. As asset specificity increases, however, the outside supplier specializes his investment relative to the buyer. This is the meaning of redeployability. As these assets become highly unique, moreover, the firm can essentially replicate the investments of an outside supplier without penalty. The firm and market production technology thus become indistinguishable at this stage.

This is illustrated in Figure 4, in which the differential production cost (ΔC) and transaction cost (ΔG) for markets and hierarchies are shown as a function of asset specificity. The curves show that markets have a large production cost advantage when asset specificity is low, but it approaches zero for high asset specificity (ΔC). For transaction costs, the market has an advantage for low asset specificity and a disadvantage for high asset specificity (ΔG).

Figure 4. Production and Transaction Costs as a Function of Asset Specificity



The implication of the asset-specificity argument, from both a transaction- and production-cost perspective, is that firms with high asset specificity will not reach the limits of size as quickly as those with low specificity. Thus, Riordan and Williamson found that "larger firms are more integrated than smaller rivals" (p. 376).

In closing, a framework based on transaction cost economics has been constructed which establishes a rationale for firm-size limits. Four factors—atmospheric consequences, bureaucratic insularity, incentive limits and communication distortion—make it difficult for firms to expand indefinitely. These negative influences can be offset by economies of scale,

and they can be moderated by the choice of an appropriate organisational form and by increasing internal asset specificity. The framework is next tested against the literature.

3.2 EVIDENCE

In general, there exists only limited research on diseconomies of scale. This is somewhat surprising, because many authorities point out that analysing the limits of firm size is critical to our understanding of the modern economy. Fortunately, the relevant literature yields fragments of evidence that not only confirm the existence of diseconomies of scale, but also explicate various features of bureaucratic failure. The composite picture derived from a review of this literature supports the theoretical framework developed in the previous section, and the hypotheses articulated later in the paper (see Chapter 4).

This section begins with a review of the literature relating to diseconomies of scale and a comparison with Williamson's theoretical framework. The following part reviews the various perspectives on the relationship between economies of scale and diseconomies of scale. Next, the section discusses the support in the literature for the moderating factors. The fourth part briefly reviews what impact, if any, the choice of industry has

on a firm's performance. Finally, the literature findings are summarised in a concluding part.

3.2.1 Diseconomies of Scale

The literature relating to firm-size limits does not follow Williamson's categorisation. Thus, the relevant studies are reviewed by general topic and author, covering bureaucracy and its negative effect on size, information loss, agency theory, and employee incentive problems. At the end of the section the arguments are summarised and related back to Williamson's four sources of diseconomies of scale.

3.2.1.1 Bureaucracy: Negative Consequences of Size

A number of sociological studies describe negative consequences of size which correlate well with Williamson's propositions in Section 3.1. Pugh et al. (1969) and Child (1973), among others, showed that size leads to bureaucracy. Large firms are usually highly bureaucratised through formalisation, and to the extent that bureaucracies breed diseconomies, this limits the growth of such firms. Williamson made a similar point: "almost surely, the added costs of bureaucracy are responsible for limitations in firm size" (1996, 266). According to Blau and Meyer the diseconomies of bureaucracy fall into three major categories: (1) excessive

rigidity, (2) conservatism/resistance to change, and (3) perpetuation of social-class differences (1987, 139-161).

Of these, the first one is relevant here because conservatism is essentially a subcategory of rigidity, and social-class differences fall outside the scope of this research. Excessive rigidity appears as organisations formalise work practices through bureaucratic procedures (Merton 1957, 197-200). Problems are solved by adding structure and the firm reaches a point at which the added structure costs more than the problem solved; Blau and Meyer referred to this as the "problem—organisation—problem—more organisation" spiral of bureaucratic growth (1987, 147). These researchers showed that factors external to the firm, such as increased number of customers or number of tasks to be performed, have little to do with increased bureaucracy. In the end, the added policies and procedures of bureaucracy stifle flexibility.

Crozier (1964) also emphasised rigidity as the most important dysfunction of bureaucracy. In fact, he viewed the bureaucratic organisational model as inherently inefficient, especially under conditions of uncertainty. Managers become increasingly insulated from reality, while lower levels of the organisation experience alienation. As Stinchcombe (1965) demonstrated, one consequence of such rigidity is that firms tend to maintain the organisation form they had when they were created.

Pondy (1969) studied administrative intensity in different industries and what causes variations in intensity. He found a positive correlation between size of administration and firm size when he included a measure of ownership-management separation. This is in line with Williamson's notion of bureaucratic insularity: the larger the organisation is, the more managers are shielded from reality, and the more distant the owners are from daily operations.

Using a demographical research approach, Carroll and Hannan (2000, 289-290) argued that older firms exhibit organisational inertia and find it increasingly difficult to adapt to external changes: "…old organizations are disadvantaged compared to younger ones in changing environments. Alternatively, accumulating rules, routines, and structures might simply impose an overhead cost that reduces the efficiency of organizations even in stable environments".

A similar logic based on institutional economics can be found in Olson (1982). His theory holds that as the institutional structure of a country ages, growth-retarding organisations such as an increasingly complex legal system, special-interest groups and nongovernmental watchdog organisations will become increasingly abundant. The theory and empiry specifically predict that older countries with stable institutions will exhibit

lower economic growth (p. 77). If this logic holds for corporations as well, then older firms will experience less growth.

3.2.1.2 Information Loss and Rigidity

A few studies from the firm-as-information-processor school of thought relate to diseconomies of scale. (Several studies within this school relate to the size distribution of firms, but do not discuss the nature of the diseconomies of scale at length. See Sutton (1997, 43-48) and Axtell (1999, 4-5) for summaries.) Arrow (1974) found that employees in large organisations tend to be highly specialised. Thus, coordination through communication becomes increasingly important. Because information flows carry a cost, organisations code (through formal or informal rules) the information available. Coding economises on resources, but it also leads to information loss and rigidity (p. 55). This means (1) that the more hierarchical levels there are, the more information loss or distortion results; and (2) the older the firm is, the higher the rigidity.

Simon ([1947] 1976) made a similar point. Based on his concept of bounded rationality—"human behavior is *intendedly* rational, but only *limited* so" (p. xxviii)—he found that information degrades as communication lines are extended. Geanakoplos and Milgrom (1991) added to this perspective by noting that there are inevitable signal delays

in an organisation. The more hierarchical levels to be traversed, the longer and more frequent the delays are. Summarising the lessons learnt during a career as a corporate executive, Barnard ([1938] 1968) argued that the size of unit organisations is "restricted very narrowly by the necessities of communication" (p. 110) and that "the size of executive organizations is limited generally by the same conditions that govern the size of unit organizations" (p. 112).[14]

Control-loss problems may contribute to diseconomies of scale as well. McAfee and McMillan (1995) argued that people in organisations exploit information asymmetries to their advantage (or in Williamson's words (1993), they are opportunistic). Dispersion of knowledge within the organisation combined with individual self-interest make conflict of interest and sub-goal pursuit inevitable. McAfee and McMillan noted, among other things, that efficiency falls as the hierarchy expands, and that "long" hierarchies are not viable in competitive industries (p. 401). Qian (1994), similarly found that in long hierarchies, employees do not contribute with a high level of effort. Employees have incomplete information about their role in the enterprise and thus suffer from a lack of motivation. Moreover, managers will need to monitor employee effort, leading to higher costs and further resistance or lack of commitment.

---

[14] That is, the mechanism which determines how large a department can be, also determines how large the firm can be.

However, Mookherjee and Reichelstein (2001) made the case that long hierarchies, under certain restrictive conditions (p. 4), do not lead to control loss: "provided the required conditions on contracting sequence, verifiability of subcontracts and unlimited liability of intermediate agents hold, our model questions the common notion that larger, more complex hierarchies are less efficient owing to 'control losses' with respect to incentives or coordination" (p. 4). It is unclear, however, whether these conditions are met by real-world firms.

3.2.1.3 Agency Theory

An early version of agency theory argued that very large firms do not strive for profit maximisation. According to Monsen and Downs, such firms need to build "bureaucratic management structures to cope with their administrative problems. But such structures inevitably introduce certain conflicts of interest between men in different positions within them. These conflicts arise because the goals of middle and lower management are different from those of top management. The introduction of these additional goals into the firm's decision-making process also leads to systematic deviations from profit-maximizing behavior" (1965, 222). Monsen and Downs furthermore found that the motives of managers differ from those of owners. Managers tend to maximise personal income, while owners maximise profits. It is

impossible for owners of large firms to control the behaviour of managers. Consequently, profit maximisation does not occur. The outcome is akin to what Williamson labelled bureaucratic insularity.

Silver and Auster (1969) argued that the "divergences of interests within the firm and the costs of dealing with them" (p. 277) mean that "the entrepreneur's time is a limitational factor" (p. 280). Employees typically "shirk their duties unless the employer takes steps to prevent this" (p. 278). As a result, senior executives will have less time for strategising and entrepreneurialism, all other things being equal. Silver and Auster furthermore made two predictions based on this argument: (1) the higher the labour content is of an industry's value added, the sooner the total cost curve will turn up, meaning such industries will be more fragmented; and (2) the more supervision employees require, the lower the industry concentration ratio.

More recently, Jensen has deepened and extended these arguments (e.g., Jensen and Meckling 1976; Jensen 1986, 1988, 1989, 2000). He defined agency costs as the sum of the monitoring expenditures by the principal, the bonding expenditures by the agent, and the residual loss. The magnitude of agency costs depends on a number of factors, including the transparency of the firm's activities and the market for managerial talent. Jensen did not, contrary to Monsen and Downs or Silver and Auster,

explicitly state that agency costs increase with the size of the firm. Jensen demonstrated, however, that managers emphasise firm size over profitability: "Managers have incentives to cause their firms to grow beyond optimal size. Growth increases managers' power by increasing the resources under their control. It is also associated with increases in managers' compensation" (1986, 323). He looked at the profitability of diversified firms, noting that they are less profitable than focused firms.

Agency theory and transaction cost economics are similar in many respects and it is not surprising that the two theories lead to the same conclusions. However, some authorities contend that agency theory is a special case of TCE and thus does not capture all the costs associated with transactions. Specifically, Williamson (1985, 20-21) and Mahoney (1992, 566) argued that agency costs correspond to the *ex post* costs of TCE. Meanwhile, TCE works with both *ex ante* and *ex post* costs.[15] Table 4 compares the two theories.

[15] In contrast, Williamson (1988, 570) argued that agency costs correspond to TCE's *ex ante* costs.

Table 4. Comparison of Agency Costs and Transaction Costs

COMPARISON OF AGENCY COSTS AND TRANSACTION COSTS		
Transaction Costs		**Agency Costs**
Ex ante	***Ex post***	
• Search and information costs • Drafting, bargaining and decision costs • Safeguarding costs	• Monitoring and enforcement costs • Adaptation and haggling costs • Bonding costs • Maladaptation costs	• Monitoring expenditures of the principal • Bonding expenditures by the agent • Residual losses

Other critics have pointed out that agency theory poorly explains the boundaries of the firm (Kumar, Rajan, and Zingales 1999, 5). Hart (1995, 20), for example, noted that "the principal-agent view is consistent with there being one huge firm in the world, consisting of a large number of divisions linked by optimal incentive contracts; but it is also consistent with there being many small, independent firms linked by optimal arm's-length contracts". For that reason, TCE provides a more nuanced foundation for the current research.

3.2.1.4 Employee Incentives and Lack of Motivation

A number of authorities have argued that job satisfaction is lower in large organisations and at large work establishments. Employees in large firms are paid significantly more than those in small firms. The reason often given for this disparity is that higher compensation makes up for a less-satisfying work environment (Brown, Hamilton and Medoff 1990, 29).

Scherer's work (1976) is representative of the extensive research conducted at the establishment level. In a review of the literature, including his own original research, he concluded that worker satisfaction was 30 per cent lower in large establishments[16] compared to small establishments (p. 109). Meanwhile, compensation was more than 15 per cent higher for equivalent job descriptions (p. 119). He argued that because establishment size is correlated to firm size, the effects of alienation in large firms appear to be significant. Later work, sponsored by the Federal Trade Commission in the United States, confirmed these findings (Kwoka 1980).

Brown, Hamilton and Medoff (1990) found that large firms pay a wage premium of 10-15 per cent over small firms when adjustments have been made for other effects such as unionisation and skill levels (p. 42). They did not conclude that this difference is necessarily related to alienation, but regardless of the cause, large firms seem to pay substantially higher wages than smaller ones.

In addition, span-of-control problems make it increasingly costly to extend incentive contracts to employees as firms grow (Rasmusen and Zenger 1990, 69). Thus, large firms favour fixed-wage contracts based on tenure rather than performance and make extensive use of monitoring to control productivity. In contrast, smaller firms link pay and performance closely

[16] More than 500 employees.

(p. 80). As a result, the larger firms have a fairly narrow spread of salaries and do not attract top talent; smaller firms may employ both superior and inferior talent, but they reward individuals accordingly. Rasmusen and Zenger's data strongly supported these conclusions, especially in functions with indivisible work, where success is dependent on joint contributions by several individuals (e.g., in research and development). The closer match between performance and pay in small firms puts large firms at a disadvantage, in line with Williamson's incentive limits as a source of diseconomies of scale. Olson (1982, 31) noted that: "in the absence of selective incentives, the incentive for group action diminishes as group size increases".

A similar argument was made by Axtell (1999), who, based on agent-based computational modelling, found that the number of free riders in a firm grows with firm size and that the limits of firm size are set at the point where the advantages of joint production (i.e., economies of scale) are smaller than the disadvantages of having many free riders in the firms whose work effort cannot be effectively monitored (p. 54): "We have interpreted firm growth and demise as a process in which agents are attracted to high-income firms, these firms grow, and once they become large get over-run with free-riders."

Many authorities point out that R&D productivity is significantly lower in large firms. Cooper (1964) surprised business leaders and academics with his article "R&D Is More Efficient in Small Companies". Based on 25 interviews with managers at large and small firms, he argued that small firms have three to ten times higher productivity in development than large firms. The key reasons: (1) small firms are able to hire better people because they can offer more tailored incentives; (2) engineers in small firms are more cost-conscious; and (3) internal communication and coordination is more effective in small firms. These reasons match three of Williamson's four sources of diseconomies: incentive limits, atmospheric consequences and communication distortion.

Later work has confirmed Cooper's anecdotal evidence both theoretically and empirically. Arrow (1983) demonstrated that large firms will invest suboptimally in development because of information loss, and that small firms have a particular advantage in novel areas of research. Schmookler (1972) found that large firms (more than 5000 employees) trail small firms in the number of patented inventions, the percentage of patented inventions used commercially and the number of significant inventions (p. 39). Yet they spend more than twice the resources per patent (p. 37).

Schmookler listed four reasons for the higher effectiveness and efficiency of small firms in R&D: a better understanding of the problem to be solved,

greater cost-consciousness, a more hospitable atmosphere for creative contributions and superior quality of technical personnel (p. 45). Thus, Schmookler quantified and confirmed Cooper's initial evidence, noting that "big firms tend to provide a haven for the mediocre in search of anonymity" (p. 43). In addition, Zenger (1989, 1994) studied employment contracts in R&D in high technology. He found that organisational diseconomies of scale overwhelm technological economies of scale in R&D. His statistical analysis of Silicon Valley firms showed that small firms attract better talent than large firms, motivate employees to try harder and tend to better tie compensation to performance (1994, 725).

Finally, leading anti-bigness ideologues have provided plenty of anecdotal evidence for such arguments, although they are lacking in formal findings. Peters (1992) supported the notion that R&D is less effective in large organisations. He argued that large firms are massively overstaffed in development and that there is little correlation between size of R&D budget and output, offering several case examples as proof. Brock (1987) argued that bigness retards technological advance because large firms are overly risk averse.

Peters, who since the early 1980s has crusaded against large firms, has discussed diseconomies of scale in several books and articles. His views were summarised in "Rethinking scale" (1992). Peters contended there

that decentralisation is necessary for large firms, but very few are as decentralised as they can and should be. Without decentralisation, they are not adaptable enough to respond to changes in the marketplace: "If big is so damn good, then why is almost everyone big working overtime to emulate small?" (p. 13). Moreover, Peters argued that any firm would be well advised to reduce vertical integration, although he does not offer evidence for why this is true. Overall, he found that the bureaucratic distortions of traditional firms lead to lower profitability and growth. In contrast, successful firms mimic the market as much as possible. These ideas are in line with Williamson's description of firm limits, except for the notion that firms should always reduce vertical integration.

Schumacher (1989, 245) identified the lack of motivation in large organisations as the key disadvantage of size, providing a useful summary: "for a large organisation, with its bureaucracies, its remote and impersonal controls, its many abstract rules and regulations, and above all the relative incomprehensibility that stems from its very size, motivation is the central problem".

3.2.1.5 Reconciliation with the "Limits of Firm Size" Framework

The above observations on diseconomies of scale do not map perfectly to Williamson's four sources of diseconomies of scale. Some are similar to his

sources, others to his outcomes. Table 5 shows that Williamson's framework is strongly supported. The most important contrary evidence is Mookherjee and Reichelstein's finding (2001) that long hierarchies do not necessarily lead to control loss, and Brown, Hamilton and Medoff's discussion (1990) of the reason for labour cost differentials between large and small firms. They noticed the differential, but found no link to motivation.

Table 5. Sources of Limits of Firm Size

SOURCES OF LIMITS OF FIRM SIZE			
Communication Distortion	**Bureaucratic Insularity**	**Atmospheric Consequences**	**Incentive Limits**
Arrow (1974): Specialisation leads to poor communication Arrow (1983): Information loss in R&D Barnard ([1938] 1968): Communication losses Cooper (1964): R&D coordination Geanakoplos and Milgrom (1991): Information signal delays McAfee and McMillan (1995): Lower efficiency Mookherjee and Reichelstein (2001): No control loss under certain restrictive conditions Simon ([1947] 1976): Processing bottlenecks	Blau and Meyer (1987): Excessive rigidity Brock (1987): Risk aversion Carroll and Hannan (2000): Firm age leads to insularity Child (1973): Insularity Crozier (1964): Rigidity Jensen (1986): Firms larger than optimum Merton (1957): Rigidity Monsen and Downs (1965): Different owner/manager objectives Olson (1982): Rigidity Pondy (1969): Increase in administration Pugh et al. (1969): Insularity from reality Schmookler (1972): Understanding market needs in R&D Stinchcombe (1965): Perpetuation of organisation form Williamson (1996): Bureaucratic rigidity	Arrow (1974): Rigidity to change Blau and Meyer (1987): Excessive rigidity Brown, Hamilton and Medoff (1990): Unexplained wage differential Child (1973): Insularity Cooper (1964): R&D cost control Crozier (1964): Alienation Kwoka (1980): Low job satisfaction in large firms Merton (1957): Rigidity Pugh et al. (1969): Insularity from reality Qian (1994): Monitoring costs/inadequate effort levels Scherer (1976): Low job satisfaction in large firms Schmookler (1972): R&D cost consciousness; Climate for innovation Schumacher (1989): Low motivation	Blau and Meyer (1987): Excessive rigidity Cooper (1964): R&D incentives Crozier (1964): Rigidity Peters (1992): Low productivity in R&D Rasmusen and Zenger (1990): Employment contracts Schmookler (1972): Quality of R&D employees Silver and Auster (1969): Limits to entrepreneurship Zenger (1989, 1994): Employment contract disincentives in R&D Williamson (1996): Weaker incentives in bureaucracies

3.2.2 Economies of Scale

This brings us to economies of scale. According to some TCE-authorities (Masten 1982; North and Wallis 1994), these should not be incorporated

into the framework because they are independent of the choice of market or hierarchy, once technological indivisibilities are captured within the firm. That is, economies of scale will be reaped regardless of whether all production is carried out in one firm or in many firms. Thus, the intuitively appealing notion that the existence of economies of scale offsets size disadvantages is, according to these authorities, incorrect. This is at odds with Riordan and Williamson's argument (1985) discussed in Section 3.1.3.

The argument has never been tested directly. However, since the 1950s, extensive research has covered the nature and magnitude of economies of scale in production costs, much of it emanating from the structure-conduct-performance school of thought. This work has been explicated in a number of books, and the findings will only be briefly summarised here. In general, the research shows that economies of scale do not play a major role in explaining firm size.

Bain pioneered this line of research in the 1950s and subsequently revolutionised the study of industry and firm behaviour with his book *Industrial Organization* (1968). "The Rationale of Concentration—Efficiency and Other Considerations" from that book reviews the scale-economies argument. Bain divided the analysis into plant- and firm-level analyses. At the plant level, economies of scale are exploited by specialising the work

force and management, and by using dedicated machinery. Each plant has a minimum optimal scale and beyond this scale few additional economies of scale can be exploited. Bain found that in a study of twenty industries (all within the manufacturing sector), only two (automobiles and typewriters) showed significant economies of scale: "in a preponderance of cases, plant scale curves tend to be at least moderately flat (and sometimes very flat)...in the bulk of cases, then, the relative flatness of plant scale curves virtually diminishes the importance of plant scale economies" (pp. 192-193). In other words, there is scant evidence at the plant level for benefits of size.

At the firm level, Bain's study showed that economies of scale derive from benefits of large-scale management, a large distribution system and purchasing power.[17] He then noted that these firm-level economies of scale are elusive, if they exist at all. His research indicated that "where economies of the multi-plant firm are encountered, they are ordinarily quite slight in magnitude...the unit costs...are typically only 1 or 2 per cent below those of a firm with one plant of minimum optimal scale". Of the twenty industries studied, Bain was able to quantify firm-level economies of scale for twelve industries. Of these twelve industries, none exhibited even moderate scale effects (p. 195).

---

[17] Bain does not mention R&D and marketing, possibly because these functions were less important in the early 1950s.

Bain (1978) later summarised his argument as follows: "It is not true that existing degrees of concentration are adequately explained simply as the result of adjustments to attain maximum efficiency in production and distribution...Industries probably tend to be 'more concentrated than necessary' for efficiency—and the larger firms bigger than necessary" (p. 94).

Scherer and Ross provided an overview of the economies of scale debate in "The Determinants of Market Structure: Economies of Scale" (1990). They underscored that it is difficult to draw simple conclusions about the relationship between size and returns. In general, they found that economies of scale are exhausted at a surprisingly small firm size.[18] In a study of twelve industries, they found that market concentration could not be explained by minimally efficient scale considerations. The largest firms in the twelve industries were between two and ten times larger than economies of scale necessitated. Scherer and Ross argued that to the extent that economies of scale accrue for large firms in those industries, they derive from savings in overhead costs (including R&D and marketing) and fixed costs in tangible assets. The economies of scale in overhead are similar to the governance-cost scale economies discussed by Riordan and Williamson (1985, 373), indicating some support for their proposition.

[18] They made the same argument at the product and plant level.

A number of theoretical studies (Ijiri and Simon 1964; Lucas 1978; Nelson and Winter 1982; Simon and Bonini 1958) have demonstrated that large firms evolve stochastically, regardless of economies of scale, for the simple reason that they beat the competition over time. Losers disappear, and winners grow at differential rates depending on how many times they won and how much time this took. Given this logic, firms are large because they are winners, not because they realise economies of scale. Based on realistic assumptions about industry growth rates, variance in firm profitability and so on, simulations have yielded firm-size distributions similar to those observed in real life. As Ijiri and Simon put it: "the observed distributions are radically different from those we would expect from explanations based on static cost curves...there appear to be no existing models other than the stochastic ones that make specific predictions of the shapes of the distribution" (p. 78).

An empirical test of the stochastic evolution model was carried out by Rumelt and Wensley (1981), who looked at whether high market share led to high profitability, or whether successful firms with high profitability, also achieve high market share. They concluded that "scale economies and/or market power are much less important than stochastic growth processes" (p. 2). Note that the stochastic-growth-process argument also implies that older firms will be more profitable than younger firms. Again,

the older firms which still exist are survivors, while younger firms include both winners and losers.

Finally, Peters argued that economies of scale do not exist any more—if they ever existed. In his words: "technology and brainware's dominance is taking the scale out of everything" (1992, 14). Adams and Brock (1986), in case studies of the steel industry, automotive industry and conglomerates, found no evidence that size leads to production scale economies at the firm level. They claimed that it is "the quintessential myth of America's corporate culture that industrial giantism is the handmaiden of economic efficiency" (p. xiii).

In sum, these studies found only slight scale effects. The evidence in the literature review is therefore inconclusive with regard to the argument made by Riordan and Williamson (1985), that economies of scale offset diseconomies of scale.

3.2.3 Moderating Factors

This section reviews the literature to validate Williamson's moderating factors: organisation form and degree of integration. It also discusses, and dismisses, a third moderating factor: financial synergies. The literature review lends strong support to Williamson's framework.

3.2.3.1 Organisation Form

Chandler has argued, in a series of well-known studies (Chandler 1962, 1977, 1982, 1990, 1992; Chandler and Daems 1980), that large firms evolve from functional structures to multidivisional structures as they grow in size and scope of activities. In Chandler's view, the functional (unitary) form is not able to achieve the necessary coordination to be successful in the marketplace; functional economies of scale are too small to make up for this deficiency. Thus, as firms became more diverse in the early twentieth century they adapted the multidivisional form pioneered by E. I. du Pont de Nemours & Company and General Motors Corporation. This line of reasoning is supported by most authorities, including Peters (1992), who found that decentralisation brings major benefits to large firms. Three important quantitative studies illustrate Chandler's argument:

Fligstein (1985, 385-386) showed that between 1919 and 1979, the number of large firms[19] with the multidivisional form went from none to 84 per cent. He estimated that the spread of the multidivisional form is mainly due to the increase of multi-product strategies, in line with Chandler's argument. Armour and Teece (1978) quantified the difference in profits between functional- and multidivisional-form firms in the petrochemical sector, and summarised as follows: "We find strong support for the M-

[19] The 131 (120) largest manufacturing firms by assets in 1919 (1979).

form hypothesis. In the 1955-1968 period the multidivisional structure significantly influenced (at better than the 99-per cent level) the rate of return on stockholders' equity, raising it on average by about two percentage points...realized by the average functional form firm" (pp. 116-117). Teece (1981) studied eighteen manufacturing industries and two retail industries. He found that the multidivisional form outperformed the functional form by an average of 2.37 percentage points (p. 188). He concluded: "the M-form innovation has been shown to display a statistically significant impact on firm performance" (p. 190). These authorities are typical of the strong support for Williamson's view that organisational structure matters and that correct organisational choices can alleviate the effects of diseconomies of scale.

3.2.3.2 Degree of Integration

There is an extensive literature on vertical and lateral integration based on transaction cost economics and other theories. Mahoney (1989, 1992) and Shelanski and Klein (1995) provide summaries. Two issues are relevant here:

- Do asset specificity, uncertainty and transaction frequency explain the degree of vertical integration?

- Does Williamson's framework extend to integration in general?

Asset specificity has repeatedly been shown to be the primary determinant of vertical integration. A number of empirical studies confirm this (e.g., Masten 1984; Masten, Meehan and Snyder 1989, 1991; Monteverde and Teece 1982; Joskow 1993; Klier 1993; Krickx 1988). Uncertainty and frequency are less important. First, they only contribute to vertical integration in conjunction with asset specificity. Second, the empirical evidence does not hold up well in statistical analyses. Walker and Weber's (1984, 1987) results are typical. They found that volume uncertainty had some impact on the decision to vertically integrate and that technological uncertainty had no impact on vertical integration. Transaction frequency has, unfortunately, not been studied explicitly, perhaps because it is not independent from various types of asset specificity. Piecemeal evidence from other studies suggests that it is even less important than uncertainty when asset specificity is part of the analysis (e.g., Mahoney 1992, 571). Finally, Holmström and Roberts (1998, 79) found that both uncertainty and transaction frequency are less important factors than asset specificity.

As for the second issue, Williamson's framework appears to extend to integration in general. Grossman and Hart (1986) and Teece (1976, 1980, 1982) illustrate the use of TCE in lateral relationships. Asset specificity influences integration from a geographic reach, product breadth, and

vertical depth point of view. Teece (1976) showed that multinational firms only exist because the combination of asset specificity and opportunism leads to moral hazard, which is difficult to contain in market transactions. Without, for example, human asset specificity, a firm could just as easily license its technology to a firm in another country, reaping the benefits of development. Tsokhas (1986) illustrated this in a case study of the Australian mining industry. Other studies have shown that market diversity reduces profitability (e.g., Bane and Neubauer 1981). Thus, there is support for Coase's 1932 view[20] that the distinction between vertical and lateral integration is without value (1993c, 40).

A number of studies of product breadth show that asset specificity plays a major role in explaining the success and failure of diversification. Rumelt (1974) found a strong correlation between profitability and human asset specificity—in this case the degree to which a firm draws on common core skills or resources (pp. 121-127). In two studies of the Fortune 500 list of American firms, he demonstrated that focused firms derive three to four percentage points higher return on capital than highly diversified firms. Subsequent studies "have merely extended or marginally modified Rumelt's (1974) original findings" (Ramanujam and Varadarajan

[20] Letter to Ronald Fowler, 24 March 1932.

1989, 539). In sum, asset specificity seems to explain integration in general, not only vertical integration.

3.2.3.3 Financial Synergies

A potential third moderating influence discussed by Williamson (1986) is the presumably efficient internal capital markets of large firms, which allows them to realise financial synergies. Bhidé (1990), however, refuted this line of reasoning and showed that the improvement in efficiency of external capital markets since the 1960s helps explain the trend away from diversification: "Investor power, which goes along with capital market sophistication, has reduced the ability of managers to preserve an inefficient organizational form". Comment and Jarrell (1995, 82-83) reached the same conclusion based on an exhaustive statistical analysis of two thousand firms listed either on the New York Stock Exchange or on the American Stock Exchange between 1978 and 1989.

There does not appear to be a strong reason to expand Williamson's framework with this moderating influence. For the purposes of the current research the financial synergies are therefore excluded.

3.2.3.4 Reconciliation with the "Limits of Firm Size" Framework

Table 6 summarises the moderating influences on diseconomies of scale. There is again strong support for Williamson's framework. The choice of M-form organisation was found to influence firm performance positively. The determinant of degree of integration has been narrowed down to asset specificity, while uncertainty and transaction frequency were found to be less important. Financial synergies do not, however, moderate diseconomies of scale—at least not in the United States where the external capital markets are relatively efficient.

Table 6. Potential Moderators of Diseconomies of Scale

POTENTIAL MODERATORS OF DISECONOMIES OF SCALE		
M-Form Organisation	**Asset Specificity**	**Financial Synergies**
Armour and Teece (1978): M-form increases ROE Chandler (e.g., 1962), Chandler and Daems (1980): M-form alleviates coordination and control problems Fligstein (1985): Multi-product coordination favours M-form Peters (1992): Decentralisation is critical to firm performance Teece (1981): M-form firms are significantly better performers than U-form firms	Bane and Neubauer (1981): Market diversity reduces profitability Coase (1993c): No distinction between vertical and lateral integration Grossman and Hart (1986), Teece (e.g., 1976): TCE applies to lateral integration Mahoney (1992), Holmström and Roberts (1998): Uncertainty and frequency not important Masten (1984), Masten et al. (1989, 1991), Monteverde and Teece (1982), Joskow (1993), Klier (1993), Krickx (1988): Asset specificity more important than uncertainty and frequency Rumelt (1974): Product diversity reduces asset specificity Teece (1976), Tsokhas (1986): Asset specificity influences geographic reach Walker and Weber (1984, 1987): Volume uncertainty is weak factor	Bhidé (1990): Internal capital markets not efficient Comment and Jarrell (1995): Financial synergies not relevant

3.2.4 Industry Influence

Finally, industry influence is not part of the TCE proposition regarding limits of firm size, except indirectly (e.g., industries with high R&D-intensity should show significant diseconomies of scale because incentive limits are important in such industries). A number of studies have shown

that there is weak correlation between profitability and which industry or industries a manufacturing firm participates in. Schmalensee (1985) suggested methods for disaggregating business-unit performance into industry, corporate-parent and market-share effects. Rumelt (1991) applied the methodology to manufacturing firms and found that industry effects accounted for 8 per cent of the explained variance in profitability. McGahan and Porter (1997) found a 19-per cent industry effect for all sectors of the economy, but only 9 per cent of explained variance in profitability for firms in the manufacturing sector (similar to Rumelt's findings). Thus, industry appears to influence profitability in the non-manufacturing sector, but only slightly in the manufacturing sector. The same appears to be true for firm growth. Hall (1986, 9) found, in an analysis of the relationship between firm growth and size in the US manufacturing sector, that the results were only marginally influenced by the use of industry dummies.

The implication for the current research is that industry influences should not be included as a variable in the analysis.

3.2.5 Conclusion

This literature review indicates that the TCE framework for firm-size limits is fairly robust. Most of the authorities support Williamson's

framework, and the mechanisms behind diseconomies of scale have been validated. The findings regarding economies of scale are somewhat inconclusive. The two transaction cost-based moderating influences on diseconomies of scale have both been validated. M-form firms outperform U-form firms, at least in the manufacturing sector. Asset specificity emerges as the most important driver of both vertical and lateral integration.

Past research indicates that the sources of diseconomies are more important in certain contexts. For example, atmospheric consequences and incentive limits are especially severe in R&D-intense industries. Communication distortion, meanwhile, is most common in diverse firms and volatile industries. It is now possible to assess how important these effects are, as well as how large a firm has to be before the effects materialise. Assessing the importance of effects is at this point necessarily qualitative, based on the collective judgement derived from the literature review for each source of diseconomies of scale, economies of scale and the moderating factors. Table 7 summarises this judgement. "Good/Poor" indicates that if, for example, a firm has no problem with incentive limits, then performance (measured as financial results) will be comparatively good. "Importance" indicates if the effect is strong or weak. The "Impact Size" parameter roughly indicates at what size (number of employees) the

effect sets in. For example, the literature review indicates that the incentive disadvantage in R&D for large firms appears to be strong for firms with more than 500 employees in the R&D function (see p. 59, above).

"Context" shows which types of firms are most sensitive to the effects of diseconomies of scale, economies of scale and the moderating factors.

Table 7. Extended TCE-Based "Limits of Firm Size" Framework

EXTENDED TCE-BASED "LIMITS OF FIRM SIZE" FRAMEWORK							
	Sources of Limits of Firm Size					Moderators	
Financial Results	**Communication Distortion**	**Bureaucratic Insularity**	**Atmospheric Consequences**	**Incentive Limits**	**Economies of Scale**	**Organisation Form**	**Degree of Integration**
Good	Low	Low	Low	Low	High	M-form	High
Poor	High	High	High	High	Low	U-form	Low
Importance	High	Fair	Fair	Fair in general; high in, e.g., R&D	Inconclusive	High	Asset specificity high; uncertainty low; frequency negligible
Impact Size: **Small (<1,000)** **Medium** **Large (>10,000)**	 Strong Strong Strong	 Weak Fair Strong	 Weak Fair Strong	 Weak Strong Strong	 Strong Fair Weak	 Strong Strong Strong	 Strong Strong Strong
Context	Diverse firms; unpredictability	Management/ board relation	R&D-intense	R&D-intense	Overhead-intense		

The table reveals, based on Williamson's framework and the literature review, that all factors (except possibly economies of scale) should have a material influence on the performance of large firms. The following chapter builds on this finding as it translates the framework into five testable hypotheses.

4. THEORETICAL FRAMEWORK AND HYPOTHESES

The previous chapter covered the theoretical and empirical studies—particularly Williamson's categorisation (1975, 117-131) of diseconomies of scale—that inform the current research. This chapter now translates the findings into five hypotheses. In the following, each individual hypothesis is first stated, and then discussed. At the end of the chapter, the hypotheses are summarised and linked.

H_1: Bureaucratic failure, in the form of atmospheric consequences, bureaucratic insularity, incentive limits and communication distortion, increases with firm size

Diseconomies of scale are bureaucratic in nature and are not easily observed. They exist because there are diminishing returns to management and because large firms cannot fully replicate the high-powered incentives that exist in the market—leading to bureaucratic failure, the opposite of market failure. Based on Williamson's categorisation, there are four types of diseconomies of scale: atmospheric consequences due to specialisation, bureaucratic insularity, incentive limits of the employment relation and communication distortion due to bounded rationality. The first hypothesis postulates that these diseconomies of scale increase with firm size.

H_2: Large firms exhibit economies of scale

The theory around economies of scale is logically broken into two parts (**H_2** and **H_4**). The second hypothesis posits that ray marginal cost decreases with firm output. This could be seen as a tautological statement, but as was shown in Chapter 3, large firms do not necessarily benefit from economies of scale. First, some authorities hold that economies of scale are exhausted at relatively small firm sizes and thus the cost curve should be flat for large firms. Second, it could be that economies of scale are available to all participants in a market. Given these two arguments, it is important to test whether economies of scale exist at all. That is, does ray marginal cost decline with increased output? The hypothesis says nothing about whether economies of scale have a material influence on firm performance, which is expressed in the fourth hypothesis.

H_3: Diseconomies of scale from bureaucratic failure have a negative impact on firm performance

The third hypothesis has four sub-hypotheses.

H_{3a}: Atmospheric consequences have a negative impact on the performance of large firms

H_{3b}: Bureaucratic insularity has a negative impact on the performance of large firms

H_{3c}: Incentive limits have a negative impact on the performance of large firms

H_{3d}: Communication distortion has a negative impact on the performance of large firms

As was shown in Chapter 2, the average size of large manufacturing firms in the United States has declined since the 1960s, relative to the total economy. Thus, as large firms have become more productive they have, on average, not been able to compensate fully for the per-unit decline in value added by expanding into new geographic markets (reach), new product areas (breadth), or by increasing vertical integration (depth). In line with Stigler's survivor principle (see p. 15, above) this indicates that the diseconomies of scale have a material, negative influence on firm performance. The four types of diseconomies are exhibited through lower profitability and/or slower growth of the largest firms relative to smaller firms, other things—such as risk—being equal.

H_4: Economies of scale increase the relative profitability of large firms over smaller firms

According to TCE, unit production and transaction costs decrease with increasing scale. However, the benefits of scale may be reaped by all participants in a market, large or small, if the market is efficient. The theoretical framework holds that this is not the case and that most economies of scale will be proprietary to the firm in which they reside. Thus, the hypothesis is that large firms have higher relative profitability than small firms, other things being equal. (Note that the theoretical framework says nothing about whether large firms grow faster than small firms.)

H_5: Diseconomies of scale are moderated by two transaction cost-related factors: organisation form and asset specificity

The fifth hypothesis has two sub-hypotheses.

H_{5a}: Large M-form firms perform better than large U-form firms

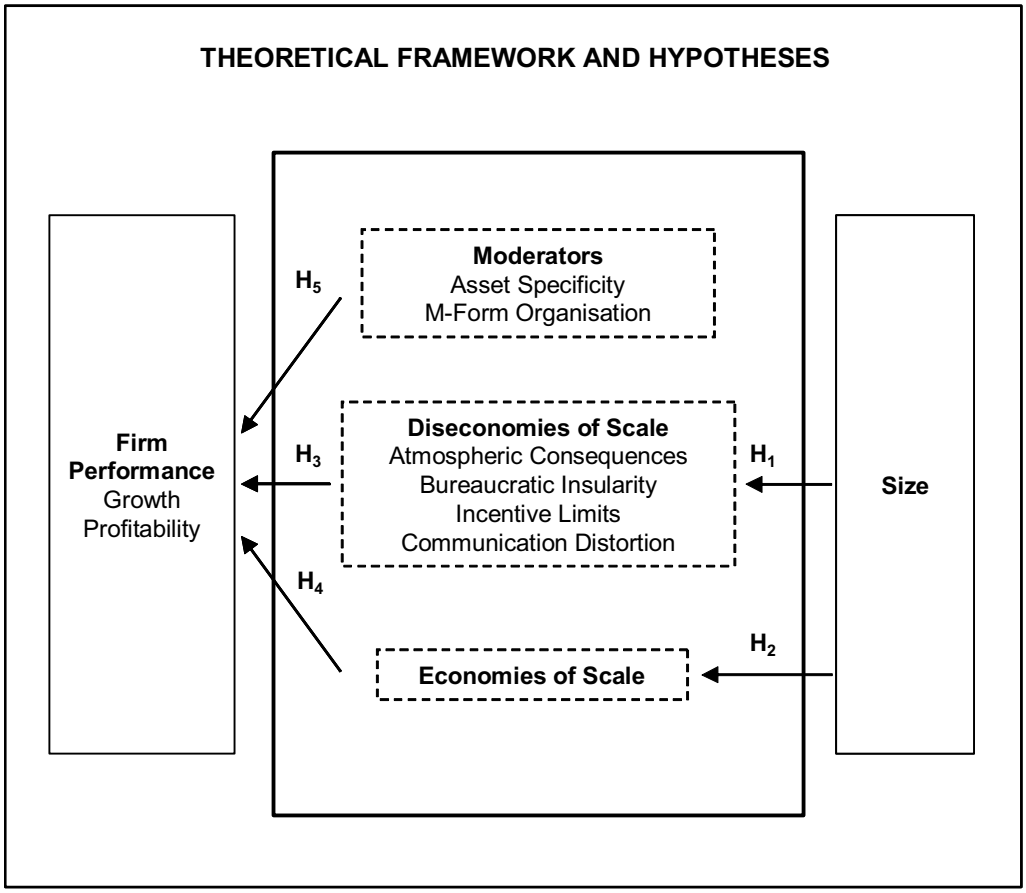
H_{5b}: High internal asset specificity affects a firm's performance positively

The theoretical framework holds that it is possible for firms to moderate the negative impact of diseconomies of scale. Transaction cost economics shows that large firms benefit from multidivisional structures, while

unitary structures impede performance. Moreover, conscious choices about the degree of integration can affect performance. In particular, firms with a high degree of internal asset specificity will outperform those with low internal asset specificity.

In sum, the performance of a firm depends on three influences. To begin with, four size-related factors contribute to diseconomies of scale and determine the firm's size limit. Second, economies of scale increase with firm size. Finally, two factors, M-form organisation and high asset specificity, can moderate the diseconomies of scale. The hypotheses are summarised in Figure 5, which also includes the theoretical framework derived from Williamson (see Section 3.1).

Figure 5. Theoretical Framework and Hypotheses



The question remains: are the hypothesised effects large enough materially to influence the performance of a large firm? Only an empirical analysis, in which the framework and hypotheses are operationalised, will answer this. This empirical analysis is not part of this paper, but can be found in "Bureaucratic Limits of Firm Size: Empirical Analysis Using Transaction Cost Economics" (Canbäck 2002).

The next chapter discusses the findings from the literature and interprets them, without the benefit of an empirical analysis. As such, the findings are highly preliminary.

5. DISCUSSION AND INTERPRETATION

Diseconomies of scale appear to be real. The literature overview discussed the theoretical underpinnings of this paper, indicating that a wide range of theoretical development and empirical research, quantitative and qualitative, supports pieces of the current theoretical predictions.

This chapter summarises and interprets the findings by linking them back to the cost curves discussed in Chapter 2 (pp. 16-18, above). It is shown that the findings are consistent with neoclassical theory and with transaction cost economics. Building on this set of modified cost curves, further implications are discussed, including the relative importance of the various factors that affect a firm's limits. The findings regarding the hypotheses are summarised in Table 8:

Table 8. Summary of Findings

SUMMARY OF FINDINGS[a]	
Hypothesis	**Literature Finding**
H_1: Bureaucratic failure, in the form of atmospheric consequences, bureaucratic insularity, incentive limits and communication distortion, increases with firm size	Confirmed
H_2: Large firms exhibit economies of scale	Confirmed
H_3: : Diseconomies of scale from bureaucratic failure have a negative impact on firm performance	Confirmed
H_{3a}: Atmospheric consequences have a negative impact on the performance of large firms	Confirmed
H_{3b}: Bureaucratic insularity has a negative impact on the performance of large firms	Confirmed
H_{3c}: Incentive limits have a negative impact on the performance of large firms	Confirmed
H_{3d}: Communication distortion has a negative impact on the performance of large firms	Confirmed
H_4: Economies of scale increase the relative profitability of large firms over smaller firms	Inconclusive
H_5: Diseconomies of scale are moderated by two transaction cost-related factors: organisation form and asset specificity	Confirmed
H_{5a}: Large M-form firms perform better than large U-form firms	Confirmed
H_{5b}: High internal asset specificity affects a firm's performance positively	Confirmed
[a] For simplicity, the word "confirmed" is used, although "not rejected" is more accurate.	

As is shown, the theoretical framework is supported by the literature. It is now possible to interpret the findings by returning to the neoclassical cost curves. First, the cost curve shown in Figure 3 is modified to reflect the characteristics of diseconomies of scale, economies of scale and the moderating factors. Second, a similar curve is constructed for firm growth. Third, these two curves are combined to show the overall impact of these two factors on firm performance.

Average cost. To begin with, the elongated U-shaped average total cost curve[21] used in neoclassical theory can be split into two parts: the average production cost curve and the average transaction cost curve. Not much evidence exists for what the relative magnitude of production and transaction costs is. However, Wallis and North (1986) attempted to quantify the relative contribution each type of cost makes to the overall economy. They found that the transaction-cost part of the economy grew from 25 per cent to 50 per cent of gross national product between 1890 and 1970 (p. 121). This suggests that an even split is a reasonable assumption.

The modified cost curves are depicted in a stylised fashion in Figure 6. The top graph shows a curve for average production cost (AC_{P}) consistent with the findings in the current research. One characteristic of the curve is important: the curve has a negative slope for all levels of firm output (Q). This agrees with the view that economies of scale can be kept proprietary to the firms that reap them.

The middle graph in Figure 6 shows the average transaction cost curve (AC_{T}). The negative slope for smaller firms, indicating bureaucratic economies of scale, is supported in the literature review. The positive

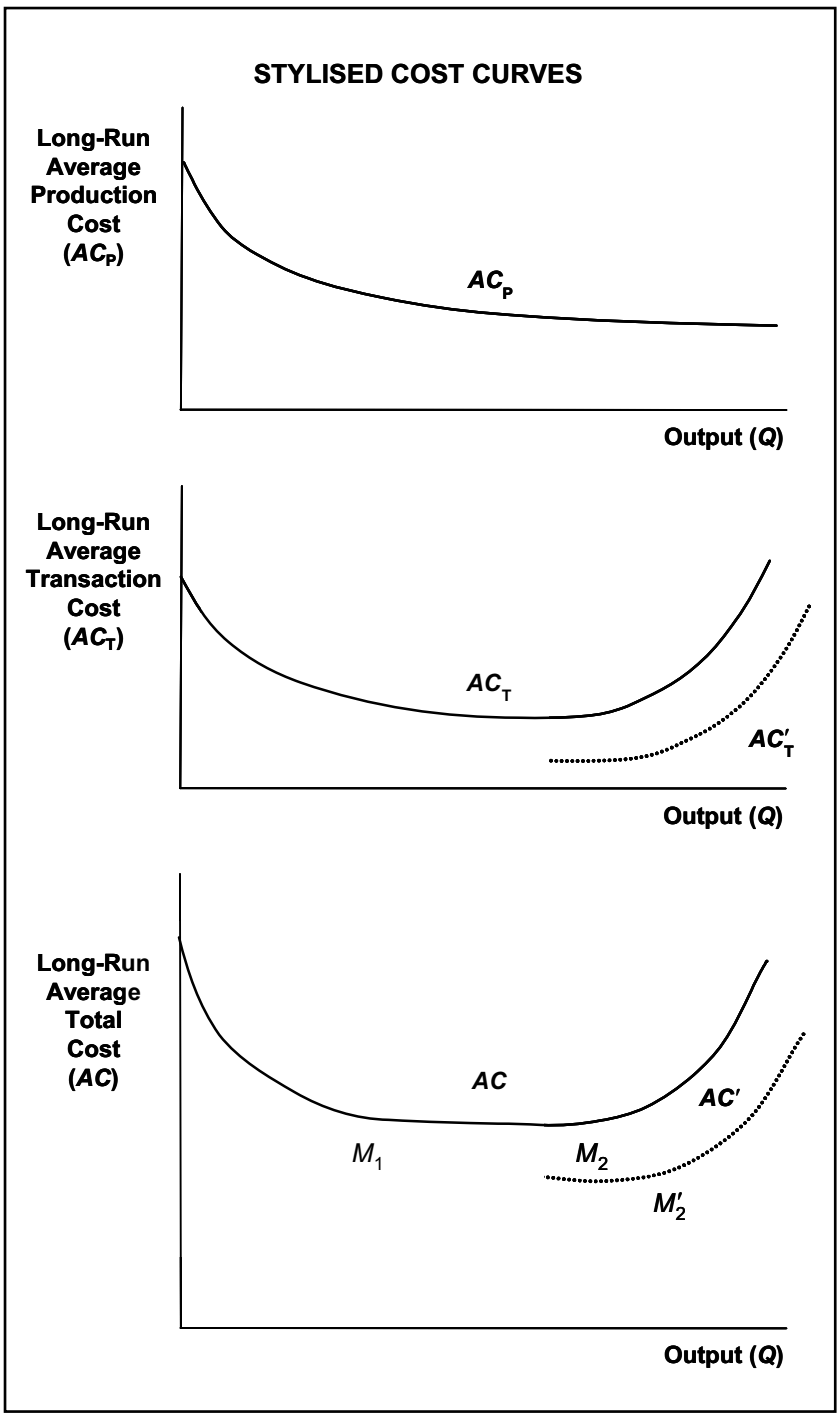
[21] It would be more stringent to talk about ray average total costs because the firms analysed are usually multi-product firms, but simplicity wins.

slope for larger firms, indicating diseconomies of scale and bureaucratic failure, is also supported by the literature.

The middle graph also shows a shifted and slightly tilted average transaction cost curve (AC'_T). The curve reflects the positive contribution from the moderating factors. AC'_T is supported by the literature.

Finally, the bottom graph in Figure 6 shows the average total cost curve (AC), with a shifted curve AC' for the moderators ($AC = AC_P + AC_T$; $AC' = AC_P + AC'_T$). The curve resembles the neoclassical curve in Figure 3. The question now is: where along this curve do firms operate? The literature may suggest that, on average, the largest firms in the sample operate at outputs somewhere close to M'_2 in the upward-sloping region of AC'. That is, they show some diseconomies of scale, but they also benefit from economies of scale and they manage to take advantage of the moderating factors.

Figure 6. Stylised Cost Curves



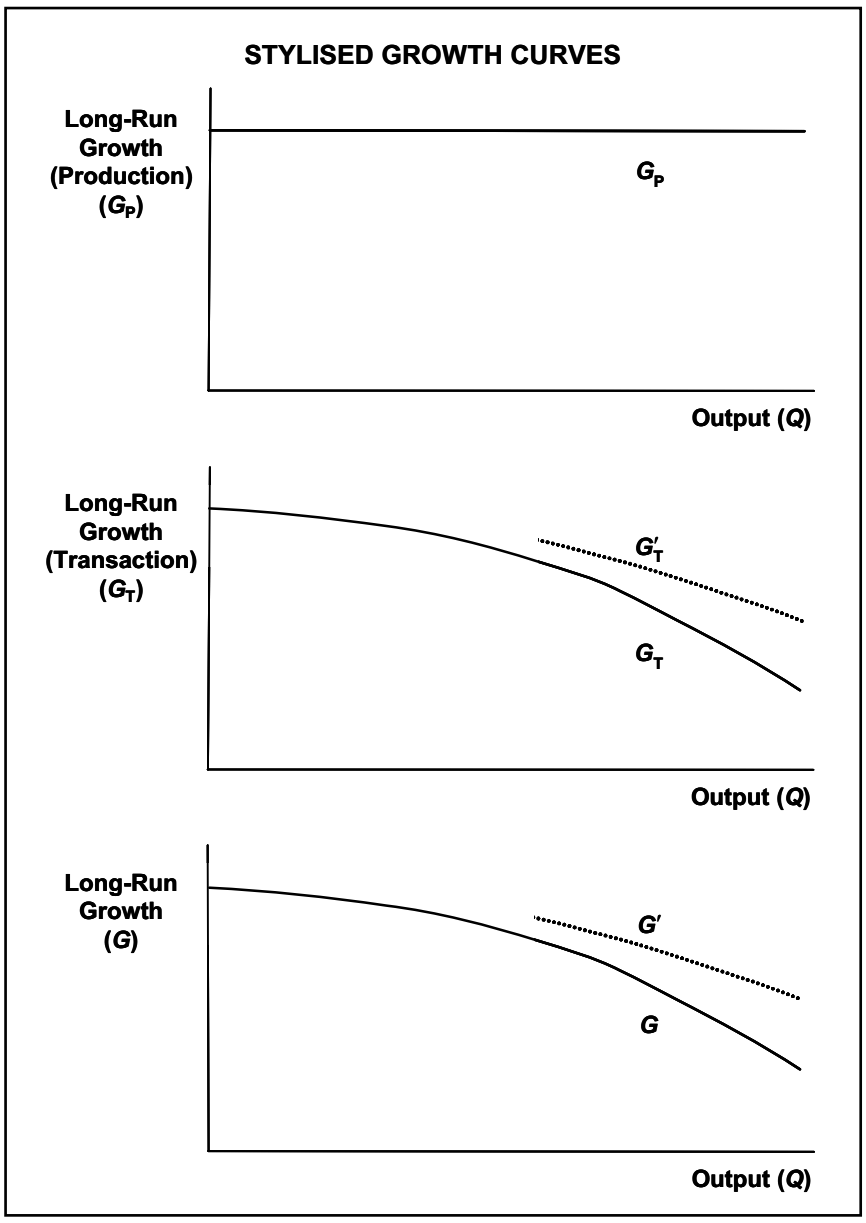
Growth. The underlying logic of the cost curves can also be applied to firm growth. Figure 7 shows the same set of graphs as above for the

relationship between firm growth and output. The top graph illustrates the relationship between growth and output, under the hypothetical assumption that firms only have neoclassical production costs (G_P). The literature does not indicate an influence and thus the graph shows a constant relationship.

The middle graph in Figure 7 portrays the growth curve resulting from bureaucratic, transaction cost-based, failure (G_T). The literature makes it fair to assume that G_T should be monotonously declining for increasing outputs. Again, the moderating influences can shift the curve, which is illustrated by G_T' in the graph.

The bottom graph in Figure 7 convolutes the production- and transaction-cost contributions to growth into overall growth (G). The graph shows that the growth capacity of firms is steadily declining as a function of output, but it can be moderated (G'). Interestingly, this interpretation of the research contradicts Gibrat's law of proportional effects (1931, 74-81), which will be discussed later in this section.

Figure 7. Stylised Growth Curves



Performance. Finally, it is instructive to combine the cost and growth curves to see how they jointly contribute to a firm's performance (Figure 8). Other factors also contribute to firm performance and the graph shows

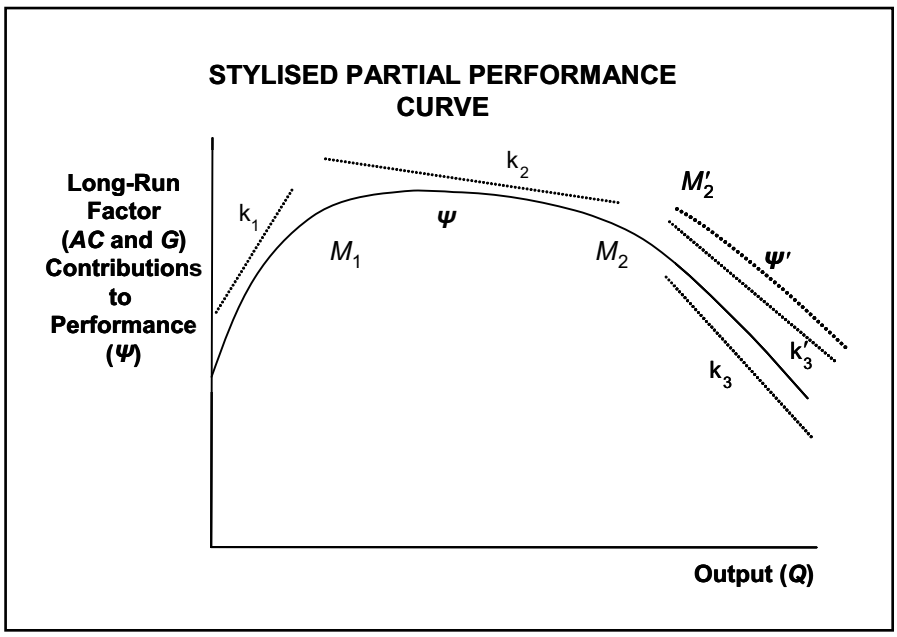
the partial contribution to performance.[22] By convoluting the average total cost (*AC*) and growth (*G*) curves, the partial performance curve Ψ results.[23] Several, perhaps speculative, interpretations can be derived from the graph: (1) Firms operating at small outputs suffer from a lack of economies of scale and this is most likely not compensated for by the higher relative growth achievable by smaller firms. Thus, the slope $k_1 > 0$. (2) There is an area where performance is fairly independent of firm size. On the one hand, economies of scale should lead to steadily lower costs. On the other hand, diminishing growth prospects reduce performance. On balance, the analyses show that $k_2 < 0$, but only slightly so. (3) As diseconomies of scale due to bureaucratic failure set in, the combined negative contribution of increasing transaction costs and lower growth far outweigh economies of scale. Thus, $k_3 < 0$. (4) The moderating factors shift the performance curve outwards from Ψ to Ψ' and $k_3 < k_3' < 0$, while $M_2' > M_2$. That is, if firms judiciously apply the moderating factors, then bureaucratic failure will set in at a larger level of output and the impact from the failure will be less severe. The four interpretations above are supported by the literature review.

---

[22] Total performance (Ψ_{TOT}) is a function of, profitability(π), growth(G), risk(β) and other factors (ε): $\Psi_{TOT} = f(\pi, G, \beta, \varepsilon) = f(TR\text{-}TC, G, \beta, \varepsilon) = f(TR\text{-}AC^{*}Q, G, \beta, \varepsilon)$

[23] The result from this convolution should not be taken for granted, but the statistical analysis showed that *AC* and *G* are reasonably independent and that they should have similar weights.

Figure 8. Stylised Partial Performance Curve



The set of curves discussed above agree well with neoclassical theory (e.g., Panzar 1989) and transaction cost economics (e.g., Williamson 1975), individually. The curves also agree with the joined perspectives on production and transaction costs expressed by, for example, Riordan and Williamson (1985) and Wallis and North (1986). What may make them interesting is the unbundling of the production cost and transaction cost contributions to firm performance, and the attempt to transform the research findings into rough estimates of the shapes of the curves.

6. CONCLUSION

Over the years, I have often been struck by how inefficient and dysfunctional large corporations can be. Yet at the same time most of them are immensely successful and deliver outstanding value to their customers, while they perform well in the stock market. I base these paradoxical comments on my interaction with large corporations, their executives and employees during almost twenty years as a management consultant at McKinsey & Company and Monitor Group. I struggled with the paradox for many years and tried privately to reconcile the advantages and disadvantages of large-scale organisation. In 1991, I happened to come across Coase's "The Nature of the Firm" (1937). After reading a twice-faxed copy of the article on a (slow) bus between the terminal and an airplane at Stuttgart airport, I became convinced that I could use transaction cost economics to improve upon my advice to large corporations, especially when working on strategic and organisational development issues. This in turn led to the ambition to do formal research on the limits of firm size. The research has confirmed many of my real-life observations. Large corporations are inefficient in many ways, but for good reasons. The benefits of large organisations are substantial, but there are inescapable drawbacks as well.

The paper demonstrates the need for research on limits of firm size, creates a framework for thinking about the problem and indicates that there are real and quantifiable diseconomies of size.

The heart of the research is a transaction cost economics-based framework which combines four distinctive aspects of Williamson's theory: (1) the sources of firm-size limits: atmospheric consequences due to specialisation, bureaucratic insularity, incentive limits of the employment relation and communication distortion due to bounded rationality; (2) the impact economies of scale have at the firm level; (3) the importance of organisational form in reducing diseconomies; and (4) the positive influence of high internal asset specificity on both transaction-cost and production-cost diseconomies. The literature survey confirms the explanatory and predictive power of the theory. As such, the research contributes to our understanding of the mechanisms behind bureaucratic failure.

There are a number of real-life implications of the research. First, strategy and structure appear to be intimately linked. Executives at large corporations have to grapple with real trade-offs when they consider expansion. Certain growth strategies are easier to execute than others, and the choice of organisation has major implications for which strategies

make sense. Indeed, structure does not necessarily follow strategy; strategy and structure inform each other continuously and forever.

Second, much of the rationale for mergers and acquisitions seems to be weak, at best. Proponents of mergers typically argue that the resulting larger entity after a merger will realise economies of scale, benefiting customers and shareholders; in addition, they claim that growth will be accelerated through the introduction of new products and services that were previously too expensive to develop. But the analysis here shows that although some economies of scale may be realised, they are likely to be offset by diseconomies of scale. Furthermore, there is no evidence that larger, merged entities innovate more and grow faster. Instead, the opposite appears to be true: innovation and growth declines, on average. This is particularly true in knowledge-intensive industries like pharmaceuticals. To be sure, mergers and acquisitions often do make sense. But executives need to think through how to minimise diseconomies of scale, as well as to maximise moderating influences, when post-merger integration is carried out.

Third, boards of directors may want to emphasise the importance of executive renewal and the elimination of rigid processes to stimulate growth. Old, large firms with entrenched management often find themselves with a fundamental dilemma. There is no indication that they

can achieve above-average, profitable growth. They must choose either to pay out excess cash flow to shareholders (as is often done) or to try to find ways to break the firm's bureaucratic insularity. Maximising the quality of governance, which is part of the board's fiduciary duties, appears to be an important lever for maximising the value of large corporations.

Fourth, firms that strive for high internal asset specificity appear to be better off than those that expand reach, breadth, or depth. This does not imply that single-product or single-geography strategies are optimal (because this reduces growth in the long run), but it does imply that any expansion strategy should strive for high asset specificity and that some firms are best off reducing their scope of activities. By and large, anecdotal and empirical evidence suggests that this has happened over the last 20 to 30 years. "Focus on the core business" and "outsourcing" have been hallmarks of restructuring programs for many years, and the current research verifies that this is often a valid strategy.

Finally, in a world in which companies increasingly try to sell solutions rather than basic products and services, incentive limits have become real and problematic. In businesses that involve team selling or large product-development efforts, attention should be paid to creating well-functioning incentive schemes for employees. The superior productivity of research and development in small firms, in which incentives are tailored to

individual performance, demonstrates why effective incentive schemes matter.

It may be that the average large firm has neither a competitive advantage nor a disadvantage when compared with small and mid-size firms. However, the individual large firm will prosper or fade depending on how well it manages diseconomies of scale.

APPENDIX: LITERATURE REFERENCES PERTAINING TO DISECONOMIES OF SCALE

Adelman, M. A. 1978. Changes in Industrial Concentration. In *Monopoly Power and Economic Performance*, edited by Mansfield, E. New York: Norton.

Arrow, K. J. 1974. *The Limits of Organization*. New York: Norton.

———. 1983. Innovation in Large and Small Firms. In *Entrepreneurship*, edited by J. Ronen. Lexington, Mass.: Lexington Books.

Axtell, R. 1999. The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distributions. Working paper, Washington: Center on Social and Economic Dynamics, Brookings Institution.

Bain, J. S. 1968. *Industrial Organization*. 2d ed. New York: Wiley.

Barnard, C. [1938] 1968. *The Functions of the Executive*. Reprint, with an introduction by K. R. Andrews. Cambridge: Harvard University Press. Original edition, Cambridge: Harvard University Press.

Blau, P. M. and M. W. Meyer. 1987. *Bureaucracy in Modern Society*. 3d ed. New York: McGraw-Hill.

Bock, B. 1978. New Numbers on Concentration: Fears and Facts. In *Monopoly Power and Economic Performance*, edited by E. Mansfield. New York: Norton.

Brock, J. W. 1987. Bigness Is the Problem, Not the Solution. *Challenge* 30 (July-August): 11-16.

Brown, C., J. Hamilton and J. Medoff. 1990. *Employers Large and Small*. Cambridge: Harvard University Press.

Carroll, G. R. and M. T. Hannan. 2000. *The Demography of Corporations and Industries*. Princeton: Princeton University Press.

Child, J. 1973. Predicting and Understanding Organization Structure. *Administrative Science Quarterly* 18: 168-185.

Coase, R. H. 1937. The Nature of the Firm. *Economica* n.s., 4 (16): 386-405.

Cooper, A. C. 1964. R&D Is More Efficient in Small Companies. *Harvard Business Review* 42 (May-June): 75-83.

Crozier, M. 1964. *The Bureaucratic Phenomenon*. Chicago: University of Chicago Press.

Florence, P. S. 1933. *The Logic of Industrial Organization*. London: Kegan Paul.

Geanakoplos, J. and P. Milgrom. 1991. A Theory of Hierarchies Based on Limited Managerial Attention. *Journal of the Japanese and International Economies* 5 (3): 205-225.

Hannah, L. 1996. Marshall's 'Trees' and the Global 'Forest': Were 'Giant Redwoods' Different? Paper read at NBER Conference, 19 October, at Cambridge, Mass.

Ibbotson Associates. 1999. Firm Size and Return. In *Stocks, Bonds, Bills, and Inflation: 1999 Yearbook*. Chicago: Ibbotson Associates.

Jensen, M. C. 1986. Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *American Economic Review* 76: 323-329.

———. 1989. Eclipse of the Public Corporation. *Harvard Business Review* 67 (September-October): 61-74.

Jensen, M. C. and W. H. Meckling. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics* 3: 305-360.

Knight, F. H. [1921] 1964. *Risk, Uncertainty and Profit*. Reprint. New York: Augustus M. Kelley. Original edition, Boston: Houghton Mifflin.

Kwoka, J. E. 1980. Establishment Size, Wages, and Job Satisfaction: The Trade-Offs. In *The Economics of Firm Size, Market Structure, and Social Performance*, edited by J. J. Siegfried. Washington: Federal Trade Commission.

Lucas, R. E., Jr. 1978. On the Size Distribution of Business Firms. *Bell Journal of Economics* 9: 508-523.

McAfee, R. P. and J. McMillan. 1995. Organizational Diseconomies of Scale. *Journal of Economics and Management Strategy* 4 (3): 399-426.

McConnell, J. L. 1945. Corporate Earnings by Size of Firm. *Survey of Current Business* 25 (May): 6-12.

Merton, R. K. 1957. *Social Theory and Social Structure*. Revised and enlarged ed. Glencoe, Ill.: Free Press.

Monsen, R. J. and A. Downs. 1965. A Theory of Large Managerial Firms. *Journal of Political Economy* 73: 221-236.

Mookherjee, D. and S. Reichelstein. 2001. Incentives and Coordination in Hierarchies. *Advances in Theoretical Economics* 1 (1): Article 4. http://www.bepress.com/bejte/advances/vol1/iss1/art4.

Mueller, W. F. and L. G. Hamm. 1974. Trends in Industrial Market Concentration, 1947 to 1970. *Review of Economics and Statistics* 56: 511-520.

Nutter, G. W. 1951. *The Extent of Enterprise Monopoly in the United States, 1899-1939*. Chicago: University of Chicago Press.

Olson, M. 1982. *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities*. New Haven: Yale University Press.

Penrose, E. [1959] 1995. *The Theory of the Growth of the Firm*. 3d ed. New York: Oxford University Press. Original edition, New York: Wiley.

Peters, T. 1992. Rethinking Scale. *California Management Review* 35 (1): 7-29.

Pondy, L. R. 1969. Effects of Size, Complexity, and Ownership on Administrative Intensity. *Administrative Science Quarterly* 14: 47-60.

Pugh, D. S., D. J. Hickson, C. R. Hinings and C. Turner. 1969. The Context of Organization Structures. *Administrative Science Quarterly* 14: 91-114.

Qian, Y. 1994. Incentives and Loss of Control in an Optimal Hierarchy. *Review of Economic Studies* 61: 527-544.

Rasmusen, E. and T. Zenger. 1990. Diseconomies of Scale in Employment Contracts. *Journal of Law, Economics, and Organization* 6 (1): 65-92.

Riordan, M. H. and O. E. Williamson. 1985. Asset Specificity and Economic Organization. *International Journal of Industrial Organization* 3: 365-378.

Robertson, D. H. 1923. *The Control of Industry*. London: Nisbet.

Robinson, A. 1934. The Problem of Management and the Size of Firms. *Economic Journal* 44 (June): 242-257.

Scherer, F. M. 1976. Industrial Structure, Scale Economies, and Worker Alienation. In *Essays on Industrial Organization in Honor of Joe S. Bain*, edited by R. T. Masson and P. D. Qualls. Cambridge, Mass.: Ballinger.

Scherer, F. M. and D. Ross. 1990. *Industrial Market Structure and Economic Performance*. 3rd ed. Boston: Houghton Mifflin.

Schmookler, J. 1972. The Size of Firm and the Growth of Knowledge. In *Patents, Invention, and Economic Change*, edited by Z. Griliches and L. Hurwicz. Cambridge: Harvard University Press.

Schumacher, E. F. 1989. *Small Is Beautiful: Economics as if People Mattered*. Perennial Library ed. New York: HarperCollins.

Silver, M. and R. Auster. 1969. Entrepreneurship, Profit, and Limits on Firm Size. *Journal of Business* 42 (April): 277-281.

Simon, H. A. [1947] 1976. *Administrative Behavior*. 3d ed. New York: Free Press. Original edition, New York: Macmillan.

Stigler, G. J. 1958. The Economies of Scale. *Journal of Law and Economics* 1: 54-71.

———. 1974. The Case against Big Business. In *Monopoly Power and Economic Performance*, edited by Mansfield, E. New York: Norton.

Stinchcombe, A. L. 1965. Social Structure and Organizations. In *Handbook of Organizations*, edited by March, J. G. Chicago: Rand McNally.

Sutton, J. 1997. Gibrat's Legacy. *Journal of Economic Literature* 35 (March): 40-59.

Williamson, O. E. 1967. Hierarchical Control and Optimum Firm Size. *Journal of Political Economy* 75: 123-138.

———. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.

———. 1984. The Incentive Limits of Firms: A Comparative Institutional Assessment of Bureaucracy. *Weltwirtschaftliches Archiv* 120 (4): 736-763.

———. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.

———. 1996. *The Mechanisms of Governance*. New York: Oxford University Press.

Zenger, T. R. 1989. Organizational Diseconomies of Scale: Pooling vs. Separating Labor Contracts in Silicon Valley. PhD Thesis, University of California, Los Angeles.

———. 1994. Explaining Organizational Diseconomies of Scale in R&D: Agency Problems and the Allocation of Engineering Talent, Ideas, and Effort by Firm Size. *Management Science* 40 (6): 708-729.

REFERENCES

Adams, W. and J. W. Brock. 1986. *The Bigness Complex: Industry, Labor, and Government in the American Economy*. New York: Pantheon.

Adelman, M. A. 1978. Changes in Industrial Concentration. In *Monopoly Power and Economic Performance*, edited by E. Mansfield. New York: Norton.

Allen, B. T. 1976. Average Concentration in Manufacturing, 1947-1972. *Journal of Economic Issues* 10 (September): 664-673.

Armour, H. O. and D. J. Teece. 1978. Organizational Structure and Economic Performance: A Test of the Multidivisional Hypothesis. *Bell Journal of Economics* 9 (1): 106-122.

Arrow, K. J. 1964. Control in Large Organizations. *Management Science* 10 (3): 397-408.

———. 1974. *The Limits of Organization*. New York: Norton.

———. 1983. Innovation in Large and Small Firms. In *Entrepreneurship*, edited by J. Ronen. Lexington, Mass.: Lexington Books.

Axtell, R. 1999. The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distributions. Working paper, Washington: Center on Social and Economic Dynamics, Brookings Institution.

Bain, J. S. 1968. *Industrial Organization*. 2d ed. New York: Wiley.

———. 1978. The General Explanation of the Development of Concentration. In *Monopoly Power and Economic Performance*, edited by E. Mansfield. New York: Norton.

Bane, W. T. and F. Neubauer. 1981. Diversification and the Failure of New Foreign Activities. *Strategic Management Journal* 2: 219-233.

Barnard, C. [1938] 1968. *The Functions of the Executive*. Revised ed. Reprint, with an introduction by K. R. Andrews. Cambridge: Harvard University Press. Original edition, Cambridge: Harvard University Press.

Berle, A. A. and G. C. Means. [1932] 1991. *The Modern Corporation and Private Property*. Revised ed. Reprint, with a new introduction by M. L. Weidenbaum and M. Jensen. New Brunswick, N.J.: Transaction Publishers. Original edition, New York: Commerce Clearing House.

Bhidé, A. 1990. Reversing Corporate Diversification. *Journal of Applied Corporate Finance* 3 (2): 70-81.

Blau, P. M. and M. W. Meyer. 1987. *Bureaucracy in Modern Society*. 3d ed. New York: McGraw-Hill.

Bock, B. 1978. New Numbers on Concentration: Fears and Facts. In *Monopoly Power and Economic Performance*, edited by E. Mansfield. New York: Norton.

Brock, J. W. 1987. Bigness Is the Problem, Not the Solution. *Challenge* 30 (July-August): 11-16.

Brown, C., J. Hamilton and J. Medoff. 1990. *Employers Large and Small*. Cambridge: Harvard University Press.

Canbäck, S. 2002. Bureaucratic Limits of Firm Size: Empirical Analysis Using Transaction Cost Economics. DBA thesis, Henley Management College, Brunel University.

Carroll, G. R. and M. T. Hannan. 2000. *The Demography of Corporations and Industries*. Princeton: Princeton University Press.

Chandler, A. D., Jr. 1962. *Strategy and Structure: Chapters in the History of the American Industrial Enterprise*. Cambridge: MIT Press.

———. 1977. *The Visible Hand*. Cambridge, Mass.: Belknap.

———. 1982. Evolution of the Large Industrial Corporation: An Evolution of the Transaction Cost Approach. *Business and Economic History* 2d series, 11: 116-134.

———. 1990. *Scale and Scope: The Dynamics of Industrial Capitalism*. Cambridge, Mass.: Belknap.

———. 1992. The Emergence of Managerial Capitalism. In *The Sociology of Economic Life*, edited by M. Granovetter and R. Swedberg. Boulder, Colo.: Westview Press.

Chandler, A. D., Jr. and H. Daems, eds. 1980. *Managerial Hierarchies: Comparative Perspectives on the Rise of the Modern Industrial Enterprise*. Cambridge: Harvard University Press.

Child, J. 1973. Predicting and Understanding Organization Structure. *Administrative Science Quarterly* 18: 168-185.

Coase, R. H. 1937. The Nature of the Firm. *Economica* n.s., 4 (16): 386-405.

———. 1993a. 1991 Nobel Lecture: The Institutional Structure of Production. In *The Nature of the Firm: Origins, Evolution, and Development*, edited by O. E. Williamson and S. G. Winter. New York: Oxford University Press.

———. 1993b. The Nature of the Firm: Meaning. In *The Nature of the Firm: Origins, Evolution, and Development*, edited by O. E. Williamson and S. G. Winter. New York: Oxford University Press.

———. 1993c. The Nature of the Firm: Origin. In *The Nature of the Firm: Origins, Evolution, and Development*, edited by O. E. Williamson and S. G. Winter. New York: Oxford University Press.

Comment, R. and G. A. Jarrell. 1995. Corporate Focus and Stock Returns. *Journal of Financial Economics* 37 (1): 67-87.

Cooper, A. C. 1964. R&D Is More Efficient in Small Companies. *Harvard Business Review* 42 (May-June): 75-83.

Council of Economic Advisers. 1998. *Economic Report of the President*. Washington: U.S. Government Printing Office.

Crozier, M. 1964. *The Bureaucratic Phenomenon*. Chicago: University of Chicago Press.

Eatwell, J., M. Milgate and P. Newman, eds. 1987. *The New Palgrave: A Dictionary of Economics*. 3 vols. New York: Stockton Press.

Farrell, C. 1998. Does Wharton's Jeremy Siegel Have Alan Greenspan's Ear? *Business Week*, July 13.

Fligstein, N. 1985. The Spread of the Multidivision Form among Large Firms, 1919-1979. *American Sociological Review* 50 (June): 377-391.

Florence, P. S. 1933. *The Logic of Industrial Organization*. London: Kegan Paul.

Fortune. 1995a. The Fortune 500, 1994. *Fortune*, 15 May, 165-227, F1-F64.

———. 1995b. The Fortune Global 500, 1994. *Fortune*, 7 August, 123-136, F1-F42.

Geanakoplos, J. and P. Milgrom. 1991. A Theory of Hierarchies Based on Limited Managerial Attention. *Journal of the Japanese and International Economies* 5 (3): 205-225.

Gibrat, R. 1931. *Les inégalités économiques*. Paris: Librairie du Recueil Sirey.

Grossman, S. J. and O. D. Hart. 1986. The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy* 94 (4): 691-719.

Hall, B. H. 1986. The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector. Working paper, Cambridge, Mass.: NBER.

Hannah, L. 1996. Marshall's 'Trees' and the Global 'Forest': Were 'Giant Redwoods' Different? Paper read at NBER Conference, 19 October, at Cambridge, Mass.

Hart, O. 1995. *Firms, Contracts, and Financial Structure*. Oxford: Clarendon Press.

Holmström, B. and J. Roberts. 1998. The Boundaries of the Firm Revisited. *Journal of Economic Perspectives* 12 (4): 73-94.

Holmström, B. R. and J. Tirole. 1989. The Theory of the Firm. In *Handbook of Industrial Organization, Vol. 1*, edited by R. Schmalensee and R. Willig. New York: North-Holland.

Ibbotson Associates. 1999. Firm Size and Return. In *Stocks, Bonds, Bills, and Inflation: 1999 Yearbook*. Chicago: Ibbotson Associates.

Ijiri, Y. and H. A. Simon. 1964. Business Firm Growth and Size. *American Economic Review* 54 (2): 77-89.

Jensen, M. C. 1986. Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *American Economic Review* 76: 323-329.

———. 1988. Takeovers: Their Causes and Consequences. *Journal of Economic Perspectives* 2 (1): 21-48.

———. 1989. Eclipse of the Public Corporation. *Harvard Business Review* 67 (September-October): 61-74.

———. 2000. *A Theory of the Firm: Governance, Residual Claims, and Organizational Forms*. Cambridge: Harvard University Press.

Jensen, M. C. and W. H. Meckling. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics* 3: 305-360.

Joskow, P. L. 1993. Asset Specificity and the Structure of Vertical Relationships: Empirical Evidence. In *The Nature of the Firm: Origins, Evolution, and Development*, edited by O. E. Williamson and S. G. Winter. New York: Oxford University Press.

Kimberley, J. R. 1976. Organizational Size and the Structuralist Perspective: A Review, Critique and Proposal. *Administrative Science Quarterly* 21: 571-597.

Klier, T. H. 1993. Transaction Cost Theory and Just-in-Time Manufacturing: A New Look at Vertical Integration in the United States Automobile Market. PhD Thesis, Department of Economics, Michigan State University.

Knight, F. H. [1921] 1964. *Risk, Uncertainty and Profit*. Reprint, New York: Augustus M. Kelley. Original edition, Boston: Houghton Mifflin.

Krickx, G. A. M. 1988. Historical Evidence on the Evolution of Vertical Exchange Mechanisms: Examples from the Computer Systems Industry. PhD Thesis, University of California.

Kumar, K. B., R. G. Rajan and L. Zingales. 1999. What Determines Firm Size? Working paper, Cambridge, Mass.: NBER.

Kwoka, J. E. 1980. Establishment Size, Wages, and Job Satisfaction: The Trade-Offs. In *The Economics of Firm Size, Market Structure, and Social Performance*, edited by J. J. Siegfried. Washington: Federal Trade Commission.

Lucas, R. E., Jr. 1978. On the Size Distribution of Business Firms. *Bell Journal of Economics* 9: 508-523.

Mahoney, J. T. 1989. Organizational Rationalization and Innovation: Vertical Integration and Multidivisional Organization. PhD Thesis, University of Pennsylvania.

———. 1992. The Choice of Organizational Form: Vertical Financial Ownership Versus Other Methods of Vertical Integration. *Strategic Management Journal* 13: 559-584.

Mankiw, N. G. 1998. *Principles of Economics*. Fort Worth, Tex.: Dryden.

Marglin, S. A. 1974. What Do Bosses Do? The Origins of Hierarchy in Capitalist Production. *Review of Radical Political Economics* 6: 33-60.

Marshall, A. [1920] 1982. *Principles of Economics: An Introductory Volume*. 8th ed. Philadelphia: Porcupine Press. Original edition, London: Macmillan.

Mas-Colell, A., M. D. Whinston and J. R. Green. 1995. *Microeconomic Theory*. New York: Oxford University Press.

Masten, S. E. 1982. Transaction Costs, Institutional Choice and the Organization of Production. PhD Thesis, University of Pennsylvania.

———. 1984. The Organization of Production: Evidence from the Aerospace Industry. *Journal of Law and Economics* 27 (October): 403-417.

Masten, S. E., J. W. Meehan, Jr. and E. A. Snyder. 1989. Vertical Integration in the US Auto Industry: A Note on the Influence of Transaction Specific Assets. *Journal of Economic Behavior and Organization* 12 (2): 265-273.

———. 1991. The Costs of Organization. *Journal of Law, Economics and Organization* 7 (1): 1-25.

McAfee, R. P. and J. McMillan. 1995. Organizational Diseconomies of Scale. *Journal of Economics and Management Strategy* 4 (3): 399-426.

McConnell, J. L. 1945. Corporate Earnings by Size of Firm. *Survey of Current Business* 25 (May): 6-12.

McGahan, A. M. and M. E. Porter. 1997. How Much Does Industry Matter, Really? *Strategic Management Journal* 18 (Summer): 15-30.

Merton, R. K. 1957. *Social Theory and Social Structure*. Revised and enlarged ed. Glencoe, Ill.: Free Press.

Monsen, R. J. and A. Downs. 1965. A Theory of Large Managerial Firms. *Journal of Political Economy* 73 (3): 221-236.

Monteverde, K. and D. J. Teece. 1982. Supplier Switching Costs and Vertical Integration in the Automobile Industry. *Bell Journal of Economics* 13 (1): 206-213.

Mookherjee, D. and S. Reichelstein. 2001. Incentives and Coordination in Hierarchies. *Advances in Theoretical Economics* 1 (1): Article 4. http://www.bepress.com/bejte/advances/vol1/iss1/art4.

Mueller, W. F. and L. G. Hamm. 1974. Trends in Industrial Market Concentration, 1947 to 1970. *Review of Economics and Statistics* 56: 511-520.

Nelson, R. and S. G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, Mass.: Belknap.

North, D. C. and J. J. Wallis. 1994. Integrating Institutional Change and Technical Change in Economic History: A Transaction Cost Approach. *Journal of Institutional and Theoretical Economics (Zeitschrift für die gesamte Staatswissenschaft)* 150 (4): 609-624.

Nutter, G. W. 1951. *The Extent of Enterprise Monopoly in the United States, 1899-1939*. Chicago: University of Chicago Press.

Olson, M. 1982. *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities*. New Haven: Yale University Press.

Panzar, J. C. 1989. Technological Determinants of Firm and Market Organization. In *Handbook of Industrial Organization, Vol. 1*, edited by R. Schmalensee and R. Willig. New York: North-Holland.

Penrose, E. [1959] 1995. *The Theory of the Growth of the Firm*. 3d ed. New York: Oxford University Press. Original edition, New York: Wiley.

Peters, T. 1992. Rethinking Scale. *California Management Review* 35 (1): 7-29.

Pondy, L. R. 1969. Effects of Size, Complexity, and Ownership on Administrative Intensity. *Administrative Science Quarterly* 14: 47-60.

Pratt, J. W. and R. J. Zeckhauser. 1985. *Principals and Agents: The Structure of Business*. Boston: Harvard Business School Press.

Pugh, D. S., D. J. Hickson, C. R. Hinings and C. Turner. 1969. The Context of Organization Structures. *Administrative Science Quarterly* 14: 91-114.

Qian, Y. 1994. Incentives and Loss of Control in an Optimal Hierarchy. *Review of Economic Studies* 61 (3): 527-544.

Ramanujam, V. and P. Varadarajan. 1989. Research on Corporate Diversification: A Synthesis. *Strategic Management Journal* 10 (6): 523-551.

Rasmusen, E. and T. Zenger. 1990. Diseconomies of Scale in Employment Contracts. *Journal of Law, Economics, and Organization* 6 (1): 65-92.

Riahi-Belkaoui, A. 1994. *Organizational and Budgetary Slack*. Westport, Conn.: Quorum.

Richardson, G. B. 1972. The Organisation of Industry. *Economic Journal* 82 (September): 883-896.

Riordan, M. H. and O. E. Williamson. 1985. Asset Specificity and Economic Organization. *International Journal of Industrial Organization* 3 (4): 365-378.

Robertson, D. H. 1923. *The Control of Industry*. London: Nisbet.

Robinson, A. 1934. The Problem of Management and the Size of Firms. *Economic Journal* 44 (June): 242-257.

Rubin, P. H. 1990. *Managing Business Transactions*. New York: Free Press.

Rumelt, R. P. 1974. *Strategy, Structure, and Economic Performance*. Cambridge: Harvard University Press.

Rumelt, R. P. and R. C. Wensley. 1981. In Search of the Market Share Effect. *Academy of Management Proceedings* 41: 2-6.

Scherer, F. M. 1976. Industrial Structure, Scale Economies, and Worker Alienation. In *Essays on Industrial Organization in Honor of Joe S. Bain*, edited by R. T. Masson and P. D. Qualls. Cambridge, Mass.: Ballinger.

Scherer, F. M. and D. Ross. 1990. *Industrial Market Structure and Economic Performance*. 3rd ed. Boston: Houghton Mifflin.

Schmalensee, R. 1985. Do Markets Differ Much? *American Economic Review* 75 (3): 341-351.

Schmookler, J. 1972. The Size of Firm and the Growth of Knowledge. In *Patents, Invention, and Economic Change*, edited by Z. Griliches and L. Hurwicz. Cambridge: Harvard University Press.

Schumacher, E. F. 1989. *Small Is Beautiful: Economics as if People Mattered*. Perennial Library ed. New York: HarperCollins.

Shelanski, H. A. and P. G. Klein. 1995. Empirical Research in Transaction Cost Economics: A Review and Assessment. *Journal of Law, Economics, and Organization* 11 (2): 335-361.

Silver, M. and R. Auster. 1969. Entrepreneurship, Profit, and Limits on Firm Size. *Journal of Business* 42 (April): 277-281.

Simon, H. A. [1947] 1976. *Administrative Behavior*. 3d ed. New York: Free Press. Original edition, New York: Macmillan.

Simon, H. A. and C. P. Bonini. 1958. The Size Distribution of Business Firms. *American Economic Review* 48 (4): 607-617.

Stigler, G. J. 1958. The Economies of Scale. *Journal of Law and Economics* 1: 54-71.

———. 1974. The Case against Big Business. In *Monopoly Power and Economic Performance*, edited by E. Mansfield. New York: Norton.

Stinchcombe, A. L. 1965. Social Structure and Organizations. In *Handbook of Organizations*, edited by J. G. March. Chicago: Rand McNally.

Sutton, J. 1997. Gibrat's Legacy. *Journal of Economic Literature* 35 (March): 40-59.

Teece, D. J. 1976. *The Multinational Corporation and the Resource Cost of International Technology Transfer*. Cambridge, Mass.: Ballinger.

———. 1980. Economies of Scope and the Scope of the Enterprise. *Journal of Economic Behavior and Organization* 1: 223-247.

———. 1981. Internal Organization and Economic Performance: An Empirical Analysis of the Profitability of Principal Firms. *Journal of Industrial Economics* 30 (2): 173-199.

———. 1982. Towards an Economic Theory of the Multiproduct Firm. *Journal of Economic Behavior and Organization* 3 (1): 39-63.

Tsokhas, K. 1986. *Beyond Dependence: Companies, Labour Processes and Australian Mining*. Melbourne: Oxford University Press.

Walker, G. and D. Weber. 1984. A Transaction Cost Approach to Make-or-Buy Decisions. *Administrative Science Quarterly* 29: 373-391.

———. 1987. Supplier Competition, Uncertainty and Make-or-Buy Decisions. *Academy of Management Journal* 30 (3): 589-596.

Wallis, J. J. and D. C. North. 1986. Measuring the Transaction Sector in the American Economy, 1870-1970. In *Long-Term Factors in American Economic Growth. Studies in Income and Wealth Series*, edited by S. L. Engerman and R. E. Gallman. Chicago: University of Chicago Press.

Williamson, O. E. 1967. Hierarchical Control and Optimum Firm Size. *Journal of Political Economy* 75 (2): 123-138.

———. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.

———. 1981. The Economics of Organization: The Transaction Cost Approach. *American Journal of Sociology* 87 (3): 548-577.

———. 1984. The Incentive Limits of Firms: A Comparative Institutional Assessment of Bureaucracy. *Weltwirtschaftliches Archiv* 120 (4): 736-763.

———. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.

———. 1986. *Economic Organization: Firms, Markets and Policy Control*. New York: New York University Press.

———. 1988. Corporate Finance and Corporate Governance. *Journal of Finance* 43 (3): 567-591.

———. 1989. Transaction Cost Economics. In *Handbook of Industrial Organization, Vol. 1*, edited by R. Schmalensee and R. Willig. New York: North-Holland.

———. 1993. Opportunism and Its Critics. *Managerial and Decision Economics* 14 (2): 97-107.

———. 1996. *The Mechanisms of Governance*. New York: Oxford University Press.

Zenger, T. R. 1989. Organizational Diseconomies of Scale: Pooling vs. Separating Labor Contracts in Silicon Valley. PhD Thesis, University of California, Los Angeles.

———. 1994. Explaining Organizational Diseconomies of Scale in R&D: Agency Problems and the Allocation of Engineering Talent, Ideas, and Effort by Firm Size. *Management Science* 40 (6): 708-729.