No. 2092

THE GROWTH OF FIRMS
IN THEORY AND IN PRACTICE

Paul A Geroski

INDUSTRIAL ORGANIZATION
THE GROWTH OF FIRMS
IN THEORY AND IN PRACTICE

Paul A Geroski

Discussion Paper No.2092
March 1999

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: http://www.cepr.org

This Discussion Paper is issued under the auspices of the Centre’s research programme in Industrial Organization. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre’s publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Paul A Geroski
ABSTRACT

The Growth of Firms in Theory and in Practice*

This paper is a reflective survey of past and recent econometric work on the growth of firms. Most of this work suggests that firm size follows a random walk; i.e. that corporate growth rates are random. The survey documents this, and shows what a strong result this is by contrasting it with several alternative (and rather obvious) models which might be used to explain corporate growth rates but which are basically inconsistent with the data. The survey also discusses complementary evidence on corporate innovation rates and adjustment costs in investment/employment decisions which is consistent with (and therefore provide some support for) these results. This particular result is striking for a number of reasons, not least because it is basically inconsistent with most theories of the growth of firms which have be developed over the years. It is also inconsistent with the recently fashionable resource based theory of the firm. The second half of this essay identified how and why these theories of growth seem to be inconsistent with the data.

JEL Classification: L1
Keywords:growth, Gibrat’s law

Paul A Geroski
Department of Economics
London Business School
Sussex Place
Regent’s Park
London NW1 4SA
UK
Tel: (44 171) 262 5050 (x3477)
Fax: (44 171) 402 0718
Email: pgeroski@lbs.ac.uk
*This paper is forthcoming in N Foss and V Mahnke (eds), *New Directions in Economic Strategy Research*, Oxford University Press, 1999. The author is obliged to Mariana Mazzucato, Stan Metcalfe, Volker Mahnke, Nicolai Foss, Giovanni Dosi, Ulrich Witt, Sid Winter, participants at the 1998 Druid Conference at Bornholm and at the 1998 ESSID summer school in Corsica for helpful comments on an earlier draft. The usual disclaimer applies.

Submitted 8 February 1999
NON-TECHNICAL SUMMARY

Very little in the theory of the firm as we know it is testable. Transaction cost based theories of the firm are usually driven by factors (such as asset specificity) which are difficult to observe, and, in any case, these theories rarely make useful predictions about the determinants of corporate growth rates. Recent resource based theories of the firm have more to say about corporate growth, but the core competencies or internal assets/skills which they are based on are also difficult to observe. Even Penrose’s famous ‘managerial limitations to growth’ hypothesis is based on coordination and team building costs which are rarely directly measured and never appear on corporate balance sheets.

To make any progress in exploring the usefulness of these different theories of the firm one needs to proceed indirectly. In principle, each of these different types of theory makes different predictions about one or more elements of corporate performance conditional of at least some observables. Direct observation of corporate performance under different conditions should, therefore, be a useful if indirect way to evaluate the different theories. As it happens, there is one thing that we know about corporate performance that is not conditional on anything that is difficult to observe. This is that corporate growth rates are random. And, as it turns out, it is difficult to reconcile most theories of the firm with this observation.

The first half of this paper sets out the basic evidence on corporate growth that has emerged from past and recent econometric work. The result that year by year changes in firm size are almost unpredictable is very robust and seems to be a feature of different samples of firms observed in different time periods. It is also a feature of quite different econometric models used to analyse the data. The proposition that corporate performance is erratic is also consistent with at least two other pieces of evidence: the production of patents and major innovations by firms is very erratic and most firms do not persistently innovate, and the establishment and enterprise level responses to investment and employment shocks is also erratic and difficult to predict.

The second half of the paper explores four basic types of theories of the growth of firms. The most common theory models a steady state firm size and then views growth as a (transitory) response to deviations from that equilibrium. A second type of model (which is now not very prevalent) is a life cycle model of corporate growth, which posits that firms follow a steady and systematic evolution through a number of distinct, identifiable stages. The third model we examine is the famous Penrose ‘managerial limits to growth’ model, while the fourth is the resource based theory of the firm (which is also to be found in Penrose’s book). In all cases, it turns out to be difficult to
reconcile these theories with the evidence discussed in the first half of the paper.

The final concluding section sketches out where one might go from here. Two observations seem pertinent: that whatever model of the firm one chooses to work with, it must have the property that firms do not necessarily follow up initial successes (i.e. that they do not always display sustained success) and it must explain not only what happens to firms but also when it happens. This second criteria matters, since the translation of a theoretical model into an empirical model must take account of the nature of the data used in empirical work (i.e. that firms are observed once every calendar year). As it happens, recent Schumpeterian models of growth display both of these properties and the paper concludes with a brief discussion of this point.
I. INTRODUCTION

The modern theory of the has mainly been concerned with devising explanations of why firms exist. In fact, there is currently a broad consensus amongst economists that: “...it remains one of the most profound challenges to identify the forces that determine whether transactions are conducted within the firm as opposed to through the market” (Moore, 1992, p. 494). Interesting as this problem is, answering it doesn't actually take us very far in understanding what we observe about firms. It is, for example, hard to get much of a theory of the growth of the firm from transactions cost or governance based theories of the firm. At best, one might argue that exogenous changes in transactions costs or governance technologies will induce changes in optimum firm size that, sooner or later, lead to changes in the size of firms. In fact, any of a number of potential determinants of optimum firm size can form the basis of a theory of corporate growth: when firms are away from their optimum, growth (or decline) occurs as competitive forces drive them towards equilibrium. However, in this view, growth is driven by exogenous events and it is a transitory phenomena which occurs only until equilibrium is restored.

Older traditions of thinking have produced rather more satisfactory (or, at least, more interesting) models than this. There have, for example, been several attempts to develop stage theories of firm growth over the years (not all of which are based on simple biological analogies). These theories are often based on a conceptual framework which has more appreciative than predictive power, and they often seem to be a little over stylized or just too simple. Most of the interesting theorizing about the growth of the firm was either developed by Penrose, 1959, or synthesized by her into her own work. She argued that the firm “...is basically a collection of resources” (p. 77), and then analysed the process of growth in terms of the speed with which firms could accumulate and assimilate such resources, and the opportunities for further growth which arise when a firm’s internal resources are under used. Following this tradition but shifting the focus slightly away from the resources themselves and towards the management
of resources, many economic historians and management scholars have argued that: “...the key concept ... (which we need to use) ... to explain ... the beginnings and growth of modern industrial enterprises is that of organizational capabilities. These capabilities were created during the learning process involved in bringing a new or greatly improved technology on steam, in coming to know the requirements of markets for new or improved products, the availability and reliability of suppliers, the intricacies of recruiting and training managers and workers.” (Chandler, 1992, p. 487).

Different as they seem, all of these various ways of thinking about the growth of firms have at least one very important feature in common, namely that they very are hard to reconcile with much of what we have learned from econometric work on the growth performance of firms. This observation applies particularly to recent resource based theories of the firm: the nature of competencies and the process by which they are accumulated is hard to reconcile with the erratic growth performance which most firms display. This observation is the subject of the paper, and documenting it forms it’s substance. The plan is as follows. In Section II, I will set out some of the basic “facts” about the growth of firms which have been uncovered in recent econometric work. In Section III, I will confront these “facts” with a set of predictions about corporate growth which have emerged from the two traditions of thinking about firms alluded to above. The main thrust of the argument is that most of these predictions are more or less inconsistent with the “facts” on corporate growth. Section IV contains some suggestions about the direction which future work on the theory of the firm might take.
II. THE GROWTH OF FIRMS IN PRACTICE

The most elementary “fact” about corporate growth thrown up by econometric work on both large and small firms is that firm size follows a random walk. If we denote the size of firm $i$ in period $t$ by $S_i(t)$, then this observation can be written concisely as

\[(1) \quad \Delta \log S_i(t) = \log S_i(t) - \log S_i(t-1) = \mu_i(t),\]

where $\mu_i(t)$ is a normally distributed iid random variable with a mean of zero and a variance of $\delta^2$.

It is worth being clear about just exactly what this “fact” means. In the first place, (1) says that increases in firm size are driven by unexpected shocks. This is not quite the same as saying that growth is driven by “mere chance” or “good luck”. An unexpected shock may occur because we do not know what will happen, but it may also arise if we know what will happen to a particular firm but are not sure when it will happen. Put another way, (1) is a compact and succinct description of a process which may be very well understood, but hard to describe or predict with any precision. Moreover, unexpected means “relative to some information set”; i.e. what we expect depends on what we know. Hence, what may seem surprising to an econometrician may not, in fact, be very surprising to the managers of the firm it happens to. In this case, (1) may turn out to be the best description of what happens to a firm that is available to outsiders.

Second and much more important, (1) says that unexpected shocks have permanent effects on the size of the firm: each idiosyncratic shock which affects the firm leaves a permanent mark. This means that corporate growth cannot be thought of as a process composed of a deterministic trend with some noise superimposed on it.¹ The trend itself is stochastic. Another way to make this point is to observe that (1) implies that
(2) \[ \log S_i(t) = S_0 + \sum \mu_i(s), \]

where the index \( s \) runs from 1 to \( t \) and \( S_0 \) is the initial value of \( S(t) \); i.e. the firm's start-up size. Thus, the size of a firm at any time is just the sum of the whole history of shocks, expected and unexpected, which it has received since it was founded in \( t = 0 \). To say that firm size follows a random walk, then, implies that corporate growth is a path dependent process. The unpredictable nature of these shocks means that it is difficult, if not impossible, to predict what a firm's size will be at any time in the future, \( t + \tau \).

Some will argue that equation (1) is an over strong characterization of what we know about corporate growth, and there may be some truth in this. There is, in fact, a large empirical literature which has explored a variant of (1), namely

(3) \[ \Delta \log S_i(t) = \alpha + \beta \log S_i(t) + \mu_i(t), \]

where the observation that \( \beta < 0 \) is taken to indicate the existence of "mean reversion"; i.e. the proposition that larger firms grow relatively slower than smaller firms. A levelling out in growth rates between large and small firms in this manner bounds the overall rise in the variance of firm sizes. Further, if \( \alpha > 0 \) and is common to all firms, then firms whose size evolves over time according to (3) will all converge to the same, common long run or steady state size of \( (-\alpha/\beta) \). The evidence on mean reversion and convergence is, however, pretty mixed. Most estimates of \( \beta \) reported in the literature are pretty small, and work using panel data almost invariably reveals that different firms display different estimated values of \( \alpha \). This means that firms (even those in the same industry) converge to different steady state sizes; i.e. that differences in firm size are permanent and not transitory. Further, direct (and arguably more powerful and persuasive) tests of convergence using long times series of data on different firms suggest that no stable or predictable differences in size or growth exist in the short or the long
run. All of this means that the evidence against the proposition that **firm sizes do not converge within or across industries** is not very strong. The simple fact is that firm size drifts unpredictably over time, and, as a consequence, predictions of $S_i(t + \tau)$ become increasingly uncertain as $\tau$ gets larger.

Two further implications of (1) are worth noting. The first is that since $\mu_i(t)$ is independently distributed across firms $i$, the growth rates of any two firms, $i$ and $j$, chosen at random are likely to be uncorrelated. That is, **corporate growth rates are likely to be idiosyncratic**. This is a surprising observation, since common sense suggests that the growth rates of most firms should rise and fall with variations in the growth rate of the economy as a whole, or at least of the industry they inhabit. In fact, the evidence is that this tendency towards common growth rates across firms in the same industry (much less across the economy as a whole) is pretty weak. Including a macroeconomic growth variable in (1) or (3) typically generates a positive and significant co-efficient that does not, however, add much explanatory power to a regression which (typically) already displays a tiny $R^2$. Further, **studies of company performance in cyclical downturns usually show that most of the effects of recessions are concentrated in a few firms; many companies are not substantially affected and some actually prosper during cyclical downturns**. The upshot of all of this is that there is apparently very little in the current or recent past growth of rival companies $j$ which can be used to help predict the current or near future growth of firm $i$; corporate growth is history dependent and every firm seems to have its own particular history.

The second further implication of (1) is the absence of any dynamics associated with lagged dependent variables. A conventional dynamic econometric model of growth might take the form of

\[ \Delta \log S_i(t) = \alpha + \theta(L) \Delta \log S_i(t-1) + \mu_i(t), \]
where \( \theta(L) \) is a polynomial in the lag operator \( L \). Although it is occasionally possible to observe significant co-efficients on \( t-1 \) and \( t-2 \) lags in the dependent variable, these terms rarely add much explanatory power to corporate growth rate regressions.\(^5\) Since the dynamics associated with the lagged dependent variable in (4) are typically interpreted in terms of adjustment costs, the fact that (4) reduces to (1) means that **corporate growth rates are not smoothed.** The implication is that firms do not appear to anticipate shocks and begin reacting before they occur; nor do they appear to be only partially adjusting to current shocks, postponing full adjustment to minimize adjustment costs.

The are two further pieces of evidence which complement the central “fact” that firm size follows a random walk. The first is that **adjustment costs seem to be fixed and not variable.** The distinction may seem a little arcane, but it has a profound impact on the times series behaviour of output, investment, employment and other choices made by firms. If **adjustment costs are variable and increasing in the size of the desired adjustment, firms will have an incentive to spread out their adjustment to shocks, responding regularly but by a small amount.** This type of behaviour will result in an equation like (4). If, on the other hand, adjustment costs are fixed, then firms have an incentive to “save up” their desired changes until it is worth incurring the fixed costs to make them, and then they will make them all in one “big bang”. This is sometimes called an \((s,S)\) policy, and even if one knows the size of the adjustment that firms wish to make, the timing of when they choose to act is likely to be unpredictable. As a consequence, firm size will follow a path which looks more like (1) than (4). In fact, the evidence is that firms typically make large but infrequent and clearly discrete changes in their operations (e.g. in employment and investment), rather than continuous but small ones. This is perfectly consistent with (1), and is presumably why (4) gives a relatively poor fit on the data.\(^6\)

The second added piece of evidence is that **most firms are erratic and irregular innovators;** that is, very few firms produce major innovations or patents on a regular basis (actually, very few firms ever innovate in this sense,
but those who do are rarely persistent innovators). This evidence is difficult to interpret, since it depends on how frequently one observes firms and on what exactly one means by “innovate”. On the one hand, most large firms regularly spend money on R&D, but, on the other, none have ever engineered a breakthrough as fundamental as splitting the atom. However, if we focus on activities which lead to noticeable technical breakthroughs which are commercially successful (i.e. “major innovations”) or patents, then the data suggests that very few firms manage to produce a regular sequence of innovative outputs. The typical pattern is that firms will innovate every once and a while, opening up what are sometimes very long periods of time between successive innovations. Erratic innovative activity is likely to mean that the growth spurts experienced by firms will be unpredictable, and this is, of course, just what (1) describes.
III. SOME THEORIES OF CORPORATE GROWTH

Although they seem to be rather simple, it turns out that the “facts”
discussed in Section II cast some doubt on the usefulness of a range of models of
or hypotheses about corporate growth which have appeared in the literature. We
begin by focusing on four rather different types of theories of growth: models of
optimum firm size, stage theories of growth, models with Penrose effects and
models of organizational capabilities. 8

(a) models of optimum firm size

Most economists look for steady state equilibrium configurations as
a foundation upon which to do comparative static exercises, and much of the
theory of the firm has been expressed in these terms. The oldest and best known
argument suggests that competition will drive firms to the bottom of their U-
shaped average cost curves. If firms have market power, then their optimum size
may differ from this minimum cost position, and, if economies of scope exist, such
differences may be more noticeable (and, of course, firms will be diversified in
this case). More recent arguments have suggested that the degree to which costs
are (endogenously or exogenously) sunk and the intensity of competition may
also be important determinants of firm size (and market structure). Further,
many believe that internal organizational factors may be as important as market
competition and technology in determining firm size. The ability of managers to
control their firms may be limited by serial errors in communicating up and
down a management hierarchy, or by other transaction costs. Although such
problems can be mitigated by restructuring the organization in various ways, the
bottom line is that sooner or later, firms are likely to get too large to control. 9
Whatever the details, these models typically yield an outcome in which optimum
firm size, $S^*$, depends on a number of exogenous variables.

The traditional way to get a model of growth out of this kind of
argument is to use $S^*$ as a benchmark, and then to interpret all observed changes
in firm size either as white noise or as part of a transitional process of convergence to $S^*$. The most commonly used model of this type is a partial adjustment model of the form:

$$\Delta \log S_i(t) = \lambda [\log S^* - \log S_i(t-1)] + \mu_i(t).$$

$\lambda$ is the interesting parameter in this model, as it governs the rate at which convergence occurs. If $\lambda = 0$, then convergence never occurs, while if $\lambda = 1$, then $\log S_i(t) = \log S^*$ always. Note that if $S^*$ is constant over time, this model is equivalent to (3) when $\alpha = \lambda \log S^*$ and $\lambda = -\beta$.

Regarded as a model of the growth of the firm, (5) has the serious disadvantage of being inconsistent with the “facts”. There are three problems. First, as we noted earlier, (3) and, therefore, (5) are at best only weakly consistent with the data: the values of $\lambda$ which have been estimated using empirical models of this type imply that convergence is, at best, implausibly slow. Further, as we noted earlier, the very poor fit of these estimated models means that most of the variation in corporate growth rates is driven by $\mu_i(t)$. Second, and probably more seriously, most of the arguments used in these models generate predictions about the size or inter-industry variation in $S^*$ which are not firm specific (e.g. technology is typically assumed to be common to all firms in the industry, serial communication costs do not depend on who communicates but on the height of the hierarchy and the span of control, and so on). This means that these models usually imply that all firms in the same industry will converge to the same size, $S^*$. This is clearly inconsistent with the data, which, as we saw earlier, suggests that there is, at best, only a very weak tendency for firms to converge towards a common size even in the same industry. Third and finally, the presumption that $S^*$ is a fixed target towards which firms tend is impossible to reconcile with the “fact” that firm size follows a random walk.

There are at least two ways to redesign this kind of model to make it consistent with (1). The basic trick is to let $S^*$ drift unpredictably over time.
Suppose, for example, that the exogeneous variables which determine log S* are altered by a large number of small, independent, firm specific shocks which occur in each period. Then, if we suppose that \( S_i(t) = S^*(t) \) (i.e. that there are no transitional dynamics of the form of (5)), and collect these shocks together into one random variable \( \mu_i(t) \), we arrive at (1). This is, of course, the standard explanation of what underlies Gibrat’s Law. A second version of this argument is the “island markets” model. The idea here is that each firm \( i \) is faced with a sequence of independent opportunities of varying sizes which arise exogeneously over time. As each opportunity arises, there is a probability \( p \) that it will be taken up by a new entrant and a probability \((1 - p)\) that the firm will colonize it. If the size of the firm has an effect on the number of opportunities that it takes advantage of, then this argument produces a model like (3); if the size of the firm has no influence on the probability of colonizing a market in any period \( t \), then this argument generates (1).

Thus, it is possible to reconcile the tradition of thinking in terms of optimum firm sizes with the “facts” by letting \( S^* \) vary unpredictably, and by assuming that firms are always at (or only depart unpredictably from) their optimum sizes. The problem with these solutions is that they effectively make corporate growth an exogeneously driven process, and this is not consistent with the many observations we have on how corporate decisions sometimes decisively alter the evolution of technology, the structure of transactions costs or the development of demand.

(b) stage theories of growth

Although they are not much in fashion now, there have been a number of attempts over the years to identify life cycles of firms, model their evolution or at least pick out identifiable stages through which they grow. For example, Greiner, 1972, argued that firms evolve through five phases, each characterized by a period of relatively stable growth. These phases (which he identified with a label which indicates the nature of the management problem
characteristic of each: “creativity”, “direction”, “delegation”, “co-ordination” and “collaboration”) are separated by four crises (of “leadership”, “autonomy”, “control” and “red tape”). More recently, Garnsey, 1998, has developed a model of corporate growth which traces out a set of phases which correspond to the development and deployment of new internal resources in young firms. Finally, in something of a different tradition, Mueller, 1972, argued that what a firm does (i.e. its propensity to maximize profits or sacrifice profits for growth) varies with age (and other factors, like investment opportunities). A strictly profit maximizing firm is likely to enjoy only a finite burst of growth associated with each innovation. However, if the innovation fuels enough growth to weaken the power of shareholders, then managers will gradually acquire some room to exercise discretion. Since they are liable to be more interested in the size or growth of the firm than in its profits, they will take advantage of this discretion to reinvest too much of the proceeds from the innovation into this or other investment opportunities. As a consequence, “too much” growth is likely to be associated with each innovation, and it is likely to go on for “too long”.

These arguments have more than a little superficial plausibility. Stage theories of growth have some basis in fact (entrepreneurial firms eventually outgrow their founders and become bureaucratic institutions, all firms mature and many decline and disappear, etc etc), and they are often a useful aid to conceptualization. However, even if all firms progress through all five of Greiner’s phases, they are likely to do so at very different rates and will probably enjoy different growth rates in each phase. Similarly, although it is not hard to believe that the goals of firms change systematically over time, it is hard to know exactly how one might see this in the data on annual growth rates which forms the basis of the “facts” discussed in Section II. The basic problem with these arguments is that they are built up around the view that there are secular (or long run), deterministic trends in the pattern of growth of firms. The “fact” is that growth rates display stochastic trends, meaning that firm size evolves in an erratic and unpredictable manner over time. Further, since the data on corporate growth rates displays very little in the way of transitional dynamics, it is hard to
draw a meaningful distinction between short and long run effects of exogenous factors on firm size. The current period shocks which propel a random walk are permanent, and, as a consequence, their effect in the long run is almost the same as it is in the short run.

(c) models with Penrose effects

The classic study of the growth of firms by Penrose contains two quite different types of arguments. One is a “resources push” theory of (endogenously driven) growth, which I will deal with when we look at models with organizational capabilities. The other argument in her book is the famous “managerial limits to growth” hypothesis, and that is our concern here.

This argument starts with the premise that management is a team effort in which individuals deploy specialized, functional skills as well as highly team specific skills that enable them to collectively co-ordinate their many activities in a coherent manner. The knowledge which underlies these specific skills is likely to be tacit, and can only be learned experientially or by direct instruction from existing managers. Hence, as the firm expands, it needs to recruit new managers and it must divert at least some existing managers from their current operational responsibilities to help manage the process of expanding the management team. Since diverting existing managerial resources to training new managers carries an opportunity cost, the faster is the planned rate of growth of the firm, the higher are these costs of growth likely to be (i.e. adjustment costs are variable and not fixed). Under these circumstances, firms are likely to smooth out their responses to current growth opportunities, sacrificing current profits but saving some of the costs of growth which they might otherwise incur to gain those profits.

Penrose argued that firms had no determinant long run or optimum size, but only a constraint on current period growth rates. This yields a model like (4): the $\theta(L)$ are determined by the size of adjustment costs (and the rate at
which they increase with increases in the rate of growth), while the lack of an optimum size means that \( \alpha \) is effectively zero (i.e. in the long run, firm size is unpredictable). It should be clear by now that this argument is not consistent with the “facts”. As we noted earlier, equation (4) frequently reduces to (1) in practice, and when it does not, estimates of \( \theta(L) \) suggest that lags higher than order 1 are almost always insignificant. This means that most adjustment costs associated with growth are incurred within a year, which is rather implausible (we know that it can take years to mould a successful management team). Further, estimates of \( \theta(1) \) are typically very small (that is, adjustment costs are not very large) and the collective contribution of all lagged dependent variables to the overall explanation of the regressions is usually extremely modest (which is to say that managerial limits to growth do not appear to explain much of what we observe). More decisively, most of the direct evidence on adjustment costs that we have suggests that they are fixed and not variable. However hard one tries, it is difficult to reconstruct Penrose's managerial limits to growth argument in a way which makes the costs of expanding management teams independent of the number of new managers to be recruited into the team.

It is possible to rescue at least some of this argument. A somewhat more modern and more formal version of it would start from the realization that a firm faced with variable adjustment costs of whatever type will have an incentive to begin adjusting to shocks which it expects to occur in the near future. This means that its optimum size will depend, inter alia, on cost and demand conditions which are expected to prevail in the near future. In a standard optimizing model, these are captured by current period expectations of future margins. Since, at the optimum, firm size depends on current expectations of future margins, growth will depend on current period changes in those expectations. It turns out that this very simple model of growth is consistent with (1) under certain not entirely uninteresting circumstances. In particular, if firms form rational expectations about future conditions, then changes in these expectations are unpredictable. Rational expectations about a future period \( t + \tau \) held at time \( t \) basically are those which use all the information available to
decision makers at time t. This, in turn, implies that current changes in expectations will be unpredictable (or, their expected value will be zero). Since, in this simple model, growth is driven by current changes in expectations, this means that growth rates are also unpredictable.\textsuperscript{11}

That a rational, profit maximizing firm might display unpredictable growth rates as a matter of course makes (1) a lot easier to accept (or, at least it makes (1) seem like less of an affront to orthodox neo-classical theory). However, expectations matter only when current actions have undoable future consequences (adjustment costs are only one example of this kind of phenomena), and this means that current actions depend on past actions. This is a feature of (4) but not of (1). The history dependence displayed in (1) arises from the fact that each burst of growth has a permanent effect on the size of the firm. The history dependence displayed in (4) is one in which the realized outcomes of current decisions depends on the realized outcomes of recent past decisions, and, as we have seen, it is inconsistent with the “facts”. These are quite different types of “history dependence” (being permanent and transitory respectively), and the “facts” sit more easily with the former than with the latter.

(d) models of organizational capabilities

Penrose thought of firms as bundles of resources bound together by a set of administrative skills or capabilities which are used to deploy them as effectively as possible. Others have followed in this spirit. Nelson, for example, has argued that: “…successful firms can be understood in terms of a hierarchy of practised organizational routines, which define lower order organizational skills and how these are co-ordinated, and higher order decision procedures for choosing what is to be done at lower levels” (1991, pp. 67-68). These routines are, of course, part of a firm’s resource base, and they define what the organization is capable of doing (or what its “competencies” are). These capabilities are almost always thought of as bundles of skills, and, indeed, they are one of the more important repositories of tacit knowledge inside firms. The fact that knowledge is the
foundation of organizational capabilities or competencies means (at least) two things: first, these competencies are not assets (and do not, therefore, appear on balance sheets, and cannot be bought and sold), and, second, that they can only be learned or maintained through use (hence, people typically talk of “practised’ routines). Each firm is likely to be born with some particular skill or knowledge base, and then develop it idiosyncratically over time as it uses what it has inherited and what it has learned to develop new skills and an augmented knowledge base. This means that each firm’s development is likely to be path dependent. Further, if, as Penrose argued, internal resources are discrete, then firms may have stocks of underutilized resources which “push” it on to further expansion.12

The basic premise of this work is that competitive advantage is based on the possession of a few key resources and routines, organizational capabilities or core competencies and, despite the proliferation of labels, there is some measure of agreement on what this means. A “core competency” (to choose the label used in most popular discussions) is something which: creates value for consumers, is unique (or at least better than that possessed by rivals), durable, generates returns which are appropriable, and it is (or should be) inimitable. There are many possible sources of inimitability: the resource may be physically scarce, its accumulation may be path dependent and there may be causal ambiguity about just how it works or what its driving feature is. If competitive advantage is based on the possession of core competencies defined in this way, then firms are likely to be heterogeneous (because competencies are unique) and realize different levels of performance (depending on the value created for consumers by using the competency and the degree of appropriability of the resulting gains) over long periods of time (because the resources which sustain competencies are durable, and the competencies themselves are difficult to imitate).13 These arguments are roughly consistent with what we know about accounting profitability: firm profitability is idiosyncratic (meaning that aggregate or industry level determinants are weak relative to idiosyncratic, firm specific determinants of profitability) and profit differences between firms persist for long
periods of time. It does not, however, seem to be quite so consistent with what we observe about company growth.

To see this last argument, consider the following simple model. Suppose that firm i has a competency which enables it to generate more revenue growth than rivals. Let \( \theta_i(t) \) be an index of firm i’s competence, measured in units so that i grows at rate:

\[
\Delta \log S_i(t) = g(t) + \theta_i(t),
\]

when it has a competence level of \( \theta_i(t) \). Firms with no particular competence (\( \theta = 0 \)) grow at rate \( g(t) \), while those who are incompetent (\( \theta < 0 \)) grow even slower. Everything that we know about competencies suggest that they evolve over time in a systematic way, with each increment in ability which is realized depending on the previous ability of the firm. An easy way to way to model this is to suppose that

\[
\theta_i(t) = \rho_1 \theta_i(t-1) + \rho_2 \theta_i(t-2) + \varepsilon_i(t),
\]

where \( \varepsilon_i(t) \) is an unpredictable change in competence. If \( \rho_1 + \rho_2 > 1 \), then the growth of firm i’s competence is explosive, increasing at ever increasing rates (there must obviously be a limit to this); if \( \rho_1 + \rho_2 < 1 \), then (7) describes a process in which competence levels gradually revert to some long run mean level. A little manipulation of (6) and (7) reveals that

\[
\Delta \log S_i(t) = \rho_1 \Delta \log S_i(t-1) + \rho_2 \Delta \log S_i(t-2) + v(t) + \varepsilon_i(t),
\]

where \( v(t) = g(t) - \rho_1 g(t-1) + \rho_2 g(t-2) \). If \( \mathbb{E}[\varepsilon_i(t)] = \alpha + \mu_i(t) \), then (8) effectively reduces to (4) with the addition of a common time trend, \( v(t) \). However, as we have seen, (4) is not a satisfactory description of corporate growth rates, and this means that the argument that corporate growth is driven by competencies is not consistent with the “facts”.

17
The basic problem should be clear: most of the literature on competencies has sprung up to explain persistent differences in accounting profitability between firms, but firms do not display persistent differences in their growth performance. Whatever it is that fuels corporate growth seems to be much more transitory than whatever it is that fuels changes in corporate profitability. Notice that (8) reduces to (1) if \( \alpha = \rho_1 = \rho_2 = 0 \); that is, if there is no trend change in \( \theta(t) \) and if changes in a firm's competencies over time are random (i.e., if \( \theta(t) \) follows a random walk). While this simplification makes the model consistent with the "facts", it is now basically inconsistent with everything we know and believe about how competencies evolve over time. To put the matter another way, if the accumulation of competencies is really what powers growth, then the kinds of competencies which we need to identify are those which either have very transitory lives, or have only very temporary effects on growth.
IV. WHERE DO WE GO FROM HERE?

There is much to be said for the modern, Coasian theory of the firm, but one thing that it simply does not do is address the growth, development and diversification of firms which we observe going on around us day in and day out. There are, of course, many arguments and theories which economists have developed over the years which do try to explain what we observe, but, as we have seen, many of them are not easily reconciled with the “facts”. The most promising recent development is probably the literature on organizational capabilities or core competences. It has the major virtues of accounting for the ubiquitous heterogeneities between firms which we observe, and it offers a plausible, history dependent story of organizational growth and development. The problem is that the literature on competences has been much influenced by empirical studies of accounting profitability which show persistent and stable differences in performance between firms. However, this is not a feature of many other measures of corporate performance. Corporate growth rates, for example, differ between firms in temporary and unpredictable ways, and it is hard to reconcile the inimitability and durability of organizational capabilities with this data. The consequence of all of this is that theorizing about competences is being driven by a correspondence with the “facts” which is, at best, partial.

The way forward seems to be reasonably clear: work on the theory of the firm needs to be redirected towards models which help to account for the uneven, erratic performance of firms over time. It is, of course, possible to argue that exogenous factors are entirely responsible for the unpredictable nature of corporate growth rates. But, this is hard to believe: many firms try to alter their environment in various ways and, further, many competitive shocks are endogenous. It is also hard to accept that growth is largely exogenous shock driven because many firms do not react quickly or well to market shocks, and others try to resist innovation. This inertia makes the timing of corporate activity difficult to predict, and, hence, it makes corporate behaviour seem erratic. Since
this is what we observe, it seems possible that models which focus more on when firms act than on what they do may be a useful step forward.

There is, in fact, a model of aggregate economic growth which has a number of properties which are roughly consistent with the data we discussed earlier, and it turns out to be relatively easy to translate that model into a theory of the firm.\textsuperscript{15} I will close out this essay by sketching out this model in a very bare bones form, and showing why it seems to fit the “facts”.

Suppose that a firm generates revenue, $y$, in each period by spending an amount of money, $x$, on labour inputs, the services of physical capital, advertising and all of that. In particular, let

\begin{equation}
\pi = y(x) - x = Ax^\alpha - x,
\end{equation}

where $0 < \alpha < 1$. The optimum choice of $x$ satisfies: $\alpha y = x$, so

\begin{equation}
\pi = (1 - \alpha)y.
\end{equation}

Innovations occur in response to expenditures on $x$, and we date each innovation by the time it arrives, $\tau$. For simplicity, assume that each innovation has the same effect on the firm's performance, namely that it raises $A$ by a factor $\gamma$; that is,

\begin{equation}
A_\tau = A_0 \gamma^\tau.
\end{equation}

Note that $\tau$ does not record calendar time; $\tau$ increases every time an innovation occurs, not every year. The implication of (11) is that

\begin{equation}
y_\tau = \gamma y_{\tau-1}.
\end{equation}
The core of this model is in the choice of how much research a firm does. For simplicity, reinterpret $x_t$ as the amount of money the firm chooses to spend on research after the arrival of an innovation in period $\tau$, and assume that the more money it spends the sooner will be the arrival date of the next innovation in period $\tau + 1$. To be more precise, suppose that the rate of innovation conditional on $x_t$ is a Poisson process, so that the length of time between the $\tau^{th}$ and the $(\tau + 1)^{st}$ innovation is exponentially distributed with parameter $\phi(x_t)$, where $\phi' > 0$ and $\phi'' < 0$. Since an innovation in $\tau$ yields profits for a period of time which is a random variable, its expected value is $V_{\tau-1} = \pi_t(r + \phi(x_t))^{-1}$, where $r$ is the discount rate. The expected value of doing research is then

$$\xi_{\tau} = \gamma \phi(x_{\tau + 1}) \pi_{\tau + 1} (r + \phi(x_{\tau + 1}))^{-1} - x_{\tau + 1}. \tag{13}$$

A profit maximizing firm will choose a sequence of values of $x_t$ which satisfy

$$\gamma \phi'(x_t) \pi_t - r - \phi(x_t) = 0. \tag{14}$$

There are two results that emerge from all of this. (14) shows the first. It basically says that $x_t$ and $x_{t+1}$ will be negatively correlated over time; i.e. if $x_{t+1}$ is expected to be large, then $x_t$ will be small. The basic idea is that if a firm knows that $x_{t+1}$ is going to be large, then it also knows that the innovation which $x_t$ will (sooner or later) generate will be short lived (because the large spending in period $\tau + 1$ will quickly generate an innovation to displace it). The innovation in $\tau$ will not, therefore, be particularly valuable, and the firm will not be willing to spend much to get it. Hence, the firm will not choose a large value of $x_t$.\textsuperscript{16} The common sense of this proposition is that firms are unlikely to sustain a regular sequence of innovations. Instead, an innovator is going to have an incentive to delay the introduction of subsequent innovations as long as possible in order to preserve returns from the first. The result will be an erratic pattern of innovative activity over time.
This is a conclusion which is consistent with much of what we observe in markets, and the pattern of behaviour it describes is not model specific. Successful innovators are often reluctant to introduce subsequent innovations which are either competence destroying or which cannibalize some of the rents which flow in from the original innovation. The consequence is that they are not quite willing or able to follow an initial innovation up with subsequent innovations unless, by chance, they are forced to (or unless these subsequent innovations are competence enhancing or do not cannibalize existing rent streams). Further, this pattern of apparently erratic behaviour is likely to be a property of models in which firms face fixed adjustment costs, and do not respond to performance deteriorations until some threshold of poor performance has been reached. Many firms do not immediately respond to cost and demand shocks. Instead, they tolerate a little deterioration in their performance until they are sure that the shocks are permanent and important. They may also need to take some time to work out just exactly how to deal with the shocks which they have experienced. The consequence is that adjustment to shocks will occur only when “enough” pressure has built up (this is basically an (s,S) strategy of response). If shocks are Poisson distributed, then the time between responses will be exponentially distributed. Further, if the firm expects to make a major initiative to improve it’s performance in period τ+1, then it is unlikely to make much of an investment in improvement in period τ. This is, of course, just the kind of behaviour that (14) describes.

The second interesting result follows from this, since an erratic pattern of innovation ought to induce an erratic pattern of growth. (12) can be written in calendar time (t, not τ time) as:

(15) \[ \log y_t = \log y_{t-1} + \epsilon_t, \]

where \( \epsilon_t = N(t-1,t)\log \gamma, \) and \( N(t-1,t) \) is the number of innovations which arrive between period \( t-1 \) and period \( t \). Our earlier assumptions imply that this is a
Poisson random variable whose expected value is $\phi$. Hence, if we define $e_t = \varepsilon_t - \phi$, then

$$(16) \quad \log y_t = \theta + \log y_{t-1} + e_t$$

where $\theta = \phi \log \gamma$. Note that $e_t$ is a normally distributed random variable with zero mean and variance $\phi (\log \gamma)^2$; i.e. (16) looks just like a classical regression model. Basically, it says that firm size follows a random walk with drift. Again, the intuition is straightforward: the natural cycle of activity from the point of view of the firms actions is the “calendar” based on innovative activity (and tracked by $\tau$), but we as econometricians observe the firm only at arbitrarily fixed yearly intervals. Since our observations are out of sync with the actions of the firm, it is hardly surprising that we are inclined to find it's activities to be erratic and unpredictable.

This conclusion springs from an observation which is sufficiently deep to be worth stressing. As econometricians, we observe firms on a regular but arbitrary basis. Financial data appears annually (say in April or October), and that is when we observe the firm and collect data on it's size (and other things). However, the important events which affect the firm's performance happen at all kinds of times, and none of these are necessarily tied in with the firms normal financial reporting cycle. Further, many of the choices (investments in R&D, competitive initiatives, and so on) which firms make do not have natural gestation periods of 12 months; nor are their consequences delivered in neat twelve month intervals conveniently dated to coincide with the release of an annual report. As a consequence, when we view firms through the lens of a twelve month financial reporting cycle, we are bound to be somewhat out of phase with the natural but irregular rhythms of it's day to day competitive life. Under the circumstances, it is not hard to believe that accounting for variations in it's performance will be difficult, and that is basically what (16) says. Notice that the sole exception to this is accounting profits, since this is the one
measureable aspect of corporate performance which is perfectly synchronized with it's annual reporting cycle.

The bottom line is that this sketch of a model of corporate displays two properties which are consistent with the facts outlined in Section II above: it generates an erratic pattern of innovative activity over time (firms in this model are rarely rapid or persistent innovators), and the resulting evolution of firm size follows a random walk. This is not, of course, the only model which displays (or can be made to display) these properties. But, it is characteristic of what I think may be the defining feature of models of the firm which try to come to terms with the facts”, namely that it is when things happens which matters and not (so much) what it is that happens.
REFERENCES


NOTES

1 This is often described as the difference between a “trend stationary” and a “difference stationary” process in the macroeconomics literature; see Nelson and Plosser, 1982. Most studies of GDP suggest that, like firm size, it is (more or less) difference stationary (i.e. follows a random walk).

2 For recent work, see Dunne and Hughes, 1994, Evans, 1987, Hall, 1987, Hart and Oulton, 1996, and others. Much of this literature is concerned with the evolution of industrial concentration, since Gibrats Law (which is what (1) is) predicts that the variance of firm sizes and, therefore, the concentration of economic activity in the hands of a few firms, will rise without bound.

3 See Geroski et al, 1997. The major caveat to these conclusions is the literature on very small, new firms, which typically finds much larger and more precisely estimated effects associated with firm size (and age); see, for example, the references cited in note #2 above, and the papers in the 1995 International Journal of Industrial Organization special issue on “The Post-Entry Performance of Firms”.

4 See Davis et al, 1996, Geroski and Gregg, 1997, and others.

5 For regression results which include one or more lags, see Geroski et al, 1997, and others. That $\theta(L) = 0$ is no surprise to those who believe that $\beta = 0$ in (3), since both simplifications yield the same outcome (namely, (1)); i.e. both are tests of Gibrats Law.

6 For some direct evidence based on establishment data taken from the Census, see Caballero et al, 1995 and 1997; Geroski and Gregg, 1997, and Davis et al, 1996 notice much the same thing in their studies of how firms respond to recessionary pressures, observing that firms are much more likely to close whole plants than to make marginal adjustments in the size of their workforce. For a survey of work on labour adjustment costs, see also Hamermesh and Pfann, 1994.

7 See Geroski, van Reenen and Walters, 1997, Cefis, 1996, Malerba and Orsenigo, 1994, and others. Cefis observes much the same pattern of innovative activity over time as Geroski et al, but interprets it slightly differently.

8 For surveys of all or part of this literature, see You, 1995, Sutton, 1997, and others. There is also a literature which examines possible trade-offs between growth and profitability, focussing on the proposition that managerially controlled firms will grow “too fast”. The “facts” discussed in Section II do not cast much light on this argument.

9 See Scherer and Ross, 1990, for a survey of many of the traditional arguments about optimum firm size; Sutton, 1991, focuses on sunk costs and the intensity of competition. Williamson, 1967, developed the model of hierarchy alluded to in the text; see also his 1970 book for early work how restructuring may mitigate some of these problems. For a recent and fairly comprehensive textbook treatment of many of
the issues originally raised by Williamson, see Milgrom and Roberts, 1992. For recent related work on transactions and governance theories of the firm, see Hart, 1995 and Williamson, 1996.

10 See Sutton, 1997 and 1998, who attributes it to Simon. Sutton is arguably not interested in the growth of firms, but, rather, in developing a lower bound on industry concentration, and this makes one reluctant to see it as a serious attempt to construct a model of corporate growth.

11 For a formal exposition of the Penrose model in terms of variable adjustment costs, see Slater, 1980, and, for an ambitious effort to extend and develop Penrose’s work, see Ghoshal et al, 1997. The argument which argues that growth is driven by changes in current period expectations about the future, and shows that current period growth rates must be unpredictable if expectations are rational is set out in Geroski, Machin and Walters, 1997. It claims some empirical support from correlations between firm growth and change in stock market value (which proxy changes in current expectations about future profits).


13 The notion of core competencies has also been extensively applied to the question of choosing the right diversification strategies for firms, the idea being that diversification should be based on a clear understanding of what any particular firm are capable of doing or designed around the acquisition of particular skills or routines. Indeed, “…mergers and acquisitions are an opportunity to trade otherwise non-marketable resources…” (Wernerfelt, 1984, pp. 1975); see also Montgomery and Wernerfelt, 1988, and, more generally, Markides, 1995.

14 See, for example, the survey in Geroski, 1998; Rumelt,1991, is frequently cited in the management literature in this context.

15 See Aghion and Howitt, 1992 and, more generally on models of endogeneous growth, their 1998 textbook.

16 This is the well known “displacement effect” which describes a situation where firms will be reluctant to introduce a second innovation which displaces returns from an earlier innovations (or cannibalized the returns from existing products). See Tirole, 1988, Chapter 10, for a discussion.