mir vol. 47, 2007/2, pp. 151 – 173

mir
Management International Review


**Peter J. Buckley/Mark Casson**

# Edith Penrose's Theory of the Growth of the Firm and the Strategic Management of Multinational Enterprises

## Abstract and Key Results

- This paper provides a formal model of Edith Penrose's Theory of the Growth of the Firm which has important implications for the strategy of multinational enterprises.
- The model provides an analysis of the trade-off between product diversification and foreign market penetration. It also can account for the speed of entry into foreign markets.
- Formalizing Penrose's Theory of the Growth of the Firm provides an account of internationalization incorporating geographical expansion patterns, sequential decision making and learning.

## Key Words

Internationalization, Multinational Firms, Global Strategic Management, Edith Penrose, Foreign Market Entry, Product Diversification

## Authors

Peter J. Buckley, Professor of International Business and Director of Centre for International Business, University of Leeds, Leeds, United Kingdom.
Mark Casson, Professor of Economics, Centre for Institutional Performance, University of Reading, Reading, United Kingdom.

Manuscript received December 2004, revised June 2005, final version received December 2005.

## Introduction

Foss (2002, p. 148) says "Penrose's work is, in the crucial dimensions, at variance with economic orthodoxy ... It should be thought of as a contribution to economic heterodoxy" Penrose and Pitelis (2002, pp. 19 et seq.), in describing Fritz Machlup as "Edith's supervisor" at Johns Hopkins says "A fascinating paradox is how Machlup, a doyen of neo-classical economics, should have been partially responsible for a work so far removed from the mainstream". Penrose has also been claimed as a feminist economist (Best/Humphries 2003). Our argument is that Penrose sought to create a theory of the growth of the firm which was logically consistent and empirically tractable. Her subsequent adoption as grandmother of the resource based view has only limited validity, based on a selective reading of her work and in defiance of its holistic qualities. See the debate between Rugman and Verbeke (2002, 2004) and Kor and Mahoney (2004) and Lockett and Thompson (2004), the latter based on Penrose's (1960) analysis of the Hercules Powder Company.

This paper presents a formalisation of Penrose's model contained in the *Theory of the Growth of the Firm* (1959) and applies it to the strategic decisions of multinational enterprises (MNEs). Previous research on Penrose and the multinational firm (e.g. Dunning 2003; Pitelis 2002, 2004) has focused on Penrose's overall contributions to strategic decisions in MNEs. This paper focuses solely on her 1959 model and compares it to the model of Buckley and Casson (1976). Interesting contrasts are found and a synthetic approach suggests that this combination is a useful basis for further theorising about the MNE and its strategic decisions.

In contrasting Penrose's theory with that of Buckley and Casson (1976), we shall see that the former's concentration on product diversification can be considered complementary to the latter's emphasis on innovation. Combining the two gives a satisfying model of the strategic management decisions within an MNE and opens up a new research agenda.

## A Simple Formal Model of Penrose's Theory of the Growth of the Firm

The key to formalising Penrose's ideas is the recognition that she reformulated the familiar cost functions used in the theory of the firm. She argued that the average cost of output is independent of the scale of production, but increases with respect to the rate of growth. Thus in so far as the average cost curve is U-shaped, it is U-shaped with respect, not to the scale of production, as commonly assumed, but to the rate of growth.

The simplest way to understand this postulate is to recognise that average costs are increased by *adjustments* in the rate of output. Changing the rate of output has a bigger impact on average cost than setting *steady state* output at a higher or lower level. Changes in the rate of output dislocate the allocation of resources. This is particularly true for human resources. Employees are usually most productive when they repeat the same routines; furthermore, when their work is repetitive, productivity may improve as a result of learning on the job. As a firm grows, the internal division of labour has to change, and this forces people to change their roles. Their previous learning of job-specific skills becomes obsolete, as they return to the start of the 'learning curve' in their new job.

Change is expensive in other ways too. Plant and machinery have to reallocated to different uses, and this process needs to be managed, creating additional demands on the management team. As the management team grows, new recruits need to be trained. Through leaving their previous employment with another firm, and joining the expanding firm, these recruits incur the same costs of retraining as those who have changed their jobs within the firm. Indeed, their training costs are greater because they know little about the institutional context of their job. To train recruits, experienced managers have to be diverted from their usual work, adding to the dislocation described above.

Costs of change may be related either to the absolute amount of change or the proportional amount of change. A case can be made for both, but in the Penrose model it is the proportional and not the absolute amount of change that matters. In alternative models of growth, however, the absolute amount of change is key.

The formalisation of Penrose's theory is based on Buckley and Casson's (2004) recent interpretation of her major work (Penrose 1959). The model analyses a firm that grows through diversification at a steady rate. The central point of the model is that the firm faces 'costs of growth' which increase, not with the size of the firm, but with its rate of growth. The key to the model is the specification of these costs of growth.

Penrose viewed the firm as a 'bundle of resources' because she saw resources – in particular human resources – as both the key to the firm's success and the principal constraint on its growth. According to her theory, the tacitness of information – on which modern resource-based theory places so much emphasis – not only protects the secrets of the firm's entrepreneurial success, but also inhibits the assimilation of the additional human resources required to sustain its growth. The growth rate of the firm reflects a balance between the entrepreneurial dynamism which drives its diversification, and the difficulty of enlarging the firm's management team to exploit the resultant opportunities.

In the basic model, the profitability of the firm is independent of its size and depends only on its rate of growth. This leads to a simple formula for the value of a growing firm, as determined by the present value of its future profit stream. Managers serving shareholder interests will maximise this value, whilst managers pursuing their own objectives – such as status – may maximise growth instead. Penrose considered both objectives; the differences are small, however, because maximising profits not only maximises shareholder value but also facilitates the internal financing of growth (see also Baumol 1967, Marris 1964).

Penrose's central point is that there is, in principle, no limit to the size to which a firm can grow.

- There are limits to the extent to which a firm can grow within a single market, which are set by the overall size of the market and the existence of competitors within it. But a firm can evade any such market size constraint by diversifying into other markets.
- It is often said that because of U-shaped average cost curves, there is a unique optimal size of firm at which average cost is a minimum. However, the logic of the U-shaped average cost curve applies to physical plant and equipment rather than to a managerial unit such as a firm. A firm that cannot expand beyond optimum plant size can expand by increasing the number of plants it operates, either through replicating plants in the same market (horizontal integration), moving into other stages of production in the same market (vertical integration) or diversifying into different markets.
- Misleading analogies have been employed to suggest that there is a limit to the size of firm. Marshall's metaphor of the firms in an industry resembling the 'trees in a forest' is misleading because it ignores the fact that firms can regenerate themselves through managerial succession. More significant still, firms can merge with each other to sustain their growth, and metamorphose into new forms, as when a small firm with highly-centralised autocratic management merges with other small firms and turns into a large highly-decentralised multi-divisional firm. As a legal and contractual entity, a firm can in principle endure for ever.
- Managerial diseconomies of scale are often said to limit the size of firms, but such limiting factors are more properly regarded as diseconomies of growth. In other words, the costs that limit the size of a firm at any time are costs that limit its continued growth, rather than costs associated with the size to which it has already grown. It is therefore more appropriate to use the concept of an optimum rate of growth of a firm than an optimum size of firm.
- In terms of international expansion, Penrose can be interpreted as considering foreign subsidiaries as autonomous companies. As such, they are beyond the reach of the firm's administrative coordination. The absence of authoritative communication would thus put them beyond the boundary of the firm. There is a potential modification of our formalisation in which the rate of growth of firms of different sizes is the same up to the extent of the reach of coordination and then zero beyond it. This is not a realistic interpretation of modern multinational firms, especially given managerial learning and technological breakthroughs in communication and control of international expansion.

When analysing growth, the natural analogue of a theory of the optimal size of firm is a theory of the optimal growth rate of the firm, and this is essentially what Penrose provides. Because size of firm does not matter, there is no reason to believe that the firm's growth rate will vary systematically over its lifetime, and so it is reasonable to postulate the existence of a steady state rate of growth. The existence of such a steady state implies that, in the limit, firms can last forever and so, with a constant rate of growth, they can eventually approach infinite size. To many observers of the corporate scene in the 1950s, it seemed as if large firms like General Motors would, indeed, continue to grow forever. Although this may now appear an extreme possibility, it was considered quite plausible at the time that Penrose was writing.

Let the size of the firm be measured by $x$. If growth proceeds continuously at a constant proportional rate $g$ then the size of the firm will increase exponentially:

$$x = X \exp(gt)$$

To simplify the subsequent discussion it is useful to normalise the size of the firm on its foundation to unity, $X = 1$. In terms of Penrose's theory, size is most naturally measured by the number of markets served by the firm; the normalisation therefore means that when the firm is newly established it operates in a single market.

All variables are non-negative unless otherwise stated. Discrete variables, such as the number of markets, are treated as though they were continuous.

For expository purposes it is useful to set out the model in three stages, with the final stage representing the model actually used in this paper.

## Stage 1

Suppose that the profit generated by a representative market is $A$, where $A$ is a constant independent of the size of the firm. As the number of markets grows at the rate $g$, total profit grows at the same rate:

$$\text{Profit at time } t = Ax = A \exp(gt).$$

If future profits are discounted at the rate $r$, and $r > g$, then the value of the firm at the time it is founded ($t = 0$) is

$$v = \int_0^\infty \exp(-rt)\, Ax\, dt = A / (r - g)$$

The factor $1 / (r - g)$, which multiplies into the profit per market, $A$, indicates the net rate at which profit is capitalised.

The first of the firm's objectives mentioned above corresponds to the maximisation of $v$. The second objective corresponds to the maximisation of $g$, subject to a constraint that the value of the firm is sufficient to support a dividend stream of acceptable size. If the dividend stream $d(t)$ grows at the same rate as the firm then

$$d(t) = D \exp(gt)$$

where $D$ is the initial level of dividend payments. To be feasible, initial dividend payments must be less than initial profit, $D \leq A$. Shareholder pressure fixes the minimum acceptable level, $D$, and managers then set dividends exactly equal to this level as part of their growth-maximising strategy. The value of the dividend stream is

$$v^d = D / (r - g)$$

and the managers maximise $g$ subject to the constraint that

$$v \geq v^d$$

If there are no costs of growth then, since $A$ is a constant, value is maximised by setting growth as high as possible. Since $g$ is constrained to be slightly less than $r$, this becomes the value-maximising rate of growth. Provided that $A \geq D$ there is no difference between growth-maximisation and value-maximisation: both objectives imply the largest possible rate of growth.

## Stage 2

To avoid this trivial result, it is necessary to introduce costs of growth. Let $B$ be a parameter measuring costs that are directly proportional to growth. It is necessary to choose the dimension of $B$ with care: it is the average cost per unit size generated per unit rate of growth. After deducting the cost of growth, net profit at time $t$ becomes $(A - Bg)x$, and the value of the firm at the outset becomes

$$v = (A - Bg)/(r - g)$$

It turns out, however, that costs which are directly proportional to growth do not fundamentally modify the trivial result above. If net profit is positive, $A > Br$, then value-maximising growth is just less than $r$, exactly as before, whereas if net profit is negative, $A < Br$, then value-maximising growth is zero. But in the latter case the firm has negative value, and so it will not be established in the first place. In the special case where profit is zero, $A = Br$, the firm is just viable and the value-maximising growth rate is indeterminate within the permitted range, $0 \leq g < r$. Thus the growth rate is either just less than $r$, indeterminate, or the firm is not established, depending upon whether profit is positive, negative, or zero.

**Stage 3**

To achieve a determinate value-maximising growth inside the permitted range it is necessary to introduce decreasing returns to the rate of growth through a quadratic cost term which reflects the decline of productivity as the rate of growth increases. This quadratic term, $Cg^2x$, represents the effects of dislocation, as described above. It is quadratic in $g$, the proportional rate of growth, rather than in $gx$, which is the absolute rate of growth, because it is the proportional rather than the absolute amount of dislocation which affects average cost.

Net profit at time $t$ becomes $(A - Bg - Cg^2)x$, and the value of the firm becomes

$$v = (A - Bg - Cg^2)/(r - g)$$

This is the fundamental formula for the value of the firm in a theory of the growth of the firm. The formula is implicit throughout Penrose's analysis.

The first order condition for value-maximisation is

$$v_g = ((A - Bg - Cg^2) / (r - g)^2) - ((B + 2Cg) / ((r - g)) = 0$$

where the subscript $g$ denotes the derivative with respect to $g$. The derivative $v_g$ measures the sensitivity of the firm's value to changes in its rate of growth, and comprises two terms. The first measures the impact of a change in growth on the capitalised value of initial profits, assuming that the profit stream remains unchanged, whilst the second term measures the cost of growth, which is equal to the value of the marginal reduction in the profit stream induced by growth, assuming that the capitalisation rate remains unchanged.

Placing both terms over a common denominator, $(r - g)^2$, and noting that the denominator is always positive over the permitted range $0 \leq g < r$, indicates that the first order condition is satisfied only when the numerator is zero. The numerator is quadratic:

$$(A - Br) - 2Crg + Cg^2 = 0$$

and its solution has two roots. These roots are real if

$$A - Br > 0; \qquad A - Br - Cr^2 < 0$$

*i.e.* if growth is profitable at the margin when growth is zero, and profit is negative when growth is equal to the discount rate, $g = r$. Only the smaller of the two roots lies in the permitted range; this root determines the optimal growth rate

$$g^* = r - (r^2 - ((A - Br)/C))^{1/2}$$

Under the assumed conditions, the numerator of the derivative always declines with respect to $g$, and since the denominator is always positive within the permitted range, $v_g$ switches from positive to negative as $g$ increases in the region of $g^*$, which guarantees that $g^*$ supports a maximum.

Differentiation of the growth equation shows that under the assumed conditions the partial derivates can be signed as follows:

$$g^*_A > 0; \qquad g^*_B < 0; \qquad g^*_C < 0; \qquad g^*_r < 0$$

Growth is higher, the higher is the profitability of the representative market, $A$, and the lower are the cost parameters, $B$, $C$. $B$ is less important than either $A$ or $C$ because its impact is mediated by the rate of cost of capital, $r$, which is normally substantially less than unity. The key insight of the growth formula, therefore, is that growth is governed by a fundamental trade-off between the profitability of the representative market, $A$, and the strength of decreasing returns to growth, as measured by $C$. Growth is also higher, the lower the cost of capital: an increase in the rate at which future profits are discounted discourages the firm from sacrificing current profit to promote future growth.

When growth is maximised subject to a dividend constraint, growth will expand up to the point where profit is zero, which implies that

$$(A - D) - Bg - Cg^2 = 0$$

This quadratic equation has two roots, and assuming that the dividend constraint is feasible, both are real. The smaller root is negative, and so the larger root must be taken, which is the root that corresponds to a maximum of growth:

$$g^{**} = ((B^2 + 4(A - D)C)^{1/2} - B) / 2C$$

Given the other assumptions, this root lies within the permitted range. Partial differentiation of the growth equation shows that:

$$g^{**}_A > 0; \qquad g^{**}_B < 0; \qquad g^{**}_C < 0; \qquad g^{**}_D < 0$$

Like the value-maximising growth rate $g^*$, the growth-maximising growth rate $g^{**}$ is heavily dependent on $A$ and $C$, but only marginally dependent on $B$, confirming that the optimal growth rate is basically determined by a trade-off between the profitability of individual markets, as measured by $A$, and decreasing returns to growth, as measured by $C$. Unlike the value-maximising growth rate, however, the growth-maximising growth rate is independent of the cost of capital $r$. The capital constraint operates through dividend commitments, $D$, instead. The higher is $D$, the less profit the firm can retain to finance set-up costs and entrepreneurial management, and hence the lower is the growth of the firm.

In view of the similarity in the results for value-maximisation and growth-maximisation, the rest of the paper considers just the value-maximising case.

## A Simple Multinational Analogue of the Penrose Model

In considering not simply the growth of the firm but also its internationalisation, it is important to note that Penrose discusses three main dimensions along which a firm can expand. While the main focus of her discussion is on product diversification, she also considers vertical integration along the supply chain, and geographical diversification which can turn the firm into an MNE. Geographical diversification can be either horizontal – the typical case where technology transfer is involved – or vertical, as exemplified by 'resource seeking' investments. In our model, as in Buckley and Casson (1976), the uni-national firm is simply a special case of a multinational firm.

Each of these three main dimensions has further sub-dimensions. Diversification can involve either different varieties of the same product, or wholly distinct products, or some combination of the two. New product varieties may substitute obsolete varieties, or complement existing varieties, or be independent of each other. Vertical integration can be either forwards – into wholesaling or retailing, for example – or backwards – into component manufacture or raw material supply. Vertical integration can also be applied to the product development process, in which the firm not only undertakes product development but integrates backwards into basic research, or forwards into quality control and warranty repairs. Finally, geographical diversification can be either local, inter-regional, international or global, depending on whether it spans regional boundaries or national boundaries, or covers the entire world.

With so many different dimensions, it is useful to focus the discussion on a single representative case. Vertical integration is therefore ignored; it is assumed that the firm concentrates on just a single stage of production. It innovates an expanding range of products which are then introduced to different national markets.

Suppose to begin with that each new product is launched simultaneously in every national market. Thus the 'markets' into which the growing firm diversifies are, in effect, global markets. Whilst the introduction of products is sequential, their introduction to individual local markets is not.

The fact that the firm serves global markets does not necessarily make it a multinational, however. The firm only becomes a multinational if its location strategy involves production in more than one country. The firm could be simply a specialised exporter which serves all its markets from the same host-country location.

In the Penrose model, an exporter can become a multinational in two possible ways. It can serve some markets through local production, and therefore become multinational as a result of its geographical expansion; alternatively, it can serve each global market from a single plant, but locate different plants in different countries, in which case multinationality is driven by product expansion instead. Multinationality driven by geographical expansion is a response to obstacles to transport and trade – connected, for example, with the perishability of the product, or the need for face-to-face delivery – whilst multinationality driven by product expansion exploits comparative advantage by specialising the production of each product on the location best endowed with the relevant non-tradable inputs.

When geography is the driver, the firm becomes a multinational whilst still a single-product firm, but when the product is the driver it becomes a multinational only when it becomes a multi-product firm. These two effects are combined in globally rationalised production, in which plants in different countries serve not only local markets but act as export platforms for particular products in the range. Case study evidence suggests that single-product firms can be multinational, which suggests that geography is a more important driver than the product. However, the interaction of the geographical dimension and the product dimension is relevant to the analysis of the globally rationalised firm.

A more refined approach to geographical expansion interprets market entry as a sequential process. Instead of entering all national markets simultaneously, the firm enters them one at a time. The basic unit of analysis is no longer an integrated global market for a product, but a particular national market for a particular product. The firm expands sequentially by first introducing one product into one market, and then introducing another product into another market, and so on.

Sequential entry into national markets for individual products can be effected in several ways. Firstly, the firm can expand initially in the product dimension, selling entirely in the home market, and then, once a critical number of products have been introduced, expand the product range into foreign markets. A second method is the converse of this: the firm expands first in the product dimension and then in the geographical dimension. The firm launches its first product into a set of key markets and only introduces its second product once this process is complete; the second product is introduced into the same set of markets, and then the third product; and so on.

A more plausible process is one in which products are innovated into the home market whilst they are in the process of diffusing through international markets. To begin with, the firm introduces its first product into the home market. If this product is successful, it introduces the product to a foreign market in the second period and, shortly afterwards, introduces a second product to its home market to capitalise on its initial success. Next, the first product is introduced to a third country, then the second product is introduced to the second country, and then a new product is introduced to the home market; and so on. The process may be termed a 'innovation-diffusion' process, since products are innovated at the same time that established products are diffusing into national markets.

The innovation-diffusion model also predicts that the rate of introduction of new products into any given market will decline over time, even though the firm is growing at a constant proportional rate. This is because the firm is extending the geographical scope of its markets as it grows. Each stage of growth is marked by the entry of a product into another national market, rather than the innovation of another new product *per se*. Thus the rate of product innovation slows as the number of national markets served increases; more effort goes into diffusion, and less into innovation.

The results generated by the innovation-diffusion model seem to be broadly consistent with evidence about the growth of MNEs in the second half of the twentieth century. Furthermore, the general approach is consistent with both Vernon's product cycle theory of foreign direct investment (Vernon 1966), and with the Upsala theory of sequential internationalisation (Johanson/Vahlne 1977). Although it goes beyond both. A further development of this approach is presented below.

**Table 1a.** Sequential Entry into National Markets with a Fixed Product Range

| | Location 1 | Location 2 | Location 3 | Location 4 |
|---|---|---|---|---|
| Product 1 | 1 | 5 | 9 | 13 |
| Product 2 | 2 | 6 | 10 | 14 |
| Product 3 | 3 | 7 | 11 | 15 |
| Product 4 | 4 | 8 | 12 | 16 |

**Table 1b.** The Innovation-Diffusion Model: Sequential Product Innovation with Overlapping Sequential Local Entry

| | Location 1 | Location 2 | Location 3 | Location 4 |
|---|---|---|---|---|
| Product 1 | 1 | 2 | 4 | 7 |
| Product 2 | 3 | 5 | 8 | 11 |
| Product 3 | 6 | 9 | 12 | 14 |
| Product 4 | 10 | 13 | 15 | 16 |

## A Contrast Between the Penrose Model and that of Buckley and Casson

A potential weakness of the Penrose approach to multinationals is that the growth of the firm is driven by product diversification rather than by technological innovation. Empirical evidence points to the crucial role of technology in stimulating the growth of multinational firms. Early post-war US foreign direct investment, for example, was heavily concentrated in technology-intensive industries.

Although Penrose recognised that large firms invested significantly in R&D, she regarded their entrepreneurial capabilities as the main driver of their growth. She believed that there were abundant opportunities for discovering new markets, irrespective of whether R&D was undertaken or not. The main constraint on the growth of the firm was not the need to finance expensive R&D, but the difficulty of expanding the firm at a sufficient rate to exploit all of the opportunities available.

In fact, decreasing returns to R&D provides a simple explanation of how the growth of a technology-intensive firm is constrained. When market opportunities drive firm growth, as in Penrose, then decreasing returns to entrepreneurship and management will constrain growth, while if technological opportunities drive the firm's growth then decreasing returns to R&D will constrain growth instead. There is, therefore, a logical parallel between the limits to growth identified by Penrose, and the limits to growth associated with R&D.

An emphasis on R&D also helps to explain why related diversifications are generally more successful than unrelated diversifications in sustaining the growth of the multinational firm. If new products are generated sequentially from a single integrated programme of research, then the cost of generating any one product can be reduced by using knowledge spill-overs from other products. Furthermore, products generated in this way may also be related in terms of the materials or components from which they are produced, thereby generating supply chain economies once they are in production. Although relatedness can also be achieved through 'bundling' products for sale, or marketing different products to the same group of customers, there is little doubt that the technological relatedness of different products has been a significant factor in boosting the profits and growth of modern multinational firms.

As it happens, the link between R&D and the growth of the firm was an important part of the model of the multinational enterprise developed by Buckley and Casson (1976). This was no accident; they introduced internalisation theory in order to explain the 'pattern of growth' of the MNE in terms of 'a long-run theory' (p. 59), and they concluded their exposition with 'a mathematical model of the growth of the research-intensive firm' (p. 62). Unlike the later work of Dunning (1977), their model did not assume that the firm possessed a given ownership advantage. Instead, it considered the generation of a stream of ownership advantages from a continuous process of R&D. The internalisation of the intangible knowledge flows generated by R&D established a link between the steady state *level* of R&D and the steady state *rate of growth* of production and sales. Just as in Penrose's model, decreasing returns to the growth of the management team restrict the rate of product diversification by an entrepreneurial firm, and hence limit its rate of growth, so in the Buckley and Casson model decreasing returns to the scale of R&D restrict the rate of product innovation by a technology-intensive firm, and hence limit its rate of growth. The similarity between the two theories is not surprising, since Buckley and Casson traced back their ideas, not only on Coase (1937), but also to Penrose (1959) (p. 36, fn 2).

According to Buckley and Casson, R&D activity has a U-shaped average cost curve. This results from the interplay of fixed costs and increasing variable costs. R&D incurs fixed costs due to the indivisible nature of laboratory facilities, and the need for a critical mass of different specialists to be combined within a research team. Beyond a certain scale, however, decreasing returns set in because the advantages of further specialisation diminish, and the team becomes increasingly difficult to manage, as individual researchers start to pursue independent lines of research unrelated to the rest of the team. By contrast, production and marketing operate with constant average costs.

Although Buckley and Casson relate decreasing returns in R&D to the *level* of R&D activity, the output of R&D activity determines the *rate of growth* of the final output of the firm. Thus an analysis of levels in R&D translates naturally into an analysis of growth in terms of output.

In the Buckley and Casson model, R&D generates a continuous stream of new products, each of higher quality than the previous one. Once introduced, each product is sold immediately in a global market; the pattern of internationalisation is summarised in Table 2. National markets are highly segmented, on account of transport costs and tariffs, and so each national market is serviced by wholly-owned production and distribution subsidiaries. In each market the firm enjoys market power, derived from the novelty of its product line, and so faces a downward-sloping demand curve. Both the scale of demand, and its price-elasticity, vary between markets, as do average costs.

The firm maximises shareholder value, which is equal to the value of the aggregate profit generated by the local markets, net of the costs of R&D. Sustained R&D leads to continuous improvement in quality, and this in turn maintains steady growth of demand in each market. The higher the level of R&D, the faster is the rate of quality improvement, the faster the growth of demand and aggregate profit, and the higher the cost of R&D. The optimal growth of the firm is determined by a trade-off between growing revenues on the one hand, and the higher costs of R&D on the other. This trade-off determines the rate of optimal level of R&D activity. This in turn determines the rate of quality improvement, and hence the rate of growth of output.

Each incremental improvement in quality yields the same incremental increase in profits. The discounted value of this additional profit is traded off against the marginal cost of generating the improvement in quality. The value-maximising strategy equates the marginal value of an increase in quality to its marginal cost, and determines a rate of quality improvement which is constant over time, and corresponds to a steady state level of expenditure on R&D. The optimal growth rate is an absolute rather than proportional rate, so that the growth of sales is linear in time.

**Table 2.** The Buckley and Casson Model: Sequential Product Innovation with Simultaneous Local Entry

| | Location 1 | Location 2 | Location 3 | Location 4 |
|---|---|---|---|---|
| Product 1 | 1= | 1= | 1= | 1= |
| Product 2 | 2= | 2= | 2= | 2= |
| Product 3 | 3= | 3= | 3= | 3= |
| Product 4 | 4= | 4= | 4= | 4= |

The optimal rate of growth is greater, the lower are marginal costs – *i.e.* the lower the salary costs of researchers, and the more slowly decreasing to returns to the scale of research set in. Growth is also greater the higher the marginal value of quality improvement – *i.e.* the more intensive the demand for the product, the lower its price-elasticity, and the lower the cost of capital.

In both the Buckley and Casson and Penrose models, the firm faces a trade-off between faster exploitation of market opportunities on the one hand, and 'costs of growth' on the other. In both cases, costs of growth increase because the productivity of human resources declines as the rate of growth they are required to sustain increases. But there are key differences too.

- In Penrose's model it is the declining productivity of the management team that limits growth, whereas in the Buckley and Casson model it is the declining productivity of scientific researchers instead.
- In Penrose's model declining productivity results from the continuous expansion of the size of the management team whereas in the Buckley and Casson model declining productivity arises from sustaining a constant rate of R&D above the scale which minimises average cost.
- The innovations made by the firm are different too. In the Penrose model, the firm diversifies into new markets, whilst in the Buckley and Casson model it upgrades the quality of a given product.
- Although the firm introduces new products in both cases, in the first case the new products are independent of existing products, whereas in the second case each new product replaces the previous version of the same product.

A weakness of the Buckley and Casson model is that it cannot be used to analyse the sequential nature of international expansion. Sequential entry is predicted only when new markets emerge as a result of the liberalisation of access, or a 'take-off' in local demand stimulated by economic growth. Thus while the Buckley and Casson model can be applied to the entry into 'emerging markets' in 'transition economies', it cannot be applied to sequential entry into mature and accessible markets. It is shown below that this weakness of the Buckley and Casson model is easily overcome.

## A Reconciliation between the Two Models

The question arises as to how far it is possible to integrate the Buckley and Casson approach with the basic Penrose model. This section presents a model which attempts to do just this. It is a two-dimensional generalisation of the one-dimensional Penrose model presented in section 2.

The model focuses on a distinctive type of firm, founded by a Schumpeterian entrepreneur (Schumpeter 1934). The entrepreneur has identified a technological opportunity to generate a range of new products. The entrepreneur is not an inventor, but an innovator; he recognises not merely a technological possibility, but also a latent demand for products that embody the new technology that he plans to exploit. Although product demand is global, national markets are spatially self-contained, and so each needs to be entered separately.

Having identified an opportunity for technological innovation, two further stages are involved in bringing a product to market. The first stage is R&D and the second is market entry – first into the home market and then into foreign markets. Multinationality arises because each national market is sourced by local production.

Each stage involves its own distinctive form of sequential diversification. The firm can only develop one product at a time, and each product can only be introduced into a new market by entering one country at a time. The two stages proceed in parallel. While R&D generates a sequence of product innovations, market entry takes each product and introduces it to a sequence of national markets. Unlike the growth-diffusion model presented above, two separate sets of resources are involved – a team of researchers responsible for product innovation, and a team of managers responsible for market entry. Each process is sequential because this arrangement makes the best use of the human resources involved.

Because of the two dimensions, there are two different growth rates: one associated with a growth in the number of products, $g_1$, and the other associated with the growth in the number of national markets served, $g_2$. The two growth rates are distinct, but related: the growth of the product range can be chosen independently of the speed of internationalisation, and *vice versa*; however, conditions which favour faster growth in the product range normally favour faster internationalisation too.

Following the general principles of the Penrose model, it is assumed that the stock of products, $x$, introduced by the firm grows at a constant proportional rate, $g$, which is chosen by the firm These products are generated exclusively by the firm's R&D; no technology is licensed in or out to other firms. It is assumed that the firm has already developed a prototype product by the time that it is founded, so that $x = 1$ and time $t = 0$. Thus

$$x = \exp(g_1 t)$$

as in the one-dimensional model. To further simplify the model, it is assumed that the prototype is not sold, so that all the firm's revenue and profit is attributable to its R&D.

The rate of product innovation depends upon the number of scientists employed in R&D. Employment in R&D is $n_0$. Each scientist receives a fixed salary $s_0$. There are constant returns to the scale of R&D; other thing being equal, doubling the employment of scientists doubles the rate of new product generation. Due to the real costs of growth, however, the average productivity of scientists declines with the rate of growth of R&D activity; as a result, the real cost of scientific output is $a_0 g$. The parameter $a_0$ measures the labour-intensity of research; it reflects the interaction between the skills of the scientists and the technological opportunities available. Thus average research productivity is independent of scale, but inversely related to growth.

When growing at rate $g_1$ with a cumulative stock of products $x$, the firm must introduce new products at a rate $g_1 x$. With real labour cost $a_0 g_1$, the firm requires $a_0 g_1^2 x$ scientists. When scientific labour is the sole input to R&D, the cost of R&D is therefore $C g_1^2 x$, where $C = a_0 s_0$.

It is assumed that each product, once produced, has the property of a public good within the firm. This has a very specific meaning in the present context; namely that, once developed, product technology can be transferred to any country in the world. Technology is not an absolutely pure public good, because there is a positive cost of transfer to each national market. But it is a public good in the sense that a product does not need to be reinvented for every national market. In other words, R&D generates global products rather than products that can be sold only in a single national market.

National markets vary in size. A large market has a large demand for all products. The operating profit generated by each product in each national market is directly proportional to the size of the market. Each product in each market earns a uniform rate of profit per unit size, $z$.

Entry into each market requires an input of managerial labour which is directly proportional to market size, $y$. No learning takes place in the 'roll out' of the product from one market to the next, and so the cost of entry is independent of the number of markets already entered.

All markets are entered at the same speed, independent of their size. The speed of entry is chosen by the firm. Fast entry is effected by employing a large number of managers for a short period of time, and slow entry by employing a smaller number of managers for a longer period of time. No learning takes place with regard to product innovation in a given market, so that the same entry same cost is incurred irrespective of how many previous products have been introduced to the market.

Speeding up entry expedites internationalisation, by bringing forward the time of entry into subsequent markets. It also reduces the productivity of managerial labour, however. With a speed of entry $h$, total managerial input is $a_1 yh$, where $a_1$

is a productivity parameter. When managers receive a salary $s_1$, the cost of entering a market at speed $h$ incurs a management resource cost $a_1 s_1 yh$.

The optimal strategy for the firm, given a positive uniform rate of profit and a commitment to sequential entry, is to enter the largest market first, and then to enter subsequent markets in descending order of size. It is assumed, therefore, that entry is sequenced in descending order of market size.

The precise speed of market entry chosen depends upon the firm's cost of capital, $r$. Once the optimal speed of market entry has been determined, using the method described below, the value, $B$, of a new innovated product can be calculated. This value is equal to the discounted value of the operating profits generated by the growing global market for the product, less the discounted value of the costs of product launch in each of the national markets. Since new products are introduced at a rate $gx$, the value of the output from R&D is $B\, g_1 x$.

The overall value of the firm is therefore

$$v = \int_0^\infty \exp(-rt)\,(B\, g_1 x - C\, g_1^2 x)\, dt = (B\, g_1 - C\, g_1^2) / (r - g_1)$$

The value $v$ is finite if and only if $g_1 < r$. The first order condition for a maximum of $v$ is a quadratic equation, which has real roots if and only if $B < Cr$. The condition for positive profit is $B > C\, g_1$, and both are satisfied simultaneously only when $g_1 < r$, *i.e.* when the firm has finite value.

Only the smaller root satisfies this latter condition; this is also the root that fulfils the second order condition for a maximum. Optimal growth is therefore

$$g_1 = r(1 - (1 - (B/Cr))^{1/2})$$

It can be seen that the optimal rate of growth increases with the value of new products, $B$ and decreases with the costs of R&D, $C$. For a given $B$, $g_1$ is a decreasing function of $r$; a high cost of capital reduces the optimal rate of growth. To complete the solution of the model, it is necessary to determine $B$.

## The Optimal Speed of Internationalization

Consider the marketing of a representative new product which has just been developed through R&D. A sequential process of internationalisation process can be modelled in the following way.

It is assumed that there is infinite number of national markets, the largest of which has size $y^+$, and the smallest size zero. They are ranked in descending order of size by a continuous index $m$. It is assumed for simplicity that market size decreases exponentially with rank, such that the size of the $m$th market is

$$y(m) = y^{+} \exp(-bm)$$

where $y^{+}$ is the size of the largest market, and $b$ is a parameter that measures the rate at which size decreases with respect to rank. The parameter $b$ indicates the inequality of market sizes, with a value close to zero indicating relative equality and a high value indicating extreme inequality.

Because there is an infinity of national markets, the internationalisation process continues indefinitely, thereby allowing the firm to expand continuously in the geographical dimension. At the same time, the decreasing size of successive markets means that the total size of the global market remains finite.

By the time the $m$th market has been reached, sales of the product have cumulated to

$$Y(m) = (y^{+}/b)\,(1 - \exp(-bm))$$

and so, as globalisation proceeds, total market size converges exponentially on the 'saturation level' of the global market, $y/b$. The size of the global market is greater, the greater the size of the largest market, $y^{+}$, and the lower the inequality in market size, $b$.

The speed of market entry is measured by the rate at which the rank of the newest market advances over time. Since the speed of entry, $h$, is the same for every market,

$$m = ht$$

Cumulative market size converges on the saturation level at a rate $bh$:

$$Y(t) = (y^{+}/b)\,(1 - \exp(-bht))$$

Because markets are entered in order of decreasing size, the proportional rate of growth of the market, $g_2$, diminishes over time, as a higher proportion of the global market is served:

$$g_2 = bh\,((y/bY) - 1)$$

As noted earlier, the cost of entry into each market is proportional to its size. If the cost of entry were independent of size, it would soon become uneconomic to enter the smallest markets, and so a rational firm would terminate internationalisation at some point before the global market was saturated. If the firm persisted beyond this point, it would incur significant costs for little benefit, and the consequent drive to minimise these costs would distort the globalisation strategy.

Given earlier assumptions, the number of markets expands at a rate $h$, and so expenditure on market entry at any time $t$ is

$$c_1 = a_1 s_1 y h^2 = a_1 s_1 h^2 y^+ \exp(-bht)$$

Net cash flow is $zY - c_1$, and so the present value of the product at the time of its launch, $u$, is

$$u = \int_0^\infty (zY - c_1)\, dt = (y^+/b)[(z/r) - [(z + a_1 b s_1 h^2)/(r + bh)]]$$

The first term, $y^+/b$, measures the size of the global market, and scales the rest of the expression. The next term, $z/r$, is the capitalised value of the profit stream from a market of unit size. The remaining two terms – $z$ and $a_1 b r s_1 h^2$ – are deductions from profit, and both are capitalised using a rate of discount $r + bh$. This rate of discount reflects the gradual nature of the build up to the saturation of the global market, which is governed by the product of the inequality parameter, $b$, and the speed of market entry, $h$. The first term reflects the deferral of profit, and the second reflects the cost of market entry.

The first order condition for a maximum of $u$ generates a quadratic equation for the optimal speed of market entry, $h$. One of the roots of the quadratic is positive and the other negative. The second order conditions show that the positive root corresponds to a maximum, $h^*$:

$$h^* = (r/b)\,((1 + (b/a_1 s_1 r^2))^{1/2} - 1)$$

It can be seen that speed, $h^*$, increases with profit, $z$, and decreases with inequality of market size, $b$, the real cost of management time, $a_1$, and the salary level, $s_1$. It also increases with respect to the cost of capital, $r$: the higher the rate of discount, the more important it is to generate profits from the smaller markets as early as possible. A faster speed translates into a higher rate of geographical growth, $g_2^*$.

Substituting the optimal value of $h$ into the valuation of the product gives the optimal value of the product at the time of its development as

$$B = (y^+ z/br)\,(1 - 2(a_1 s_1 r^2/bz)^{1/2})$$

whence the optimal rate of growth $g_1^*$ becomes

$$g_1^* = r(1 - (1 - (y^+ z/a_0 s_0 b r^2)(1 - 2(a_1 s_1 r^2/bz))^{1/2})$$

It follows that growth is an increasing function of maximum market size, $y^+$, and profitability, $z$, and a decreasing function of the inequality of market size, $b$, the real cost of researchers, $a_0$, the real cost of management, $a_1$, and the respective salary levels, $s_0$, $s_1$.

A high cost of capital, $r$, reduces the optimal rate of growth, $g_1$* . The present value of the prospective revenues is reduced – especially the revenues generated by the smaller markets which are the last to be entered. Although a higher cost of capital encourages internationalisation to be speeded up, this speed-up is at the expense of higher entry costs, which further reduce the net value of the net profit stream. Thus a higher $r$ reduces the value of $B$, and thereby reduces the incentive for growth, $g_1$*.

It should be noted that the overall growth of the firm's sales depends on the interplay between the two growth rates $g_1$*, $g_2$*. In its early phase of expansion the firm expands quickly in both the product and geographical dimensions, but gradually converges to an overall growth rate $g_1$ as the internationalisation process matures.

## Synthesis

While Penrose's dismissal of scale effects is plausible in the context of many production processes, it is not particularly plausible where R&D is concerned. Researchers are often motivated by personal curiosity which leads them away from their intended line of research, while the introspective nature of research activity makes it difficult to monitor. Control loss is therefore a genuine concern in large teams. Furthermore, the notion that all research facilities have to start small and then grow as fast as possible is difficult to reconcile with the way that many research establishments are founded on a substantial scale. Many successful firms are able to grow large without continually expanding the sizes of their research facilities.

But while the notion of incremental expansion does not fit well with research, it fits much better with the internationalisation process, although even here the speed with which some firms internationalise has been quite high. Nevertheless, the notion that the speed of internationalisation is constrained by the difficulty of enlarging the management team sufficiently fast is more plausible than an assumption that internationalisation is instantaneous.

It appears, therefore, that while the Penrose model offers a superior account of internationalisation, the Buckley and Casson model offers a superior account of innovation and R&D. This suggests that the two models should be combined by taking the analysis of internationalisation in section 6 and combining it with the approach to innovation summarised in section 4.

Following the Buckley and Casson model, let the costs of R&D be

$$C = C_0 + C_2 q^2$$

where $C_0$ is the fixed cost of the R&D facility, $C_2 = a_0s_0$ is a parameter governing variable costs, and $q$ is the output of R&D. Output is measured by the rate of product innovation that it sustains. The cost function corresponds to a U-shaped average cost curve with a minimum at $q = (C_0/ C_2)^{1/2}$.

The value of an innovation exploited through an optimal internationalisation strategy is $B$, as derived above. The value of the firm is

$$v = \int_0^{\infty} \exp(-rt)(Bq - C_0 - C_2q^2)\, dt \quad = (Bq - C_0 - C_2q^2)/r$$

The first order condition for a maximum of value determines the optimal rate of research output, $q^*$

$$q^* = B/2C_2$$

The second order conditions for a maximum are always satisfied. Substituting for $B$ expresses $q^*$ as an increasing function of market size, $y^+$, the profitability of sales, $z$, and a decreasing function of the inequality of market size, $b$, the real costs of R&D and market entry, $a_0$, $a_1$, the salaries of scientists and managers, $s_0$, $s_1$, and the cost of capital, $r$:

$$q^* = (y^+z/2a_0s_0br)\,(1 - 2(a_1s_1r^2/bz)^{1/2})$$

In contrast to the exponential growth of the Penrosian model, the firm now grows in the long run at a constant absolute rate $q^*$. Sales grow faster in the early stages of internationalisation, but converge on this level as the internationalisation process matures. Since the rate of growth is constant in absolute terms, it declines in relative terms, which is consistent with the lifetime growth pattern of a typical firm.

This model is simpler than the pure Penrosian model, and offers a more realistic account of the R&D process. Because it is simpler, if offers greater scope for further development, and can therefore be recommended as a starting point for further research on this subject.

## Conclusion

This paper has shown that Penrose's *Theory of the Growth of the Firm* provides a tractable formal model which has important implications for the strategy of MNEs. Its analysis of the appropriate modes of internationalisation can be integrated with a satisfying account of the trade off between (product) diversification and foreign market penetration. The account of speed of entry is an advance on current theories of internationalisation including "stages" and product cycle based models. There is much profitable work to be done in extending Penrose's implicit model of multinational enterprise. For now, we can recognise the contribution of Penrose's book to the analysis of geographical expansion patterns, sequential decision making and learning in the MNE as key factors in international strategic management.

## Acknowledgements

The authors would like to thank Christos Pitelis for his comments on earlier versions of this article.

## References

Barney, J. B., Strategic Factor Markets, *Management Science,* 32, 10, 1985, pp. 1231–1241.
Baumol, W. J., *Business Behavior, Value and Growth,* New York: Harcourt, Brace, Jovanovich 1967.
Best, M. H./Humphries, J., Edith Penrose: A Feminist Economist?, *Feminist Economics*, 9, 1, 2003, pp. 47–73.
Buckley, P. J./Casson, M., *The Future of the Multinational Enterprise*, London: Macmillan 1976.
Buckley, P. J./Casson, M., The Theory of the Growth of the Firm: Extending the Penrose Model, *mimeo* 2004.
Caves, R. E./Porter, M. E., From Entry Barriers to Mobility Barriers: Conjectural Variations and Contrived Deterrence to New Competition, *Quarterly Journal of Economics*, 91, 3, 1977, pp. 241–262.
Coase, R. H., The Nature of the Firm, *Economica*, 4, 4, 1937, pp. 387–405.
Demsetz, H., Industrial Structure, Market Rivalry, and Public Policy, *Journal of Law and Economics*, 16, 1, 1973, pp. 1–10.
Demsetz, H., *Efficiency, Competition and Policy*, Oxford: Basil Blackwell 1982.
Dunning, J. H., Trade, Location of Economic Activity and the Multinational Enterprise: The Search for an Eclectic Approach, in Ohlin, B. et al. (eds.), *The International Allocation of Economic Activity,* London: Macmillan 1977.
Dunning, J. H., The Contribution of Edith Penrose to International Business Scholarship, *Management International Review*, 43, 1, 2003, pp 3–19.
Foss, N. J., *Resources, Firms and Strategies*, Oxford: Oxford University Press 1997.
Foss, N. J., Edith Penrose, Economics and Strategic Management, in Pitelis, C. N. (ed.), *The Growth of the Firm: The Legacy of Edith Penrose*, Oxford: Oxford University Press 2002.
Johanson J./Vahlne, J. E., The Internationalisation Process of the Firm: A Model of Knowledge Development and Increasing Market Commitments, *Journal of International Business Studies*, 8, 1, 1977, pp. 23–32.
Kor, Y. Y./Mahoney, J. T., Edith Penrose's (1959) Contributions to the Resource-based View of Strategic Management, *Journal of Management Studies*, 41, 1, 2004, pp. 183–191.
Lippman, S. A./Rumelt, R. P., Uncertain Immutability: An Analysis of Inter-firm Differences in Efficiency under Competition, *Bell Journal of Economics*, 13, 3, 1982, pp. 418–438.
Lockett, A./Thompson, S., Edith Penrose's Contributions to the Resource-based View: An Alternative Perspective, *Journal of Management Studies*, 41, 1, 2004, pp. 193–203.

Marris, R. L., *Economic Theory of Managerial Capitalism*, London: Macmillan 1964.
Penrose, E. T., *The Theory of the Growth of the Firm*, Oxford: Blackwell 1959.
Penrose, E. T., The Growth of the Firm: A Case Study: The Hercules Powder Company, *The Business History Review*, 34, 1, 1960, pp. 1–20.
Penrose, P./Pitelis, C. N., Edith Elura Tilton Penrose: Life Contribution and Influence, in Pitelis, C. N. (ed.), *The Growth of the Firm: The Legacy of Edith Penrose*, Oxford: Oxford University Press 2002.
Peteraf, M. A., The Cornerstones of Competitive Advantage: A Resource-based View, *Strategic Management Journal*, 14, 3, 1993, pp. 179–91.
Pitelis, C. N. (ed.), *The Growth of the Firm: The Legacy of Edith Penrose*, Oxford: Oxford University Press 2002.
Pitelis, C. N., Edith Penrose and the Resource-based View of (International) Business Strategy, *International Business Review*, 13, 4, 2002, pp. 523 – 532
Rugman, A. M./Verbeke, A., Edith Penrose's Contribution to the Resource-based View of Strategic Management, *Strategic Management Journal*, 23, 7, 2002, pp. 769–80.
Rugman, A. M./Verbeke, A., A Final Word on Edith Penrose, *Journal of Management Studies*, 41, 1, 2004, pp. 205–217.
Schumpeter, J. A., *The Theory of Economic Development*, Cambridge, MA: Harvard University Press 1934.
Teece, D. J., Economics of Scope and the Scope of the Enterprise, *Journal of Economic Behavior and Organization*, 1, 3, 1980, pp. 223–233.
Vernon, R., International Investment and International Trade in the Product Cycle, *Quarterly Journal of Economics*, 80, 2, 1966, pp. 190–207.