Go Back

Simulation of the business firm using several different pricing approaches

Paper to be presented at the
19th International Conference of the System Dynamics Society
Emory Hotel and Conference Centre, Atlanta,
23-27 July, 2001

Malcolm Brady

Dublin City University Business School
Glasnevin
Dublin 9
Ireland

PhD Student at
Lancaster University Management School
Lancaster
England

Tel. +353 1 7005188
Fax +353 1 7005446
Email malcolm.brady@dcu.ie
Web site: http://webpages.dcu.ie/~bradym

Abstract

This paper examines the behaviour of firms over time using different economic models as a basis for setting price. A simulation model was created in order to examine the behaviour of the economic models over time. The paper examines the behaviour of a firm using two standard neoclassical economic models: a model built on the principle of marginalism, ie. the firm produces at the point where marginal cost equals marginal revenue, and a model built under the assumption of average cost pricing. The paper also examines the behaviour of a firm over time using a variant of the average cost model developed by the author. The results of the simulation show that the long run behaviour of the firm is significantly different depending on the pricing model chosen.

This work is part of the author's PhD research and represents ongoing rather than completed work. Please do not quote without prior permission.

Keywords: firm, dynamic, behaviour, growth, pricing, average cost

Introduction

This paper represents part of the author's ongoing PhD research into the growth and decay of firms over time. Growth of firms has long been a subject of research in managerial and economic research. Baumol (1962) introduced a model of firm behaviour based on

maximising the growth rate of the firm. Marris (1963) introduced a model of the behaviour of the managerial firm based on the competing objectives of firm growth and security from takeover or bankruptcy. Penrose (1980) in her theory of the growth of the firm suggested that pressure for expansion arose due to human resources being underutilised and, ironically, suggested that growth was limited by the capabilities of the management team. Higgins (1977) developed a model of sustainable growth whereby firm financial policies do not conflict with growth policy.

Economic growth and its sustainability in the broader context of the world economy has also been a subject of research. Douthwaite (1992:20) suggests that 'in our present system, the choice is between growth and collapse, not growth and stability. No wonder people want growth so badly…It is easy to see why businesses and governments constantly strive to create growth, since the alternative is debt, depression, unemployment and commercial disaster'. The Club of Rome examined the consequences of the then exponential growth in the world economy using a system dynamics modeling approach and concluded that if those trends continued a limit to economic growth would be reached within a hundred years (Meadows et al, 1972:23).

This particular paper examines the long run behaviour of firms using several different approaches to pricing products. Two of the approaches are well known and are founded in neoclassical microeconomic theory. In the first approach the marginalist principle is used to set price: the firm produces at the point where the marginal revenue gained from the sale of the product is equal to the marginal cost of the product. The marginalist approach is founded on the laws of calculus and is an optimising approach; it is also largely driven by the demand function faced by the firm. In the second approach the average cost pricing principle is used: the firm sells its product at cost plus a margin for profit. The average cost approach is founded on observation of the practice of real firms. In the average cost approach behaviour of the firm is largely driven by supply ie. the production of goods or services. Finally, a third approach, developed by the author is examined. Here, the behaviour of the firm is examined taking both supply and demand into account. This model is referred to as 'average cost including demand' model for the purpose of this paper. The results of this model are examined and compared with the results of the first two approaches.

The economic models consist of a set of equations comprising both identities and behavioural equations. The identities are definitional equations and are based on standard economic or accounting practice eg. profit = revenue - cost. The behavioural equations are based on some assumptions. These assumptions may be based on standard economic theory, for example production is modeled as a Cobb-Douglas function, or may be based on well-known rules of thumb, observation or personal experience.

This economic model of the firm is shown as a system in figure 1 as a block diagram (Shearer et al, 1971). Here the behavioural equations are shown as square boxes and the circular junction points represent how the outputs of these functions are combined. The functions f(t), c(t), d(t), p(t), m(t) represent production, cost, demand, price, and marketing/ administrative/ depreciation expenses respectively. The functions I(t), T(t), and D(t) represent the interest, tax and dividend respectively. The model is circular in that retained earnings are combined with original capital to give assets. The dotted line between m(t) and p(t) indicates that marketing activity may influence demand. The flows in the model are represented by arrows and refer to cash. Several cash takeoff points are shown on the diagram: to labour and suppliers, to bondholders, to government, to shareholders, and to the firm by way of retained earnings.

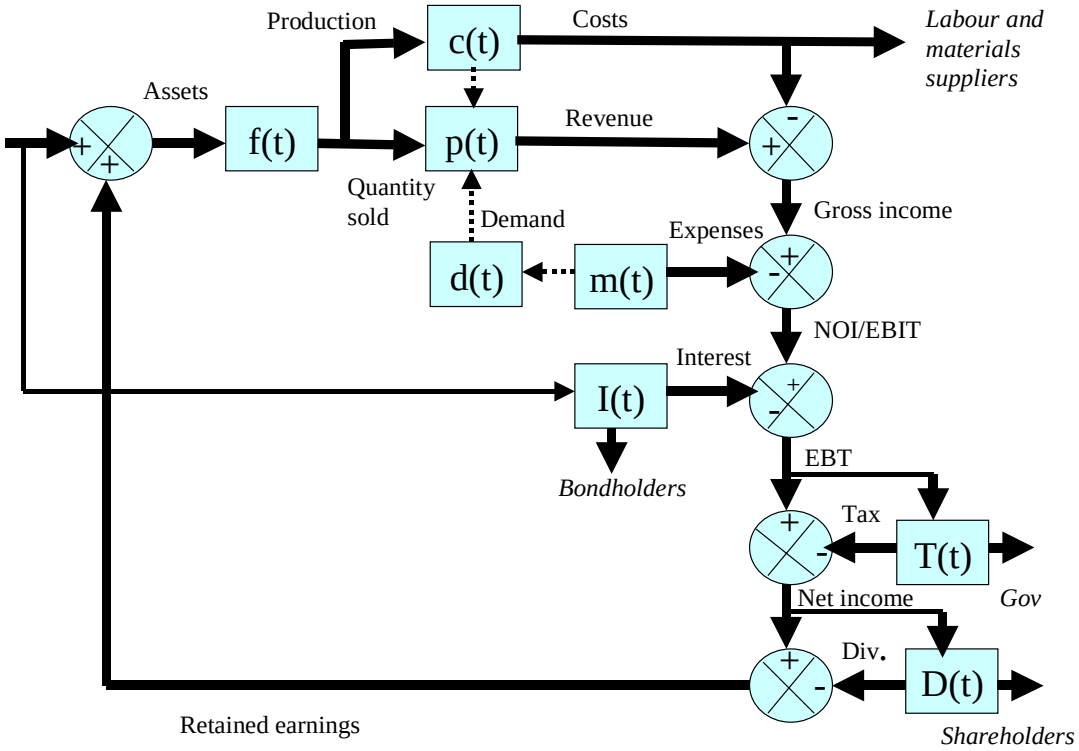


Figure 1. Block diagram of the firm as a system.

The three modeling approaches examined in this paper differ only in how the pricing function p(t) operates. In all other respects the three models are identical. The average cost approach determines price from cost, represented by the dotted line between c(t) and p(t), and can be regarded as a push system in that output pushes sales. The marginalist approach determines price according to the marginalist principle and can be regarded as a pull system: sufficient assets and productive capacity are pulled into the system to meet the required demand. The third approach - average cost including demand - takes cost into account when determining price and demand into account when determining quantity sold at that price.

The economic models are transformed into dynamic models broadly following system dynamics principles:

identify a problem;
develop a dynamic hypothesis explaining the cause of the problem;
build a computer simulation model of the system at the root of the problem;
test the model to be certain that it reproduces the behavior seen in the real world;
devise and test in the model alternative policies that alleviate the problem;
implement this solution;

(System Dynamics Society, 2001).

The 'problem' under examination is the growth and decay of the business firm. Initially the feedback mechanisms inherent in the model were identified and then explicitly modeled. Variables and parameters were identified and variables classified as level, rate, and auxiliary variables. A simulation model was created by entering these variables, together with their descriptions, into the Powersim simulation software package. The relationships between variables were developed and recorded. Finally, the level and rate equations were developed using the defined variables and relationships. This paper presents the results of testing the model using different pricing policies. As this is a research project implementing a solution is not relevant and that phase of the methodology will not be carried out.

The simulation was carried out using a simulation software package (Powersim, 2000). The economic model was based on prior work by the author using a spreadsheet model (Brady, 1999) and using a model created using the C++ programming language (Brady, 2000). These models examined different aspects of the growth of a firm using a variant of the average cost pricing approach developed by the author.

In the following sections the average cost model, the marginalist model, and the average cost including demand model are described in detail. The results of the simulation using each model are then described. Finally, the results and the experimental process are discussed and some tentative conclusions drawn.

The average cost model

The average cost model is based on the average cost principle ie. price is set at the average cost of producing the product plus a margin for profit (Koutsoyiannis, 1979:271). The structure of the model is shown as a diagram in appendix one and is described in detail below. The operation of the simulation is outlined and then the results of the simulation are discussed.

Structure

Production is determined using the Cobb-Douglas function, often used by economists to model production due to its useful properties. This function is expressed mathematically as:

$$Q_p = a\, L^{b0}\ K^{b1} \qquad \ldots (1)$$

The function assumes that both labour, L, and capital, K, are factors of production. In the models used in this paper initial values for labour and capital are set at 100 labour units and $1,000,000 respectively.

The parameter a in the Cobb-Douglas function can be regarded as representing the entrepreneurial and organisational efficiency of the firm. In the models used in this paper the parameter a is represented by a productivity factor with value 1.6 and a capacity utilization factor with value 0.6; the product of these two factors is 0.96 and is the value used for parameter a.

It can be shown that if $b_0+b_1>1$ then economies of scale exist in production and if $b_0+b_1<1$ then diseconomies of scale exist (Koutsoyiannis, 1979:78). In the models used in this paper b_0 was set to 0.6 and b_1 was set to 0.7; therefore $b_0+b_1 = 1.3$ and some economies of scale in production exist.

The level of production for period one can be easily determined by substituting the above values in equation 1 yielding a total of 241,141 units produced. As the values of labour and capital change over time the level of production will change in later periods. A nonnegativity condition is built into the model to ensure that the level of production cannot become negative; this is very unlikely to happen but is theoretically possible should the value of labour or capital become negative.

In earlier versions of the model (Brady, 1999 and 2000) a simpler function was used to determine production: $Q = aK$. This is equivalent to the Cobb-Douglas function with b_0 set equal to 0 and b_1 set equal to 1.

Total variable costs are assumed to follow the law of variable proportions, also known as the law of eventually decreasing returns. This is approximated in the model by a cubic polynomial equation (Koutsoyiannis, 1979:114):

$$TVC = dQ_p - eQ_p^2 + fQ_p^3 \qquad \ldots (2)$$

Careful choice of coefficients in this equation yields the inverted S shape typical of total variable cost curves. The first term represents the portion of total cost that increases linearly with the number of units produced. The parameter c can be regarded as the unit variable cost of production. The second term represents the gaining of economies of scale as production increases. As this term contains production to the power of two economies of scale become proportionately more significant as production increases. The parameter d is set to a much lower value than parameter c. The third term represents the incurring of diseconomies of scale as production increases even further. To ensure that these diseconomies of scale are not incurred until very high levels of production the parameter e is set to a very low value. However, because this term is a cube of production once diseconomies of scale start to become significant they then increase very rapidly.

In the models discussed in this paper parameters d, e and f have values 7, 0.000001 ($1x10^{-6}$), and 0.00000000000005 ($5x10^{-14}$) respectively. Using these values we can determine that the total variable cost for one unit is $7, for 10m units of production is $20m, and for 25m units of production is $331m. This means that the average variable cost per unit produced is $7 for one unit of production, decreasing to $2 for 10m units of production as economies of scale are gained and increasing to $13.3 for 25m units of production as diseconomies of scale are incurred.

A nonnegativity condition is built into the model to ensure that total variable costs cannot become negative. While in practice it is impossible for total variable costs to become negative, in a simulation model it is possible should a large value be selected for parameter d. This would mean that the total variable cost curve is not S-shaped with a single inflection point but the more typical cubic shape with two turning points: a minimum and a maximum; should the minimum one drop below the horizontal line this would give negative total variable cost for a particular range of production. In this model given a unit variable cost of $7 (parameter d) the values for parameters e and f were determined by trial and error so that a reasonably shaped total variable cost curve was defined.

Total cost is determined by adding total fixed costs and total variable costs. Total fixed costs in the models used in this paper are set at $0.5m increasing to $1.5m after 50 periods, achieved using the STEP function provided by Powersim. Average total cost is then determined by dividing total cost by the number of units produced ie.

$$ATC = (TVC + TFC) / Q_p \qquad \ldots (3)$$

Price is determined by the average cost pricing (mark-up) rule ie.

$$P = ATC * (1+ net\_margin) \qquad \ldots (4)$$

Net margin is assumed to be a given ie. the firm chooses net margin so that it can sell all its product; in the models used in this paper net margin is set at 10%.

The average cost model is a supply-driven model: the model assumes that the firm chooses its price so that it can sell all it produces. According to Koutsoyiannis (1979:273) it does this in two steps. The first step is to determine price according to equation 6. The second step is to determine the entry deterring price ie. the price below which new entrants will not be attracted into the industry. If this is below the price given by equation 6 then the lower, entry

deterring, price is used by the firm. Competition among firms already in the industry is assumed to be resolved by the mechanisms of tacit collusion or price leadership and so does not need to be taken explicitly into account. This model does not attempt to separately determine the value of the entry deterring price; in effect it assumes that the price determined by equation 6 is less than or equal to the entry deterring price.

Growth in production occurs by way of growth in the individual factors of production: ie. growth in capital K and growth in labour L. This mechanism of growth is the primary dynamic element in the average cost simulation model. In dynamical systems terminology it is an example of reinforcing or positive feedback.

Growth in capital K is entirely organic ie. retained earnings increase the level of capital available to the firm during each period in which retained earnings are positive. This growth in assets primarily by way of internally generated funds is in accordance with the pecking order theory from the field of corporate finance (Brealey and Myers, 1996:501). Expressed as an equation this internal growth in assets is represented as:

$$K_{n+1} = K_n + \text{retained\_earnings} \qquad \ldots (5)$$

Losses are modeled as an increase in debt; interest is then paid on the debt in following periods. This too is in accordance with the pecking order theory: if internally generated funds are not available the firm prefers taking on debt rather than issuing further equity.

In this model once losses are incurred by the firm they remain with the firm forever. In earlier versions of the model losses were modeled by reducing the level of assets. However this meant that production was reduced in the period following the loss and this was felt to be unlikely behaviour by the business firm. For the purposes of this paper it has therefore been assumed that firms will incur debt rather than reduce producing assets. Having said that, persistent losses due to overproduction will eventually lead management to reduce the level of productive assets ie. managers will close down facilities and let staff go. However, this model does not yet include this kind of relatively sophisticated managerial decision-making capability; this may be an area for further research work. Also, in later profitable periods firms may choose to pay down debt rather than retain all earnings. This too is a managerial decision and, again, this level of sophistication has not been included in the model; it also may provide an opportunity for further research.

An alternative approach would be to model producing assets and non-producing assets (eg. cash) separately as suggested by Marris (1963). In his view, liquid assets are non-producing and therefore non-earning and are kept primarily for reasons of security. Losses could then be paid for by reducing the level of non-producing assets while retaining the level of producing assets. This is not an unreasonable approach: firms often retain large amounts of cash for various purposes including this one; for example in 1991 Microsoft held $689m of its £1.6b assets in cash (Microsoft, 1991). This approach would entail modeling the asset side of the firm's balance sheet in more detail; for example, current assets and fixed assets could be separately modeled. Losses could be paid for out of current assets; fixed assets would act as the basis on which production is modeled. Some new decision rules would also be required: for example, a ratio in which retained earnings are apportioned to fixed and current assets would need to be specified, as would a target liquidity ratio of current to fixed assets.

Labour growth is assumed to be in proportion to retained earnings ie. labour is increased in periods where retained earnings are positive and labour is decreased in periods where retained earnings are negative. The model used in this paper uses the nonlinear function

$$\begin{aligned} \delta L &= fL && \text{for retained earnings} \geq 0 \quad) \\ \delta L &= -gL && \text{for retained earnings} < 0 \quad) \end{aligned} \qquad \ldots (6)$$

to determine change in labour. Both parameters f and g have value 0.000005 ($5x10^{-6}$). However, they could be given different values if the firm being modeled had a greater propensity to take on staff in good times than it does to shed staff in poor times, or vice versa. Using the above parameter values retained earnings of $1m would lead to an increase in labour of 5 units; a loss of $1m would lead to a reduction in labour of 5 units.

In contrast to growth in capital, growth in labour is not founded on any formal theoretical construct. However, examination of the business press shows that the number of staff on the payroll of a firm is an item regularly mentioned in firm annual reports and that firms do take on staff in good times and reduce staff numbers when trading conditions become difficult. However, the basis for taking on and reducing staff numbers is not clear-cut. In this model the assumption has been made that change in labour is a function of retained earnings. Other variables could have been used as an approximate basis for determining change in labour eg. absolute value of or change in revenue, production, or assets. Revenue was not considered as a basis for determining change in labour because it is possible for a firm to have increasing revenue and still have losses; in this situation it is would be unlikely behaviour for a firm to automatically increase labour. Similarly, a firm could increase production while still incurring losses; again it is unlikely that a firm in this situation would automatically increase its labour. Change in assets could have been used; however, as the firms being modeled grow organically change in assets occurs as a direct result of retained earnings, so linking labour change to retained earnings was chosen as a better approach. Finally, retained earnings has the advantage that by its very nature it is a measure of change and can be used directly to determine change in labour; change in revenue, production, or assets would have to be determined by subtracting the appropriate values for two consecutive periods.

Growth in labour could be tackled in a more sophisticated manner. For example, change in net income or retained earnings rather than their absolute value could be used. This is not inconsistent with reality: firms do not always wait until losses are incurred before laying off staff; they will often lay off staff once profits begin to stagnate or reduce. At the time of writing many examples of layoffs resulting from both downturns and losses are being reported: Dell and Gateway announced a reduction in staff numbers due to declining demand for personal computers (O'Clery,2001b; Brennan, 2001); Lucent Technologies announced a reduction of 10,000 in staff worldwide following a loss of $1b for the quarter ending in December 2000 (O'Clery, 2001a); Nissan are laying off 21,000 staff due to ongoing losses during the nineties (Business Eccountant, 2001).

Another level of sophistication would be to include a delay or lag in the introduction of productive new labour; this represents the real world situation whereby it may take some time between making the decision to take on new labour and that new labour being productive. However, no lag or delay in the introduction of new labour was incorporated into this model; this is consistent with the modeling approach taken in this paper where delays or lags have in general not been included.

Revenue in this model is determined by the following equation:

$$R = Q_p * P \qquad \ldots (7)$$

ie. revenue is equal to units produced multiplied by price. The underlying premise of the average cost model is evident here: the firm is assumed to sell all that it produces.

From the revenue figure a series of amounts are taken off in succession. In this paper North American practice and terminology for the profit and loss account or income statement is used (Weston and Copeland, 1992:26). The take-off amounts are in order: cost of goods sold,

administration and marketing expenses, depreciation, interest, tax, and dividends. The residual is retained earnings and is kept by the firm as an addition to its assets.

Firstly, gross income is determined by subtracting total cost of goods sold from revenue ie.

$$\text{Gross income} = R - TC \qquad \ldots (8)$$

where total cost (TC) is equal to total fixed costs (TFC) plus total variable costs (TVC). Revenue and total costs have been discussed above.

Next, net operating income (NOI) is obtained by subtracting administrative and marketing expenses and depreciation expenses from gross income ie.

$$\text{NOI} = \text{Gross income} - \text{Admin\&Marketing expense} - \text{depreciation} \qquad \ldots (9)$$

In this model administration and marketing expenses are set to a percentage of revenue; this percentage is set to zero ie. marketing and administrative expenses are zero and cost of goods sold represents all costs to the firm.

Depreciation is estimated at an amount equal to a percentage of assets. In this model the rate of depreciation of assets is assumed to be 10% of assets per period. The initial investment of $1m would therefore be fully depreciated after ten years. This is broadly equivalent to straight-line depreciation over a ten year period (Weston and Copeland, 1992:137) although it is carried out on all assets of the firm and for the full life of the firm. A nonnegativity condition is built into the model: depreciation is set at zero should capital become negative (although it is recognised that there would be little hope for the firm should this occur).

Depreciation is modeled as an expense and reduces the profit of the firm. In the model depreciation is not reflected in the balance sheet ie. the value of the assets is not reduced by the amount of depreciation in a trading period. Unlike accounting depreciation in this model depreciation is a cash flow: the depreciation amount is assumed to be invested by the firm in maintaining assets at the end of the trading period as they were at the beginning. For example, if assets at the beginning of the period were $1m, the firm spends $100,000 during the period maintaining the assets so that at the end of the period the assets are still worth $1m. In effect the model assumes that depreciation in each period is used for capital investment; it is recognised that in reality capital investment occurs in a lumpy rather this smooth manner.

In this model we are considering assets in aggregate whereas accounting depreciation is applied to separate physical assets such as individual pieces of plant, machinery, and equipment. The depreciation figure in a firm's annual accounts is an aggregate of these individual depreciations. This model therefore is not operated strictly in line with accounting rules of depreciation; however, it would complicate the model greatly and provide little additional benefit to keep a record of the age of each asset, depreciate each separately, and then determine the aggregate depreciation. Depreciation as used in this model is in keeping with the spirit of depreciation - smoothing the maintenance of assets over time - and represents a more general obsolescence or wearing out of producing assets.

Depreciation conventions also vary according to national tax laws, industry norms, firm practice, and asset life; within this context each firm forms its own approach to depreciation. The model described in this paper is a general model of the business firm and the structure of the model aims to be independent of firm, industry or country specific practices.

Finally, depreciation in this model represents an actual cost to the firm. It is not the same as an opportunity cost or a cost of capital. The model does not include the opportunity cost of

capital or aim to determine economic value added; these may provide areas for future research.

In this model it is assumed that no source of income other than operations exists; examples of possible other income are royalties or interest earned on marketable securities. Therefore earnings before interest and tax (EBIT) is taken to be equal to net operating income, a common assumption (Weston and Copeland, 1992:666).

Earnings before tax is computed by subtracting interest expense from net operating income. Interest expense is computed by multiplying losses by the interest rate. (As discussed above, losses are assumed to be paid for by taking on debt). The interest rate used is 7%, an arbitrary but reasonable figure.

Net income is computed by subtracting tax from earnings before tax. Tax is computed by multiplying earnings before tax by the taxation rate. The taxation rate used in this model is 40%. Again, this is an arbitrary but reasonable figure: Brealey and Myers (1996:475) suggest using a 35% corporate tax rate in the US; at the time of writing the corporation taxation rate in the UK is 30% (Lymer and Hancock, 2001:ch8,p18) and in Ireland is 24% (TCA, 1997: section 21) having steadily reduced from 40% a number of years ago. A non-negativity condition is built into the model: no tax is paid if earnings before tax are negative.

Retained earnings is computed by subtracting the dividend paid to shareholders from net income. Shareholder dividend is computed by multiplying net income by the dividend retention rate. The dividend retention rate used in this model is 50%, again an arbitrary but reasonable rate; Lintner (1956) found 50% to be the payout ratio most commonly used by firms. A non-negativity condition is built into the model: dividends are not paid if net income is negative (ie. if the firm incurs losses). It should be noted that not all real firms behave like this: some firms maintain their dividend payment even when losses are made to avoid the reduction in the firm's stock price that can follow announcement of a dividend cut (Brealey and Myers, 1996:422). For example, Black & Decker maintained its dividend during the several years it suffered severe financial pressure resulting from its acquisition of Emhart Corporation (Thompson and Strickland, 1999:C-432).

It is recognised that this approach to modeling dividends is somewhat simplistic. The dividend amount is generally regarded as being 'sticky' ie. the amount paid out on each share is more or less constant. This behaviour of firms is due to the perceived information value of dividends: by altering the amount of the dividend firms send a signal to the stock markets that they may prefer not to send. This model uses a constant dividend rate rather than a constant dividend amount; this can lead to fluctuations of dividend amount if profits significantly rise or fall. However, the focus of this research is firm growth and to model dividend as a constant amount could lead to an unrealistic result if after a number of time periods profit increased substantially but dividends were to remain constant. Setting dividend to a proportion of net income ensures that the effects of growth are automatically taken into account.

Lintner's (1956) research shows that dividends are determined using a combination of current earnings and previous dividends. While firms have a target payout ratio then do not necessarily issue this payout in dividends; instead they move to this target payout ratio over a number of years according to the firm's adjustment rate. Lintner's findings show that while firms move their dividends upwards when earnings increase they are reluctant to move them downwards when earnings decline.

Growth in dividend could be modeled under the constant dividend assumption but some basis for determining dividend growth would still be required. A possible approach is that dividend growth be in proportion to growth in net income or revenue. This is still a crude

approach but would have the disadvantage of making the model more complex. Should future research choose to focus more sharply on dividend policy and its effect on firm growth then this extra complexity in the model may be warranted.

Retention ratios, or their inverse: dividend payout ratios, are widely used in the academic literature. Marris (1963) includes the dividend retention ratio as a parameter in his balanced growth model of the firm. Higgins (1977) includes the target retention ratio as a key policy parameter in his sustainable growth model. Lintner (1956) uses the payout ratio as a variable in his study of dividend policy.

As discussed above retained earnings increase capital when they are positive (ie. when profits are made) and increase debt when they are negative (ie. when losses are incurred). This increase in capital increases the productive capability of the firm and drives the growth of the firm under the average cost assumption.

Operation

Variables and parameters are defined within the Powersim simulation package as level, rate or auxiliary variables in keeping with system dynamics convention. Four of the variables in the model are defined as level variables (square shaped in Powersim diagrams): capital, debt, labour and demand. Rate variables are (circle shape with valve: circle to signify a variable and valve to signify rate):, retained earnings, retained loss, change in labour, and change in demand. Other variables are defined as auxiliary variables (circle shaped). Parameters are defined as constants (diamond shaped).

Having defined the structure of the model, the equations specifying how each variable is determined are entered into the variable definitions. Initial values for level variables and parameter values are also entered. Once all elements of the model have been defined, the Powersim software package is used to simulate the behaviour of the firm for the chosen number of periods of time. Changing the number of periods for which to simulate is an easy matter using the simulation setup item on the menu bar.

Results of the simulation are displayed as a number of graphs and tables. Graphs produced are timelines and scatter diagrams. Timeline graphs are: units produced and units demanded over time, revenue and total cost over time, retained earnings over time, capital and accumulated loss over time, labour over time, unit price and average variable cost over time. Scatter diagrams are: total cost against units produced, and average variable cost against units produced. Three tables are also produced that list the values over time of most of the variables used in the model.

Results

Under the assumptions of average cost pricing the firm will grow indefinitely. No constraint exists within the model to limit the growth of the firm. The firm merely adds its mark-up to its average cost and is assumed to sell all that it produces. The growth path of the firm traces out a price curve above the average cost curve by an amount of profit determined according to the margin percentage. The average cost curve increases monotonically to the right and so therefore will this price curve, and hence the growth path of the firm, as shown in figure 2.

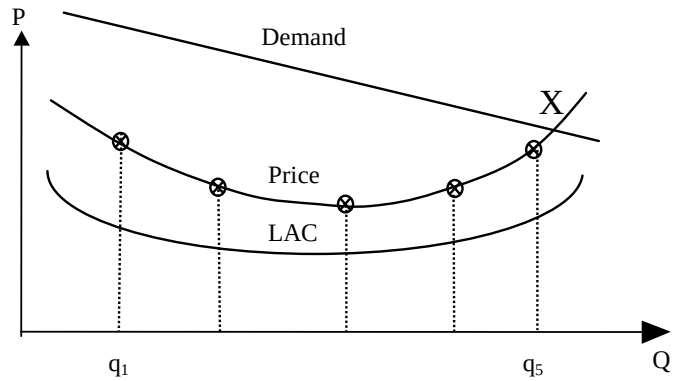


Figure 2. Growth path of the firm under the average cost pricing assumption.

A major flaw exists in this approach. When the firm reaches production levels such that diseconomies of scale occur its costs begin to rise rapidly (represented by the cubic term in the cost function). Consequently, the price of the firms products will also rise rapidly. It is impossible to believe that the demand for the firms product will not begin to decline at this point. However, as demand is not formally taken into account under the average cost assumption there is no way for this decline to be shown to take place within the model. Instead the model shows ever increasing costs, prices, revenue, and - due to the mark-up - ever increasing earnings. It is entirely counterintuitive to expect that the firm will perform better when diseconomies of scale occur and costs increase disproportionately.

A limit to the growth of the firm is to introduce a demand constraint as shown in figure 1. This limits the growth of the firm at point X on that figure. Beyond point X demand for the firms product will decline and all product will not be sold.

From a management point of view, the average cost pricing model is also unsatisfactory as it provides no incentive for managers to ensure that the firm operates efficiently: all costs are passed on to the buyer so reducing costs leads to no advantage. Indeed there is an incentive to increase costs as, with a fixed margin rate, higher costs lead to high amounts of profit; for example a 10% markup on $100 cost gives $10 profit whereas a 10% markup on $150 cost gives $15 profit.

The marginalist model

Developing a marginalist model (Koutsoyiannis, 1979:158) required a considerable change to the average cost model, both in structure and in the chain of cause and effect. The structure of the marginalist model is shown as a diagram in appendix two and is described in detail below. Much of the structure is identical to that of the average cost model: for example, the calculation of total cost, revenue, gross income and the various forms of earnings is identical. The main areas of difference are the determination of price using the demand function, the inclusion of a growth function for demand, and the exclusion of the production function. The operation of the model is briefly outlined and the results of the simulation then discussed.

Structure

In the marginalist model demand is the primary driver of firm behaviour. Demand is modeled as a linear demand curve ie. demand is assumed to decrease linearly with price:

$$P = a - bQ_d \qquad \ldots (10)$$

This equation gives the familiar demand curve that slopes downwards from left to right. The parameter a can be understood as the maximum price paid for the product or the price that would be paid if only one unit of the product was available on the market. It is the intercept on the vertical axis. The parameter b is the slope of the demand curve and has a positive value; this ensures that the demand curve always slopes downward. The variable Q_d represents quantity demanded. The demand function essentially says that as quantity available increases price reduces in proportion to the quantity available. In this model the initial value of parameter a is set at 25 and the initial value of parameter b is set at 0.0001. This means that price is $25 when one unit is available, $24 when 10,000 units are available, $15 when 100,000 units are available, and $5 when 200,000 units are available. Price would become zero when 250,000 units are available. While this behaviour is clearly unrealistic linear demand is a common assumption; replacing the linear demand curve with a nonlinear curve would avoid price going to zero; this was not done in this research but could provide an opportunity for further research.

Growth in demand is modeled as a shift upward in the demand curve; this is generated by the growth of parameter a in equation 10 above as follows:

$$a_{n+1} = a_n + (r_1 a_n - r_2 a_n^2) \qquad \ldots (11)$$

This is the logistic growth function with r_1 being the growth parameter and r_2 the decline parameter. The value of a will initially grow in an exponential manner but the rate of growth will decrease when a becomes large. The decline parameter r_2 may be set to zero to give compound growth ie.

$$\begin{aligned} a_{n+1} &= a_n + r_1 a_n \\ &= a_n (1+r_1) \qquad \ldots (12) \end{aligned}$$

However, compounding represents explosive growth and will not be sustainable in the long run. The logistic function mirrors the product life cycle (S-shaped) curve producing a more realistic demand growth function.

This growth function has the effect of increasing price from the levels of price in the previous time period for all given levels of quantity available. It also means that for a given price the quantity demanded will increase from period to period. It is this latter sense that is of interest to this research: we are assuming that the firm largely keeps its price constant but increases the amount of units sold in each period. In effect we are relaxing the *ceteris paribus* assumptions on which the shape of the demand function rests: for example, word of mouth about the quality of the product during one period could serve to increase demand for the product during the next period. Both of these effects are equal: should word of mouth serve to increase demand but quantity available remain constant then price will increase. The parameter r_1 is set at 5%. This represents moderate annual growth for a business firm; fast growth firms can achieve 20 - 25% or greater compound annual growth. For example, the average compound growth in revenues between 1969 and 1989, discounted for inflation, for W.L. Gore and Associates is estimated to have been greater than 18% per annum (Thompson and Strickland, 1999:C-510).

Change in demand for the firm could also be modeled by changing the value of parameter b of the demand function. Reducing b would make the slope of the demand curve less steep meaning that for a given price the firm would sell a greater quantity of product. The firm would continue to grow over time until its demand curve, and hence also its marginal revenue curve, becomes horizontal with price equal to the value of parameter a. Output of the firm is

determined by the point at which the marginal cost curve crosses the now horizontal marginal revenue curve; this point represents a limit to the growth of the firm.

There are some difficulties however with this approach. Firstly, it may be unrealistic to assume that the firm could maintain its price and at the same time increase the quantity sold of its product. In response to this we could argue that the firm is moving its strategic approach from differentiated to cost leadership. For example, Coca Cola has consistently increased its output of soft drinks while keeping its price relatively constant. Although Coca Cola is market leader, a prestige product and has built up considerable brand loyalty, the price of a can of Coca Cola is generally the same as the price of competitor products. Similarly, although MacDonalds is market leader in the fast food industry and its brand is widely known it charges the same price for burgers as do its direct competitors. It could be argued that this is a relatively risky strategy to pursue as the firm is constantly increasing the price elasticity of demand for its product making customers more sensitive to an increase in price.

Other changes in demand could be modeled. For example, parameter b could be reduced simulating a differentiation strategy. Or, parameters a and b could be changed simultaneously leading to several different growth paths for demand.

As in the average cost model total cost is assumed to vary according to the law of variable proportions and is modeled by the cubic equation:

$$C = c + dQ - eQ^2 + fQ^3 \qquad \ldots (13)$$

Marginal cost is determined by differentiating the cost function as follows:

$$MC = d/dQ\ (C)$$

$$= d/dQ\ (c + dQ - eQ^2 + fQ^3)$$

$$= d - 2eQ + 3f\ Q2 \qquad \ldots (14)$$

Revenue is given by the identity:

$$R = P * Q \qquad \ldots (15)$$

Combining this with the demand function gives:

$$R = Q(a - bQ)$$

$$= aQ - bQ^2 \qquad \ldots (16)$$

Marginal revenue is determined by differentiating the revenue function as follows:

$$MR = D/dQ\ (R)$$

$$= d/\ dQ\ (aQ^2 - abQ)$$

$$= a - 2bQ \qquad \ldots (17)$$

This result is familiar to us: the marginal revenue line has twice the slope of the demand line.

Profit is given by the identity:

$$\Pi = R - C \qquad \ldots (18)$$

The marginalist principle, derived from the laws of the calculus, states that for maximum profits the firm should produce at a level of output given by the equation MR = MC.

Substituting equations 14 and 17 above for MC and MR gives:

$$a - 2bQ = d - 2eQ + 3fQ^2$$

$$\Rightarrow \; 3fQ^2 + (2b - 2e)Q + (d - a) = 0$$

Solving this quadratic equation for Q, and taking the positive value of the square root, we get:

$$Q = (-(2b - 2e) + \sqrt{((2b - 2e)^2 - 4(3f)(d - a))}) / 6f \qquad \ldots (19)$$

Having obtained a value for Q, price can now be determined from the demand function above, total cost from equation 13, and revenue from equation 15, and profit from equation 18.

Note that production plays no part in this model. The marginalist approach basically assumes that sufficient labour and capital resources already exist or can be immediately obtained to meet the required demand. This is a significant assumption as in reality it may take considerable time to recruit and train new staff or to purchase, install, and commission plant, machinery or equipment.

Operation

The model under the marginalist approach is operated in exactly the same fashion as it is under the average cost approach. The set of calculations is performed in each period. At the end of each period demand will grow as discussed above. The simulation is then repeated for this new level of demand and so on. As described above, it is the change in demand that drives the dynamic behaviour of the model.

Once the model has been developed, the simulation package software can be used to execute the model for a variety of different parameter values and various timescales.

Results

Under the marginalist principle the simulation results show the firm to be an eternal money making machine for as long as demand continues to increase. Revenues show exponential growth.

As demand grows the demand curve shifts outwards parallel with the original demand curve. The marginal revenue line similarly shifts outwards parallel with the original marginal revenue line. Each new marginal revenue line cuts the marginal cost curve above and to the right of the previous marginal revenue line. The growth path therefore traces out the path of the marginal cost curve. Over time the firm produces output at a point that follows its marginal cost curve upwards towards infinity as shown schematically in figure 3.

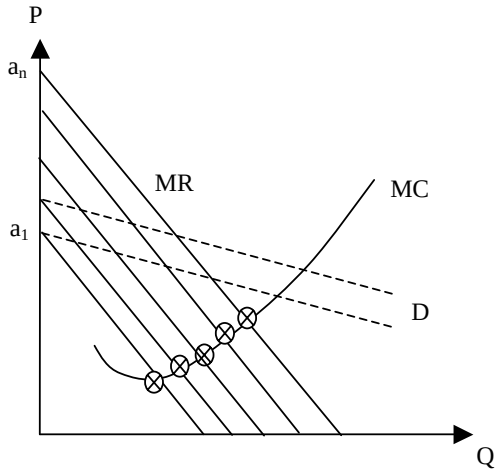


Figure 3. Growth of firm using the marginalist approach

This behaviour pattern is also unrealistic. Demand will only increase exponentially to infinity if it is modeled as compound growth. If demand is modeled as a logistic function, by setting the demand delay factor to a positive non-zero value, at some point demand will begin to stagnate or decline. The firm will stop growing at the point on the marginal cost curve at which this occurs. Results for a logistic demand function show asymptotic behaviour for revenue and costs are shown in figure 4 (note that the demand delay factor was set to 0.0005 for this simulation).

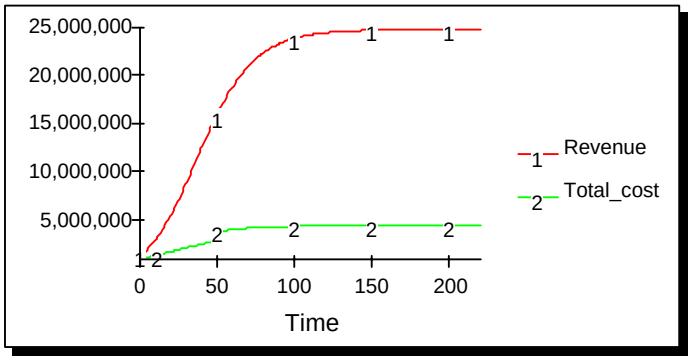


Figure 4. Behaviour of firm under logistic demand growth

Under the marginalist principle the firm grows by raising its price as demand increases. This is a consequence of the marginal cost curve increasing monotonically to the right. Under the marginalist principle the firm will raise prices more quickly than under the average cost principle as the marginal cost curve rises more steeply than the average cost curve (the square term in the marginal cost curve has factor 3*f whereas the square term in the average cost curve has factor 1*f).

While in theory the marginal cost curve will always cut a downward sloping demand curve in practice one must question the realism of modeling buyer behaviour over time as a set of parallel demand curves. It may be more realistic to model the demand behaviour by altering both the intercept parameter a and the slope parameter b rather than just the intercept parameter alone. The growth path would still lie along the marginal cost curve but would not necessarily trace out a monotonically increasing path. It is even possible for the growth path

to trace out a portion of the marginal cost curve in a downward rather than in an upward fashion as shown in figure 5.

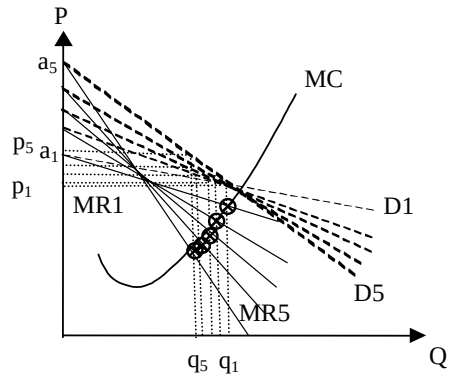


Figure 5. Growth in demand where parameters a and b change over time.

It is interesting to examine the behaviour of the model if demand does not grow. Here - provided that we set the depreciation rate to zero - revenue, cost, and retained earnings remain constant throughout the simulation (and indeed for all time). The accumulator or level variables - capital and labour - show a constant increase over time. However, if we set the depreciation rate to a non-zero value, then retained earnings reduce asymptotically to zero as shown in figure 6. This is because the depreciation amount is based on the value of capital and as capital increases so does depreciation. With fixed revenues and fixed costs but with increasing depreciation retained earnings must reduce over time.

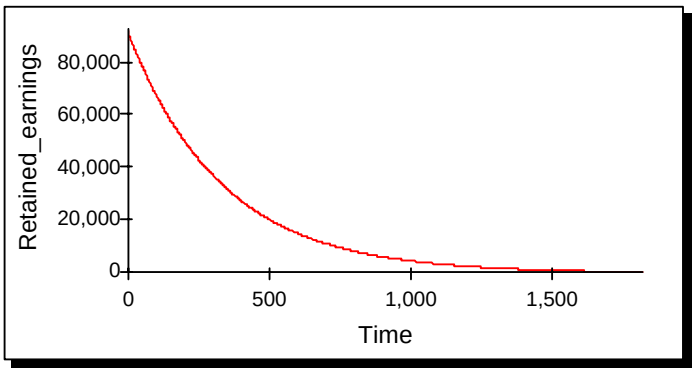


Figure 6. Asymptotic decrease in retained earnings with zero growth in demand

It could be argued however that with constant demand there is no need for retained earnings to be used for productive purposes; the earnings retained would likely be kept as liquid assets, not suffer wear and tear, and therefore not be in need of depreciation. This argument supports separating assets into productive and liquid categories as suggested earlier in this paper.

Koutsoyiannis (1979:107) suggests that fixed costs typically include depreciation costs. That poses a difficulty for the model: should depreciation be modeled as a fixed amount (and therefore modeled as a fixed cost) or should it be modeled as a proportionate amount of productive assets as it is in the current model. The question boils down to: is depreciation a fixed or a variable cost? and if variable, according to what does it vary? The effect of varying depreciation did not manifest itself when demand is growing because retained earnings were

also growing and this masked the impact of depreciation. However, when growth in demand is held constant, varying depreciation quickly showed its effect as shown in figure 6 above.

Average Cost including demand model

This model is very similar to the average cost approach except that demand is also taken into account. The model is driven neither by supply nor demand alone but by a combination of the two. Once again we will examine the structure of the model, outline its operation and discuss the results of the simulation.

Structure

The firm is assumed to have resources to produce a certain amount of product according to the Cobb-Douglas function as described above under the average cost model. The mechanism of the cost curve adjusted by the margin gives the price at which this product can be sold, again as described above under the average cost approach. The demand function then gives the amount of product that can be sold at that price. This amount may be equal to, greater than, or less than the amount produced. The basis of the model is that the amount of product sold cannot be greater than either the amount produced or the amount demanded; the lower of the two amounts therefore is used as the amount sold.

This may mean that the firm produces more than it can sell; this excess production goes into inventory or perishes (as we do not attempt to model inventory in this model the excess production perishes). Alternatively, as under the average cost model, the firm may not meet its demand and so leave an unsatisfied demand. Except in the rather unlikely event that the firm meets its demand exactly, the firm will either have excess capacity or unsatisfied demand. Over time the firm may have extended periods of excess capacity, extended periods of unsatisfied demand, or may alternate between the two.

Note that excess production cannot occur under either the average cost or marginalist pricing approach. Under the marginalist approach the firm obtains just enough factors of production to meet demand. Under the average cost approach the firm sells all that it produces. However, in reality firms do overproduce. Producing for inventory is a form of overproduction. Retaining staff when there is a downturn in demand is a form of overproduction. Firms do this for valid business reasons eg. firms build inventory to smooth out varying demand for its product, firms retain staff during lean times to keep up morale and to hold on to capable people in anticipation of a resurgence in demand.

On the other hand firms may deliberately or inadvertently undersupply the market. Brandenburger and Nalebuff (1996:112-120) give the examples of Nintendo undersupplying the market for videogames and De Beers undersupplying the market for diamonds and outline how both of these firms gained significant market power as a result.

The detailed structure of this model is based on both the marginalist and the average cost models as described above. The production and cost elements of the model are as for the average cost model. The demand elements of the model are as for the marginalist model. The elements of the model below gross income are the same for all three models. The model differs from the other two models primarily in its calculation of revenue. Price is determined in the same manner as in the average cost model. However, quantity sold is assumed to be quantity produced or quantity demanded, whichever is the minimum. Revenue is then calculated by multiplying price by quantity sold.

We can take the model one step further. Should the demand for the product be higher than its supply it could make sense for the firm to seek out a higher price for its product than that set

by the average cost pricing rule. This higher price can easily be determined from the demand function, this time solving for price given the quantity produced. Selling the same quantity at a higher price would clearly give the firm greater profit and therefore would be attractive to the firm. However, to do this the firm must know the exact nature of its demand curve, something that the firm generally does not know; indeed the average cost pricing approach is used in the first place because the firm does not know its demand curve.

In the approach outlined above the firm does not need to know the exact nature of the demand curve; it need only know if its price/quantity point lies above or below the demand curve. This is easily found out in practice: if its sells all its product it lies below, if it does not it lies above. Indeed in this model the firm does not really need to know anything about the demand function. It merely senses that it has exceeded its demand function if it is left with unsold goods.

A model including both supply and demand could also work in another manner, at least in theory. Supply can be determined as outlined above using the Cobb-Douglas function. Rather than determining price using the average cost markup rule, price could be determined using the demand function. The firm would then sell the quantity produced at the price determined. This may even lead to higher revenue for the firm. However, this approach would require full knowledge of the firm's demand curve, something that firms generally lack.

Growth in this model is largely driven by growth in productive capacity of the firm. As in the average cost model retained earnings increase assets which in turn increases production. However, in this model demand is also assumed to grow in the fashion discussed under the marginalist model. It is the interplay between growth of demand and growth of supply that gives the variety of behaviour shown by the model.

Operation

This model is operated in the same fashion as the previous two models.

Results

Under the average cost including demand model the growth path of the firm will be as shown in figure 7. Initially the growth path follows the average cost curve to the right until it crosses the demand curve at point X; this is where supply first outstrips demand. Further growth takes place by following the demand curve to the left as shown. Point X is a bifurcation point representing a significant switch in the behaviour pattern of the firm; it can be seen in a different form in Brady (2000).

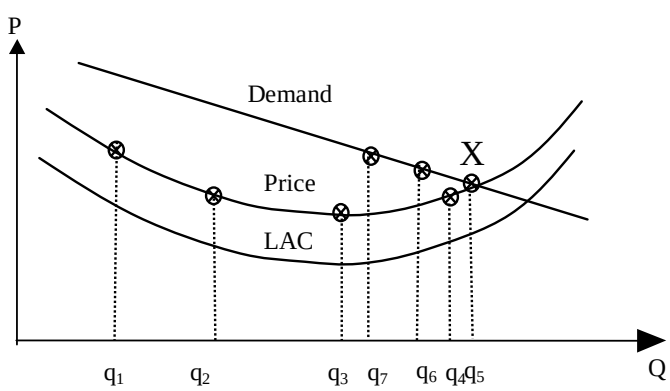


Figure 7. Growth path under average cost including demand

To see why this behaviour is so we need to look in more detail at what happens around point X as shown in figure 8. Up until that point demand for the firms product exceeds supply and the firm sells all that it produces at its marked up price. However, at some point, say q_n, the earnings retained will allow extra production such that the firm in the next period supplies a quantity to the right of point X. As firms do not usually know their demand curve, the firm will not be aware that it has exceeded its demand and will set its price following its usual markup rule. However instead of selling q_s, the quantity produced, the firm will sell only q_d, the quantity demanded at that price. The firm will therefore receive less revenue and make less profit that it anticipates. It will expect to take in revenue represented by the outer rectangle $opbq_s$ covering areas A+B+C+D. It will actually take in revenue represented by rectangle $opaq_d$ covering areas A+D. It will incur costs represented by the rectangle oc_scq_s covering areas C+D. The firm's profit is represented by A minus C. Clearly if rectangle A is greater than rectangle C the firm makes profit; if A is less than C the firm makes losses. The growth path of the firm will therefore follow the price curve to the right until point X is reached and will then follow the demand curve to the left until it reaches a point where the size of rectangle A equals the size of C. At this point the firm will stop growing organically as it is no longer generating retained earnings.

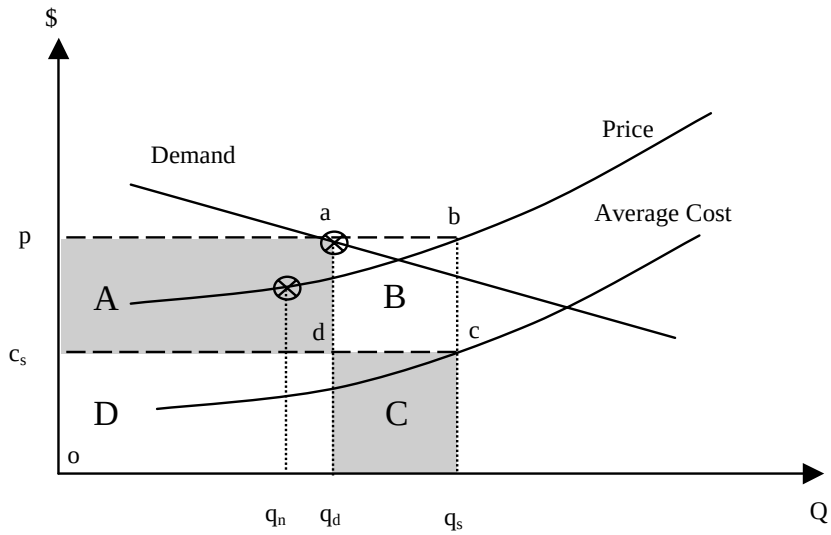


Figure 8. Behaviour at bifurcation point

In reality managers will take action to overcome this difficulty. This action may take several forms. Firms could attempt to reduce costs which, because of the markup, will in turn reduce price. This has the effect of shifting both the price and the cost curve downwards and shifting point X to the right. This action will not eliminate the problem but will put it off until some future point in time when point X is once again exceeded. Indeed, any change to costs so as to shift the cost curve downwards or reduce the slope after it cuts the demand curve would have the same effect. Other actions that could achieve this are gaining economies by moving up the learning curve, or changing technology to achieve greater economies of scale.

Alternatively, firms could reduce the markup rate thereby reducing price. This has the effect of shifting the price curve downwards and again shifting point X to the right. The problem will again be put off until later; however, once point X is exceeded, because of the reduced margin, rectangle A will be smaller than before whereas rectangle C will be the same as before; the firm will therefore find itself more quickly in a loss-making situation. Another avenue open for managers to tackle this problem is to shift the demand curve upwards using marketing activity.

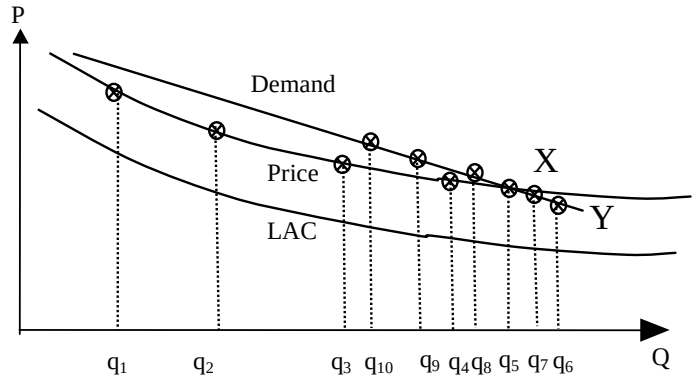


Figure 9. Growth path (where economies of scale still exist)

So far we have discussed the growth path when diseconomies of scale occur. The case where the price curve has positive slope when it cuts the demand curve - ie. where economies of scale still exist - is slightly more complex as shown in figure 9. Here the growth path follows the price curve until point X as before. From there it follows the demand curve to the right until it reaches point Y which is horizontal to the minimum point on the price curve. The growth path then turns back on itself and follows the demand curve to the left; figure 10 shows the area around point X in more detail. This case therefore has two bifurcation points - X and Y - where growth behaviour changes significantly.

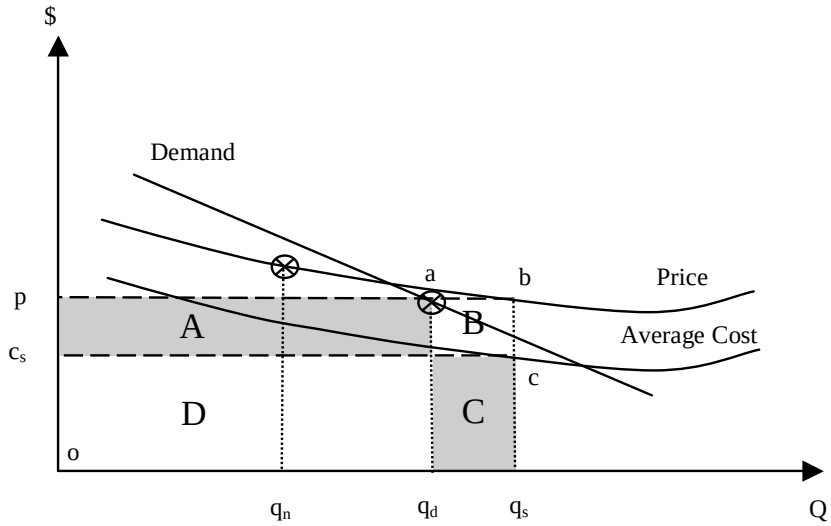


Figure 10. Behaviour at bifurcation point (where economies of scale still exist).

So far in the discussion demand is assumed to have remained constant. Clearly the behaviour will be more complex if demand is allowed to vary either by shifting upwards, changing slope, or both. It is possible for the growth path to follow the price curve to the right, then follow the demand curve, and then follow the price curve once again if the demand curve has shifted outwards.

Under the average cost including demand approach it does not appear possible for the firm to grow indefinitely without managerial intervention. This is not inconsistent with reality where it is rare for firms to survive for a long time and where those firms that do survive are the ones that best respond to a changing environment (De Geus, 1997). De Geus has found the life expectancy of the average corporation to be 'well below twenty years'.

Decline or stagnation of the firm under the average cost pricing approach is due to one of several causes: variable costs exceed revenues and the firm makes losses; production exceeds demand to such an extent that the firm makes losses; or demand begins to decline.

Variable costs at some level of production will become very large due to the effect of diseconomies of scale (modeled by the cubic term in the variable cost function). These large costs will eventually make the firm unprofitable if it continues to produce at that level of output. This is similar to the situation discussed above where the average cost model includes a demand constraint.

It is also possible that due to the growth of the firm production may exceed demand. Retained earnings increase the amount of productive assets available to the firm and this in turn increases production. It can easily occur that production becomes so great as to exceed demand. At this point, while the average cost of goods produced may not change, the average cost of goods sold increases because not all goods that are produced are sold. The excess of goods produced over goods sold either perish or are put into inventory. As mentioned above, this simulation does not model inventory; it assumes that the product being produced is a service or some other non-inventoriable or perishable product. Because standard economic models assume supply equals demand they need only refer to a single average cost variable; they do not separate average cost as two variables as in this model: average cost of goods produced and average cost of goods sold. The interesting behaviour occurs because while average cost of goods produced may remain constant the average cost of goods sold may be increasing and it is goods sold, not goods produced, that determine revenues.

It is also possible that the rate of growth of demand decreases over time - the product life cycle effect - and this in turn may lead to, at some point in time, production exceeding demand with the same consequence as discussed in the above paragraph. In all three cases the simulation demonstrates that losses become very large very quickly.

Discussion

The results outlined above show that the behaviour of the firm under the average cost pricing model, the marginalist pricing model, and the average cost including demand model is qualitatively different. Under the marginalist principle the firm continues to grow for as long as demand continues to grow, under the average cost pricing assumption the firm continues to grow unless it meets a demand constraint, whereas under the average cost including demand approach the firm eventually reaches a limit to growth even if demand continues to grow.

The models differ also in the nature of their causal drivers. The marginalist model is driven by change in demand, the average cost model is driven by change in supply, and the average cost including demand model is driven by both supply and demand. To underline how little impact the supply side has in the marginalist model we note that a production function is not even required for the model to operate. Demand is determined as outlined above and production is assumed to meet this demand. Factors of production, labour and capital, are assumed to be available as necessary in order to produce the quantity demanded. Similarly under the average cost model a demand function is not necessary for the model to operate. Goods are produced according to the production function, priced according to the average cost rule, and it is assumed that all goods are sold.

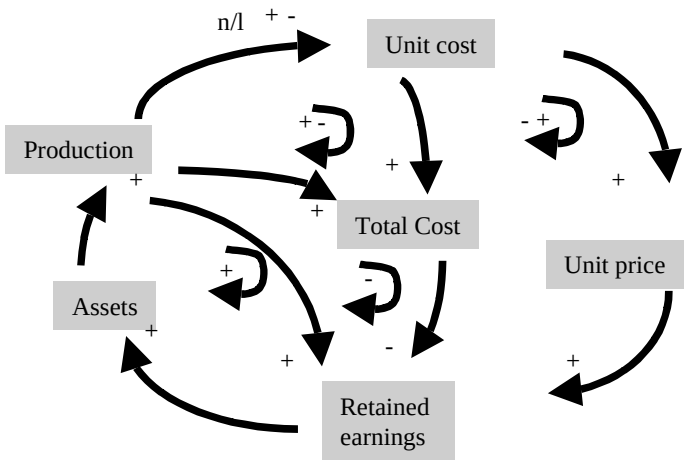


Figure 11. Causal loop diagram under average cost approach.

A causal loop diagram for the average cost model, showing essential variables only, is given in figure 11. Four loops are evident in this diagram. The primary positive or reinforcing loop is from assets to production to unit cost to unit price to retained earnings (loop 1). The primary negative or balancing loop is that from assets to production to unit cost to retained earnings (loop 2). However, at high levels of production, when diseconomies of scale arise, both of these loops change sign. Two minor loops remain: the negative or balancing loop from assets to production to total cost to retained earnings (loop 3) and the positive or reinforcing loop from assets to production to retained earnings (loop 4). Both of these loops are necessary because production is a factor in computing both total cost and retained earnings.

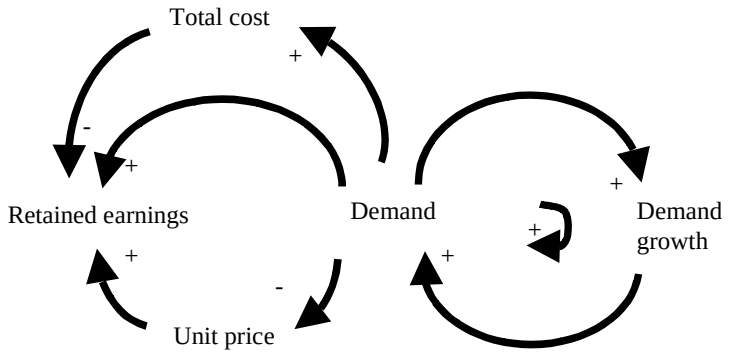


Figure 12. Causal loop diagram under marginalist approach.

A causal loop diagram, showing essential variables only, for the marginalist approach is given in figure 12. Here a single reinforcing loop, governing change in demand, drives the dynamic behaviour of the firm. While causal relationships exist between the variables on the left hand side of the demand variable, no loop exists. Given this much simpler causal loop diagram we would expect that the dynamic behaviour of the firm under the marginalist approach will be less complex than that under the average cost approach.

In contrast to the marginalist and average cost models, the average cost including demand model is driven by both supply and demand as shown in the causal loop diagram in figure 13. This diagram shows six loops. It contains the same single reinforcing loop as the demand

model (loop 1). It also contains three of the same loops as the average cost model (loops 2 to 4). Loop number four in the average cost model does not exist here because production now does not influence retained earnings directly but only via quantity sold. It is replaced in this model by the loop: assets, production, quantity sold, retained earnings (loop 5). This however is not a simple reinforcing loop but instead contains a nonlinearity: quantity sold is only increased by an increase in production when production is less than demand; when production is greater than demand an increase in production has no effect on quantity sold. The sixth loop is: assets, production, unit cost, unit price, demand, quantity sold, and retained earnings. This loop is positive when the relationship between production and unit cost is negative, ie. when economies of scale occur, and negative when diseconomies of scale occur. Also, the relationship between demand and quantity sold is nonlinear: an increase in demand increases quantity sold only when demand is less than production; if demand is greater than production then quantity sold is not altered by a change in demand. Hence, sometimes loop six is active and sometimes loop five is active; at any point in time only one or the other loop is active.

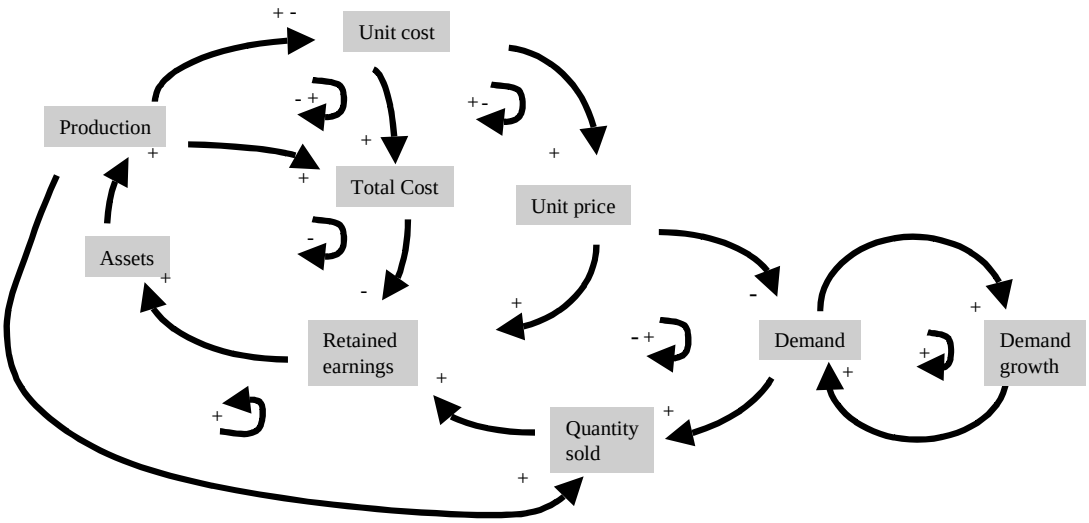


Figure 13. Causal loop diagram (average cost including demand approach).

This model assumes that quantity sold is the minimum of that produced and that demanded. This minimum function is a non-linear function and can lead to a variety of behaviours, including intermittency (profit-loss-profit-loss…), on the part of the firm. Interestingly, the simulation results for this model show that limits to growth exist even when demand continues to increase.

The growth mechanism is different in the three models. In the marginalist model growth of the firm results from growth in demand; in the average cost model growth of the firm results from retained earnings funding further production; in the average cost including demand model growth is influenced both by demand growing and by retained earnings funding further production.

The average cost including demand model is also different to both the marginalist model and the average cost model in that it allows the possibility of excess production. Under the marginalist model no excess production is possible: firms simultaneously determine price and output and then produce exactly that quantity of output. Under the average cost model all production is assumed to be sold and by definition no excess production can occur. However, under the average cost including demand model production is firstly determined, then average cost and price; price then determines demand. It is possible for this level of demand to be less

than the quantity produced giving excess production. As discussed above, this excess production may lead the firm into a loss-making situation.

Both the average cost model and the average cost including demand model will allow underproduction ie. will allow unmet demand to exist. The firm under marginalist assumptions by definition always fully meets its demand. This unmet demand implies that firms adopting average cost or average cost including demand assumptions could always, in theory at least, increase price (of course, only during periods when a demand constraint is not operating). This assumes genuine unmet demand and not simply that the firm is selling all it produces due to extensive marketing activity carried out by the firm.

Under the marginalist approach the firm can survive and prosper without managerial intervention. The growth path under the marginalist approach follows the marginal cost curve to infinity. Under the average cost approach, in theory at least, the firm can also survive and prosper without managerial intervention. Growth under the average cost approach follows the price curve, which lies above the average cost curve by the markup amount, until infinity. However, in practice it is unlikely that a firm can simply forever sell all it produces, especially as diseconomies of scale are incurred and its price rises significantly. In practice a demand constraint would be reached, even if that demand curve is not known to the firm. At this point the firm cannot sell all that it produces and incurs the extra cost of overproduction; the behaviour of the firm will then be similar to that shown by the average cost including demand model.

Under the average cost including demand approach the simulation results show that a firm reaches a limit to its growth, and possibly stagnation and decline, unless managerial action is taken. This occurs because at some point in the firm's evolution a demand constraint is reached and the firm begins to incur the cost of overproduction. Several reasons for reaching this demand constraint are possible: success of the firm leading to increased production capability and hence goods supplied; a natural decline in demand due to the product life cycle effect; or increasing diseconomies of scale causing the firms costs and hence its prices to rise significantly. Managerial action to ameliorate this situation could be: the selective selling of productive assets to reduce overproduction, triggered for example when output exceeds demand by a threshold amount; or increasing marketing activity to increase demand; or reengineering the firms processes to reduce costs and reduce diseconomies of scale.

A study of firm failure has been carried out by Argenti (1976) who lays the blame for failure squarely on the shoulders of top management. Argenti points out that firms take a long time to fail: he found that defects of myopic top management, lack of financial information, and inadequate response to change, often coupled with an economic shock, lead to a decline in profits. This decline in profits in turn prompted the strategic mistakes of excessive leverage, overtrading or taking on a too-large project; this finally lead to a 'nosedive' in cash flows and failure of the firm. His suggestion to firms that are failing is to change the top management, install adequate financial control systems, and put in place a corporate planning activity. This research appears to support some of Argenti's conclusions: simulated firm's often fail slowly at first and then 'nosedive'; overtrading, situations where revenue growth is not matched by profit growth do sometimes occur in the simulation (increasing leverage and taking on a major project are not options currently allowable in the simulation).

It could be inferred from this research that the models that form the theory of the firm are anti-growth or at least do not formally take growth into account. Both the marginal cost and the average cost theories of the firm are predicated around the two variables: price and quantity. Output grows as firms seek to achieve economies of scale; economies of scale are the driving force behind this increase in output not growth itself. For example, in the marginalist model once all long run economies of scale are achieved the firm will stop growing: it would remain for all time at the level of output corresponding to the minimum

point on its long run average cost curve assuming no further change in its demand or cost functions. Marris's model does consider growth by diversification but does not examine the growth of the single product firm.

And yet observation of the real world shows that many well known single business firms are driven by growth objectives and these growth objectives are not predicated simply on gaining economies of scale (although of course that would not refuse them); for example the following excerpts from latest annual reports of Coca-Cola, Wal-Mart, and McDonalds all emphasise the importance that these firms attach to growth as an end in itself:

> 'The ultimate objectives of our business strategy are to increase volume, expand our share of worldwide nonalcoholic ready-to-drink beverage sales, maximize our long-term cash flows and create economic value added by improving economic profit.'
>
> (CocaCola, 1999)

> 'Q. Wal-Mart is the largest retailer in the world, with sales of $165 billion last year. With that in mind, how will you continue to grow sales and profits into the future? Where will the growth come from?
> WALTON: Over the next five years, 60 to 70 percent of our growth in sales and earnings will come from the domestic markets with our Wal-Mart stores and Supercenters, and another 10-15 percent from SAM'S Club and McLane. The remaining 20 percent of the growth will come from our planned growth in the international markets. This means we have a great opportunity to drive our growth doing the things that we do best today in the U.S. market.'
>
> (WalMart, 2000)

> 'To achieve our vision, we are focused on three worldwide strategies:
> - be the best employer for our people in each community around the world;
> - deliver operational excellence to our customers in each of our restaurants; and
> - achieve enduring profitable growth by expanding the brand and leveraging the strengths of the McDonald's System through innovation and technology.'
>
> (McDonalds, 1999)

Research by Marris (1963) and Baumol (1962) indicates that managers lay great emphasis on growth in the size of firms and the rate of growth. Marris suggests that the reward package of managers - salary, power and prestige - is more strongly correlated with growth of the firm than with profitability of the firm.

Growth has been modeled above as a change in the demand function, in particular by a change in parameter a of the demand function causing the demand function to shift outwards. Several different possible growth paths can be followed by the firm. As discussed above, under the marginalist principle the growth path of the firm will be relatively steep with an emphasis on increasing price rather than increasing output as illustrated in figure 14.

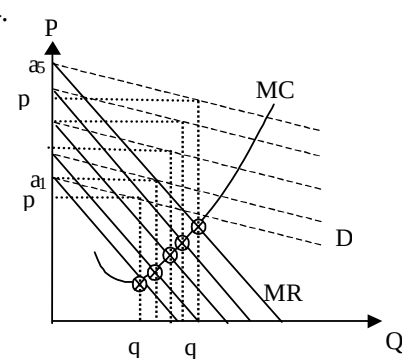


Figure 14. Marginalist growth path

The real firms given as examples above appear to follow a growth path where price is maintained at a relatively constant level but demand is significantly increased by marketing activity. A growth path that matches this situation is shown in figure 15. Real firms would therefore appear to follow a strategy not of maximising profit but of maximising market share. Assuming diminishing returns such firms must at some point trade where all economies of scale have been gained and diseconomies of scale begin to occur. Firm profit margins will increase as economies of scale are gained but decrease as diseconomies of scale are incurred; firms will eventually incur losses when they pass the output point where price is below the average cost curve (this occurs at point X in figure 15). This provides a very real limit to the growth of the firm under an assumption of constant price.

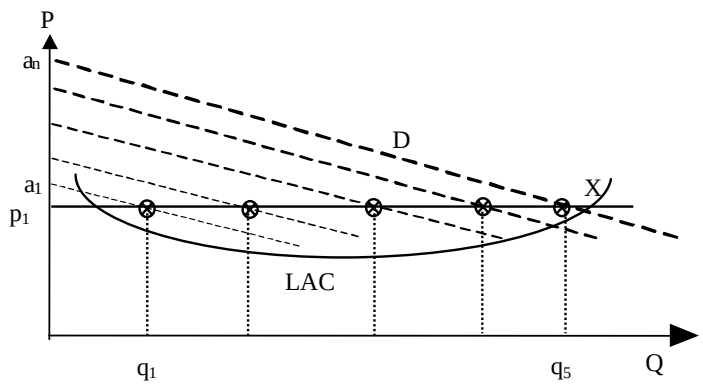


Figure 15. Suggested real firm growth path

The above discussion focuses on marketing activity that sets out to increase demand by increasing parameter a of the demand curve. Marketing activity can also set out to increase brand loyalty by increasing parameter b of the demand curve ie. by steepening the slope of the curve and thereby reducing the elasticity of demand making buyers less sensitive to changes in price. The growth path of firms following this approach is to follow the marginal cost curve downwards and to the left as depicted in figure 16.

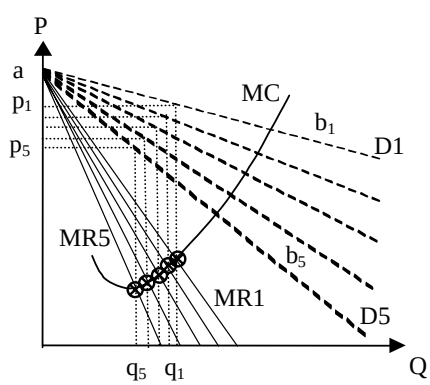


Figure 16. Growth path of firms building brand loyalty.

This figure shows that output of firms that follow this path will decline over time as marketing activities develop brand loyalty and the marginalist principle drives the firm to a new equilibrium point at lower output. Growth will naturally come to a stop at the point

where the marginal revenue line is tangent to the average total cost curve (not shown in figure 16).

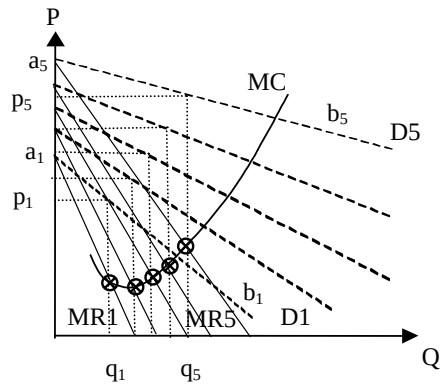


Figure 17. Growth path for firm increasing demand and reducing brand loyalty.

An alternative growth path is shown in figure 17. Here demand for the firms product is increasing but brand loyalty is decreasing, a situation that occurs as a once differentiated product becomes a commodity. Assuming marginalist principles the growth path will follow the marginal cost curve to the right. This is a similar situation to that shown in figure 6 except that here growth will take place faster due to the demand curves becoming less steep.

The average cost including demand model has an advantage over both the marginalist and the average cost models in that it includes both demand and supply as causal elements in the model, whereas the other two models respectively include only demand or supply. It also avoids the somewhat unrealistic assumption of equating demand and supply for the firm during each trading period and instead mirrors real firm behaviour by suggesting that goods sold is the minimum of goods demanded and goods produced. The model allows the real-world behaviour of overproduction and undersupply of goods; the average cost model only allows undersupply, the marginalist model allows neither. The model also allows alternating behaviour ie. from overproduction to undersupply or vice versa, again behaviour not allowable under either of the other two models.

Future Research

Two aspects of the models could be elaborated further: the way in which the models handle losses; and the way in which growth in demand is modeled.

Losses are assumed to be met by the firm taking on debt; this debt accumulates as losses mount; these debts are paid for by interest payments in subsequent periods. However, in reality when faced with a loss-making situation firms may choose to sell off assets rather than take on debt to cover losses. Also, when making profits firms may choose to pay down debt rather than retain it. The model could be enhanced to include the opportunity to make these two distinct managerial decisions: meeting losses from retained earnings of previous periods rather than taking on debt; paying down debt during profitable periods.

An interesting extension to the model would be to model the effect of marketing activity on demand. While it would not make a difference in the average cost approach (because demand is not taken into account) it would likely make a significant difference to the other two approaches. It would also provide another source of feedback in the model which could make the dynamics more interesting and possibly more realistic. Modeling the marketing expense as a proportion of revenue is consistent with the industry practice whereby marketing expenditure is often quoted as a percentage of sales.

The difficulty in modeling the marketing activity of the firm lies in determining its effect on demand. A simple function could be postulated eg.

$$\delta a = \alpha * \text{Marketing expense}$$

where a is the intercept parameters of the demand function and α is a new parameter to represent the effect of marketing on the demand function. This function implies that marketing activities increase the level of demand by shifting the demand function outwards.

Alternatively, marketing activity could be used to increase brand loyalty rather than to increase volume. This could be modeled as:

$$\delta b = \beta * \text{Marketing expense}$$

where b is the slope parameter of the demand curve and β is a new parameter to represent the effect of marketing on the demand function. Increasing the value of b increases the price elasticity of demand meaning that consumers are less sensitive to change in price of the good. This quasi-monopoly position results from the firm's product being differentiated from competitor products due to, for example, brand loyalty. Koutsoyiannis (1979:204) uses the term 'fancied' to distinguish this type of product differentiation, generated through marketing activities, from 'real' differentiation generated through production or service activities. The ability of firms to differentiate their products through building up brand loyalty should not be underestimated. Klein (2000) has examined the marketing activities of firms and their ability to create significant loyalty for a brand within their customer base.

An interesting area for further exploration, especially in the current climate of electronic commerce, is that of increasing returns to scale. Arthur (1990, 1996) has examined industries where increasing returns to scale apply, in particular high technology industries; he points out that the additional cost to companies such as Microsoft of producing a software package on CD is very little, the vast bulk of the costs of software are fixed and upfront. This case of no diseconomies to scale can be modeled by allowing the cubic term in the total variable cost function to tend to zero. Under the average cost approach the average cost per unit produced tends to zero and therefore also does price; this occurs because economies of scale are continually achieved and cost per unit continually reduced as production volume increases. This also means that although the firm continues to grow in revenue its rate of growth decreases due to decreasing prices for its products.

However, it is hard to believe that firms could sustain this kind of growth ad infinitum; it would appear inevitable that some sort of diseconomies of scale must eventually be incurred; growth loops such as this are explosive in nature and therefore unsustainable in the long run.

It is interesting to note that a related phenomenon is being predicted in e-commerce marketplaces. Analysts are suggesting that as buyers appropriate more and more of the transaction economies being wrung from supplying firm value chains - reconfigured due to use of the internet - it will become harder for firms to maintain their revenue streams in an ecommerce environment (Hagel, 2000).

Finally, it would be interesting to extend the model to include inventory. This will include additional delays in the structure of the model and will likely increase the complexity of the dynamic behaviour being modeled. Also, it will be interesting to examine how the use of inventory hides the cost of overproduction. Inventory represents goods retained by the firm from one period to another; it is a means by which the firm balances out peaks and troughs in the demand for its product. In accounting convention, cost of goods sold represents the cost

of producing the goods that were sold during a period; it does not represent the cost of goods produced during the period. In the case where goods may be retained in inventory, revenue less cost of goods sold does equal the margin of the firm. Margins are not reduced due to overproduction as modeled in this paper.

In recent years, mainly due to the influence and success of Japanese manufacturing techniques, firms have been waging a war on inventory. Ohno (1988) use the analogy of a lake: in the same way that the water in a lake hides many obstacles underneath its surface, so too does inventory hide many costs to the firm; once inventory is reduced or eliminated the other forms of waste are allowed become visible. The recent JIT (just-in-time) movement in manufacturing is aimed at eliminating all waste, including inventory. This movement has highlighted several distinct forms of waste, and has declared the waste due to overproduction as its primary target (Imai, 1986; Robinson, 1991).

Bibliography

Argenti, John (1976). 'Corporate planning and corporate collapse', *Long Range Planning*, December, pp.12-17.

Arthur, W. Brian (1990). 'Positive feedbacks in the economy', *Scientific American*, 262, pp.92-99, Feb.

Arthur, W. Brian (1996). 'Increasing returns and the new world of business', *Harvard Business Review*, July-August.

Arthur, W. Brian (1989). 'Competing technologies, increasing returns, and lock-in by historical events', *The Economic Journal*, March, pp.116-131.

Baumol William J (1962). 'On the theory of expansion of the firm', *American Economic Review*, v.52, n.5, pp.1078-1087

Brady, Malcolm (1999). 'Patterns of growth: an investigation into the dynamics of firm performance', paper presented at *Young Operational Researchers Forum*, 22-24 March, University of Sheffield.

Brady, Malcolm (2000). 'Simulation of a business firm', *YOR11: Young Operational Research Conference*, 28-30 March 2000, Fitzwilliam College, Cambridge

Brandenburger, Adam M. and Barry J. Nalebuff (1996). *Co-opetition*, Currency Doubleday.

Brealey, Richard and Stewart Myers (1996). *Principles of corporate finance*, McGraw Hill (5th edition).

Brennan, Ciaran (2001). 'Gateway staff rocked by news of staff losses', *Irish Times*, Jan 13th.

Business Eccountant, The (2001). 'Selected company profile: Nissan', 45th ed., 25th Feb.

CocaCola (1999). CocaCola Annual Report for 1999, available at http://www.cocacola.com/annualreport/1999/financials (accessed 26 March 2001)

De Geus, Arie (1997). 'The living company', *Harvard Business Review*, March-April, pp. 51-9.

Douthwaite, Richard (1992). *The growth illusion: how economic growth has enriched the few, impoverished the many, and endangered the planet*, Lilliput.

Hagel, John (2000). 'Whither E-commerce?', *Knowledge at Wharton: Managing Technology*, Available at: http://knowledge.wharton.upenn.edu/print_version.cfm?articleid=219&catid=14

Higgens, (1977). 'How much growth can a firm afford?', *Financial Management*, Fall, pp. 7-16.

Imai, Maasaki (1986). *Kaizen: the key to Japanese competitive success*, Random House.

Klein, Naomi (2000). *No logo*, Flamingo.

Koutsoyiannis, A. (1979). *Modern microeconomics*, Macmillan, 2nd ed.

Lintner, John (1956). 'Distribution of incomes of corporations among dividends, retained earnings and taxes', *American Economic Review*, May, pp. 97-113.

Lymer, Andy and Dora Hancock (2001), *Taxation policy and practice: 7th ed. 2000/2001*, Thompson Learning.

Marris, Robin (1963). 'A model of the managerial enterprise', *Quarterly Journal of Economics*, v.77, n.2., pp.185-209.

McDonalds (1999). McDonald's Corporation annual report for 1999, available at http://www.mcdonalds.com/corporate/investor/financialinfo/annualreport/online99/index.html (accessed 26 March 2001).

Meadows, Donella, Dennis Meadows, Jorgen Randers, William Behrens (1972). *The limits to growth*, Potomac Associates.

Microsoft (1991). 'Bill Gates and the management of Microsoft', Case no. 9-392-019, Harvard Business School.-

O'Clery, Conor (2001a). 'Lucent Technologies to make 10,000 redundant', *Irish Times*, Jan 25th.

O'Clery, Conor (2001b). 'Job cuts at Motorola and Dell drive prices lower', *Irish Times*, Feb 10th.

Ohno, Taaichi (1988). 'Toyota production system: beyond large scale production', in Robinson (1991).

Penrose, Edith (1980). *The theory of the growth of the firm*, Basil Blackwell (2nd edition, first published in 1959).

Powersim (2000). *Powersim Constructor Lite*, (simulation software package), available at www.powersim.com accessed on 31 January 2001.

Robinson, Alan (ed.) (1991). *Continuous improvement in operations*, Productivity Press.

Shearer, J. Lowen, Arthur T. Murphy, Herbert H. Richardson (1971). *Introduction to system dynamics*, Addison-Wesley.

System Dynamics Society (2001). 'What is system dynamics?', available at http://www.albany.edu/cpr/sds (accessed 23 March 2001).

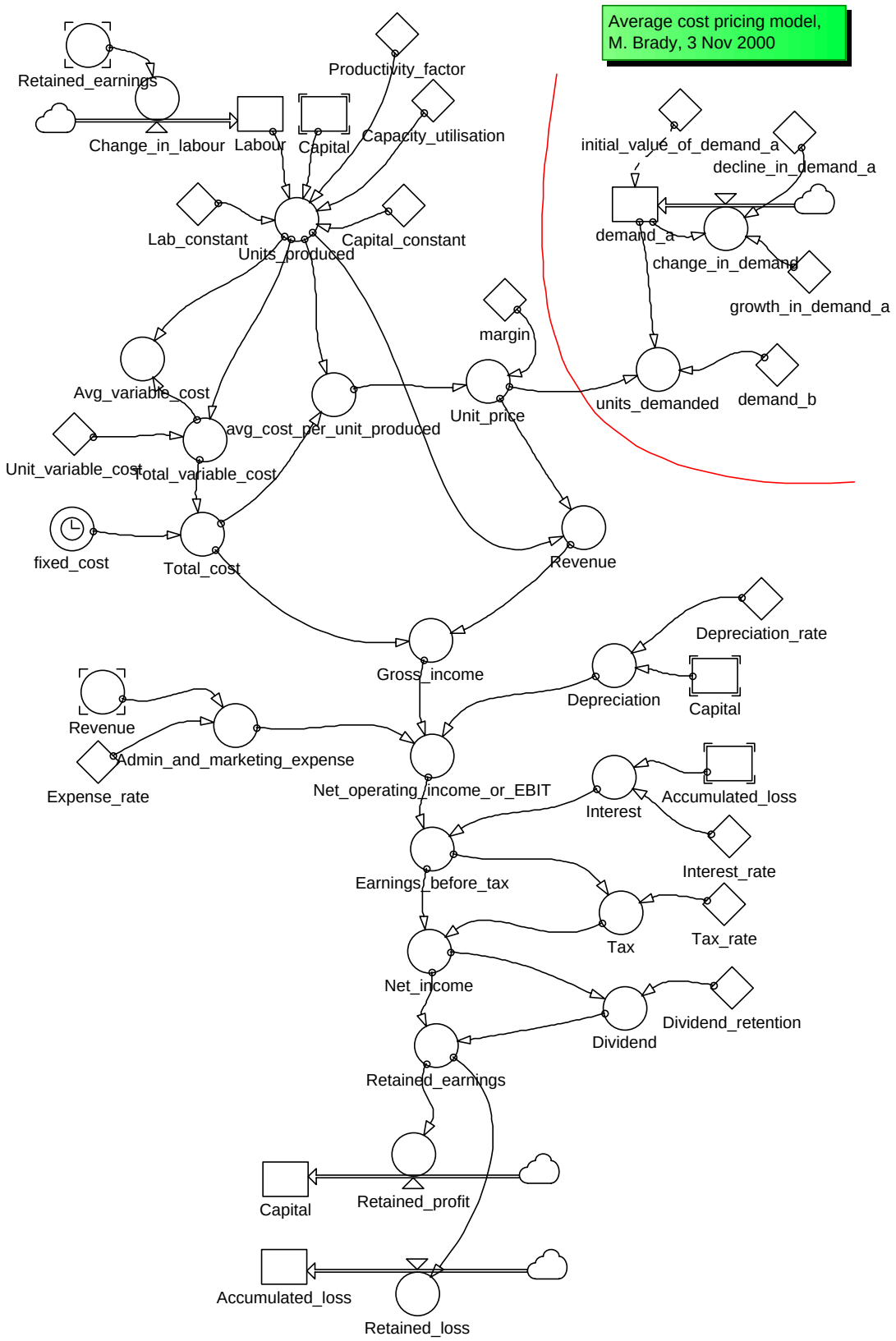
WalMart (2000). WalMart Annual Report for year 2000, (question one from section 2: Question and answer session), available at http://161.165.194.32/newsstand/news_spl_anrpt.shtml (accessed 26 March 2001)

Weston, J. Fred and Thomas E. Copeland (1992). *Managerial finance*, Dryden Press (9th ed).

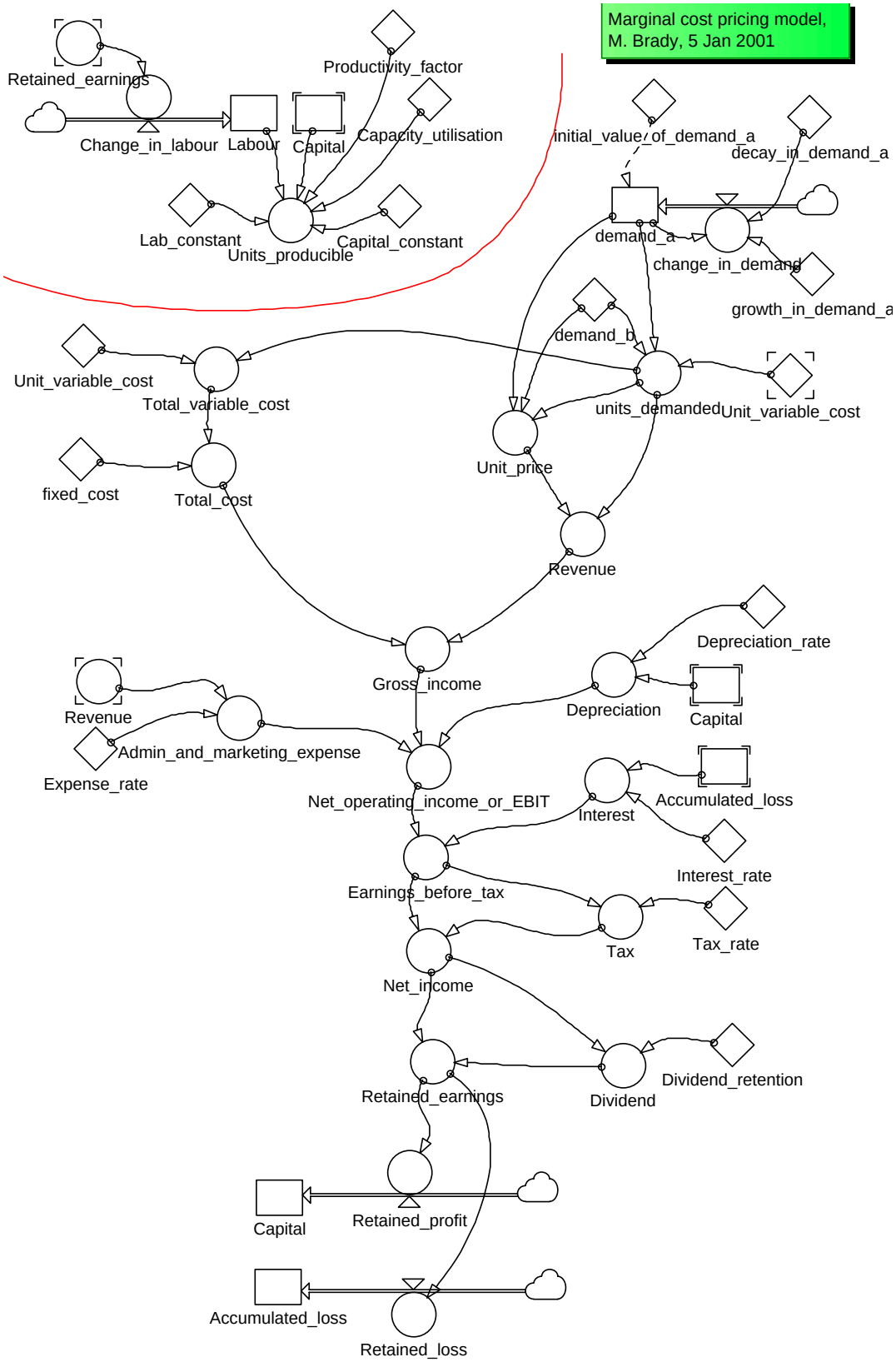
TCA (1997). *Taxes Consolidation Act, 1997,* Government of Ireland.

Thompson, A.J. and Arthur Strickland (1999). *Strategic Management: concepts and cases*, Irwin McGraw-Hill, (11th edition).

APPENDIX ONE



APPENDIX TWO



APPENDIX THREE