INSURANCE PROFITABILITY: AN ECONOMIC PERSPECTIVE

By Lee M. Smith

Introduction

> The evil results of the failure to emphasize the theoretica character of economic speculation are apparant in every field of practical economics.... The problem of profit is one way of looking at the problem of the contrast between perfect competition and acutal competition.... The key to the whole tangle will be found to lie in the notion of risk and uncertainty....[1]

These introductory remarks to Professor Knight's classic treatise on profit provide useful perspective for a contemporary venture into this most difficult subject matter.

Profit is an economic concept, and cannot be meaningfully understood in any other context. In a free enterprise economy profit plays a key role in the determinations as to production and consumption patterns. Profit levels are a determinant in investment patterns which determine and are determined by consumption patterns. The profit levels which bring forth the allocation of resources which maximizes societal welfare are the levels which are to be sought. Determination of criteria for maximum societal welfare, of course, would occur prior to any decision regarding profit levels.

Unfortunately, there is little consensus regarding the appropriate definition or role of profits in a market economy. Classical economists following Adam Smith saw profits as a class of income for the capitalist-entrepreneur. As such, profits were somewhat indistinguishable from interest income. Subsequent writers focusing more specifically on profits separated the function of the entrepreneur, considered responsible for profit, from that of the capitalist to whom interest income accrued.

Clark[2] divided total income into wages, interest, and profit. The latter was seen as dynamic income resulting from the coordination function of the entrepreneur. This amounted to a distribution of residual income based on the productivity of the various factors.

Hawley[3] disagreed with Clark's approach, preferring to look at the ownership function primarily as one of risk-taking rather than coordination. This was a precursor of the risk theories of profit. It could be categorized as a disutility division of ultimate price.

Willett[4] saw both the activity and reward aspects of profit. Profit was seen as a motivating force in economic activity. Conversely, risk was seen as a deterrent. Profit was a temporary gain to be ultimately eliminated by competition.

Knight[5] saw profits as resulting from incorrect estimates about future returns. The idea of profit as payment for service or sacrifice was not a satisfactory one. Profits arise because of limited competition which is a result of uncertainty. Non-economic factors such as power and ability can also give rise to profits.

Many contemporary debates on profitability focus on technical issues of measurement and criteria for optimal level. By not considering underlying economic factors, the combatants often get lost in arcane abstrusities. Disagreements over profit measurements often amount to differences in attitudes as to the role profits should play. That role has been variously associated with the ideas of capital, entrepreneurship, uncertainty, dynamism, and monopoly. Only when viewed in an economic context can profitability be dealt with in the only meaningful way, as a mechanism in attaining the goal of optimal resource allocation.

Economic Theory

The subject of economics is usually dichotomized into macroeconomics and microeconomics. Macroeconomics deals with variables which represent aggregated economic concepts. Microeconomics deals with individual economic units. Profit enters into both of these subject matters.

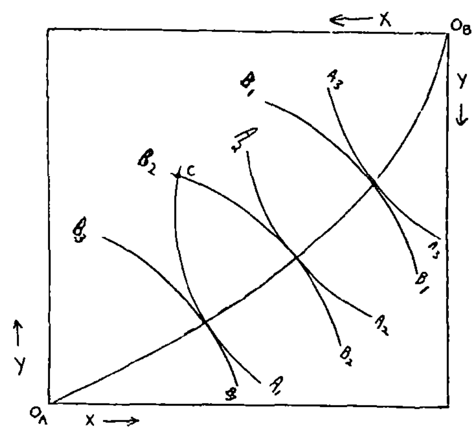
Interest rate, a variable of prime importance in macroeconomic analysis, is an element involved in the determination of societal investment levels. The relationship of anticipated profit from various projects to anticipated interest rates is a major element in general and specific investment levels. Firms presumably add to their capital stock as long as such additions are expected to increase rate of return relative to other investments. Profit levels, then are a determined and determining variable in macroeconomic analysis.

It is not in the area of macroeconomics, however, where we will find guidance in terms of criteria by which to evaluate profit levels in a given industry. The existence of profit in an industry gives rise to a particular output of goods for that industry. The apropriateness of that level of output, and of the profit level giving rise to it, can be evaluated only with tools available within the subject matter of microeconomics. A profit level for an industry is appropriate if it gives rise to a level of output which most efficiently optimizes societal welfare as defined by the decision maker.

In a free enterprise economy in which pure competition exists in all markets, the role of profit is fairly clear. Supply and demand interact to set output and price levels. Long run profit levels are fully determined at the level of opportunity costs of alternative uses of the profit generating factors. In a mixed economy with many industries dominated by a few firms and with social costs of production being a greater and greater proportion of total costs, the role and appropriate level of profit for a firm or industry may be less clear. Development of mechanisms which move profit levels in such an economy toward the optimal levels they would attain in a purely competitive economy with no externalities is a herculean task.

In a purely competitive market, economic forces tend to bring about an optimal allocation of resources relative to the normative Pareto optimality standard. An allocation is Pareto optimal if production and distribution cannot be reorganized so as to increase the utility of one or more individuals without decreasing the utility of others. This is the easiest standard to work with, because other standards require development of interpersonal utility evaluations.

A distribution of consumer goods is Pareto optimal if every possible reallocation of goods that increases the utility of one or more consumers results in a decrease in utility for others. Each consumer's utility must be at a maximum given the utility levels of all other consumers. An Edgeworth box-diagram is often used to illustrate this idea in a simplified economy consisting of two individuals and two commodities.

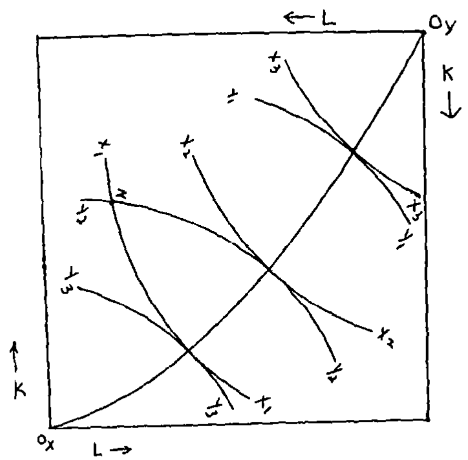


Each point within the box represents a particular distribution between individuals A and B of the available goods in the economy. The A_i's represent indifference curves for A, the B_i's represent B's indifference curves. Indifference curves show the various combinations of two goods which provide equal satisfaction to a consumer.

If the initial distribution of goods is such that it falls at a point where the individuals' indifference curves are not tangent, there is a basis for mutually advantageous exchange. For example, if the initial point is C, A will be willing to give up more of good Y for a unit of X that is needed to induce B to give up a unit. This can be seen by comparing the horizontal distance one moves along A's curve for one vertical unit to the distance along B's curve. At point C the marginal rate of substitution between the goods differs for A and B, and a redistribution of goods can increase utilities of both.

It is at points of tangency of the indifference curves where the marginal rate of substitution between the goods is equal for both individuals and Pareto optimality is attained. At these points the utility of one individual cannot be increased without decreasing the utility of another. Joining all these points defines what is known as a consumption contract curve, 0_A0_B. The simple economy is in equilibrium in the consumption sector when on this curve.

A fully Pareto optimal allocation of resources requires equilibrium in the production sector as well. Such equilibrium results when the marginal rate of technical substitution between factor inputs in the production of all goods is equal. Again considering a simple economy consisting of two goods and two factors of production, we can represent equilibrium positions in a box diagram.

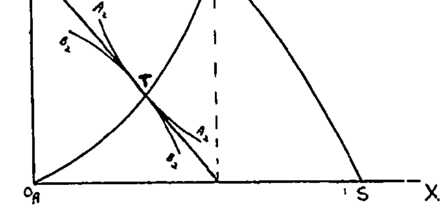


Each point within the box represents production levels of X and Y given the total available levels of factor inputs. Product X's isoquants are given by the X_i's, product Y's by the Y_i's. An isoquant shows different combinations of labor and capital with which a firm can produce a particular level of output of a good.

If the initial level of output is such that it does not fall at a point of tangency of the isoquants of the two goods, the economy will not be maximizing its output given the available factor inputs. For example, if the initial level of output falls at point Z, some capital can be transferred from production of X to Y and some labor from Y to X and output of Y will be increased without decreasing output of good X.

At point Z, the slope of the X isoquant (i.e., the marginal rate of technical substitution) exceeds that of the Y isoquant. The marginal rate of technical substitution is the amount of one factor a firm can give up by increasing the amount of the other factor a unit and remain at the same production level. In such a situation a factor input can be transferred to production of another good and used more efficiently. This is related to the fundamental economic idea of diminishing returns.

Joining the tangency points gives the production contract curve O_XO_Y. This curve shows the various combinations of X and Y the economy can produce by fully utilizing its factors of production and using the best available technology. This curve can be mapped into the XY plane which shows efficient output combinations. In that plane it is known as a product transformation curve.



The product transformation curve gives the various combinations of X and Y the economy can produce when in equilibrium. Its slope at a particular point in time gives the marginal rate of transformation of X for Y at that point. This tells us how much the economy must reduce its output of one good to release enough factors of production to produce one more unit of the other. At this point it should be noted that if profit levels for a firm or industry are not appropriate it may not be producing at optimal levels or with an optimal factor mix, and production will be distorted.

General equilibrium of production and exchange can now be described for our simple economy. By taking a point on the transformation curve we define a particular (X,Y) combination. The consumption contract curve for this combination can be constructed for the two individuals in the economy. We will see, then, that once production levels are determined by available productive resources, general economic equilibrium can be described. It is this point toward which all economic activity should be aimed, assuming Pareto optimality is the goal.

General equilibrium occurs at the point where $MRT_{XY}=(MRS_{XY})_A=(MRS_{XY})_B$. Each point on the transformation curve RS corresponds to a point of equilibrium production. If the output of the economy is at point OB, an Edgeworth box corresponding to that point can be constructed. Every point on the consumption contract curve 0_A0_B is a point of exchange equilibrium. Simultaneous equilibrium occurs at point T where the marginal rates of substitution of the individuals for the two goods equals the marginal rate of transformation of the goods.

At any point on the contract curve where the MRS of the individuals does not equal the MRT of the products, the individuals would be willing to give up more of one commodity for another than production facilities require. For example, if $(MRS_{XY})_A= (MRS_{XY})_B=2$ and $(MRT_{XY}) = 1$, the consumers would be willing to give up two units of Y to get a unit of X. To produce the additional unit of X, however, would require giving up only one unit of Y. Only when all marginal rates are equivalent will there be no incentive to change production or consumption levels.

A consumption contract curve can be constructed for each point on the product transformation curve. An equilibrium point can be derived for each such point as was done for point OB above. The set of equilibrium points can be mapped into a utility space from the output space by considering the utility curves of the individuals in the society. The resulting curve represents Pareto optimal utility levels for the individuals, and is called the Grand Utility Possibility curve. To choose a point of optimal utility from this curve would require development of a social welfare function which could only be done if interpersonal utility comparisons were to be made.

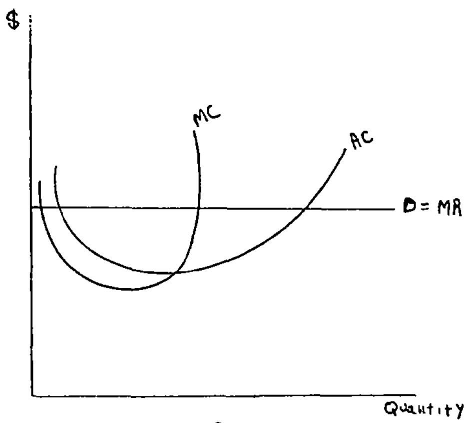
These principles for attainment of Pareto optimality relate to more familiar elements of microeconomic theory. In a particular market for a particular good, say insurance, equilibrium is attained when supply and demand for the good are equal. Consumer tuility curves underlie market demand curves. Consumers are presumed to be in equilibrium when the marginal utility gained by the last dollar spent on each commodity they consume is the same. Demand curves generally exhibit an inverse relationship with price.

Behind supply curves are the factor costs involved in production. Optimal capital-labor ratios for given production levels can be determined by tangency points of isoquants and isocost curves. A firm's short run marginal cost curve is generally considered to be its supply curve. The slope of the long run cost curves depend upon economies of scale.

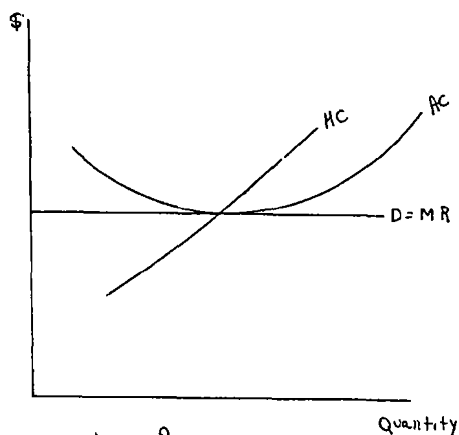
Equilibrium between supply and demand depends to a large degree on the nature of the market for the good in question. In a perfectly competitive market the price and quantity of the commodity are determined solely by the intersection of the market supply and demand curves. Each firm faces a horizontal demand curve and produces to the point where its marginal revenue, i.e. market price, equals its marginal cost in producing the last unit. It should be noted that market price allows for what is known as normal profit, and in the long run this is the profit level which leads to optimal resource allocation.

When there are monopolistic elements in the marketplace for a commodity, the firm's marginal revenue curve will tend to lie below its demand curve. By selling at the profit maximizing point, where marginal revenue equals marginal cost, a monopolistic firm generates some pure profit and a smaller than optimal output. One view of the regulatory responsibility in such a situation is to bring the situation back toward the results in a competitive market.

To see how these economic ideas come together in a demonstration of conditions for Pareto optimal resource allocation, it is useful to consider equilibrium conditions arising under different market structures. The graphs below illustrate equilibrium conditions in the long and short run in perfectly competitive markets.



Short Run



Long Run

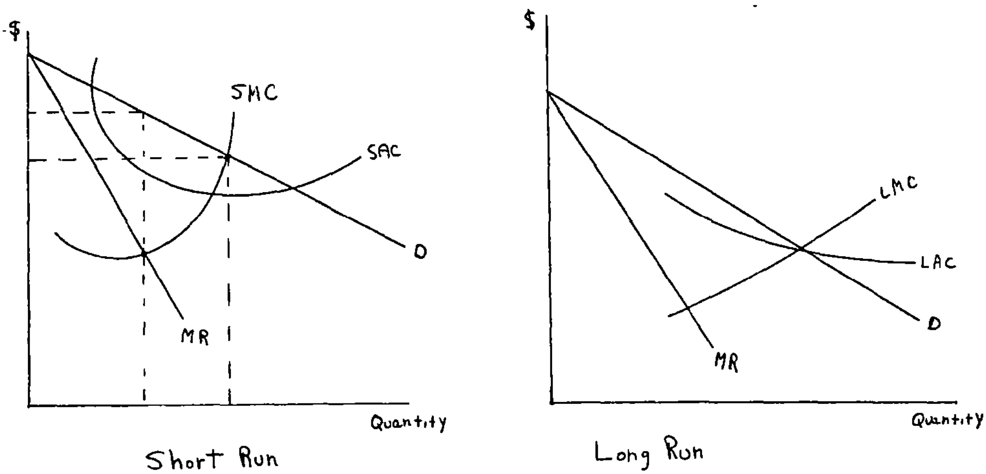
The key distinguishing feature of equilibrium conditions in a perfectly competitive market is that produ ion stops at the point where price equals marginal cost. This is because a profit oriented enterprise always produces to the point where the revenue resulting from the last unit of production is just equal to the cost of producing that unit. When markets are perfectly competitive, a firm has no influence on price and as a result the marginal revenue of each unit sold is the market price. Setting marginal cost equal to marginal revenue, then, results in the production level at which marginal cost equals price.

In the short run abnormal profits can arise in a perfectly competitive market because firms may not be operating at their most efficient levels. A profit per unit equal to the difference between average cost and price will be earned. In the long run competition will squeeze out excess profit as remaining firms move to most efficient production modes. As can be seen on the long run graph, equilibrium occurs where the long run marginal cost curve intersects the long run average cost curve. This occurs at the lowest point on the average cost curve.

This equality of price with marginal cost is essential to attainment of Pareto optimality. To see this, it is helpful to consider equilibrium conditions under an alternative market structure, monopoly. When the market for a good is dominated by a single firm the firm is the industry, and it thus faces a downward sloping demand curve. To increase sales a monopolist must reduce price.

As a result, the monopolist's marginal revenue curve lies below its demand curve.

Equilibrium conditions in a monopoly market are illustrated in the graphs below:



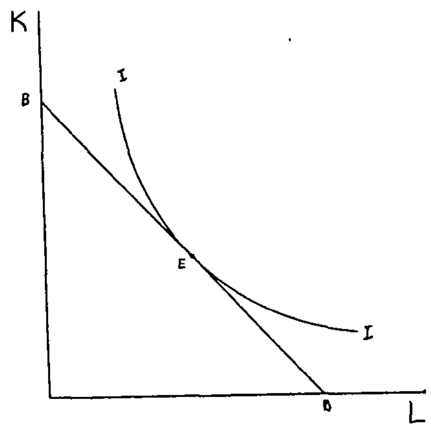
In equilibrium, as was the case in a perfectly competitive market, a monopolist produces to the point where marginal cost equals marginal revenue. Unlike the perfectly competitive firm, however, a monopolist does not have equality between price and marginal revenue. Thus, in a market with monopolistic tendencies, price and marginal cost are not equal in equilibrium.

It should be noted that while a monopolist is presumed to operate at the most efficient scale in the long run, output and price are not ideal from the point of view of Pareto optimality. As shall be shown shortly, only when price equals marginal cost can resource allocation be said to be Pareto optimal. One approach to regulating rate of return, then, is to set the price in a regulated industry at the point which would be attained in a perfectly competitive market; i.e., where price equals marginal cost.

Demonstrating the critical element of equilibrium in a perfectly competitive market to be the equality of price with marginal cost is the key to a demonstration that perfectly competitive markets lead to Pareto optimality. The demonstration that competitive markets lead to Pareto optimality is the key to a demonstration that the economic standard for rate of return in a regulated industry is the rate which would arise in a perfectly competitive market. That economic standard for rate of return is the hypothesized starting point of any development of criteria for profitability for a firm or industry.

As we have seen, there are three conditions which must be met for production and distribution in an economy to be Pareto optimal. The first is that the marginal rate of technical substitution (MRTS) between factor inputs be equivalent for all products. The second is that the marginal rate of substitution (MRS) between commodities be equivalent for all individuals. The third is that the marginal rate of transformation (MRT) between goods in production equal the marginal rate of substitution of the same goods between individuals.

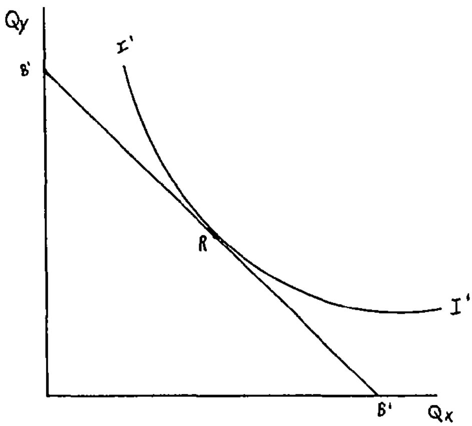
The first condition is met when markets are perfectly competitive. From production theory, we know that equilibrium occurs at the point where output is maximized given the budget constraint. This occurs at the point of tangency of the isocost line which shows combinations of factor inputs in a firm can produce within its budget and an isoquant which show combinations of factor inputs which can be used to produce a particular level of output. Point E in the graph below represents such an equilibrium point.



Since the slope of the isoquant equals the marginal rate of technical substitution of the factors, and the slope of the isocost line is the price ratio, we have the following relationship at equilibrium: $MRTS_{LK} = MP_L/MP_K = P_L/P_K$, where (L,K) represent factor inputs. Thus, $MP_L/P_L = MP_K/P_K$ or the marginal product of the last dollar spent on each factor input is equal.

When there is perfect competition in factor markets the cost of factor inputs is the same in each industry. Thus P_L/P_K is the same for all products. As a result, $(MRTS_{LK})_x = (MRTS_{LK})_y$ and the first condition for Pareto optimal allocation is met. It should be noted that in an industry like insurance where cost curves are influenced significantly by things other than factor inputs demand analysis takes on more importance.

Turning to demand theory, we see that the second condition for Pareto optimal resource allocation is also met when there are perfectly competitive markets. Consumption equilibrium for an individual occurs when utility is maximized given available income. This occurs at the point of tangency of the individual's budget line and indifference curve; i.e., where the individual allocates income so as to reach the highest indifference curve, and therefore highest utility. Point R on the graph below represents such an equilibrium point.



The slope of the indifference curve equals the marginal rate of substitution between the goods and the slope of the budget line equals their price ratio. Thus, MRS_{xy} = MUx/MUy = Px/Py in equilibrium. Consumption equilibrium, then, occurs where the marginal utility derived per dollar spent on each commodity is the same: MUx/Px = MUy/Py.

When there is perfect competition in product markets, the price of a good is the same for all consumers. Thus Px/Py is the same for all consumers. As a result, $(MRS_{xy})_A = (MRS_{xy})_B$, and the second condition for Pareto optimality is met in a perfectly competitive market.

The final condition for Pareto optimal resource allocation is that the marginal rate of transformation in production for each set of goods equal the marginal rate of substitution between the goods for each individual. Now the marginal rate of transformation between goods is the amount by which production of one must be reduced to release enough factors to produce a unit of the other; i.e., MC_x/MC_y. Under perfect competition, $MC_x = P_x$ and $MC_y = P_y$. Thus, $MC_x/MC_y = P_x/P_y = MRT_{xy}$. But we saw that in consumption equilibrium $P_x/P_y = MRS_{xy}$. Thus, when there is perfect competition in all markets $MRT_{xy} = MRS_{xy}$ and Pareto optimal resource allocation results.

The above review of elementary economic principles reveals the richness of that subject matter and its usefulness in dealing with an issue like profitability. The economic function of profit is to assure society's resources are used efficiently and result in optimal consumption patterns. It performs that function by rising to the level in each industry which brings forth the supply of goods which optimally satisfies consumer demand. While value judgments play a significant role in determining what optimal consumption is, economic laws determine how that consumption will be satisfied once it is established.

The Insurance Profitability Debates

Putting abstract economic theory to work in resolving a very concrete problem like profitability in the insurance industry would be quite a formidable undertaking. Luckily, there is a precedent for such activity as a result of the fact that certain

American industries have been subject to regulation for many years. Appropriate rate of return for public utilities is an issue of continuing analysis and a large body of knowledge has arisen therefrom.

Before considering some of the elements of utility regulation which are helpful in developing approaches to insurance regulation, however, we should review some of the very significant analysis which has been done on the specific subject of insurer profitability. In the late 1960's a combination of economic and social forces caused persons involved in the process of regulating the insurance industry to consider evaluating profit levels. Rate levels for certain insurance coverages were rising at what then seemed very steep rates, and reasons were being sought. One target in such a situation is the profit levels of the firms involved. Thus began the insurance profitability studies.

The National Association of Insurance Commissioners (NAIC) developed a report capsulizing what was learned from the studies. This report was made part of the Proceedings of the NAIC in 1970. The various problems related to the profitability issue were discussed in that report. Among those problems were the unique accounting methods used in the insurance industry, the various possible ways of measuring profitability, and the treatment of investment income in ratemaking.

A major problem introduced by insurance industry accounting practices is that financial records have been kept on what is called a statutory basis. This basis treats income and outgo relating to insurance transactions on an inconsistent basis. Premiums are treated as income only as earned or accrued, but certain expenses are charged off immediately. One way to improve this situation from the point of view of profitability measurement is to go to a GAAP method of accounting which treats expenses on a basis consistent with premiums.

Considering the use of investment income in ratemaking proved to an emotional, but not very enlightening, exercise. There seems to be little theoretical basis for ignoring investment income in ratemaking, but there seems to be little one can do to incorporate it until one has developed a profitability standard. The discussions about good and bad aspects of using investment income are inconsequential if they don't go the question of resultant profitability level and its appropriateness.

Thus we come to the measurement of profitability and the development of criteria by which to evaluate the resultant figures, as the key to the situation. The starting point of the discussion on this issue was a study done by Arthur D. Little, Inc. (ADL). Using a measure of profitability equal to total income from all sources divided by total investable funds, they found no evidence of excessive profitability relative to the risk insurers face. The study also indicated the existence of an adequate degree of competition to assure monopoly profits would not arise.

NAIC staff felt the ADL measure was defective because insurance companies have access to large amounts of interest-free capital supplied by policy holders. When profitability is measured as total income relative to total investable funds, insurance companies have significant funds in the denominator which contribute no income in the numerator. This results in a deflated looking return when compared with industries which don't have this interest-free captial.

Bob Bailey[7] suggested that the conceptual problem with the ADL approach was that it ignored the "interest" policy holders receive in the form of lower rates in return for the use of their funds. Robert Ferrari[8] developed a model relating underwriting and investment income to net worth which showed where the ADL measure differed from some alternatives. There was little doubt that the ADL measure introduced some problems in making inter-industry comparisons.

Other ways of measuring profitability were discussed. Particular attention was paid to variations of a method originally proposed by Russell Goddard. This method essentially involves developing a measure of income relative to a corresponding net worth figure. The NAIC staff took the original proposal and made some modifications which they felt led to a reasonable measure of return on net worth. This measure could then be used in inter-industry comparisons.

Norton Masterson, hired as a consultant by the NAIC, developed a model relating net profit to earned premium and net worth. The profit formula was designed to allow for net after tax profit sufficient to maintain the company's net worth at a level allowing for adequate growth. Profit was also to be adequate for a return on net worth to stockholders comensurate with risk and competitive investments. In addition, premium to surplus ratios were to be kept within reasonable limits.

Other approaches considered included one developed for a Virginia proceeding, and a discounted cash flow approach. The former was developed by a Dr. Schotta who felt that principles of public utility ratemaking could be appropriate for the insurance industry. The discounted cash flow approach evaluated the present value of income and outflow from an insurance transaction. Problems involved in determining appropriate discount factor, payment patterns, inflation and tax effects made the latter approach of limited practical value even for estimating investment earnings.

The NAIC study concluded that there were many problems associated with determining profitability standards for the insurance industry, and that a good deal of theoretical work was needed. Standards for reasonable profit levels would include levels in similar industries, levels needed to attract and retain capital, and levels which would preserve the financial integrity of companies. In the interim, a series of six measures of return have been calculated and published annually.

Financial Analysis

Many of the issues arising in the insurance profitability debate are issues in utility regulation as well. Thus, an understanding of utility regulation should prove valuable in our search for criteria for insurance profits. Because techniques of financial analysis have been a significant element in utility rate of return regulation, an introduction to that subject would be useful. Financial analysis originally fell within the overall domain of accounting. As socioeconomic forces became more and more important in determining financial activity, however, the subject matter of economics became more and more important to financial analysts. Financial analysis is currently well-steeped in economic theory.

Expected rate of return rather than past earnings streams have become prime data used by financial analysts. Evaluation of expected price of a company's stock primarily involves estimating future earnings of the company. This estimation is sometimes done by looking at past earnings and projecting forward. It can also be done by determining variables which have been correlated with company earnings and projecting based on expected movements in those variables. Determination of the value of a company's common stock is a critical element of profitability analysis because one standard for appropriate rate of return is the rate of return which leaves stock price unchanged.

Once future earnings are estimated, stock values must be projected therefrom. One way is to estimate the expected stream of dividends to arise from the stock and calculate the present value of that flow. Present value can be calculated at the capitalization rate investors in the stock are expected to require. The rate will depend on alternative investment opportunities and the relative riskiness of the company being evaluated.

Determination of security values, then, involves determination of a number of variables of a firm. Among the values of a company which are needed for a full evaluation of stock values are rate of return and portion paid out in dividends, expected growth rate, and yield required by investors. The first place one would normally go to derive data for these variables would be the financial statements of the firm in question.

One time-honored approach to determination of historical rate of return is to calculate the product of profit margin (profit/sales) and turnover ratio (sales/assets). Unfortunately, this measure does not provide insight into the impact of financial structure (debt/equity ratio) on return. A profit measure which separates interest on debt from return on assets would resolve that problem.

Define the following terms: Π = profit; t - tax rate; r = pre-tax, pre interest return on assets; A = asset level; i = interest rate on borrowed funds; L = liabilities; E = equities. A measure of profit which would allow evaluation of effect of financial structure on return would be:

$$\Pi = (1-t)\ (rA-iL) = (1-t)\ [r+(r-i)\ (L/E)]\ E$$

The capital structure (L/E) is explicitly recognized in this formula and the effect thereof on rate of return can be evaluated over time.

The rate of return on equity, then, is $\Pi/E = (1-t)[r+(r-i)(L/E)]$. To maximize return, management would set the debt/equity ratio at the level where the derivative of the equation equals zero: $(1-t)\ (r-i)$. The implication of this is that the ratio of debt to equity should be expanded until the rate of return on assets equals the interest rate paid on borrowed funds. Unfortunately, since r and i were treated as constants, they cannot move together.

To determine an optimal relationship between rate of return on equity and capital structure, then, a relationship between r or i and (L/E) must by hypothesized. Assuming i is a rising function of debt/equity ratio, for example, will allow solution for maximum rate of return at some relationship between return on assets and interest rate. Such considerations play an important role in deliberations over proper rate of return in a regulated industry.

One way of determining the appropriate price of an asset is to calculate the present value of the income stream expected to arise from that asset. A formula for the present value of a continuous flow of funds from an asset is given by: $P = \int_0^{\infty} rA_0\, e^{-dt}\, d\tau$ where r is the rate of return earned on the asset, A_0 the asset's original value, and d the discount rate applied. Solving for P, the asset value, we get: $\frac{rAo}{d}$, the familiar value of a perpetuity of amount rA_0 at interest rate d. By setting purchase price of the asset, A_0, equal to the value of this flow, set $P = A_0$ and solve for r. We see that r = d which means the discount rate converting the cash stream of an asset into a present value should equal the rate of return earned on the asset, all other things being equal.

One simplifying assumption of the above analysis is that earnings of a company over time are expected to remain constant. Incorporating growth into the income stream does not produce much complication. Assuming an original dividend of D_0 and growth rate g, the value of the dividend at time t is D_0e^{gt}, assuming continuous growth. At a discount rate of d, the present value of a stock with initial dividend D_0 expected to grow continuously at rate g is: $P_0 = \int_0^{\infty} D_0e^{gt}e^{-dt}dt$. Solving for P_0, we get: $D_0/(d-g)$. Thus, the price of a security equals its return capitalized at a rate equal to the difference between its discount and growth rates.

This model can be made more generally applicable by incorporating determinants of the dividend, growth rate, and discount rate. Assuming that at any point in time, r is expected to remain constant and b is the proportion of earnings retained by the company, we arrive at $D_t = (1-b)rA_t$. Solving for the value of the stock at time t, $P_t = (1-b)rA_t/(d-g)$. Thus, we now have a measure of stock price at a point in time based on the values which determine dividend levels.

The rate of discount is generally considered to be a function of the riskless rate of return plus a loading for risk. The risk component of the discount rate can be expressed as s Var(g) where s represents risk-aversion preferences of investors and Var(g) the variance in the growth rate. Thus, the discount rate d can be expressed as a+s Var(g) where a is the riskless rate of return.

To understand the implication of this relationship for discount rate in the problem of determining share price, it is necessary to develop the idea of a certainty equivalent function. For rate of return, a certainty equivalent function is a function that indicates combinations of risk and return to which an investor is indifferent. A commonly accepted behavioral proposition for this function is that a certainty equivalent of a particular expected return, E(r), is a function of that expected value and the variance in return from that value. In particular, $CE(r) = E(r) - s\sigma^2 r$ where s is the index of risk aversion introduced above. Now we have the price of a share of stock for a firm evaluated as: $P_o = \frac{Do}{d-g}$. We have seen that d, the yield required by investors, can be defined as a+s Var(g). Thus,

$$P_o = \frac{Do}{[(a+s\ Var(g))-g]} = \frac{Do}{(a-[g-s\ Var(g)])} .$$

The denominator on the right hand side, then is equal to the riskless rate of return less the certainty equivalent growth rate. The rate at which investors capitalize their income streams, then, is the riskless rate of return less the certainty equivalent growth rate of dividends.

A complete stock valuation model containing all variables the impact of which we want to assess can now be formulated. Defining growth rate as rate of return times proportion of earnings retained (rb), we get

$$P_o = \frac{(1-b)rA_o}{a+s\ Var(rb)-rb} .$$

This equation can be manipulated to determine a number of relationships which will prove to be important in determining criteria for rate of return.

One example would be to consider discount rate as a constant and solve the equation for r:

$$r = \frac{dP}{b(P-A)+A} .$$

From this equation we can determine the effect of changes in rate of return (set by the regulatory agency) and retention rate (b) on stock price. This equation can also be used to determine the retention rate (or its complement the dividend payout rate) which will maximize share price.

Solving for P:

$$P = \frac{(1-b)rA}{d-rb}\text{, and taking the first derivative, } \frac{\partial P}{\partial b} = 0,$$

we find r = d. But at r = d, P = A, i.e. market value of the stock equals book value. While this is neither a maximum nor minimum, it has been used as a standard by regulatory agencies, i.e. the appropriate rate of return is that which keeps the market value of the firm's stock equal to its book value.

This, then, is an introduction to some of the variables which influence share price and which must be considered in evaluating alternative approaches to rate of return. One additional subject to be considered before considering specific approaches used in regulated industries for rate of return in portfolio theory. The basic tool we want to derive from portfolio theory is a way to measure needed rate of return for one investment versus another.

The portfolio problem in its simplest sense is to combine a group of securities so as to maximize return for a given level of risk. The portfolio analyst determines weights to apply to various securities given their individual risks and returns. Risk is measured by variance of return which basically amounts to variance of expected dividend growth rate. The variance of a portfolio depends on the covariance between pairs of securities.

Portfolio theory characterizes the probability distribution of a portfolio's rate of return by its mean and variance. The indifference curves for various combinations of risk and return can be drawn for various investors. These indifference curves can be used to determine the portfolio appropriate to a particular decision maker.

The rate of return on a portfolio is a weighted average of the rates of return of its component securities. Relationships among rates of return of various securities may be given in terms of correlation coefficients and covariances. Standard deviation of return is the measure of risk normally applied to a portfolio.

The element of portfolio theory with which we are most concerned is the evaluation of expected return (required yield) for a specific security. The formula most commonly used is that of the Capital Asset Pricing Model associated with Sharpe[9], Lintner[10], and Mossin[11]. In that model,

$$E(Ri) = Rf + \frac{E(Rm)-Rf}{6(Rm)} \cdot \frac{COV(Ri,Rm)}{6(Rm)}, \text{ where}$$

E(Ri) is the rate of return on the security in question, Rf is the risk free rate of return, and E(Rm) is the expected rate of return on the market portfolio. In other words, the reward for bearing risk on a security, [E(Ri)-Rf], equals a constant times the covariance between the security's rate of return and that of the market as a whole. Thus, the required yield or cost of capital used as a benchmark in rate of return regulation can be specifically derived or hypothesized from market variables.

Utility Rate of Return Regulation

As was indicated earlier, it is in the area of utility regulation where most of the existing theoretical rate of return work has been done. Public utilities have been subject to regulation for many years. As a result, there exists a body of knowledge on the general subject of rate of return regulation. The primary reasons utilities have been singled out for such regulation include the essential nature of the services they provide and the fact that the production of their product leads to economies of scale and a natural monopoly market.

Economic principles have evolved as the basis for rate determination. The basic principle is that rates should be adequate to produce a return adequate to allow for capital attraction and control of demand but that they should be such as to provide incentives for managerial efficiency. By meeting these criteria they assure the economic work to be done is done in an optimal manner.

A reasonable rate is one which resolves conflicts between investors to whom the rate is a source of income and consumers to whom it is an expense. To an economist an optimal rate is one that covers the marginal cost of the last unit produced. The Hope[12] case defined the two main functions of utility prices to be capital attraction and use rationing.

The idea is to assure adequate kinds and amounts of services are produced economically. The cost of service standard for a rate performs the consumer rationing and capital attraction functions, but not management incentive. Also economic costs seldom equal social costs which are harder to determine. In addition, the issue of actual versus reproduction costs must be resolved.

Using actual costs, rate levels should be sufficient to cover actual operating costs plus a fair return on the depreciated cost of plan and equipment. Under a reproduction cost standard, rates should cover actual operating costs plus a fair return on depreciated cost. Under either method determination of the appropriate time period over which to equate rates with costs is a problem.

An alternative to a cost of service standard is a value of service one. Unfortunately, no way of measuring value of service to the population or its members has been devised. Approximations thereto such as the price allowing maximum profit or that which clears the market are fairly crude.

Attempting to duplicate results which would arise in a perfectly competitive market also causes problems. In a dynamic setting supply-demand interaction may cause prices to be above or below normal production costs. In the short run capacity may be inadequate and excess profits can arise within a competitive price. Only in long run equilibrium does price equal marginal cost which equals the lowest possible average cost.

The basic standard used by regulators in determining an appropriate and fair rate of return is cost of capital. This standard allows for capital attraction and credit maintenance. The formula normally used separates capital into long term debt, preferred stock, and common stock. It can be based on an actual or preferred capital structure.

The primary issue in evaluating cost of debt capital is whether to use current capital market rates or charges on existing securities. Cost of equity capital is more difficult to determine because dividends depend on earnings which depend on allowed rate of return which is the subject of a rate case. The current cost test looks at rate of return on new common stock capital.

Combining a current cost measure on equity capital with an actual cost measure on debt capital is often done. While this would appear to be inconsistent, many economists argue that it provides the most meaningful measure. One method used in determining current cost of equity capital is to look at actual earnings-price ratios and assume they provide an estimate of required yield. The major alternative to a cost of capital standard is a comparable earnings standard. The latter pegs allowed rate of return at a level equivalent to that of investments involving similar risks.

One of the more controversial issues in utility rate of return regulation is that of the effect of leverage on earnings and required yield. An article written in 1958[13] began a debate which still rages 20 year later. Arguing that an investor is indifferent between leverage in their personal and corporate accounts, Modigliani and Miller went on to show that the yield required on a share of stock of a levered corporation would be: d+(d-a) (L/E). Share price would rise as a result of the leverage if required yield remained the same. This formulation of the model, however, sees required yield increasing to offset the increased earnings brought by the leverage. The end result is to determine that the cost of debt capital is equivalent to that of equity capital, i.e. the yield at which the stock would sell in the absence of leverage.

The behavioral implications of this model are quite significant. If correct, it means common stock should always sell at book value and the nature of a firm's capital structure is irrelevant to its stock value. These results are somewhat at odds with what most analysts feel is the true nature of capital markets, and many writers have attempted to detect where the M-M analysis breaks down.

One of the more promising alternatives is to add constraints imposed by financial and product markets. Expansion in the product market can lead to price reductions which would negatively affect rate of return. Expanding in financial markets, i.e., increased borrowing, can lead to a higher interest rate. These constraints can affect assumptions about cost of capital from various types of financing.

For example, share price is assumed to be a function of four variables: 1) rate of return; 2) retention rate; 3) risk free interest rate; and 4) debt/equity ratio. If we assume constraints in the product and financial markets, however, management loses two degrees of freedom. It can set the debt/equity ratio, but then the financial markets will determine the risk free interest rate. Similarly, once rate of return is set, the product market will determine retention rate which will leave share price unchanged.

Before becoming too immersed in some of the specific controversies of utility rate of return regulation, it may be helpful to review the general perspective within which these issues percolate. The regulatory objective is to determine the price providing the lowest rate of return (R of R) on capital consistent with the investment level required by the public interest. Because the objective of management is assumed to be maximization of share price, allowed R of R should be set such that the desired investment level at that rate of return will maximize share price.

Gordon[14] discusses the issue of determining cost of capital, i.e. appropriate rate of return, from the perspective of traditional theory. One model which has evolved for evaluating cost of capital is the perfectly competitive capital markets (PCCM) model. According to this approach, required yield is independent of dividend and stock financing rates, but increases with leverage. Using the Modigliani-Miller assumption regarding effect of leverage on required yield, this theory finds the cost of capital is the leverage free cost regardless of the amount of debt in the financial structure. When allowed R of R is set at the leverage free required yield, stock price is independent of investment, and desired investment levels are reached.

The traditional cost of capital theory differs from PCCM only in assuming required yield does not rise as rapidly with leverage. As a result, cost of debt capital is less than cost of equity capital and total cost of capital is a decreasing function of leverage rate. The practical implication of this, obviously, is that it leads to a lower allowable R of R than PCCM to the extent there is debt in the capital structure.

Both theories assume that cost of capital is the appropriate rate of return to be allowed, i.e., the rate that leaves stock prices unchanged. One result of these approaches is to keep the market price of the stock of a regulated firm at book value. Thus, by these theories a regulator could use the relationship of market and book value of a firm's stock to determine if allowed rate of return is appropriate.

Another way of considering the difference between the PCCM and traditional theories is to look at formulations for cost of capital under each. According to PCCM, cost of capital is: (Pk+Li)/(P+L), where P and L are market values of equity and debt, i is current bond yield, and k is required yield on equity. According to the traditional theory, cost of capital is: (Ek+Bc)/(E+B), where E and B are book values of equity and debt, and c is coupon rate of the company's bonds. Thus, the PCCM theory leads to a current level of return, the traditional theory a historical level.

In actuality, the setting of rates becomes a process of compromise. There is no scientifically correct rate of return, but only a zone of reasonableness within which the correct rate likely lies. The bottom limit on the zone is the level needed to attract capital. The upper limit is the estimate of what capital invested in enterprises of similar risk is earning.

From an economist's point of view, the correct rate of return is that which covers the marginal cost of capital. Unfortunately, such cost is impossible to measure. Also, it is a static measure and allowed rate of return must fulfill a dynamic institutional function. It must provide the incentives needed to bring forth optimal production levels most efficiently. Ironically, the very existence of a regulatory constraint may provide disincentives to cost minimizing behavior.[15]

It should be noted that even if a measure of fair rate of return can be developed, the issue of rate base to which the rate should apply remains open. The two most commonly accepted rate bases are market and book value of assets. Methods of computing either basis are quite complex with one of the major problems being determination of appropriate depreciation charges. Once the rate base is calculated, the problem of applying a consistent rate of return remains.

Thus we see that while an enormous amount of work has been done in the area of utility rate of return regulation, there are almost as many unresolved issues as there are in areas which have not had the benefit of such intense study. Debate over effect and relevance of leverage rate, dividend policies, risk, rate base, and so on goes forward with no indication that any generally accepted approaches are in the offering. The implications of this for those of us in an industry considering a similar plunge into rate of return regulation should not be overlooked.

Conclusion

We have seen that economic principles provide the key to any evaluation of appropriate profit level in an industry. The use of economic and financial analysis tools by regulators of utilities have been examined. Studies conducted on the specific subject of insurer profitability have been reviewed. What useful conclusions about approaches to the issue of profitability can be gleaned from all that?

For one thing, it is clear that determining criteria for proper profit levels and measurement thereof is an enormously complex undertaking. Even when accomplished, the end result is a very uncertain and debatable figure. In the insurance industry there are additional complications in deciding upon proper rate of return.

Insurance ratemaking is done on a prospective basis. Historical financial data provides only a crude measure of appropriate rate level. Even if an agreed upon measure of profit could be established and its level calculated, appropriate rate of return at any point in time for a particular type of coverage in a particular jurisdiction would generally be quite uncertain. This can be seen in many of the current disputes over needed rate level in which both parties use the same profit load.

Insurance ratemaking in inflationary times for coverages in which losses do not settle quickly is quite sensitive to assumptions about the way loss costs will move in the future. A typical rate increase may be almost entirely based on such assumptions. When potential error in trend and development factors can amount to as much as ten to twenty percent, a difference of one or two percent in the profit load can be virtually ignored.

Concentrating on overall profit load can lead to neglect of another very important issue--the way rate level for a coverage is spread to various types of policy holders. The classification issue is emerging as one of paramount importance, and evidence has been developed sharing that enormous inequities may exist. Spending enormous amounts of time fine turning overall rate levels when individual policy holders may be overcharged by 100% or more might seem to be straining on a gnat while swallowing a camel.

The National Association of Insurance Commissioners has taken the very responsible approach of developing several alternative measures of profitability for interested parties to evaluate. The existence of these measures provides everyone with useful information without tying anyone's hands as to particular approaches. Until a full review of alternative regulatory activity is completed, this may be the ideal way to handle the profitability issue.

FOOTNOTES

[1]Knight, Frank H., Risk, Uncertainty and Profit, Houghton Mifflin Co., 1921.

[2]Clark, John B., "Insurance and Business Profit," Quarterly Journal of Economics, Vol. VII, p. 46.

[3]Hawley, "The Risk Theory of Profit," Quarterly Journal of Economics, Vol. VII, p. 472.

[4]Willett, The Economic Theory of Risk and Insurance, New York City, 1901.

[5]Knight, op. cit.

[6]Proceedings, National Association of Insurance Commissioners, Vol. IIA, 1970, p. 719.

[7]Bailey, Robert A., "A Review of the Little Report of Rate of Return in the Property-Liability Insurance Industry," PCAS, 1969, p. 133.

[8]Ferrari, J.F., "The Relationship of Underwriting, Investments, Leverage, and Exposure to Total Return on Owners' Equity," PCAS, 1968, p. 295.

[9]Sharpe, William F., "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," Journal of Finance, September, 1964, p. 425.

[10]Lintner, John, "The Valuation of Risk Assets and the Selection of Risky Assets in Stock Portfolios and Capital Budgets," Review of Economics and Statistics, February, 1965, p. 13.

[11]Mossin, Jan, "Equilibrium in a Capital Asset Market," Econometrica, October, 1966, p. 768.

[12]Federal Power Commission et al. v Hope Natural Gas Company, 320 U.S. 591 (1949) at 603.

[13]Modigliani, Franco and Morton Miller, "The Cost of Capital, Corporation Finance, and the Theory of Investment," American Economic Review, 1958, p. 261.

[14]Gordon, Myron J., The Cost of Capital to a Public Utility, Michigan State University Public Utilities Studies, 1974.

[15]Averch, Harvey and Leland L. Johnson, "Behavior of the Firm under Regulatory Constraint," American Economic Review, 1962, p. 1053.