

ECONOMIES OF SCALE IN  
THE U.S. LIFE INSURANCE INDUSTRY:  
AN ECONOMETRIC ANALYSIS

by

THOMAS V. DAULA

U.S. Military Academy, B.S., 1974

SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE  
OF MASTER OF SCIENCE

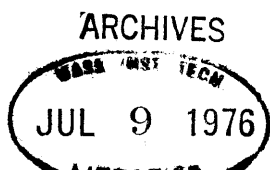
at the  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1976

Signature of Author *Thomas V. Daula* .....  
Department of Economics, May 1, 1976

Certified by *[Signature]* .....  
Thesis Supervisor

Accepted by .....  
Chairman, Departmental Committee on Graduate Students



Abstract

This thesis provides new evidence on the existence of economies of scale in the life insurance industry. New insights in this area are obtained by employing a composite output measure in our estimation of long-run average cost curves for this industry. The use of such a measure enables us to avoid pitfalls which produced inconsistent estimates in prior studies. In addition to introducing consistent estimation techniques, we also utilize more efficient estimation procedures than previous analyses. Our findings indicate that the introduction of these improved procedures does not alter the conclusion of past studies that statistically and economically significant economies of scale are present in the U.S. life insurance industry. Finally, we note in our concluding remarks avenues of future research in this area.

### Acknowledgement

I would like to express my appreciation to Professor J. Hausman for his advice and helpful comments on earlier drafts, and for teaching me whatever skills in econometrics which I may now possess. I am also grateful to Professor R. Cohn for suggesting the topic and for aiding me in my initial research, and to R. Geehan for allowing me to use his composite measure of life insurance output. In addition, I would like to thank all of my fellow students, and in particular Andy Abel, Louis Beleza, Jorge Desormeaux, Yves Balcer, and Dave Smith, for the help and encouragement that they provided me over the last two years. A final note of appreciation belongs to Vicki Elms who typed this final draft.

Table of Contents

	page
Section I: Introduction .....	1
Section II: Economies of Scale, Cost Function Theory, and General Estimation Problems .....	3
Section III: Survey of the Literature .....	18
Section IV: The Model .....	32
Section V: The Data .....	42
Section VI: Results .....	51
Section VII: An Analysis of Market Structure .....	66
Section VIII: Conclusion and Recommendations for Further Research .....	74
Bibliography .....	76

List of Tables

	page
Table 1: Houston and Simon: Regressions .....	23
Table 2: Average Cost by Size of Firm .....	26
Table 3: Unit Costs (Weights) of Life Insurance Activities .....	45
Table 4: Regressions: Houston & Simon's Model .....	53
Table 5: Regressions: $AC = a + bQ + cOWN + dMS + fA + e..$	55
Table 6: Regressions: $AC = a + bQ + cQ^2 + dMS + fA + e...$	56
Table 7: Regressions: $AC = a + b\log(Q) + cOWN + dMS +$ $fA + e -.....$	57
Table 8: Concentration Ratio of Ten Largest Firms .....	72
Table 9: Concentration Ratio of "Medium" Size Firms .....	72
Table 10: Company Statistics .....	73

List of Figures

Figure 1: Linear Regression Results .....	61
Figure 2: Quadratic Regression Results .....	62
Figure 3: Logarithmic Regression Results .....	63

## I. Introduction

Information concerning economies of scale has important influence on the analysis of the structure and functioning of an industry. This information is valuable not only to managers of a given industry, but also, in the case of a regulated industry such as life insurance, to the administrators of various regulatory commissions who depend on such information in evaluating the consequences of their actions. The range and the depth of this issue has been astutely summarized by Joskow in the following passage:

Substantial cost advantages for very large scale operations could indicate substantial barriers to entry and the possibility that large firms could set prices substantially above marginal cost without providing competitive entry. At the same time the presence of a large number of small, high cost fringe firms may indicate that large companies ...have succeeded in keeping prices above the competitive level, thus protecting inefficient producers.<sup>1</sup>

In this paper, we will evaluate the case for the existence of economies of scale in the U.S. life insurance industry. Although various previous authors have consistently discovered such economies in their analyses, their results are open to question on both procedural and theoretical grounds. By introducing a composite output measure, and by utilizing superior estimation techniques, this study is able to shed new insights on the robustness of past results. In so doing, we hope to bring the current body of knowledge a step closer to providing a definitive answer to this problematic question.

I Introduction

1. Joskow, p. 384.

## II. Economies of Scale, Cost Function Theory, and General Estimation Problems

In this study, we will attempt to determine if there exists valid empirical evidence that economies of scale are present in the U.S. life insurance industry. Returns to scale is a technological concept whereby such economies are said to be present when a proportional increase in all factors of production produces an even greater proportional increase in output. That is, if we specify the production function as,  $Q = f(X_1, X_2, \dots, X_m)$ , and expand all of our inputs by some constant factor  $\lambda$  so that  $\alpha Q = f(\lambda X_1, \lambda X_2, \dots, \lambda X_m)$ ; then decreasing, constant, or increasing returns to scale are said to be present as  $\alpha \begin{matrix} < \\ = \\ > \end{matrix} \lambda$ .

Although returns to scale is a technological concept which is defined in terms of the production function, we will employ a cost function approach in our attempt to determine if such scale economies are present in the U.S. life insurance industry. The decision to utilize a cost function methodology is based on the difficulties encountered in trying to adapt production theory to an industry such as insurance. The insurance industry is a service industry which produces a multitude of products. Because of this fact, it is extremely difficult to merely identify the output, let alone specify a functional relationship which would describe how the various outputs are produced from the many factors employed by an insurance firm.

While the use of a cost function greatly simplifies the nature of the problem, it may result in the clouding of the interpretation of the results. In particular, we must be careful not to mistake pecuniary



returns to cost, which stem from induced factor price changes, for the economies of scale which we seek to measure. While such factor price differentials are not likely to be a significant source of cost savings for large insurance firms because of the lack of specificity of the majority of the factors employed by these firms, we must recognize the possibility that such an occurrence will influence the shape of the average cost function, and thereby distort our results. A second possible source of distortion is the change of mistaking returns to substitution, in which factor proportions change as output varies, for scale economies. The occurrence of either or both of these factors will cause a downward slope to the average cost curve, which, as we will see below, is also indicative of scale economies. For this reason, Kamerschen and Pascusci (1969) recommend the use of the term cost economies for empirical studies which utilize cost functions.<sup>1</sup> In order to avoid confusion, however, this report, while recognizing the above caveats, will continue to employ the term economies of scale in its analysis.

To better understand the theory underlying our use of cost functions to determine the existence of scale economies, we will first review a rigorous statement of the relation between the occurrence of returns to scale and the shape of the average cost curve recently made by Igmarr Sandmo (1970). Let  $X$  denote the quantity of output, and let  $v_i$  represent the amount of factor  $i$  ( $i = 1, \dots, m$ ) utilized in the production of  $X$ . Thus we have that

$$X = f(v_1, \dots, v_m). \quad (1)$$

Consider a proportional increase in each factor by 100k%, so that

$$k^{\epsilon t} X = f(kV_1, \dots, kV_m) \quad (2)$$

where  $t$  is the scale elasticity of output. Differentiation with respect to  $k$  gives

$$\epsilon k^{\epsilon-1} X = \sum f_i V_i. \quad (3)$$

Now  $k^{\epsilon-1} X = f/k$ , implies that

$$\epsilon = k \sum f_i v_i / f \quad (4)$$

Without loss of generality let us set  $k = 1$ .

Total costs are given by

$$C = \sum w_i v_i \quad (5)$$

where the  $w_i$  are the factor prices. If we assume profit maximization, which in turn implies cost minimization for each level of output, we have that each firm maximizes (5) subject to (1). Forming the corresponding Lagrangian and differentiating produces the following first order conditions for cost minimization

$$w_i - f_i \lambda = 0 \quad (i = 1, \dots, m) \quad (6)$$

Equations 1 and 6 provide us with  $m+1$  equations to determine  $v_i$  and  $\lambda$  as functions of  $w_i$  and  $X$ . If we take factor prices as given, this allows us to write cost as of function of output, i.e.

$$C = C(X).$$

Differentiating equations 5 and 1, we get that,

$$dC = \sum w_i dv_i$$

$$dX = \sum f_i dv_i$$

or upon substitution from (6) that

$$dC = \lambda \sum f_i dv_i = \lambda dX$$

Thus, we get that the multiplier gives us the marginal cost at X, i.e.

$$\frac{dC}{dX} = \lambda \tag{7}$$

Now define average cost to be denoted  $b(X) = C(X)/X$ . Differentiating with respect to X gives

$$\frac{db}{dX} = \frac{1}{X^2} \left[ \frac{dC}{dX} X - C \right]$$

which implies that

$$\frac{db}{dX} \underset{|| \vee \wedge}{=} 0 \quad \text{as} \quad \frac{dC}{dX} \underset{|| \vee \wedge}{=} \frac{C(X)}{X}$$

Substituting from (7), (6) and (1), we get

$$\frac{db}{dX} \underset{|| \vee \wedge}{=} 0 \quad \text{as} \quad \lambda \underset{|| \vee \wedge}{=} \frac{\sum f_i v_i}{f} \quad \text{or as} \quad \epsilon \underset{|| \vee \wedge}{=} 1.$$

Therefore, we have as our result that an average cost curve has a negative, positive or zero slope, as the returns to scale are increasing, decreasing or constant.<sup>2</sup>

It should be noted that the above result does not hold when the firm exercises some degree of monopsonistic power in factor markets, since there is then no direct connection between costs and<sup>3</sup> the purely technological concept of returns to scale;

but that it does remain true "irrespective of whether the firm is a price taker or not in the market for its output."<sup>4</sup> We should also remember that in obtaining our result we had to assume cost minimization, and that the firm produced a single homogeneous output.

Three general reasons are usually listed for the possible existence of economies of scale for low ranges of output in any given industry. These are: first, the greater ease of dealing with large quantities; second, the spreading of risks and the resultant reduction in the costs of uncertainty; and third, the existence of indivisibility in both men and capital equipment.<sup>5</sup> Because of the personal nature of the insurance product, it is doubtful that the first rationale is very significant for the insurance industry. While the second factor is certainly present at the very small end of the output range, the fact that risk spreading occurs quite rapidly for the insurance product itself, and that investment risks may be reduced efficiently through the use of financial intermediaries such as mutual funds indicates that this factor's influence on average cost would become insignificant above the very short range of the output spectrum. The third reason for the existence of size economies is the one most often mentioned in the literature, in the sense that expanded size would better enable the use of cost-reducing electronic data processing techniques. As the electronics industry progresses, however, these same techniques are becoming more

accessible to smaller firms. A fourth reason for cost economies, which has been given for the life insurance industry in particular, is that larger firms may be able to attract better management through the payment of higher commissions or salaries, thereby attaining greater operating efficiency.<sup>6</sup> Because no unique management expertise is required by life insurance firms, these firms may draw upon a large pool of trained administrators. For this reason, we doubt that this final source of size economies would prove to be significant for the life insurance industry. Thus, we see that there is no strong a priori basis for the existence of substantial size economies in insurance firms.

In conducting empirical studies in which the estimation of cost functions is required, certain obstacles to that estimation must be overcome, or at least accounted for. For the insurance industry, one such obstacle is the definition of its output. As we shall see below, a method which has been employed in the past is to use net premiums as a proxy variable for output. Employing net premiums as a measure of output, which is the same as measuring output as net sales, is appropriate, however, only if the product is homogeneous, and sold at the same prices by all companies. As we will report in the next section, available evidence indicates that both of the above preconditions fail for the life insurance product. Realizing that net premiums is a poor measure of output, Allen in a recent paper considered three methods of obtaining an appropriate measure of output for this industry. These methods are: first, to construct a composite output measure; second, to include proportions of each insurance line in the estimating

equation as explanatory variables along with net premiums; and third, to select a sample for study in a manner that provides for premium comparability.<sup>7</sup> While Allen chose the latter method as the best procedure available to him, we will employ a composite figure based on unit costs of the various lines of output as our measure of the insurance product in this study.<sup>8</sup>

A second obstacle to empirical cost studies is the problem of defining costs. The usual practice in these studies, and the one which we will also employ here, is to use an accounting definition of cost. Accounting data, however, contains several flaws as a measure of cost in the economic sense. Often mentioned in such criticisms of accounting cost is that the distribution of depreciation of an asset over its life cycle is usually determined by the taxation authorities rather than by economic criteria, and that the valuation of capital services occurs on the basis of historical price rather than at replacement cost.<sup>9</sup> Neither of these deficiencies, however, are of major importance to insurance data since only a small portion of total costs are determined by capital usage. Another problem with accounting cost, which does remain important for insurance firms, because of the multi-product nature of their output, is the often arbitrary allocation of overhead costs by this technique. Finally, in a cross-section study, accounting practices may vary from firm to firm, which may affect reported costs.

The appropriate definition of cost also includes a return to entrepreneurial initiative or risk. Accounting profit, however, also includes any monopoly profits which may be present as well as

legitimate entrepreneurial return or costs. As a result, following common practice, accounting profit is not considered a cost in this study. Such an omission of all accounting profits from cost, probably causes an underestimation of cost. Despite this shortcoming, as Walters points out, "one does not know whether this bias invalidates the shape of empirical cost curves."<sup>10</sup> While this potential source of distortion must be recognized, a determination of the most efficient method of correcting for this shortcoming in studies based on accounting data is beyond the scope of this paper.

In empirical studies utilizing cost functions, a single equation approach may not be employed unless it is assumed that output is exogenously determined with respect to individual firms, and firms minimize cost. As Benston points out, these necessary assumptions appear to be valid for regulated financial institutions.<sup>11</sup> For life insurance firms, those costs, such as advertising, which would affect the demand function facing a firm, represent a very small fraction of the total expenses incurred by any firm during the course of a year. Furthermore, insurance is the type of product whose demand is determined primarily by such factors as the population and income of the area in which it is marketed. For these reasons, it seems appropriate to assume that the insurance industry, and each firm therein, faces a stable demand curve on the basis of which its output is exogenously determined.

In our study, the use of an aggregate output measure based on unit costs may introduce a spurious dependence between cost and output.<sup>12</sup> If such a dependence is introduced in this fashion, our

aggregate output measure will lose its exogenous qualities. As a result, even if we may assume that the "output" of an insurance firm may be considered an exogenous variable, the output measure we employ in our study may not possess this necessary attribute. Because of this fact, we will employ an instrumental variable technique in our estimation procedure to correct for any inconsistencies injected into our analysis by the non-exogeneity of our output measure.

In conducting an empirical test for economies of scale, one may employ either cross-sectional or time series data. The use of time series data, however, may introduce several sources of bias into the results. For example, prices at which costs and output are evaluated may change over time. Alternatively, the technological process employed by a firm may also change over time. Since we have predominantly experienced inflationary price movements in recent years, the impacts of the former source of distortion has probably been to impart an upward bias on long-run average cost curves based on time series data. Technological progress, on the other hand, would have a negative effect on the long-run average cost curve. Although the effect of price movements over time may be corrected for to a large extent, the net effect of the remaining price and technology changes would result in an ambiguous bias in our empirical estimate of a long-run average cost curve.<sup>13</sup>

In this paper, we will conduct our empirical analysis on the basis of cross-section data. While the use of cross-section data avoids the problems associated with changing conditions over time which



distort time series studies, it presents our analysis with a different set of problems. In attempting to determine the existence of size economies on the basis of a long-run average cost curve, we must assume that all the firms in our cross-section are using the same technology or production function to produce their output. This assumption appears to be plausible for a service industry such as insurance. Furthermore, minor differences between firms, which may cause variations about the cost-scale line, are unlikely to impart a systematic bias to the shape of an average cost curve produced by cross-sectional analysis.<sup>14</sup>

The use of cross-section data has been objected to by Friedman on the basis of two observations by him. The first is that in a competitive framework, with no specialized factors of production, the average cost curve would be the same for all the firms in the cross-section, and each firm would be producing the same level of output. Friedman went on to argue that in such a context, differences in output between firms would only be attributable either to mistakes or the existence of specialized factors of production unique to individual firms. If the capital market and accounting system were operating efficiently, the result of this situation would be firms operating at different output levels but displaying constant average cost. Such an occurrence could mistakenly be taken for evidence of constant returns to scale.<sup>15</sup>

Johnston has correctly pointed out that this criticism by Friedman of cross-section studies rests upon the assumption of a perfectly competitive industry.<sup>16</sup> When reality diverges from perfect competition, this flaw no longer applies. Such divergence, in fact, seems to be the case in the U.S. life insurance industry. In a former study of the life

insurance industry, Belth discovered a great deal of price variance in this industry, and "concluded that this degree of price variance is not consistent with a perfectly competitive market."<sup>17</sup>

The most important criticism levelled by Friedman against cross-section methods is the so-called "regression fallacy". According to Friedman's view, a firm's output at any given moment in time is likely to be the combination of an expected or normal component of output, and a transient or random component. Furthermore, firms with output

above the industry mean at a particular point in time are more likely to be experiencing a positive random component, than firms with output below the industry mean. If average variable cost were constant for these firms within their individual range of variation due to the transient components of their output, then lowered unit costs would be expected for firms with positive transient components. As a result, a negative bias could be imparted to the slope of the long-run average cost curve.

As Johnston has pointed out, in making his original argument, Friedman ignored the role of variable cost.<sup>18</sup> If instead of the case of constant variable outlined above, the variable costs rose sharply as a result of the random components of output, then a positive bias to the slope of the long-run average cost curve would result. Furthermore, it has been argued that if output is truly subject to random fluctuations, then the expected cost curve, and not one influenced by random fluctuations in output, would be the relevant cost curve for decision-making purposes. Since accounting data usually encompasses several economic periods, a cross-sectional study utilizing this data would best approximate this underlying expected cost curve.<sup>19</sup>

As a final point with respect to the role that Friedman's regression fallacy may play in influencing the results of this paper, it should be noted that our sample contains firms which differ in scale by a factor of 5,000 on the basis of our aggregate output measure. The presence of such a variance in firm size would seem to rule out the possibility of making any meaningful statement on the sign of the random component from the observation of the relation of a firm's size to average firm size in the sample.

Therefore, in view of the arguments made in the preceding paragraph and the wide latitude in firm size in our sample, it is unlikely that the regression fallacy is a serious source of distortion in our results.

Theory II

1. Kamerschen, p. 494.
2. A different proof of this proposition is also given in McElroy (1970).
3. Sandmo, p. 152.
4. Ibid.
5. Walter, p. 40.
6. Hammond, etc., p. 182.
7. Allen, p. 101.
8. Allen, in making his choice, stated his belief that at this time net premiums are the only output measure available for research on this topic.
9. Walters, p. 43.
10. Ibid, p. 42.

11. Benston, p. 317.
12. Johnston (1958), p. 346.
13. Randy's Proposal.
14. Johnston (1958), p. 348.
15. Walters, p. 44.
16. Johnston (1958), p. 347.
17. Launie, P. 284.
18. Johnston (1958), p. 350.
19. Walters, pp. 48-49.

### III. Survey of the Literature

Although there have been many studies concerning the existence of economies of scale in the general category of financial institutions, there have been only two serious attempts to determine if such economies exist in the life insurance industry. These studies were conducted by Houston and Simon in 1971, and by S. Travis Pritchett in 1973. Before reviewing the results of these investigations, however, we will first review some of the work done on the more general topic concerning size economies in other forms of financial institutions. The authors of these works faced problems in characterizing their models and testable hypotheses which are endemic to investigations concerning the structure of the financial industry, and to multi-product, service industries in general. Thus, a look at the results of these studies should be enlightening to the reader from a methodological viewpoint, and in addition provide him with a feel for what result he may expect from the study on the basis of those achieved with respect to similar institutions.

George J. Benston produced a prototype paper on economies of scale in savings and loan associations.<sup>1</sup> Because of the problems associated with defining output in a multi-product, service producing firm, he chose to employ a cost function approach. In particular, he advocated the use of the following general specification of the cost function in studies of this type:

$$c = f(Q, G, P, U)$$

where:

c = operating costs per period

Q = rate of output per period

G = output homogeneity variables that account for the fact that Q is not a homogeneous measure

P = differences in factor prices, organizational structure, and management ability of firms

U = other unspecified factors.

In choosing this form, he is explicitly taking account of the fact that one is often forced to employ a proxy variable for the measure of output. The variable chosen is often related to sales or a general measure of services performed. Such a proxy will yield statistically consistent results, however, only if the product is homogeneous in nature, and sold at the same price across firms in the industry. While the latter precondition is assumed, Benston recommends correcting for the former through the use of "homogeneity variables". Benston also notes that the use of this single-equation, cost function approach requires that we assume that output is exogenously determined and that firms minimize cost.<sup>2</sup>

Benston decided that costs are incurred as a result of services performed, and so measured output as the number of business transactions. Using this measure of output, and a variation of the above general form of the cost function, he found that savings and loan associations exhibited consistent and significant economies of scale for all years studied.



There have been several studies conducted to determine if economies of scale exist in property and liability insurance companies. We will consider the results of three recent papers on this topic. The first study which we will consider was conducted by Hammond, Melander and Shilling in 1971. In that paper, they examined the relationship between expense ratios and size based on net premiums. From the results of their analysis, they concluded that economies of scale do exist for major non-loss costs of property and liability insurance companies, and that the majority of insurance companies are of sizes below which the economies disappear. They found no evidence of a U-shaped average cost curve, but decided instead that the appropriate form was L-shaped. From these results, they inferred a possible justification for the merger of smaller firms.

Two later studies on this same topic of size economies in property and liability insurance companies differed from Hammond, et al., in finding either no economies, or economies which disappeared after a relatively small rise in the level of output. In his 1974 paper, Robert Allen recognized that Hammond's use of net premiums as measure of output was inappropriate since the product is neither homogeneous, nor sold at the same price by all companies. In order to avoid this problem, Allen chose to pattern his sample in a manner which would provide inter-company comparability of output, and to use policyholders' surplus as the measure of company size.<sup>4</sup> With these alterations incorporated into his model, Allen found scale economies to be unimportant beyond relatively small output levels. In particular, whereas Hammond concluded that size

economies existed up to the range of \$300-600 million in annual premiums, Allen concluded that they extended no further than for companies writing from \$30-60 million in annual net premiums.<sup>5</sup>

An even stronger indictment than Allen's of Hammond's results can in fact be found in the form of an earlier article by Joskow, written in 1973. In that article, he looked at the relation between expense ratios, and size as measured by premium volume. He also corrected through the use of dummy variables for the methods employed by the companies in writing the insurance (i.e., whether they employed direct agents, or sold through agencies). On the basis of his results, Joskow concluded that the production of property insurance is characterized by constant returns to scale. Thus, we see that while the evidence is mixed on this issue, some of the most recent studies produce the result that economies of scale are either absent or of little significance for property-liability insurance companies.

The first major attempt to statistically test for the existence of size economies in the life insurance industry was a study performed by Huston and Simon (1970). In their model, they considered three possible specifications for the long-run average cost function. In addition, they sought to control for differences in insurance "products" by including three variables which were defined as the ratios of industrial, group and new premiums to total premium, respectively. A variable defined as the fraction of the insurance lapsed during the year was included to allow for the extra costs associated with the unrecovered initial expenses of young policies which were lapsed. They also included dummy variables

which were intended to correct for the possible influence of corporate concentration (for this variable both a regular and an interaction dummy variable were employed, each of which took a value of one for the ten largest companies). The results of their tests are reproduced in Table 1. The reader should note that the market structure variable was omitted in their published report on the basis of its being statistically insignificant in their earlier results.

A review of Houston and Simon's results shows that all of the regression coefficients of the size variables were of the "right" sign for their hypothesis; that is, they all indicated decreasing average costs with increasing premiums. On the basis of these results, these authors decided that economies of scale, as defined, are present in the U.S. life insurance industry.

A critical analysis of their paper, however, soon uncovers certain deficiencies in their model which make their results something less than unequivocal. In particular, two fundamental assumptions which are required for their model to produce consistent estimates appear to be incorrect on the basis of available information regarding the life insurance industry. The first of these assumptions is that life insurance may be viewed as a homogeneous product. Although the authors seek to correct for this deficiency in their model through the use of three output ratio variables, this attempt contains several obvious inadequacies. The product of the life insurance industry is far too heterogeneous to enable the assumption of premiums as a proxy

Table  
(Source: Houston and Simon, 1970, p. 860)

Regressions

C	f(P)	B	B f(P)	IP/P	GP/P	NP/P	LR	R <sup>2</sup>	Sy
<u>f(P)=log(P)</u>									
1.25	-.0137 (.0016)		.018 (.007)	.174 (.083)	-.186 (.042)	.413 (.056)	.335 (.141)	.56	.154
<u>f(P)=1/P</u>									
.248	45,400 (6,300)				-.184 (.043)	.437 (.056)	.688 (.121)	.53	.159
<u>f(P)=1/log(P)</u>									
-.605	6.19 (.686)	.105 (.053)		.170 (.081)	-.182 (.042)	.398 (.055)	.335 (.137)	.57	.151

Legend

P= premiums

IP/P = Industrial Premium Ratio

GP/P = Group Premium Ratio

NP/P = New Premium Ratio

LR = Lapse Ratio

B = "Bigness" Dummy

for a homogeneous output to be acceptable after such a simple correction device. Their method fails to correct for the basic dichotomy in the field of ordinary insurance between term and ordinary policies where, in the former case, premiums represent payment for the buying of a pure insurance plan, while for the latter type of policy a major portion of the premium represents an investment option. Also neglected was any adjustment for the portion of the total premium income which came from annuity considerations.

The second major failing of the underpinnings of their model is the necessity to assume equal pricing by all firms in their sample. The inaccuracy of this assumption was demonstrated in a study conducted by Belth (1966). The results of this study concerning price differentials has led to the conclusion by many that the present market for individual life insurance is characterized by massive price ignorance.<sup>7</sup> This price ignorance is so prevalent in fact that Belth has stated in one article,

My reference to price ignorance in the market is intended to encompass more than price ignorance among life insurance buyers. It is intended to include price ignorance among life insurance agents... (and) to encompass price ignorance among life insurance companies.

Thus, we see that while the first of their assumptions raises doubts concerning the validity of their model and the results stemming from it, the second assumption's wide divergence from fact undoubtedly interjects inconsistencies into their regression results. The exact degree and nature of the biases emanating from these sources with respect to the regression coefficients obtained, however, cannot be ascertained without a detailed examination of the data which Houston and Simon employed in reaching their results.

In two different studies, Pritchett (1971, 1973) looked for evidence of size economies in the life insurance industry. In both studies, the author employed the very simple statistical technique of comparing mean expense ratios among four arbitrarily chosen size categories of firms. The expense ratio employed was defined to be the ratio of total actual general expenses and commissions to a standard expense determined on the basis of the volume of business as measured by the outputs of several products such as the number of new policies, the amount of first-year premiums, etc. In conducting his study, Pritchett chose his sample so as to provide for comparability of the output for the different companies. The basic results of his studies are summarized in Table II.

As we observe, Pritchett's results also imply the existence of size economies. A test of the equality of the four average expense ratios was rejected at the 5% confidence level.<sup>9</sup> On the basis of these results, Pritchett concludes that "the relationship between the size of expense ratios and the volume of ordinary business is essentially L-shaped with a negative slope continuing on the bottom segment".<sup>10</sup> In his second study, Pritchett did find evidence, however, that diseconomies of size were present in home office expenses for the largest insurers.<sup>11</sup> Thus, we see that we are again left with the result that economies of scale exist in the life insurance industry at least up through the range of medium to large scale insurance firms. This conclusion, however, is still not firmly emplaced due to the crude methods employed in conducting this study.

Table 2

Average Cost by Size of Firm

(Source: Pritchett, 1971, p. 560)

Size Category	Average Cost (Arithmetic Mean)	Standard Deviation
Small	1.1876	.1352
Medium	1.0835	.1412
Large	1.0354	.0789
Very large	1.0069	.0534

The final study which we will look at is one of the general insurance industry of Canada conducted by Halpern and Mathewson (1975). This study is most interesting from a methodological viewpoint. In particular, these authors note that since inputs (costs) and outputs are determined simultaneously by profit-maximizing firms, outputs are stochastic. Their use as explanatory variables in ordinary least-squares regressions by Benston, Houston and Simon and others gives rise to biased and inconsistent estimates of regression coefficients. Anything short of price or output directly fixed by regulators produces outputs with this property.<sup>12</sup>

In order to avoid such problems, Halpern and Mathewson develop a model from the first-order conditions of profit maximization. Their final equation avoids the use of cost information altogether, and thus, represents a reduced form of the cost-output generating system.

A general form of their model may be represented as

$$R \cdot N = \frac{1}{\gamma_R} - \frac{\gamma_C}{\gamma_R} Z^C - \frac{\partial_F}{\partial_R} Z^F + \epsilon$$

where  $R$  = premium revenue per policy  
 $N$  = number of policies  
 $Z^C$  = an exogenous cost variable  
 $Z^F$  = an exogenous financial cost variable.

Economies of scale are said to exist if  $\gamma_R < 0$ . Note that by avoiding the use of cost data, this model bypasses any problems associated with Friedman's regression fallacy, which we described above, as well as correcting for possible simultaneous equation bias. Note also, however, that in addition to the assumption of profit maximization,



the use of this model, as formulated and employed by Halpern and Mathewson, requires the assumption that the specification of the marginal revenue and marginal cost functions as linear functions is appropriate.<sup>13</sup>

The results of their regressions, which employed several expanded variations of their basic model, prompted two conclusions by Mathewson and Halpern. First, Canadian joint stock insurance firms realize diversification economies in the sense of lower marginal costs resulting from writing the same dollar amount of premium revenue over more lines. Second, Canadian mutual companies realize economies of size in the sense of reduced marginal costs through the joint expansion of automobile and fire/personal property insurance. In the latter instance, the economies for the mutual firms apparently stem from the use by these firms of the marketing technique of writing insurance directly as opposed to using the agency system.<sup>14</sup> Thus, we see that through the use of their technique, Halpern and Mathewson have reestablished some basis for the belief in the existence of some economies of scale in the property and liability insurance industry, after this view had been severely shaken by Joskow's and Allen's results.

While the technique employed by Halpern and Mathewson is an interesting method of approaching the problem of measuring size economies, it is doubtful that it represents a major improvement over previous methods. As we discussed above, it is not unreasonable to assume that the insurance output is very nearly exogenous from the individual firm's viewpoint. Furthermore, in the course of formulating the model, which they actually employed in estimation, Halpern and Mathewson were

forced to make several questionable assumptions in deriving their exogenous variables (such as, the linear separability of all cost factors of each insurance line, and that premiums by line represents a good proxy variable for output). They also employed certain demographic measures as exogenous variables which would be difficult for researchers on this topic to construct. Therefore, in view of the compromising assumptions which also are entailed under Halpern and Mathewson's model, and because of the greater relative difficulty in its formulation, we believe that the traditional cost function approach, in conjunction with an instrumental variable procedure, represents a superior method of correcting for possible simultaneous equation bias. This view is bolstered by the fact that both methods employ single equation techniques, and the use of instrumental variables under the cost function approach would also correct for possible errors in variables which remain uncorrected under Halpern and Mathewson's specification.

As we have seen, on the basis of past work in this field, the question of whether economies of scale exist in the insurance industry in general, and in the life insurance industry in particular, is still unresolved. While investigations of the life insurance have thus far produced consistent evidence that economies of scale do exist, as opposed to the contradictory results that the examinations of the property and liability insurance industry have brought forth, the past work with respect to the former industry has been at best rudimentary. For this reason, further studies of the life insurance industry, such as the one we are attempting here, seems to be called for at this time.

Survey III

1. Benston (1972).
2. Ibid, p. 317.
3. Allen, 1974.
4. The choice of policyholder surplus as his size measure stemmed from the fact that because of the method of state regulation this variable represents a limiting condition on the output of these companies, and thus is a measure of their capacity to produce insurance. Allen, p. 101.
5. Allen, p. 103.
5. Joskow, p. 388.
7. Kamerschen, et al., p. 496.
8. In Kamerschen, et al., p. 496, from Belth, "Price Disclosures in Life Insurance," Business Horizons, Vol. XI, No. 3, June 1968, pp. 32-33.

9. Pritchett (1971), p. 560.

10. Pritchett (1973), p. 158.

11. Ibid, p. 165.

12. Halpern and Mathewson, p. 204.

13. Ibid, p. 204.

14. Ibid, p. 205.

#### IV. The Model

In testing for economies of scale in the life insurance industry, we will use three general specifications of the average cost function. These are:

$$1) \quad AC = a + bQ + \varepsilon$$

$$2) \quad AC = a + bQ + cQ^2 + \varepsilon$$

$$3) \quad \log (AC) = a + b \log (Q) + \varepsilon$$

where  $Q$  = aggregate output measure (as defined below)

$AC$  = average cost = total costs/ $Q$  (with total costs as defined below)

$\varepsilon$  = error term

Equation (1) reflects the hypothesis that average cost is linearly related to output. Equations (2) and (3) provide for the possibility that A.C. is non-linearly related to output. Equation (2) represents the functional form of the U-shaped A.C. curve which is commonly referred to in the literature. Equation (3) is the average cost function analogue to the Cobb-Douglas form of the production function.<sup>1</sup>

We showed above that the presence of economies of scale are indicated when the A.C. function is downward sloping. For equations (1) and (3) this property is indicated by a negative value for the  $b$  coefficient.<sup>2</sup> Thus, it is the value of this expression which determines whether economies of scale are present at various ranges of outputs for the second functional form.

In running our empirical tests we will also consider three other variables or factors which might influence the determination of

average cost. The first of these factors is the market or organizational structure of the individual firms. In particular, we will explore the possible effect upon a firm's administration of whether it is a publicly owned stock company, whose managers are accountable to its owners, or a mutual company, in which case managerial accountability is much more ambiguous. In an effort to account for any differential in managerial efficiency stemming from this corporate structure dichotomy, we include a dummy variable (M.S.), which takes a value of 1 for mutual companies, and of 0 for stock companies. The use of an interaction dummy is rejected on the grounds that it is not economically plausible for this dichotomy to result in the use of different technologies by companies. While certain inefficiencies might result from mutual ownership which would cause a slight upward shift in the intercept term, it is unlikely that this factor would result in different techniques being employed by these firms, and thereby causing a different slope for the cost function.

The second factor which we will consider as having a possible effect on the average cost curve is the company age. In particular, we sought to test the hypothesis that insurance companies face initial start-up or organizational costs which shift up the average cost curve for young firms. In order to accomplish this test, we include a dummy variable (A) which will take a value of 1 for firms with less than or equal to ten years in business, and of 0 for firms of greater age.

About 25% of the firms in our sample are owned by larger insurance companies. This set up has been established by these companies in order to keep the strict New York Insurance Regulations from governing

the workings of the entire corporation. While the organizational and accounting structures of these firms are entirely separate from the parent company's, it is possible that some benefits accrue to these firms for being part of a larger organization. In order to correct for cost savings which might result from this relationship, we include a dummy variable (OWN) in our equations which takes a value of 1 for these owned companies.

As a result of the inclusion of these variables, the final specification which we will employ in our analysis will contain the following equations:

$$(1a) \quad AC = a + bQ + c MS + dA + f OWN + \epsilon$$

$$(2a) \quad AC = a + bQ + cQ^2 + dMS + fA + g OWN + \epsilon$$

$$(3a) \quad \log (AC) = a + b \log Q + cMS + dA + f OWN + \epsilon$$

where the variables are defined as above.

Because of the great variation in the size of the firms in our sample, with the largest being over 5,000 times bigger than the smallest firms, heteroscedasticity is a serious problem in our data. Although the presence of heteroscedasticity will not affect the consistency of our coefficient estimates, it will cause a loss in the efficiency of our regression results. This lost efficiency could have serious consequences with respect to the interpretation of the results of a study such as this, in that inflated standard errors may alter the outcome of significance tests for various coefficients. That is, a failure to reject the null hypothesis that economies of scale do not exist could result from imprecision in our coefficient estimates rather than the true absence of such economies.

The correction for the presence of heterocedasticity in our data will be made by running weighted least squares on several of our equations. In particular, we assume that the variance of the error term ( $\epsilon$ ) for each observation in the first two equations is proportional to the size of the individual firm or observation raised to some power alpha ( $\alpha$ ). Thus, we have for these equations that  $\text{var } \epsilon_i = \sigma^2 Q_i^\alpha$ , where  $\sigma^2$  is a scalar component of the variance which is common to all observations. In the case of the third equation, we assume instead that the variance of the error term is proportional to the logarithm of the size of the firm raised to some power alpha. That is, we have for this equation that  $\text{var } \epsilon_i = \sigma^2 (\log Q_i)^\alpha$ , where  $\sigma^2$  is again a common component of the variance. On the basis of these assumptions, we know that the proper method by which to correct for heterocedasticity is achieved by dividing each variable by the appropriate measure of size raised to the  $\alpha/2$  power for each observation, and running OLS on this transformed body of data. This correction reestablished the scalar covariance matrix necessary for least squares to produce the best linear unbiased estimation.

To estimate a value for alpha for each equation, we will utilize a simple three step method. This method consists, first, of estimating the original equation by OLS. The second step entails the calculation of the residuals from the regression for each observation. In the third step, you regress the square of these residuals on the variable to which the variance of the error term is hypothesized to be related. In our case, this third step is accomplished by running



OLS on the equation:

$$\log \epsilon_i^2 = c + \alpha \log (f(Q_i)) + u,$$

where  $\epsilon_i^2$  = the square of the  $i^{\text{th}}$  residual

$f(Q_i)$  =  $Q_i$  for equations 1 and 2, and  $\log Q_i$  for equation 3.

From this regression, we obtain an efficient estimate of  $\alpha$  which enables us to run the weighted least squares procedure that we described above.

In addition to the presence of heteroscedasticity, the stochastic specification in our equations may diverge from the specification necessary to make OLS the appropriate procedure to obtain our coefficient estimates for two additional reasons. The first of these reasons, which we mentioned above, is the possible occurrence of simultaneous equation bias. The second, and more likely problem which we will face in our data, is the presence of errors in our measurement of output. By this we mean, that our aggregate output measure, which will be defined below, remains a proxy variable which we are using in place of the true, but unobservable, output measure,  $Q^*$ . It is a natural assumption that the proxy will differ from  $Q^*$  by a random error term,  $v$ . That is,

$$Q^* = Q + v$$

where we assume that  $E(v) = 0$  and  $\text{var}(V) = \sigma^2 I$ .

The presence of these biases represents a departure from the stochastic specification necessary to invoke the Gauss-Markhov result that OLS produces BLUE coefficient estimates. The departure in this case carries with it far more serious consequences to our analysis than

the presence of heteroscedasticity. Whereas heteroscedasticity merely resulted in reduced estimation efficiency, the occurrence of either simultaneous equation bias or errors in variables will cause inconsistent coefficient estimates to result from the naive usage of OLS to estimate our average cost equations. By this we mean, that even as our sample grows infinitely large, our coefficient estimates will fail to approach the "true" parameter values in these equations, but will instead approach some other value in the limit with probability approaching one.

In order to correct for the presence of either of these sources of inconsistencies in our parameter estimates, we will employ an instrumental variable procedure. It is a well known fact that such a procedure, when correctly utilized, will produce consistent parameter estimates for an equation in the face of either, or both, of these problems.<sup>4</sup> A word of caution is warranted here in that when employing the method of instrumental variables care must be taken in the choice of the candidates for the instrument. Notably, these candidates must be correlated with the variable causing the inconsistencies, but remain independent of the error term in the equation, for spurious correlation between the right hand side variable and this error term is the underlying source of the original inconsistencies.

It should also be noted that the use of instrumental variables serves two purposes in our analysis. In addition to correcting for possible inconsistencies in our parameter estimates, it also represents a means of obtaining evidence of the presence of these biasing influences.

If we simplify our specification of equations 1 and 3 by assuming that we may lump our dummy variables into the constant term (thus leaving us with a two variable case), it is easy to show that the OLS coefficient estimate of the variable measured with error is biased toward zero in the probability limit.<sup>5</sup> On the other hand, if simultaneous equation bias is present in this simplified model, the OLS coefficient estimate, is biased toward 1.<sup>6</sup> Since the true coefficient of the size variable in both of these equations is certainly less than or equal to 1, we see, therefore, that these sources of bias will act in the same direction unless the scale coefficient is greater than zero. It is unlikely, therefore, that the direction of movements which this parameter estimates takes following the introduction of instrumental variables into the estimation procedure will provide any evidence as to which of these sources of inconsistencies is causing any biases which we observe in our OLS results. Only the general presence of inconsistencies may be established from this knowledge.

The instrumental variable candidates, which we will employ in our estimation, are members of a set of such candidates suggested by Wald, Durbin and others.<sup>7</sup> All of these candidates are based on the rank ordering of the variable responsible for producing inconsistencies in the OLS estimates. The Wald candidate is formed by assigning a value of plus or minus 1 to each element position, according to the corresponding element in the data matrix for the variable under consideration is greater or less than the median value of that variable.

Alternative candidates are based on the assignment of values to elements according to the quantiles in which their corresponding element in the original data matrix falls with respect to a rank ordering of the variable being replaced. Probably, the most efficient of these candidates is one which Durbin has suggested. This candidate is formed by merely replacing the original elements of a variable by their rank order.

Instrumental variables derived from the above class of candidates, which we will henceforth refer to as Wald instruments or candidates, were chosen for use in this study for one primary reason. Simply, they were the only logical choice of candidates available. In particular, while it is logical to assume that instruments formed on the basis of these candidates possessed the necessary prerequisites to achieving consistent estimators, no other candidate could be found which would fulfill these conditions. Although the instruments from these candidates are certainly correlated with our size variable, it is unlikely that they would be correlated with the error term in the equations in the plim (i.e.,  $\text{plim } \frac{1}{T} w' \varepsilon = 0$ )

After choosing from this list of candidates to form our instruments, our instrumental variable estimator then becomes,

$$\hat{b}_{IV} = (w'x)^{-1}w'y$$

where  $w$  is defined to be the instrumental variable  
 $x$  is the original data matrix  
 $y$  is the dependent variable.

Although our instrumental variable procedure provides consistent parameter estimates, it will not provide efficient estimates unless the underlying heterocedasticity in our data is also corrected. To accomplish this we will employ a method which basically represents running weighted least squares in conjunction with instrumental variables. Under this approach, our estimator becomes:

$$b_{IV} = (W'Q^{*-1}Q^{*-1}X)^{-1} W'Q^{*-1}Q^{*-1}y)$$

where W is the instrument

X is the original data matrix

$Q^*$  is a diagonal matrix with  $Q_i^{\alpha/2}$  as the diagonal elements

y is the dependent variable.

Model IV

1. For proof of this proposition see Henderson and Quandt, 1970, pp. 77-84.
2. Ibid.
3. See Johnston (1972), pp. 278-291 and pp. 380-384.
4. Johnston (1972), p. 282.
5. Ibid, p. 344.
6. Ibid, p. 284-285.

V. The Data:

The data for this study was gathered from the Annual Report of the Superintendent of Insurance for New York for year ended December 31, 1970.

Our sample consisted of sixty companies doing business in New York state. The data was drawn from the annual statements of these companies which were published in the above report.

A brief comparison of our sample with the total U.S. life insurance industry indicates that the distinguishing characteristics of our sample compare favorably with those of the industry in general. While real estate, policy loans and mortgages accounted for 3, 7.8 and 35.9 percent respectively of the total assets invested by all U.S. life insurance companies in 1970, the corresponding figures for our sample were 8.3, 10.3 and 31.6 percent. The average size of an ordinary life insurance policy over all U.S. life insurance companies was \$6,110 in that year. The average size of an ordinary policy in our sample was \$5,844. As far as the percentage of the market represented by our sample, our sample accounted for 61.2 and 78.3 percent of the total ordinary and group life insurance sold by U.S. firms respectively.<sup>1</sup> Thus, we see that on the basis of these comparisons our sample is a good representation of the life insurance industry as it existed in 1970. The only factor which would prevent the drawing of assertive conclusions with respect to the structure of the life insurance industry from our empirical results is our failure to use random sampling techniques when we chose our sample.

The fact that this study is being conducted on a regulated, financial industry means that the problems associated with doing empirical analysis on the basis of accounting data, which we discussed in Section II, are less severe than they would be if other types of industries were being studied. Because they are regulated and produce very similar outputs, these institutions tend to record their output and costs uniformly. Also, because there are few externalities and only small amounts of inventories or capital equipment are required, with most expenses coming in the form of wages or supplies, accounting costs for these firms corresponds closely to economic cost.<sup>2</sup> Due to these properties, we expect that cost curve analysis of this industry has significant economic meaning.

The definition of cost which we employ is that cost is given by the sum of commission fees and general insurance expenses. Unfortunately, while output is being measured only with respect to the life insurance and annuity business of these firms, these costs include expenses attributable to the accident and health insurance written by these firms, as well as their life insurance output. As a result, our dependent variable contains a spurious error component, which possesses a nonzero mean.

It is a well known econometric fact that error in the measurement of the dependent does not affect the unbiased nature of OLS results, as long as this error has mean zero. In order for us to invoke this property of least-squares estimation, therefore, we must transform our cost variable so that its error component has mean zero. This transformation is accomplished by multiplying our cost



variable by the fraction which the premiums stemming from the business incorporated into our output measure represents of the total premiums received by the firm. We assume that the residual error in our measurement of the dependent variable following this transformation has mean zero. As long as this assumption is correct, the effect of the measurement error in the dependent variable is only a reduction in the efficiency of our estimator, and does not influence any of its consistency properties.

As we have described in prior sections of this report, the definition of output which we use in our estimation is a composite measure. This composite measure is formed on the basis of a weighted sum of various products provided by these firms. The weights used in this sum are unit cost measures taken from a study performed by Geehan and Heishhorn (1975). The unit costs for the various products which were formulated in that study were derived from the report of the Expense Committee of the Canadian Institute of Actuaries for 1971. These unit cost weights are reproduced in Table 3.

In order to use the unit costs given in Table 3 as the basis for creating our aggregate output measure, we must assume that these unit costs are also relevant for the firms in our sample. Although the original weights were derived from Canadian data, while our sample consists entirely of U.S. companies, this assumption does not appear to be too restrictive. We feel justified in making the claim because of the great similarities between the Canadian and American markets, and because of the wide range of contacts

Table 3: Unit Costs (Weights) of Life Insurance Activities

<u>Output Category</u>	<u>Unit Cost</u>
Ordinary Insurance	
✓First Year	\$100.00 per policy
Term and Temporary Additions	\$ 10.92 per \$1,000 new effected
Basic Insurance	\$ 28.22 per \$1,000 new effected
✓Renewal	\$ 7.10 per policy
Term and Temporary Additions	\$ 0.94 per \$1,000 in force
Basic	\$ 2.03 per \$1,000 in force
Group Insurance	
✓First Year	\$100.00 per policy
	\$ 1.35 per \$1,000 new effected
✓Renewal	\$ 0.65 per \$1,000 in force
Ordinary Deferred Annuities	
✓First Year	<u>\$100.00 per policy</u>
	<u>\$ 31.25 per \$1,000 annual</u>
	<u>payment, new effected</u>
	80% of first year premium income
✓Renewal	<u>\$ 6.60 per policy</u>
	<u>\$ 1.67 per \$1,000 annual payment,</u>
	<u>in force</u>
	4.5% of renewal premium income
✓Single Premium Annuities	5.0% of single premium income
	(net dividends to policy-
	holders)
Group Annuities	<u>\$100.00 per policy</u>
✓First Year	20% of premium income
✓Renewal	10% of renewal premium income
✓Single Premium	3% of single premium income (net
	of dividends to policyholders)
Total Disability Waiver of Premium	<u>\$ 0.12 per \$1,000 in force</u>
Disability Insurance	<u>\$ 0.12 per \$1,000 in force</u>
Vested Annuities (ordinary and group)	\$ 12.00 per policy
✓Mortgage and Real Estate Assets	0.32% of assets (\$)
✓Policy Loans	0.64% of assets (\$)
Segregated Funds	0.18% of assets (\$)
✓Balance of Ledger Assets	<u>0.12% of assets (\$)</u>

Notes: All dollar series are deflated by an appropriate price index; see text. The unit costs in this Table are derived from the report of the Expense Committee of the Canadian Institute of Actuaries (1971). Data for time series of each activity are taken from the annual reports of the Superintendent of insurance for Canada, Volumes I and III. Data limitations prevented us from extending the time series back prior to 1955.

between the firms operating in these two countries.

The "products" that we include in our output measure are the following: ordinary insurance (all classes), group insurance (all classes), ordinary and group deferred annuities (all classes), mortgage and real estate assets, policy loans, and the balance of ledger assets. These items are denoted by a check next to their titles in Table 3. The reader should note, however, that these products do not comprise all of the output or service provided by the firms in our sample. The variables which we do not include because of insufficient data, or because they represent only a very insignificant part of the total output of these firms are: total disability waiver of premium, disability insurance, vested annuities, and segregated funds. In addition to these variables, a portion of the annuity business that we listed above as being incorporated into our output measure, is also missing from our composite because of the unavailability of data. This portion is the output measure corresponding to the number of annuity policies and the amount of annual payments credited to these policies. All of these missing variables are indicated in Table 3 by the underlining of their respective weights.

The portion of the theoretical composite output which is missing in the output measure that we utilized represents a specification error in the model we used in our estimation. Specifically, the specification error is that of a misspecified explanatory variable, and the result will be inconsistent estimation. It should be noted also that the inconsistencies which are introduced in this way are separate from those which stem from simultaneous equation bias,

or errors in variables. Furthermore, the specification inconsistencies may not be corrected through the use of instrumental variables.

The impact of the specification error which we described in the preceding paragraphs should not be overstated however. In particular, available evidence indicates that the output measure that we employ in our estimation procedures accounts for the vast majority of the theoretical composite output. In the case of the missing portion of the annuity business, data from the Canadian industry indicates that this part represents only 8% of their annuity output, which in turn makes up only about 12% of the total output for these companies. Thus, we see that the annuity part of the business, which we have left out of our composite measure, results in our exclusion of roughly only 1% of the total, theoretical composite output. Similar calculations for the other portions of our "missing" output variable imply that all of these components put together do not represent more than an additional 3% of the total output. The output variable which provides the basis for our empirical results, therefore, captures at least 95% of the total, theoretical composite output.

The fact that we know the form of the misspecification error allows us to predict the direction of the resulting bias in our OLS size coefficients. It is easily shown that the expected values of these coefficients in the presence of misspecification error in the  $K^{\text{th}}$  variable of these coefficients are given by

$$E(b_0)_h = \beta_h + p_{hK} K \quad h=1, \dots, K-1$$

$$= p_{KK} \beta_K \quad h=K$$

where  $\beta_h$  is the  $h^{\text{th}}$  element of the true coefficient vector;

$b_{o_h}$  is the  $h^{\text{th}}$  element of the OLS coefficient vector;

and the  $p$ 's are the coefficients of the following auxiliary regression:

$$x_{\alpha K} = \sum_{h=1}^{K-1} p_{hK} x_{\alpha K} + p_{KK} x_{\alpha K} + \epsilon$$

$x_{\alpha K}$  and  $X_{\alpha K}^*$  being the  $(\alpha, K)^{\text{th}}$  element of the true and misspecified data matrix respectively.<sup>3</sup> Since, as we saw above,  $x_{\alpha K}$  is approximately equal to  $1.053 X_{\alpha K}^*$ , we expect that  $p_{KK}$  would be positive in sign. If the above relation between  $X_{\alpha K}$  and  $X_{\alpha K}^*$  held exactly, the plim of  $p_{KK}$ , in fact, would be 1.053. Thus, we would expect a slight upward or no bias in the absolute magnitude of the size coefficient in our results on the basis of this analysis.

The biases occurring in the case of the dummy variable coefficients, however, are less clearcut. Although the plim of these biases would be zero if the above relation held exactly, this is unlikely to be true, especially in the sample. Furthermore, we have no basis for a priori prediction of the direction that these biases may take (i.e., for the signs of  $p_{hK}$ ) given our sample. Thus, the principle of insufficient reason dictates that we assume no biases to be present in our estimates.

In the case of our instrumental variable estimates, it is not practical to attempt the type of analysis which we invoked in the preceding paragraphs. This realization arises from the fact

that the asymptotic bias in our instrumental variable estimator due to specification error is given by<sup>4</sup>

$$\text{plim } (b_{IV} - \beta) = \beta_1 (w'X)^{-1} w'x$$

where  $\beta_1$  is the coefficient of the misspecified variable;

$x$  is the data vector of the misspecified variable.

Since the direction of bias for all coefficients in this case depends upon the value of  $w'x$ , which is unknown, no prediction of the value or direction of this bias is possible for this estimator.

Data V

1. The figures describing the total U.S. insurance market were taken from the Life Insurance Handbook for 1970, published by the Life Insurance Institute in New York.
2. Benston, pp. 316-318.
3. Theil, p. 553.
4. Ibid, p. 552.

## VI. Results:

Before examining the results we achieved in our estimation, we will first reiterate the basic rationale for this study, so as to better put our results into perspective. This study represents a reexamination of the presence of economies of scale in the life insurance industry. The reappraisal of this issue is necessary, because available evidence indicates that past conclusions with respect to this topic were based on the results of biased and inconsistent estimation procedures. This paper represents a technical improvement over this prior work because it employs a more theoretically acceptable measure of output, and because various other sources of bias are considered and corrected for whenever possible. Finally, as we indicated above, we approached this issue with the hypothesis that inconsistent estimation had produced the prior evidence of the existence of economies of scale in the U.S. life insurance industry; and that sounder estimation procedures would reverse this finding. This hypothesis was based on the fact that, in our opinion, no good theoretical justification for these size economies could be formulated.

Since a major motivational factor for conducting this study was the belief that past work on this topic was in error, we began our empirical analysis by rerunning a modified version of Houston and Simon's model on our body of data. In the model we estimated, we dropped their dummy variable for "bigness" from the list of right hand side variables, and reinstated a dummy variable for



market structure. The results of this estimation are contained in Table 4.

A comparison of our results with those obtained by Houston and Simon, which were reproduced in Table 1 above, shows that they yield substantially the same conclusions. In particular, both sets of results imply the presence of significant economies of scale.<sup>1</sup> The sign of the coefficients for new premiums and group premiums also agreed with Houston and Simon's finding and a priori expectations. The coefficients for the industrial premium and the lapse ratio failed to be statistically significant in our results, but their respective point estimates were of the appropriate sign.<sup>2</sup>

The only coefficient which contradicts Houston and Simon's findings is that for the market structure dummy. To attain correspondence with Houston and Simon's regressions, we ran our regressions with this dummy defined to have a value of 1 for mutual companies and 0 for stock companies. Thus, our prior expectations would assign a positive sign to this variable's coefficient. The fact that the opposite result attains very significantly for our weighted regressions is a puzzling situation. Since Houston and Simon never report a regression which contains this variable, although they discuss theoretical reasons for its presence, we cannot determine whether their point estimates suffered from the same problem. While we offer no explanation for the occurrence of this apparent sign reversal, we note that it may be indicative of the presence of specification errors in their model, with this variable acting as a proxy for an excluded

Table 4

(Regressions: Houston & Simon's Model)

C	f(P)	IR	LR	MS	GP	NP	R <sup>2</sup>	Sy	Log-L.
<u>f(P) = log(P)</u>									
OLS									
.822	-.0284	.153	.285	-.035	-.194	.141	.66	.078	67.94
(.128)	(.0068)	(.131)	(.229)	(.0242)	(.0461)	(.083)			
(6.44)	(-4.19)	(1.17)	(1.24)	(-1.44)	(-4.19)	(1.71)			
<u>Weighted L.S.</u>									
.608	-.0178	.251	.333	-.0599	-.168	.293			
(.102)	(.0052)	(.130)	(.220)	(.0183)	(.0325)	(.0761)			
(5.95)	(-3.42)	(1.94)	(1.51)	(-3.27)	(-5.17)	(3.85)			
<u>f(P) = 1/P</u>									
OLS									
.419	81370	.092	.031	-.086	-.288	.262	.67	.084	63.8
(.024)	(22685)	(.134)	(.229)	(.025)	(.04)	(.085)			
(17.7)	(3.59)	(.69)	(.14)	(-3.5)	(-7.15)	(3.09)			
<u>Weighted L.S.</u>									
.267	107539	.219	.391	-.043	-.205	.328			
(.0154)	(31586)	(.131)	(.221)	(.019)	(.0315)	(.0768)			
(17.4)	(3.4)	(1.66)	(1.77)	(-2.29)	(-6.49)	(4.27)			
<u>f(P) = 1/log(P)</u>									
OLS									
-.179	8.75	.169	.195	-.035	-.189	.149	.68	.075	70.0
(.101)	(4.75)	(.126)	(.226)	(.023)	(.044)	(.079)			
(-1.77)	(4.75)	(1.34)	(.861)	(-1.50)	(-4.27)	(1.88)			
<u>Weighted L.S.</u>									
-.0726	6.49	.247	.270	-.06	-.163	.286			
(.086)	(1.65)	(.126)	(.218)	(.0178)	(.0317)	(.074)			
(-.839)	(3.93)	(1.96)	(1.24)	(-3.4)	(-5.14)	(3.87)			

Legend

See Table 1.

MS = Market Structure Lummy

variable.

The estimation results for our model are contained in Tables 5-7.<sup>3</sup> A review of these tables shows that for each specification, while the scale coefficient is insignificant for the least squares and instrumental variables regressions which were uncorrected for heteroscedasticity, the scale coefficient for the weighted regressions, both least square and instrumental variables, is statistically different from zero at greater than a 97.5% confidence level in each case. The implication of this result is that prior studies were essentially correct in finding scale economies for the life insurance industry.

A closer examination of the results shows them to be theoretically sound in content. In each case, the change to statistical significance of the coefficient of the size variable arose from a marked reduction in the standard error of the estimate rather than a large change in its absolute value, in accordance with the theoretical prediction for the effect of such a heteroscedasticity correction. In view of the great size differentials present in our example, it is not surprising, in fact, that the correction for heteroscedasticity brought about such dramatic changes in the standard errors of our estimates.

As we recall from our analysis in the preceding section, the effect of the mispecification in our output measure is to cause a slight upward bias in the absolute value of the size coefficients for our OLS estimator, and a bias in determinable sign for our instrumental variable estimator. In view of our determination of the slight extent of our specification error, and the rather small impact that it is likely to have or

Table 5

(Regressions:  $AC = a + bQ + cOWN + dMS + fA + e$ )

C	Q	OWN	MS	A	R <sup>2</sup>	Sy	Log-L.
<u>1. OLS</u>							
1.03	-.254E-9	-.279	-.154	-.153	.205	.228	417.9
(.065)	(.214E-9)	(.088)	(.078)	(.106)			
(15.8)	(-1.19)	(-3.17)	(-1.95)	(-1.44)			
<u>2. Weighted L.S.</u>							
1.02	-.195E-9	-.351	-.148	-.164			
(.07)	(.075E-9)	(.082)	(.076)	(.163)			
(14.6)	(-2.6)	(-4.27)	(-1.94)	(-1.0)			
<u>3. Instrumental Variables(IV)*</u>							
1.04	-.464E-9	-.274	-.135	-.158		.23	
(.066)	(.373E-9)	(.089)	(.083)	(.107)			
(15.7)	(-1.24)	(-3.06)	(-1.61)	(-1.47)			
<u>4. Weighted IV*</u>							
1.02	-.346E-9	-.323	-.111	-.166			
(.069)	(.154E-9)	(.084)	(.081)	(.144)			
(14.7)	(-2.25)	(-3.83)	(-1.38)	(-1.16)			

\*Lurbin's candidate was used in forming the instruments.

Table 6

(Regressions:  $AC = a + bQ + cQ^2 + cOWN + dMS + fA + e$ )

C	Q	Q <sup>2</sup>	OWN	MS	A	R <sup>2</sup>	Sy	Log-L.
<u>1. OLS</u>								
1.041	-.912E-9	.943E-18	-.263	-.136	-.168	.21	.22	417.8
(.0057)	(.718E-9)	(.982E-18)	(.09)	(.081)	(.107)			
(15.8)	(-1.27)	(.960)	(-.292)	(-1.69)	(-1.56)			
<u>2. Weighted L.S.</u>								
1.0324	-.912E-9	.898E-18	-.321	-.105	-.192			
(.068)	(.333E-9)	(.406E-18)	(.081)	(.076)	(.163)			
(15.1)	(-2.73)	(2.21)	(-3.95)	-1.37)	(-1.17)			
<u>3. I.V.*</u>								
1.00 <sup>2</sup>	2.47E-9	-6.05E-18	-.342	-.132	-.094		.385	
(.115)	(2.97E-9)	(5.6E-18)	(.165)	(.139)	(.189)			
(8.73)	(.833)	(-1.08)	(-2.08)	(-.94)	(-.50)			

\*Lurbin's candidate and Wald's candidate, based on a 20<sup>th</sup> quantile breakdown, were used in forming the instruments.

Table 7

(Regressions:  $\log AC = a + b \log(Q) + cOWN + dMS + fA + e$ )

C	$\log(Q)$	OWN	MS	A	$R^2$	Sy	Log-L.
<u>1. OLS</u>							
-6.37	-.0350	-.309	-.1098	-.317	.23	.291	-10.99
(.385)	(.0243)	(.117)	(.105)	(.146)			
(-16.6)	(-1.44)	(-2.65)	(-1.04)	(-2.16)			
<u>2. Weighted L.S.</u>							
-6.23	-.0439	-.350	-.0712	-.365			
(.313)	(.0187)	(.107)	(.101)	(.199)			
(-19.9)	(-2.34)	(-3.26)	(-.704)	(-1.83)			
<u>3. I.V.*</u>							
-6.59	-.0209	-.328	-.133	-.283		.2915	
(.398)	(.0251)	(.117)	(.106)	(.148)			
(-16.6)	(-.83)	(-2.79)	(-1.26)	(-1.92)			
<u>4. Weighted I.V.*</u>							
-6.29	-.0399	-.356	-.0803	-.354			
(.324)	(.0195)	(.108)	(.102)	(.200)			
(-19.4)	(-2.05)	(-3.30)	(-.788)	(-1.77)			

\*Lurbin's candidate was used in forming the instruments.

our estimated coefficients, we reject this as a possible explanation of our finding significant economies for all specifications for both weighted least squares and weighted instrumental variable estimators.

The only other variable, besides the constant and the output measure, which attained statistical significance on the basis of a two-tailed t-test at a 95% confidence level, is the dummy variable that accounted for ownership by a larger firm. The coefficient for this variable took the appropriate sign in accordance with theoretical expectations for each specification. The large value of this coefficient is somewhat surprising, however. Our results indicate in fact, that the downward shift in the average cost curve due to being a subsidiary is of the same magnitude as the reduction in cost between our smallest and largest firms caused by the presence of scale economies. Thus, it appears that these firms are able to take advantage of whatever size economies are present in the controlling firm, even though their operations are ostensibly independent in nature.

Although the coefficient of the market structure, dummy variable is not statistically significant in any of our regressions, its point estimate attained the appropriate sign in all cases.<sup>4</sup> This is an interesting result since it approaches statistical significance in several regressions, and because of the inappropriate sign that it attained under Houston and Simon's model. Our results indicate, therefore, that the appropriate sign for this coefficient is indeed negative (positive for Houston and Simon's model).

The coefficient of the dummy variable which sought to correct for organizational costs that might be present in new firms (A) did not achieve statistical significance in any of our weighted regressions. Its point estimate, nonetheless, differed in sign from what theory predicted. The fact, however, that it only neared significance for the logarithmic specification indicates that this does not represent evidence of any major failings in our specification. Furthermore, the occurrence of this incorrect sign may result from the fact that this variable is positively correlated ( $p=.25$ ) with the ownership dummy. Thus, this statistically insignificant deviation from theory may be due to spurious correlation between two included variables.

In the two instances that instrumental variables produced reasonable results, the direction in which the size coefficient estimates shifted for the weighted instrumental variable estimator, with respect to their original values under weighted least square, differed between the two specifications. Specifically, the coefficients increased in absolute value for the linear specification, and decreased for the logarithmic specification. As we recall from our prior analysis, the coefficients should have shifted upward in value in the presence of either simultaneous equation bias, or errors in variables. Although no rigorous statistical test was conducted, we feel that, on the basis of this divergence in coefficient movement, and because of the relatively small magnitudes of the movements, it is unlikely that such a test would yield the result that the instrumental variable estimates were significantly different from the least squares estimates. This



inference indicates that the above sources of bias were not significant in our data.

In addition to the question of the presence of statistically significant economies of scale in the life insurance industry, the issue of fundamental importance which underlies this study is whether such economies, if present, are economically important. This transition in the analysis to the economic interpretation of statistical correlations, as Johnston shrewdly points out, is always hazardous.<sup>5</sup> In our results, although the size coefficient is uniformly small in magnitude, the implied size economies are economically meaningful. In Figures 1-3 appear graphs of derived cost functions. These figures represent depictions of equations 4, 2 and 4 from Tables 5, 6, and 7 respectively, and are included as an aid to the reader in determining the extent of the size economies that are implied by the algebraic expressions. In each case, the graphs are drawn for mutual companies, which are neither new, nor subsidiaries of larger firms (i.e., all dummy variables take the value zero).

As the reader can ascertain from these graphs, the differentials in average costs stemming from scale economies over the range of firm sizes in our sample represent reductions in average cost of roughly 25% between the largest and smallest firms. Differentials of this magnitude are indeed significant in any industry.

In closing this section, we would also like to note that the results from our quadratic specification indicate the a U-shaped average cost curve apparently fits our data as well as the other specifications. Furthermore, two of our firms lie on the

Figure 1: Linear Regression Results

NO. 340-LS10 DIETZGEN GRAPH PAPER  
EUGENE DIETZGEN CO.  
MADE IN U.S.A.  
SEMI-LOGARITHMIC  
5 CYCLES X 10 DIVISIONS PER INCH

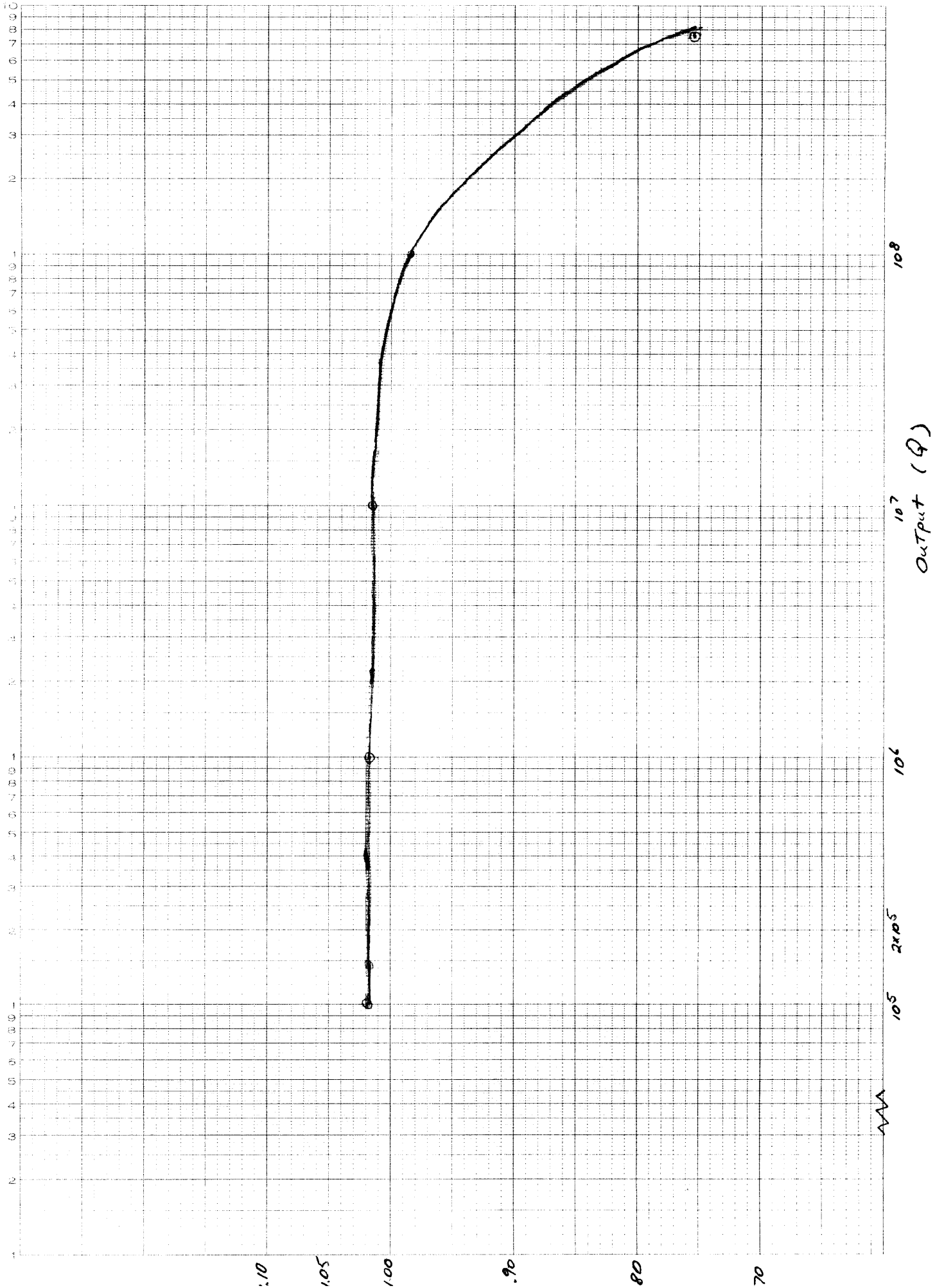
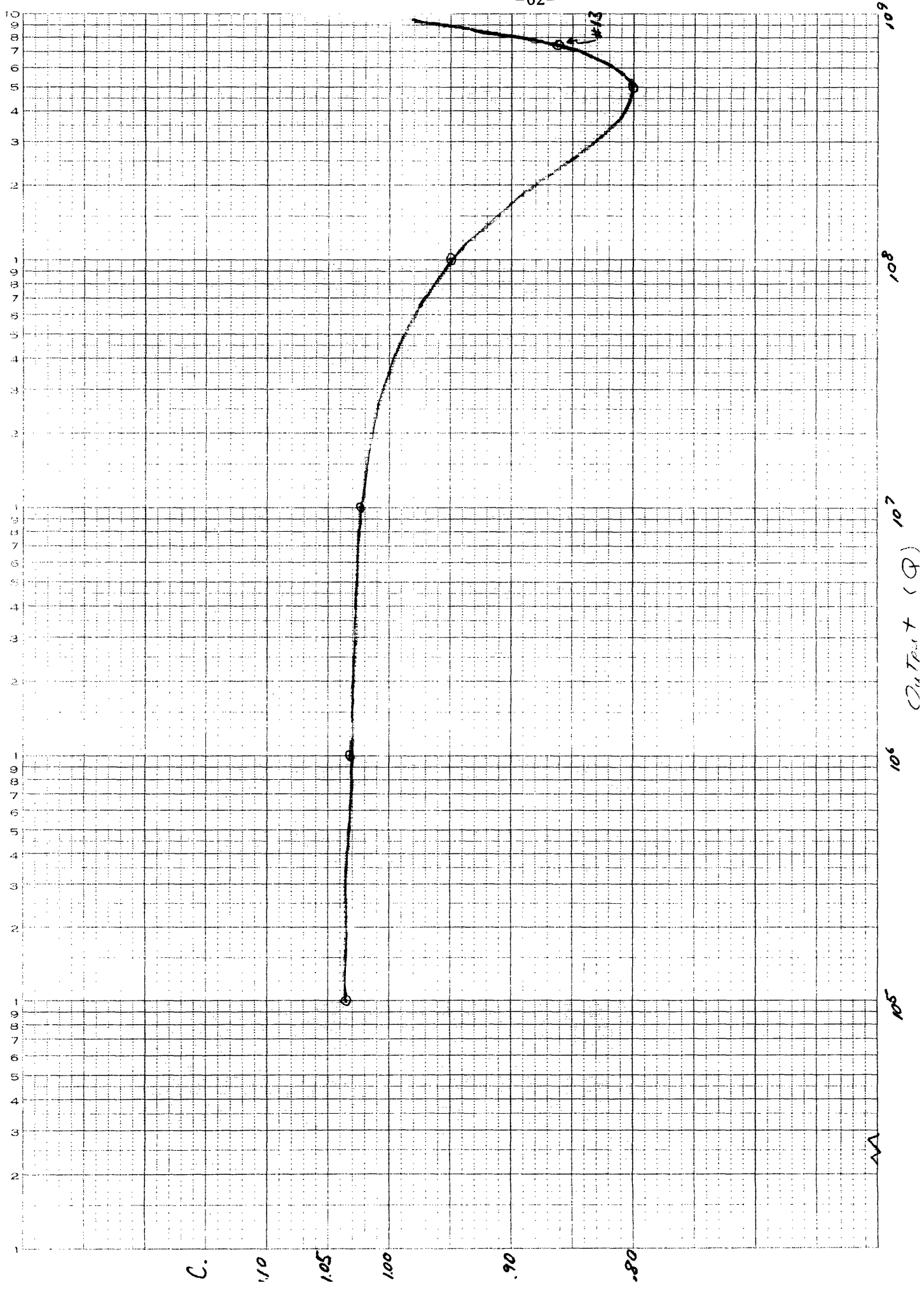
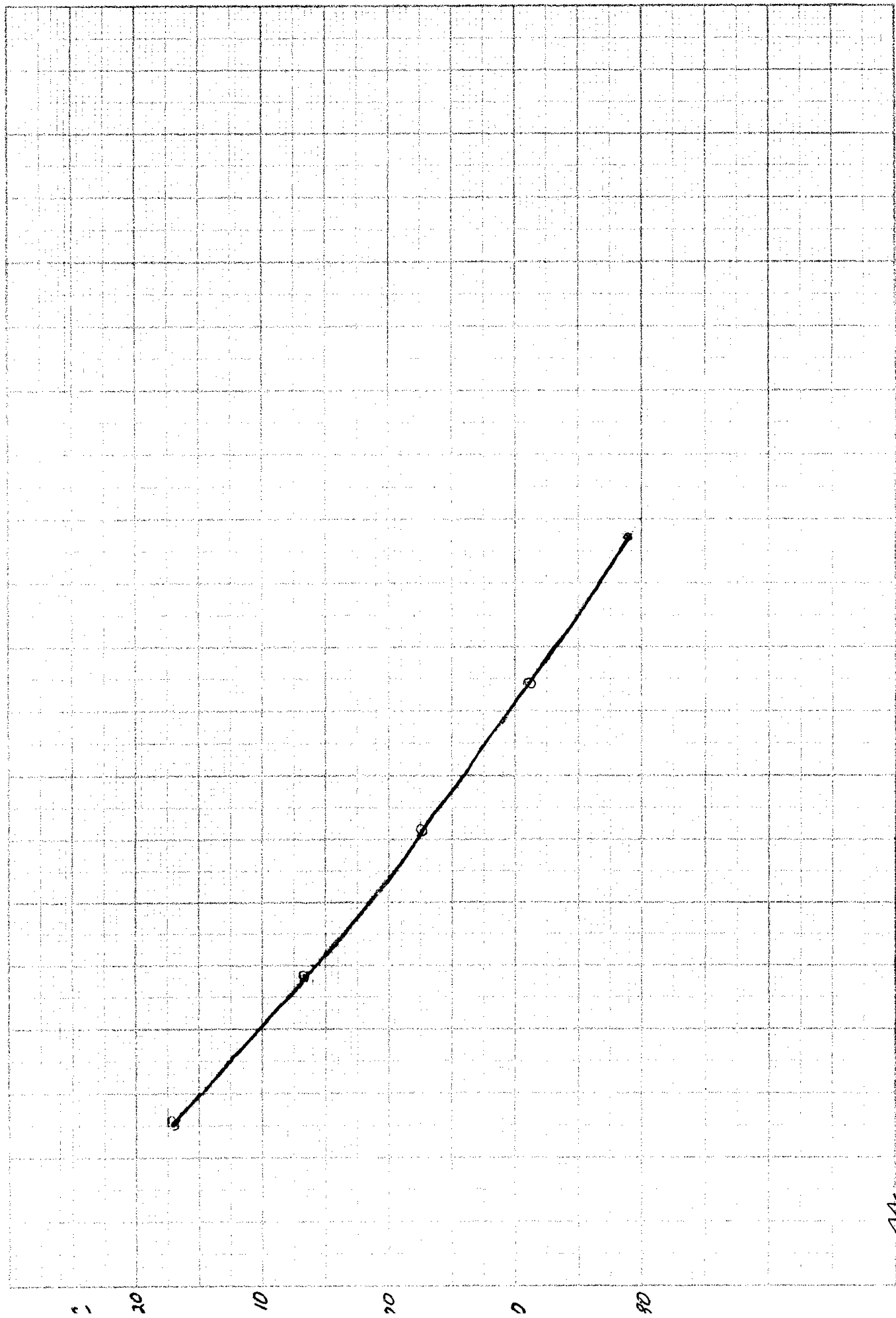


Figure 2: Quadratic Regression Results

NO. 34C-LS1C DIETZGEN GRAPH PAPER  
SEMI-LOGARITHMIC  
5 CYCLES X 10 DIVISIONS PER INCH  
EUGENE DIETZGEN CO.  
MADE IN U. S. A.



WAVE 20720 TO THE INCH 4L 194P  
FOR THE INCH 4L 194P  
NEUTRAL 6.8882 02



W 11 12 13 14 15 16 17 18 19 20 21

upward sloping portion of the implied curve. This is interesting in that empirical cost studies generally imply L-shaped cost functions, as opposed to the U-shaped curves which are implied by theory. No claim may be made, however, from our results that such a U-shaped curve is actually the appropriate one for the U.S. insurance industry.

Results VI

1. The relevant t-statistic which enables the rejection of the null hypothesis that the coefficient of the size variable is zero is  $t \geq 1.68$ . In the case of the other variables, since two-tailed tests are required for them, the appropriate t-value is  $t \geq 2.01$ .
2. See footnote 1.
3. These tables provide the results for only the preferred specification with respect to our choice of an instrumental variable candidate. Various other candidates were tried during preliminary regressions, but the Durbin candidate provided the most efficient estimates in all cases.
4. Note that this dummy value is defined in our model to take the value one for stock companies and zero for mutual companies. A negative coefficient is, therefore, what theory would predict.
5. Johnston (1958), p. 348.

## Section VII. An Analysis of Market Structure

As we noted in the previous section, the magnitude of the size economies implied by our point estimates of the slopes of our various cost function specifications would be economically significant for an industry such as insurance. Accordingly, our results contain implications for the evolution and shape of the structure of the life insurance industry. Thus, an indirect test of the soundness of our statistical results may be performed by investigating the recent behavior of the U.S. life insurance industry to see if these expected features are present. To such an analysis we now turn our attention.

A possible implication of our results concerns the evolution of the insurance industry. That is, if our results are correct and significant economies of scale are present, we would expect to observe a trend toward increasing market concentration. In order to test the presumption, we constructed market concentration ratios on the basis of total premium receipts and insurance in force for the past two decades (for the ten largest life insurance companies). These concentration ratios are tabulated in Table 8. Although simple concentration ratios are an extremely inaccurate measure of market concentration, the figures in Table 8 are quite disturbing. Instead of the rising concentration trend we expected, we observe a dramatic reduction in market concentration over these years. Furthermore, while the movement seems to be abating as measured by insurance in force, the decline seems to be continuing on the basis of premium receipts.<sup>1</sup>

In light of our market concentration figures, we are faced with the problem of apparently contradicting results. Since concentration ratios are very much summary statistics, their inconsistency with our more detailed and elaborate statistical investigation of the insurance industry need not necessarily be damning to our previous results. In particular, there are four possible explanations why such a conflict might be observed. These are:

- 1) behavior motivated by monopoly profits on the part of large firms;
- 2) the effect of entering companies, 3) the rise of insurance, and
- 4) the spatial distribution of firms.

It is certainly plausible that a profit maximizing company which possesses a considerable amount of potential market power based on economies of scale might choose not to exercise their ability to drive smaller competitive firms out of the industry. Instead, in accordance with dynamic limit pricing, they might choose to charge prices which are greater than their marginal costs, thereby earning monopoly profits, at the cost of allowing competition by medium sized companies.<sup>2</sup> If this were the case, the market concentration ratios of these very large firms could decline over the years as the ranks of the medium size companies expanded, or as the result of greater competition by medium size companies which are larger than the marginal scale implied by the prices dictated by the largest firms.

Although dynamic limit pricing represents a plausible explanation for our results, it is difficult in practice to test for the existence of such behavior by large firms. In particular, accurate information concerning



the economic profits earned by insurance companies is not readily obtainable, so that a direct test of our hypothesis, by relating unit profits to firm size, is not possible. In an effort to check the implications of this hypothesis with respect to medium size firms, however, we calculated market concentration ratios for the ten firms which ranked from 65th to 75th in size over these years. The results are recorded in Table 9. While the movement is less dramatic and more uneven than the decline for the largest firms, the predicted upward trend is certainly apparent. The rise, however, is small and the spectrum considered is narrow, so that no definitive statement can be made on the basis of these figures on whether the decline in market concentration by very large firms could be accounted for by an expansion in market power by infra-marginal medium size firms. Thus, we see that our evidence is too crude to enable any meaningful conclusions with respect to the existence of dynamic limit pricing behavior on the part of very large firms.

A second explanation for the inconsistencies in the evidence in this field may be the rapid growth experienced by this industry in recent years. That is, while the long run implications of the presence of scale economies point toward a rise in market concentration, the transient influx of a large number of small firms may have reversed this trend over our observed period. In addition, we note that this factor need not be thought of as independent of our first explanation. In particular, the existence of dynamic limit pricing on the part of firms would serve as an impetus for these small firms to suffer through several initial lean years in the expectation of recouping their losses through future excess profits once they have achieved the requisite size.

The growth of this industry over the past two decades has been very substantial. Half of the 1465 companies operating in 1963 were less than 10 years old. By 1968, an additional 300 companies had been formed. In light of the magnitude of this proliferation of insurance firms, the impact of these new firms could easily have been a significant factor in the observed downward trend in market concentration. A word of caution, however, should be made in that our figures are only suggestive, and that a more thorough analysis in the future should be made to examine whether the prediction of dynamic limit pricing of an initial "lean" period for entering firms is substantiated, and to determine the actual impact of these new firms on market concentration.

With respect to the final two rationalizations, our reasoning is that in the present of reinsurance, the smaller firms may be thought of as in effect representing agents for larger firms, and that the observed concentration ratios may be misleading if this is not accounted for. As far as spatial distribution is concerned, it might be that the cost reductions from size economies are significant only when operating in larger urban markets. As a result, the large companies may have avoided high cost rural areas, leaving them to small regional firms. A disproportionate rise in rural life insurance could then account for part of the observed decline in market concentration. Available evidence, however, indicates that reinsurance is not a significant factor in the life insurance industry. Furthermore, the argument for differential emphasis by large firms between communities of various sizes is compelling neither on theoretical grounds, especially since independent agents may be used, nor on the basis of empirical observation, since all the largest firms are chartered in all 50

states and employ thousands of field representatives.<sup>3</sup> (See Table 10)

In summary, we find that the economic implications of our cost function estimations seem to be at variance with the actual experience of the life insurance industry in recent years. In an effort to explain this fact, while maintaining the correctness of our basic results, we considered four hypotheses. Of these, only the presence of dynamic limit pricing behavior by the largest firms, and the impact of the proliferation of many new firms during this period appear to be reasonable explanations at this time. These hypotheses show, however, that conclusions concerning the validity of our cost function estimates may be misleading when based on simplistic, aggregate market measures. In addition, while the behavior of market concentration ratios in recent years is not able to repudiate our findings by itself, it does demonstrate the necessity of examining your statistical results in light of independent information that is relevant to your analysis. Accordingly, we note that our findings must be regarded as tentative awaiting further market investigation and refined statistical estimates.

1. While both measures are fraught with problems, premium receipts is a superior basis for measuring market concentration since it is better able to capture the diverse nature of these companies' "output".

2. See Gaskings (1971).

3. It should be noted that this empirical evidence is weak in that a finer breakdown between metropolitan and rural sales effort according to size is the measure actually desired. Such conformation, however, is not presently available.

Table 8

Concentration Ratio of Ten Largest Firms (%)

	On Premium Basis	On Basis of Ins. in Force
1954	59.88	
1956	58.87	
1958	57.66	55.23
1960	55.77	53.25
1962	53.94	51.90
1964	52.55	49.96
1966	49.98	47.46
1968	49.34	45.72
1970	48.2	45.85
1972	45.86	46.4

Table 9

Concentration Ratio of "Medium" Size Firms (On Basis of Ins. in Force)

1958	.0202
1960	.0206
1962	.0211
1964	.0219
1966	.0223
1968	.0212
1970	.0214
1972	.0221

Table 10  
Company Statistics

Company (In Order of Size, 1975)	# of States Covered	# of Field Representatives
(1) Prudential	50	(1590 offices)
(2) Metropolitan	"	+25,000
(3) Equitable	"	6,600
(4) John Hancock	"	10,900
(5) Aetna	"	2,400

### VIII. Conclusions and Recommendations for Further Research

In summary, the findings of this study demonstrate that past authors in this topic were correct in finding economically significant economies of scale in the U.S. life insurance industry. While fault may be found with their methodology in approaching this question, their basic conclusions have withstood the scrutiny of our analysis. Neither the introduction of a composite output measure, nor the use of superior estimation techniques, succeeded in altering this theoretically perplexing conclusion.

With respect to future research, our analysis has demonstrated that certain key issues remain unresolved. The outcome of our study raises the question, if economies of scale of the magnitude implied by our estimates are present, why are there over 1,700 insurance companies operating in the American market? Is it because of market imperfections with respect to pricing policies? In an effort to gain some insights into the nature of the life insurance market and deal with these questions, we looked at the evolution of market concentration for this industry. As we saw, however, this superficial investigation failed to provide any meaningful results. A much more detailed analysis of the life insurance market and industry organization is necessary to place our empirical findings in a proper perspective. Unfortunately, such an analysis is beyond the scope of this report.

In another direction, future studies could look at insurance companies in a more refined light. These analyses could break the

insurance firms into sub-components, such as the investment component of the firm, to determine in exactly which of these components economies of scale are arising. Ideally, a production function for the insurance industry could be worked out. Not only would such a function provide an answer to the question of which components possess economies of scale; but it would also provide less ambiguous conclusions since that approach would avoid the theoretical complications associated with utilizing cost functions to investigate a technological phenomena. Through investigations such as these, a good theoretical justification for what is now merely an empirical observation for the insurance industry may be found.



Bibliography

- Allen, Robert F., "Cross-Sectional Estimates of Cost Economies in Stock Property-Liability Companies," Review of Economics and Statistics, February 1974, pp. 100-104.
- Belth, Joseph M., The Retail Price Structure in American Life Insurance, Bloomington, Indiana: Bureau of Business Research, Graduate School of Business, Indiana University, 1966.
- Benston, George, "Economies of Scale of Financial Institutions," Journal of Money, Credit and Banking, May 1972, pp. 312-341.
- Borts, G., "Economies of Scale of Financial Institutions: Comments," Journal of Money, Credit and Banking, May 1972, pp. 419-421.
- Cummins, J., Denenberg, H. and Scheels, W., "Concentration in the U.S. Life Insurance Industry," Journal of Risk and Insurance, June 1972, pp. 177-199.
- Gaskins, D. W., "Dynamic Limit Pricing: Optimal Pricing Under Threat of Entry," Journal of Economic Theory, 3, 1971, pp. 306-322.
- Halpern, P. J. and Mathewson, G., "Economies of Scale in Financial Institutions--A General Model Applied to Insurance," Journal of Monetary Economics, 1, 1975, pp. 203-220.

Hammond, J., Melander, E., and Shelling, N., "Economies of Scale in the Property and Liability Insurance Industry," Journal of Risk and Insurance, June 1971, pp. 181-191.

Henderson, J. and Quandt, R., Microeconomic Theory: A Mathematical Approach, New York: McGraw-Hill, Inc., 1971.

Houston, D. and Simon, R., "Economies of Scale in Financial Institutions: A Study in Life Insurance," Econometrica, Vol. 38, No. 6, November 1970, pp. 856-864.

Institute of Life Insurance, Life Insurance Fact Book 1975, New York: Institute of Life Insurance, 1975.

Johnston, J., "Statistical Cost Functions: A Reappraisal," Review of Economics and Statistics, Vol. 40, 1958, pp. 339-350.

Johnston, J., Econometric Methods, New York: McGraw-Hill, Inc., 1972.

Joskow, P., "Cartels, Competition and Regulation in the Property-Liability Insurance Industry," The Bell Journal of Economics and Management Science, Vol. 4, No. 2, Autumn 1973, pp. 375-427.

Kamerschen, D. and Pascucci, J., "The Retail Price Structure in American Life Insurance: Comment," Journal of Risk and Insurance, Vol. XXXVI, No. 4, September 1969, pp. 493-495.

Launie, J., "Price Differentials in Life Insurance," Journal of Risk and Insurance, June 1968, pp. 283-288.

McElroy, R., "Returns to Scale and Cost Curves: Comments," Southern Economic Journal, October 1970, pp. 227-228.

112 Annual Report of the Superintendent of Insurance for Year Ended, December 31, 1970, Vol. 1, New York.

Pritchett, S.T., "An Intercompany Expense Comparison For Thirty Life Insurers," Journal of Risk and Insurance, Vol. 38, No. 4, December 1971, pp. 553-562.

Pritchett, S.T., "Operating Expenses of Life Insurers, 1961-1970, Implications for Economies of Size," The Journal of Risk and Insurance, Vol. 40, No. 2, June 1973, pp. 175-165.

Sandmo, A., "Returns to Scale and the Average Cost Curve," Swedish Journal of Economics, 1970, pp. 149-152.

Theil, H., Principles of Econometrics, New York: John Wiley & Sons, Inc., 1971.

Walters, A. A., "Production and Cost Functions: An Econometric Survey," Econometrica, January-April 1963, pp. 1-66.