NAME OF AUTHOR: Charles Frederick Riley

TITLE OF THESIS: An Examination of the Economic Literature Pertaining to the Behaviour of Costs in General Acute Hospitals

UNIVERSITY: of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED: MHSA

YEAR THIS DEGREE GRANTED: 1974

Permission is hereby granted to THE NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(Signed) Charles F. Riley

PERMANENT ADDRESS:

#10-2697 Whitely Court
North Vancouver, BC

DATED April 11 1974

NL-91 (10-68)

THE UNIVERSITY OF ALBERTA

AN EXAMINATION OF THE ECONOMIC LITERATURE PERTAINING TO THE BEHAVIOUR OF COSTS IN GENERAL ACUTE HOSPITALS

by

(C) CHARLES FREDERICK RILEY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF HEALTH SERVICES ADMINISTRATION

DIVISION OF HEALTH SERVICES ADMINISTRATION

DEPARTMENT OF COMMUNITY MEDICINE

EDMONTON, ALBERTA

FALL, 1974

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled AN EXAMINATION OF THE ECONOMIC LITERATURE PERTAINING TO THE BEHAVIOUR OF COSTS IN GENERAL ACUTE HOSPITALS submitted by Charles Frederick Riley in partial fulfilment of the requirements for the degree of Master of Health Services Administration.

R. H. McFlau ....
Supervisor

Dennis M. Baker ....

Lawrence Hartman ....

Date 10 april, 1974 ....

ABSTRACT

Health care is one of the largest and fastest growing activities in the Canadian economy and hospital services form the largest part of the national expenditure on health care. Hospital services are consuming a growing proportion of Canada's resources and these are limited.

It was observed that a small but significant body of hospital oriented economic literature has been developing. Generally, the literature concerned itself with the analysis of hospital cost data for the purpose of determining the level at which a hospital's average cost reached its lowest level; i.e. the point at which hospital size was optimal. In light of the rapidly increasing costs of hospital services, it was considered that a review of the literature which would consolidate the existing body of knowledge regarding hospital costs would be a valuable contribution. This thesis represents a review of the pertinent literature.

The theories of production and cost provided the empirical background against which the economic literature has been examined. The literature was surveyed and many specific criticisms have resulted. An overall commentary on the literature as it exists and directions for future research have been described.

The primary conclusions of this thesis are:

1) that there is no concensus in the literature that one particular size is optimal for a hospital, although it is agreed that small hospitals are less efficient than medium and large size hospitals;

2) that considerable research effort is required to develop the

tools of economic analysis so that they can reflect hospital activities more accurately; and

3) that 'micro' economic analyses at the level of the individual hospital are required so that each community will be able to maximize its return for each health dollar spent.

ACKNOWLEDGEMENTS

I would like to express my appreciation to Profess[illegible] H.M[illegible]Plain for his assistance and patience in the preparation of this thesis.

I am sincerely grateful to my wife, Rachelle, for her determination that this thesis be completed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

Since 1957 the manner in which health care services have been provided in Canada has substantially changed. The introduction of government hospitalization plans was completed in all of the provinces in 1961 when Quebec's plan became operational. In the early months of 1971 Prince Edward Island became the last province to begin to operate a plan insuring medical services; i.e. medicare. Health services, which once operated as part of the private sector of the economy, are now largely part of the services provided by government. During this period the variety and the comprehensiveness of health services made available to Canadians have greatly expanded. "Health care has, for some time, constituted one of the largest and fastest growing activities in the economy."[1]

[1] Economic Council of Canada, Seventh Annual Review: Patterns of Growth, (Ottawa: Queens Printer, 1970), P.38.

Trends in Health Care Expenditures[2]

National Trends

The rapid expansion in quantity and type of health services offered in Canada during recent years has been accompanied by a rapid increase in the cost of providing these services. (see Tables I-1 and I-2) The amount spent for personal health care has increased from $1,047,403,000 in 1957 to an estimated $4,387,028,000 in 1970. This represents a total increase of 418.8 per cent or an average annual rate of increase of 11.6 per cent. The extent of the increase is even more dramatic when expressed in terms of expenditure per capita. Between 1957 and 1970 the Canadian population grew from 16,677,000 to 21,406,000 people; - an increase of 128.4 per cent. During the same period the expenditure per person on personal health care rose from $62.81 to $204.94 or 326.3 per cent.

The largest segment of the expenditure for personal health care has been in the area of hospital services. Expenditures on hospital services have increased from $587,370,000 to an estimated $2,787,356,000 between 1957 and 1970. This represents an increase of 474.6 per cent

[2] The discussion in this section is based upon information obtained from the following sources: R.H.M. Plain, Resource Allocation in the Medical Service Market in Alberta, Section I, An unpublished report submitted to the Alberta Health Care Insurance Commission, April, 1971; and the Department of National Health and Welfare, Research and Statistic Memos, as listed below:
"Expenditures on Personal Health Care in Canada, 1957-1969", (October, 1970), "Expenditures on Personal Health Care in the Provinces of Canada 1957-1969", (November, 1970), and "Expenditures on Health Care in Canada, 1960-1970", (December, 1971), (Ottawa: Research and Statistics Directorate).

TABLE I-1

EXPENDITURE[a] ON PERSONAL HEALTH CARE IN CANADA, 1957 TO 1970

Year	Hospital Services: General and Allied Special	Hospital Services: All Hospitals	Total Personal Health Care
1957	422,913	587,370	1,047,403
1958	462,305	640,608	1,144,888
1959	534,728	735,626	1,290,468
1960	640,587	847,573	1,443,324
1961	722,057	949,014	1,587,626
1962	811,848	1,054,181	1,722,778
1963	909,762	1,174,887	1,921,856
1964	1,015,148	1,300,228	2,114,623
1965	1,144,479	1,461,916	2,367,051
1966	1,319,048	1,668,768	2,665,016
1967	1,523,035	1,916,296	3,029,129
1968	1,789,968	2,218,395	3,478,450
1969[b]	1,997,516	2,464,142	3,875,407
1970[b]	2,278,654	2,787,356	4,387,028

[a] In Thousands of dollars.

[b] Preliminary estimates.

Compiled from Department of National Health and Welfare, Research and Statistics Memos, (Ottawa: Research and Statistics Directorate).

1957 to 1967 data from "Expenditures on Personal Health Care in Canada 1957-1969", (1970), Table 1.

1968 to 1970 data from "Expenditures on Personal Health Care in Canada 1960-1970", (1971), Table 1.

TABLE I-2

PER CAPITA EXPENDITURE ON PERSONAL HEALTH CARE IN CANADA, 1957 TO 1970

Year	Hospital Services: General and Allied Special	Hospital Services: All Hospitals	Total Personal Health Care
1957	25.36	35.22	62.81
1958	27.00	37.42	66.87
1959	31.03	41.98	73.65
1960	35.77	47.33	80.59
1961	39.52	51.94	86.89
1962	43.61	56.63	92.55
1963	47.97	61.95	101.34
1964	52.53	67.28	109.42
1965	58.16	74.29	120.29
1966	65.79	83.23	132.92
1967	74.51	93.75	148.19
1968	86.17	106.80	167.46
1969[a]	94.72	116.84	183.76
1970[a]	106.45	130.21	204.94

[a]Preliminary estimates.

Compiled from Department of National Health and Welfare, Research and Statistics Memos, (Ottawa: Research and Statistics Directorate).

1957 to 1967 data from "Expenditures on Personal Health Care in Canada, 1957-1969", (1970), Table 3.

1968 to 1970 data from "Expenditures on Personal Health Care in Canada, 1960-1970", (1971), Table 3.

or an average annual rate of increase of 13 per cent. During this period the pattern of spending on health care has shifted. Whereas in 1957 56.1 per cent of the total expenditure on personal health care was for hospital services, it is estimated that in 1970 the expenditure for hospital services represented 63.5 per cent of the total. During the 1957 - 1970 period expenditures per person on hospital services rose from $35.22 to $130.21 or 369.7 per cent.

The increased expenditures for personal health care and for hospital services consume a growing proportion of Canada's resources. This trend is illustrated in Table I-3. Total expenditures on personal health care amounted to 3.18 per cent of G.N.P.[3] in 1957 and 5.19 per cent in 1970; - an increase of 163.2 per cent. Expenditures on hospital services represented 1.78 per cent of G.N.P. in 1957 and 3.30 per cent in 1970 - an increase of 191.0 per cent.

Provincial Trends

The rising trend of expenditures for personal health care is one which has been experienced in every Canadian province between 1957 and 1970 (see Table I-4).

The Province of Alberta has not remained isolated from this trend of increasing expenditures for personal health care and for hospital services. (Refer to Tables I-5 and I-6). Between 1957 and

---

[3] G.N.P. is a number which represents the total current market value of all final goods and services produced within the nation in a given year.

TABLE I-3

EXPENDITURE ON PERSONAL HEALTH CARE, CANADA,
1957 TO 1970, PERCENTAGE OF GROSS
NATIONAL PRODUCT AT MARKET PRICES

Year	General and Allied Special Hospitals	All Hospitals	Total Personal Health Care	G.N.P. ($,000)
1957	1.29	1.78	3.18	32,907
1958	1.36	1.88	3.36	34,094
1959	1.50	2.03	3.56	36,266
1960	1.70	2.24	3.82	37,775
1961	1.85	2.43	4.06	39,080
1962	1.92	2.49	4.07	42,353
1963	2.00	2.58	4.23	45,465
1964	2.04	2.61	4.25	49,783
1965	2.08	2.66	4.31	54,897
1966	2.15	2.72	4.34	61,421
1967	2.32	2.92	4.62	65,608
1968	2.51	3.11	4.87	71,388
1969	2.54	3.14	4.93	78,560
1970	2.70	3.30	5.19	84,468

Compiled from Department of National Health and Welfare, Research and Statistics Memos, (Ottawa: Research and Statistics Directorate).

1957 to 1967 data from "Expenditures on Personal Health Care in Canada, 1957 to 1969", (1970), (a) Table 4, (b) Table 8.

1968 to 1970 data from "Expenditures on Personal Health Care in Canada, 1960 to 1970", (1971), (a) Table 4, (b) Table 45.

TABLE I-4

EXPENDITURE ON PERSONAL HEALTH CARE, CANADA AND PROVINCES, 1961 - 1970, PERCENTAGE INCREASE OVER PRECEDING YEAR

Province	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
Newfoundland	10.5	8.6	10.3	11.5	15.8	11.4	20.6	21.1	12.1	10.4
Prince Edward Island	12.9	4.6	9.4	7.6	9.8	10.5	13.8	8.0	12.3	11.7
Nova Scotia	12.3	6.8	8.1	12.3	9.4	10.9	14.1	10.8	9.9	16.4
New Brunswick	7.7	6.5	9.9	6.9	8.2	6.4	13.4	13.0	7.6	5.9
Quebec	13.4	12.5	14.9	13.1	16.8	17.5	9.7	12.5	9.8	12.3
Ontario	10.1	8.8	11.5	10.1	11.2	10.1	16.5	17.0	12.0	15.2
Manitoba	11.9	6.5	8.0	5.1	8.9	8.8	9.6	14.1	13.3	14.0
Saskatchewan	4.1	0.1	17.2	7.3	5.7	7.9	7.7	8.7	6.8	8.1
Alberta	9.7	10.0	8.7	7.9	10.5	14.6	16.7	18.0	10.4	12.7
British Columbia	5.9	4.1	6.8	7.9	9.0	13.8	16.0	14.6	15.4	12.9
Territories	26.8	6.5	3.0	1.3	11.4	10.5	10.2	5.5	14.4	11.7
CANADA	10.2	8.5	11.6	10.0	11.9	12.6	13.7	14.8	11.4	13.2

Source: Department of National Health and Welfare, "Expenditures on Personal Health Care in Canada, 1960-1970", Research and Statistics Memos, (Ottawa: Research and Statistics Directorate, 1971), Table 25.

1970 expenditures in Alberta for personal health care have risen from $82,863,000 to an estimated $346,732,000;- an increase of 418.4 per cent. Between 1965 and 1970, Alberta experienced an average annual increase in expenditure of 13.8 per cent which was the second highest rate in Canada of any of the provinces.

The population of Alberta rose to 1,604,000 in 1970 from 1,144,000 in 1957 which represents an increase of 140.2 per cent. During this period the expenditure per person on personal health care rose from $70.88 to $216.17 or 305 per cent.

Alberta spent the largest proportion of its health care dollars on hospital services, which reflects the national experience. Alberta spent $47,126,000 for hospital services in 1957 and $211,301,000 in 1970;- an increase of 448.4 per cent. Between 1957 and 1970 expenditure per person for hospital services in Alberta rose from $40.31 to $131.73 which is an increase of 326.7 per cent.

Between 1957 and 1970 total spending on personal health care in Alberta increased at about the same rate as for the nation as a whole; 418 per cent. The total increase experienced during the same period in hospital services for all of Canada exceeded the increase experienced by Alberta. While Alberta did spend more per person in 1970 for both health care and hospital services than was expended nationally, over the 1957 - 1970 period, the rates of increase in Alberta's per capita expenditure for both health care and hospital services has been

TABLE I-5

EXPENDITURE[a] ON PERSONAL HEALTH CARE IN ALBERTA, 1957 TO 1970

Year	Hospital Services: General and Allied Special	Hospital Services: All Hospitals	Total Personal Health Care
1957	34,532	47,126	82,863
1958	38,379	52,927	92,532
1959	43,593	59,899	102,521
1960	48,994	65,913	112,911
1961	55,246	74,465	123,899
1962	62,529	81,728	136,242
1963	69,479	89,953	148,143
1964	76,519	98,337	159,828
1965	84,163	107,772	176,680
1966	99,351	127,137	202,418
1967	118,144	148,552	236,201
1968	137,374	170,754	278,743
1969[b]	153,032	189,604	307,596
1970[b]	173,366	211,301	346,732

[a]In Thousands of dollars.

[b]Preliminary estimates.

Compiled from Department of National Health and Welfare, Research and Statistics Memos, (Ottawa: Research and Statistics Directorate).

1957 to 1967 data from "Expenditures on Personal Health Care in the Provinces of Canada, 1957-1969", (1970), Table A9.

1968 to 1970 data from "Expenditures on Personal Health Care in Canada, 1960-1970", (1971), Table 14.

TABLE I-6

PER CAPITA EXPENDITURE ON PERSONAL HEALTH CARE IN ALBERTA, 1957 TO 1970

Year	Hospital Services: General and Allied Special	Hospital Services: All Hospitals	Total Personal Health Care
1957	29.54	40.31	70.88
1958	31.69	43.71	76.41
1959	34.82	47.84	81.89
1960	37.80	50.86	87.12
1961	41.38	55.78	92.81
1962	45.54	59.53	99.23
1963	49.38	63.93	105.29
1964	53.47	68.72	111.69
1965	58.00	74.27	121.76
1966	67.82	86.78	138.17
1967	79.13	99.50	158.21
1968	89.85	111.68	182.30
1969[a]	97.91	121.31	196.80
1970[a]	108.08	131.73	216.17

[a]Preliminary estimates.

Compiled from Department of National Health and Welfare, Research and Statistics Memos, (Ottawa: Research and Statistics Directorate).

1957 to 1967 data from, "Expenditure on Personal Health Care in the Provinces of Canada, 1957-1969", (1970), Table A20.

1968 to 1970 data from, "Expenditure on Personal Health Care in Canada 1960-1970", (1971), Tables 26, 30 and 34.

considerably less than those for Canada as a whole.

Reaction to the Increasing Expenditures

During 1970, the Economic Council of Canada estimated that if the increases in expenditures for health care and education continued into the future at the same rates as during the period 1965 to 1970, that before the year 2000, "these two areas of activity alone would absorb the entire potential national product".[4]

Increases in health costs have aroused a considerable degree of concern in the governments of Canada, especially at the provincial level. Generally, expenditures for health care represent the second largest segment of a provincial government's budget after education. In Alberta the provincial government allocated approximately 25 per cent of its budget for the provision of health services in 1971.[5] The Government of Alberta, as have other provincial governments, has opted for a policy of cost control in the provision of health services in an effort to restrain the annual rate of increase in health care expenditures.[6]

---

[4] Economic Council of Canada, Seventh Annual Report, P. 38.

[5] J.D. Henderson, Health Care Delivery in Alberta: Future Prospects, a paper presented to the Health Services Administration Program, University of Alberta, (January, 1971), P. 2.

[6] J.D. Henderson, Health Care Delivery in Alberta, P. 2.

The concern of government about rapidly increasing expenditures in health arises from the fact that provincial and national resources are not limitless. Casual observation indicates that the proportion of its resources which a nation is prepared to allocate to the provision of health services appears to be a great deal less than 100 per cent. It may now be that the current government interest in rising health costs indicates that health expenditures are beginning to approach their maximum acceptable level in proportion to the nation's total current resources. Perhaps future increases in health care expenditures may be only as great as any increase in the nation's total resources. It may now be the case that, "new or expanded services can only be anticipated as the result of greater efficiency within the overall (health) system."[7]

Government Actions

The rapidly rising cost of health and hospital care is stimulating a great deal of governmental and academic study and analysis of the problem. In November, 1968, the Conference of Ministers of Health of Canada established the Committee on Costs of Health Services. The Committee appointed seven task forces composed of people from the health field to investigate specific problem areas. Four task forces reviewed various aspects of the costs of hospital services and three reviewed medical services. The reports of the task forces are summa-

---

[7] J.D. Henderson, Health Care Delivery in Alberta, P.3.

rized in the Committee Report of November, 1969.[8]

The Task Force Report represents the first attempt in Canada to discuss the problem of rising health and hospital costs. It has been observed about the reports that: "They identify a great many specific sources of industry malfunction. Unfortunately the economic innocence of the Report disables it from developing any consistent theory or explanation as to why the decision-makers in the (health) industry behave as they do and thus greatly reduces its ability to recommend effective improvements".[9]

The Provinces of Ontario and Quebec have both recently conducted reviews and analyses of their total health service systems with a view to ensuring the delivery of comprehensive health services at reasonable costs.[10] The Ontario report includes economic analyses of various parts of the health care sector[11] but neither report specifically discusses the economic factors operative in the hospital sub-sector.

---

[8] The Task Force Reports on The Cost of Health Services in Canada, (Ottawa: Queen's Printer, 1969).

[9] R.G. Evans and W.D. Walker, Public Policy Problems in the Canadian Health Services Industry, (Vancouver: U.B.C. Department of Economics, 1970), P.3.

[10] Report of the Ontario Committee on the Healing Arts, (Toronto: Queen's Printer, 1970); and Report of the Commission of Inquiry on Health and Social Welfare, Part 2, Vol.IV, "Health", (Quebec City: Quebec Official Publisher, 1970).

[11] R.D. Fraser, Selected Economic Aspects of the Health Care Sector in Ontario, A Study for the Committee on the Healing Arts, (Toronto: Queen's Printer, 1970).

In Alberta the government has demonstrated its concern over the rapidly escalating costs of health by merging the Departments of Health and Social Development so as to achieve a co-ordinated and integrated government administration of these areas. The Alberta Hospital Services Commission was created in order to plan and administer the hospital system in the province. One of its main functions is to retard the rapid growth rate of government expenditures made to provide hospital services.[12]

The health care sector (and the hospital sub-sector) of the economy has only recently interested economists as an area for empirical study. During the past decade the number of economic studies in the health sector has substantially increased. Academic interest developed as health expenditures annually consumed more of the nation's resources and governments became more determined to interject economic considerations into the process by which resources are allocated in the health sector.

The Purpose of the Thesis

This thesis attempts to shed light upon the question: Does there exist a unique hospital size which can be said to be optimal? The purpose of this thesis is to consider the pertinent, available economic literature related to this question and to evaluate it with

[12] Dr. J. Bradley, Chairman, Alberta Hospital Services Commission, private interview, Edmonton, August 1971.

regard to its behavioural assumptions, methods and results. Finally, the portent for public policy of the conclusions in the literature shall be considered.

Format

Chapter II consists of a discussion of the theories of production and cost. Chapter III surveys the economic literature dealing with cost relationships in general acute hospitals. In Chapter IV an analytical discussion of the literature is presented. The final chapter, Chapter V, contains a brief summary, some observations and conclusions as well as recommendations for future research directions.

CHAPTER II

THE THEORIES OF PRODUCTION AND COST

This Chapter consists of a brief review of the Theories of Production and Cost.

The Theory of Production

The term 'production' in economics refers to any process which changes a commodity or commodities into a different commodity.[1] Two commodities are different if they are not regarded by consumers as being completely interchangeable. A production process generally requires more than one input. These are usually of different qualities depending upon the nature of the final good or service.

The Production Function

A production function refers to, "the physical relation between a firm's inputs of resource and its output of goods or services per unit of time, leaving prices aside".[2] It is defined as "a schedule

---

[1]Kelvin Lancaster, Introduction to Microeconomics, (Chicago: Rand McNally & Co., 1964), P.59.

[2]Richard H. Leftwich, The Price System and Resource Allocation, (Hinsdale, III.: Dryden Press Inc., 1970), P. 116.

(or table or mathematical equation) showing the maximum amount of output that can be produced from any specific set of inputs, given the existing technology."[3]

A production process may be fully expressed by the production function of the form:

$$Q = f\ (K,\ L,\ M)$$

where Q represents output, K all capital inputs, L all labour inputs and M all other inputs. There are a variety of differentiated inputs included within each of these three classifications. Therefore, the general form of the production function can be expressed:

$$Q = f\ (X_1, X_2, X_3, \ldots, X_n)$$

Q is the output and $X_1, \ldots, X_n$ refers to all of the various inputs.

It is improbable that in the real world any production process exists in which there is only one input as in the case:

$$Q = f\ (L)$$

Most production functions require a variety of inputs. It is highly unlikely that all of the inputs will vary simultaneously and in like proportions. Inputs generally vary at different rates over time. For example, it is usually faster to increase the labour input of a production process than it is to construct a new physical plant when an increase in output is required. Within some lengthy period of

---

[3] C.E. Ferguson, Microeconomic Theory, (Homewood Ill.: Richard D. Irwin, Inc., 1969), P.118.

time, however, it may be expected that all inputs will vary; - no matter how unalterable they may appear to be. That period of time within which every input in a particular production process is variable is referred to as the 'long-run'. Any period of time within which one or more inputs is fixed - (does not vary) - is referred to as the short-run.[4]

The Short-Run

In planning the short-run operation of most production processes, planners face a set of production curves similar in shape to those depicted in Figure 1. These curves represent the output produced when there are two or more inputs but only one of these factors of production is allowed to vary within the period under consideration; the other inputs remain constant. The production function in this situation is:

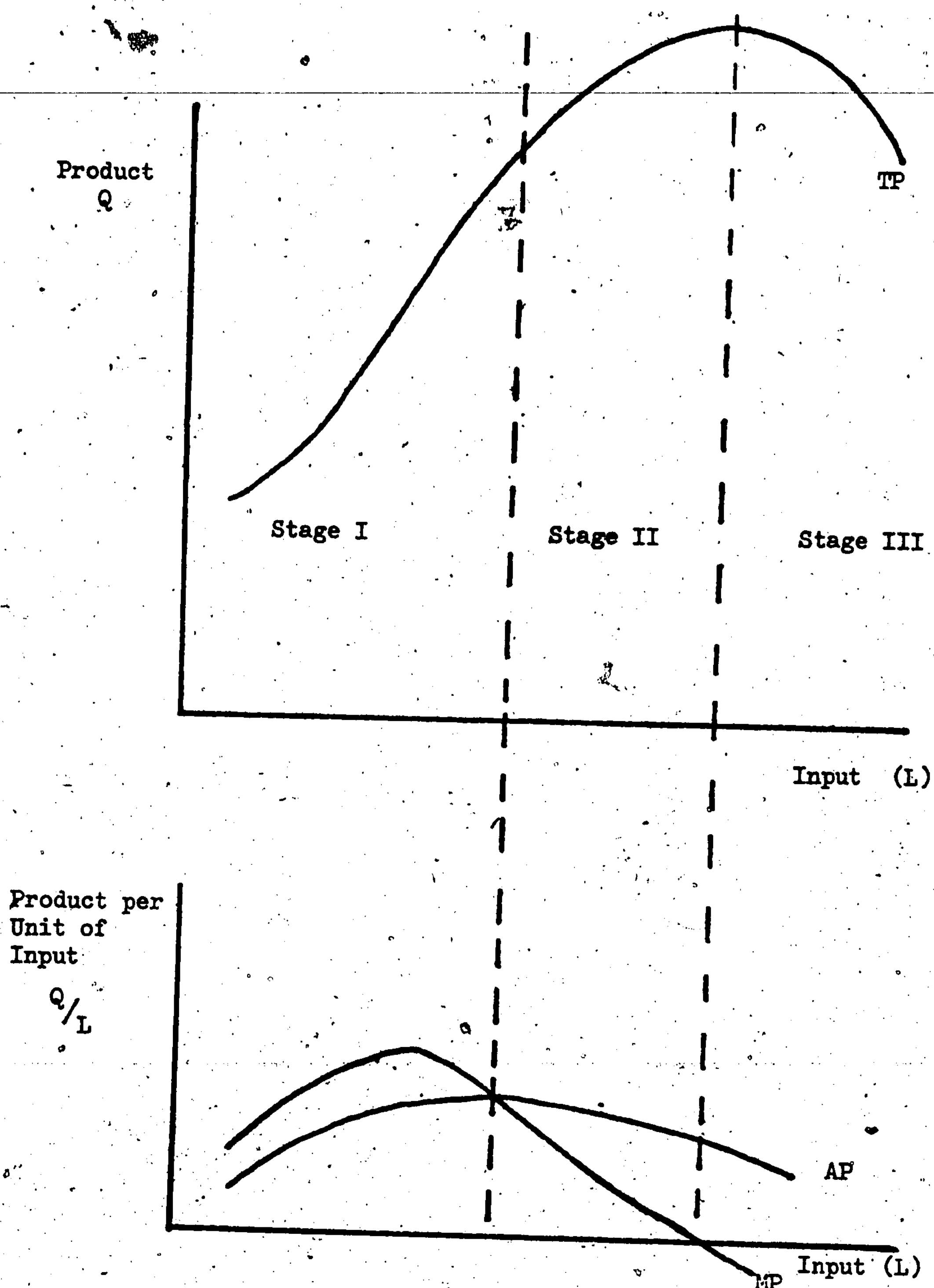
$$Q = f\ (L,\ K = \overline{K},\ M = \overline{M})$$

where capital (K) and the remaining inputs (M) are considered to be constant. This production function reflects the economic 'law' of Diminishing Marginal Returns which is an axiom that states that: "If equal increments of an input are added, the quantities of other inputs held constant, the resulting increments to product will decrease beyond some point: that is the marginal product of the input

[4]Edwin Mansfield, Microeconomics, (New York: W.W. Norton & Co. Inc., 1970), P. 117.

FIGURE 1

PRODUCT CURVES AND THE STAGES OF PRODUCTION



will diminish."[5] (The law assumes that the level of technology does not change and that there are at least two inputs.)

Figure 1 illustrates the short-run situation. It shows the total product initially increases rapidly as increasing amounts of the variable input are combined with the fixed inputs. Eventually, the law of diminishing marginal returns begins to take effect and the total product begins to increase at a declining rate, finally decreasing as more units of the variable input are added. Figure 1 illustrates the relationship between the total, average and marginal products of the variable input. The marginal product of an input is always equal to its average product at the maximum level attained by the average product. When the marginal product of an input becomes less than zero, then the total product begins to decrease.

The average product represents the average amount of output derived from using a certain amount of variable input in the production process. It is calculated by dividing the total product by the number of units of input required to produce it ([illegible] $= \frac{Q}{L}$). The marginal product represents the change in output which results from a change in the variable input. It is calculated by dividing the change in the total product by the associated change in the variable input ($MP = \frac{\Delta Q}{\Delta L}$).

The Stages of Production

"The relations among total, average and marginal products are

[5] *Ibid.*, P. 123.

used to define three stages of production."[6] Figure 1 illustrates these three stages.

In stage I the average return to a unit of variable input (i.e. average product) increases as more units of the variable input are added.[7] The average product reaches its maximum at the level of production represented by the division line between stages I and II. A rational producer would not operate in stage I because the average product of each additional unit of the variable input is greater than the one before.

In stage III, if additional units of the variable input are added, the result is a decrease in the total amount of output produced. The variable input is combined in uneconomically large proportions with the fixed input(s). In stage III the marginal product of the variable input is negative and total product is declining. Even if the cost of the variable input(s) was zero, the rational producer would not operate anywhere in stage III because to do so would result in declining output.[8]

The rational producer in the short-run operates somewhere within stage II; between the point of maximum average product and zero marginal product of the variable input. Before one can predict the specific level at which a production process will operate in the short-run, the costs of the various input factors need to be known.

---

[6] Ferguson, Microeconomic Theory, P. 130.

[7] Ibid.

[8] Ibid., P. 131.

The Long-Run

In the long-run the shapes of the product curves are determined by the effects of Economies and Diseconomies of Scale. In the long-run every input into a production process can be varied and the producer may arrange them in optimal proportions.

Isoquants

"An isoquant is a curve showing all possible (efficient) combinations of inputs that are capable of producing a certain quantity of output".[9] A number of isoquants are shown in Figure 2 and each of q_1, q_2, q_3 and q_4 represents a different output rate. Figure 2 shows the different combinations of two resources, labour and capital, with which a firm can produce equal amounts of output. For example, points along isoquant q_1 show the combinations of labour and capital, which will produce q_1 units of output. Production will be q_1 units both when K_1 units of capital and L_1 units of labour are used and also when K_2 units and L_2 units of capital and labour are used. A greater amount of output is represented by a higher isoquant such as q_4.

In isoquant mapping it is assumed that the most efficient technology available is used in the production process. Two isoquants, therefore, will not intersect. If two isoquants were to intersect it would mean that two different quantities of output could be produced by the same combination of resources, which is impossible under the assumption.[10]

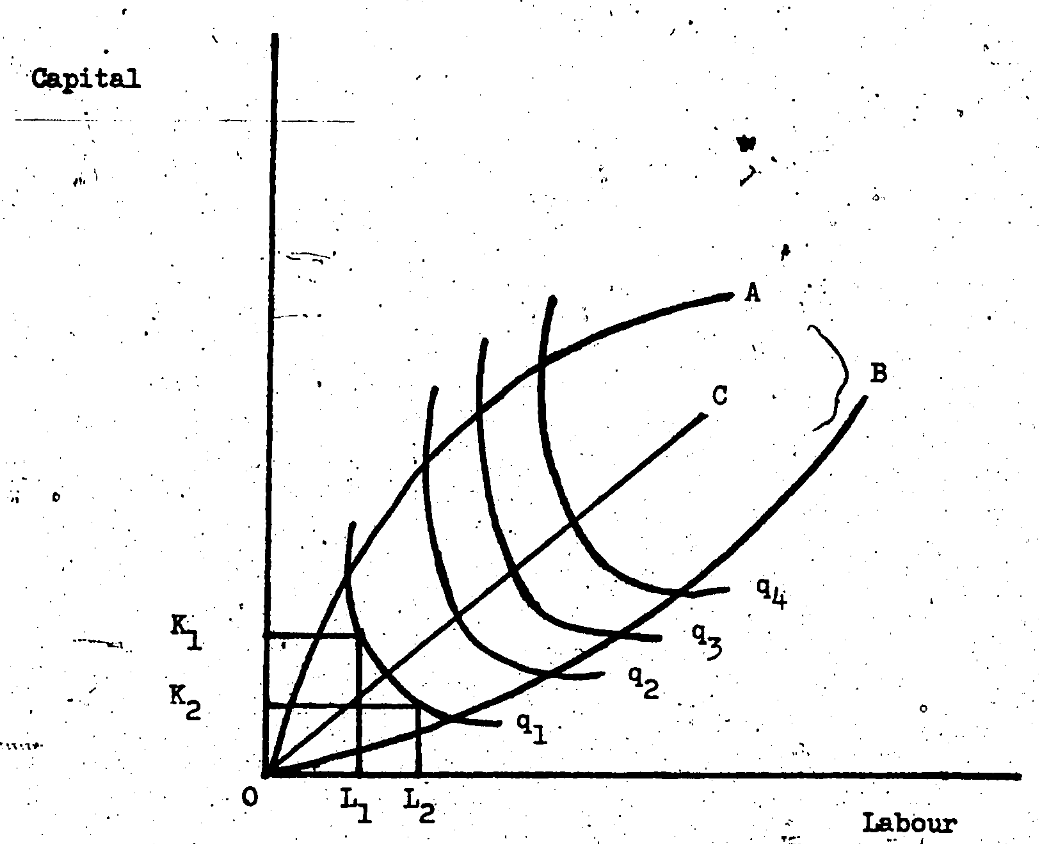
---

[9]Mansfield, Microeconomic Theory, P. 132.

[10]Richard H. Leftwich, The Price System and Resource Allocation, (New York: Holt, Rinehart & Winston, 1961), P. 128.

FIGURE 2

FULL ISOQUANT MAP AND THE RELEVANT RANGE OF PRODUCTION



Isoquants slope downward to the right and they are generally convex to the origin.[11] The downward slope to the right results because resource inputs can be technical substitutes for one another in a production process. When resources are technical substitutes, and less of one is used, then more of the other input must be used to compensate for the loss and so maintain the level of production. Isoquants tend to be convex because resources are generally not perfect technical substitutes for one another. For example, the more labour and the less capital used in a production process, the more difficult it becomes to substitute additional labour for capital. This reflects the Principle of the Diminishing Marginal Rate of Technical Substitution.

In Figure 2 the Line OC is a 'ray'. Such a ray from the origin defines a constant ratio of the inputs for a production process.[12] In moving along the ray from O towards C, the level of output changes but the input ratio remains constant.

The Economic Region of Production

Generally, isoquants possess negative slopes but they may possess positively sloped segments. In Figure 2 the segments of the isoquants above OA and below OB have positive slopes which indicate that increases in both capital and labour are necessary to maintain a certain level of production. A rational producer would not operate in these two regions because the same output can be produced with less of both inputs and so

[11] *Ibid.*

[12] Ferguson, *Microeconomic Theory*, P. 153.

for less cost. The zone between OA and OB has been referred to as the "economic region of production"[13] and it corresponds to stage II of the three stages of production, discussed above.

Technology and Technological Change

Technology has been defined as "Society's pool of knowledge regarding the industrial arts."[14] The level of technology available in a society sets a limit upon the amounts and the types of commodities which can be produced from a given amount of resources. Part I of Figure 3 depicts the limit which the level of technology imposes on a production process. Isoquant B in Part I illustrates the maximum level of output that a particular production process may achieve, no matter how the combinations of labour and capital are varied when there is no change in the technology employed in the process and the total amount of resource inputs used does not change. The higher level of production represented by isoquant C is unobtainable unless there is a change in the level of technology or an increase in the amount of capital and/or labour utilized in the production process.

". . . One of the most important ways that economic welfare is increased is through the alteration of production functions due to technological change".[15] Technological change results from advances in knowledge. It takes such forms as:

a) new methods of producing existing products;

---

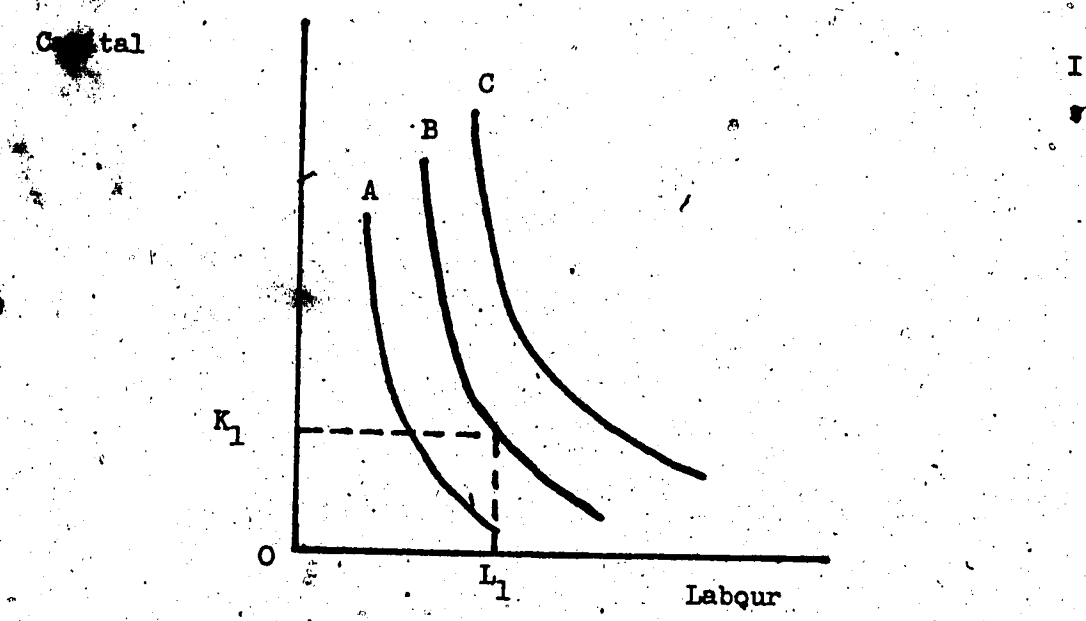
[13]Mansfield, Microeconomics, P. 135.

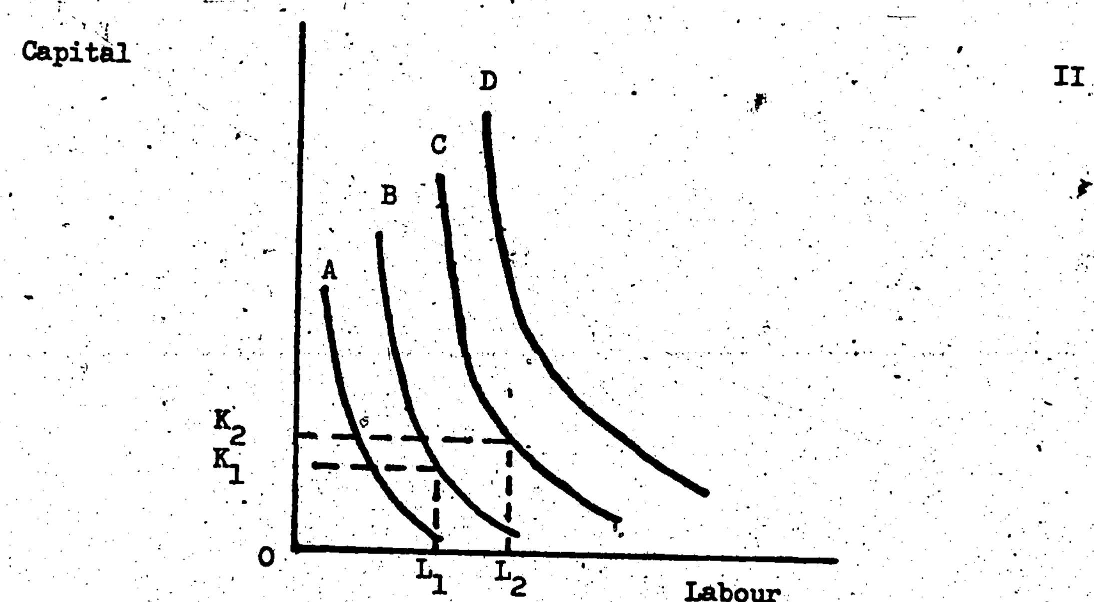
[14]Ibid., P. 10.

[15]Ibid., P. 440.

FIGURE 3

NEUTRAL TECHNOLOGICAL CHANGE





b) new techniques of organization and management; and

c) new products[16]

The effects of technological change (otherwise referred to as increases in the level of technology) are illustrated by comparing Parts I and II of Figure 3. Part I shows the level of production possible before a change in technology is developed. Isoquant B represents the highest level of output which can be obtained from the capital and labour combinations given the constraints of technology and limited resources. Part II demonstrates the effect of technological change upon the levels of production. The isoquants have shifted toward the origin. Isoquant B in both Parts I and II represents an identical level of production, however, in Part II, after technological change has occurred, less input of resources is required to achieve isoquant B. If the identical amount of combined resources continued to be expended in production after the increase occurred in the level of technology, then total possible output would rise to isoquant C. Isoquant C represents the new maximum level of output possible from the given resources at the new level of technology.

The technological change depicted in Figure 3 is a neutral one. It is described as neutral because the ratio of capital to labour remains unchanged where there is an improvement in the technology. In other words, the marginal products of both capital and labour have increased by the same percentage.[17] In Figure 3, Part II, the distance

---

[16] *Ibid.*, P. 441.

[17] Ferguson, *Microeconomic Theory*, P. 387.

between the isoquants remains the same throughout for each level of output.

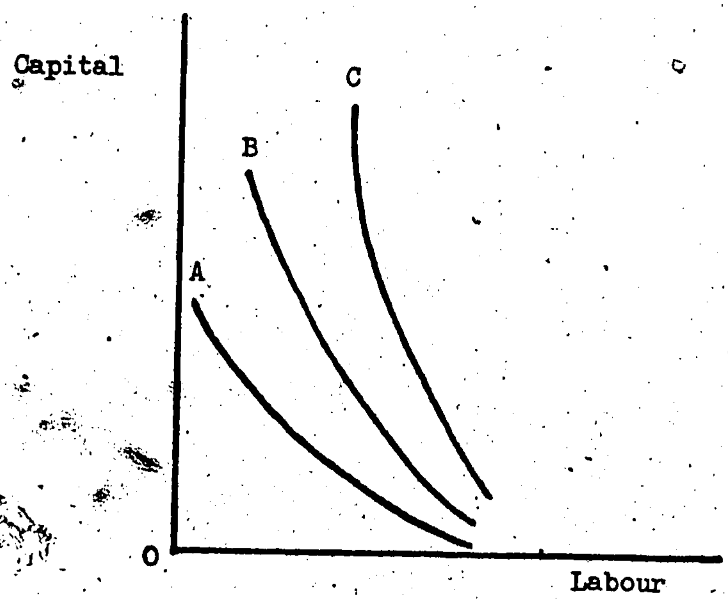
Technological change may be biased in that it may be described either as capital-using or labour-using.[18] The two parts of Figure 4 illustrate each of the cases of biased technological change. Capital-using technological change occurs when, at a constant capital-labour ratio the marginal product of capital increases relative to the marginal product of labour. Such an instance may be represented by the situation in which a mechanized and automated process is altered with the result that total output is increased. As Part I of Figure 4 shows, in the instance of capital-using technological change, the slope of the isoquants decreases as one moves toward the origin. This implies that there is a decline in the marginal rate of technical substitution; ($MRTS_{K \text{ for } L} = \frac{MP_L}{MP_K}$).

Similarly, Part II of Figure 4 illustrates the isoquants in cases of labour-using technological change. Improvement in plant layouts and the use of higher level management skills are examples of the factors which may contribute to an increase in the marginal product for labour relative to that for capital in instances of labour-using technological change. In this case the slope of the isoquants increase as one moves toward the origin and this reflects an increase in the marginal rate of technical substitution.

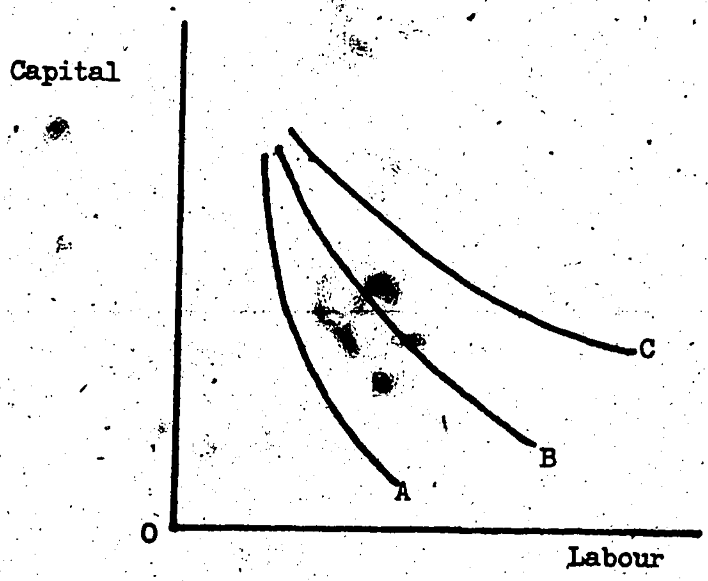
[18] Ibid.

FIGURE 4

BIASED TECHNOLOGICAL CHANGES



I Capital-Using Technological Change



II Labour-Using Technological Change

Change in the level of technology is an ongoing process. The development of new, more efficient and more desirable products is as continuous as the increase of productive capacity. Over the long run, isoquants tend to shift to the left as changes in the level of technology are implemented.

Returns to Scale

At a given level of technology the maximum level of output per unit of time which a firm can achieve is determined by the scale of the plant used in the process. "The quantities of fixed resources used determine the size of the firm's plant, or its scale of plant".[19] The scale of the plant can be altered in the long run. The change which occurs in the level of production as a result of change in the size of plant is referred to as a return to scale.

There are three types of returns to scale which may be encountered as the scale of the plant is altered. Figures 5, 6 and 7 depict the three types for production processes in which there are two resource inputs: capital and labour. If the amount of both inputs must be doubled in order to double the output then "constant returns to scale" are present. Since the quantity of output may be measured by the distance of an isoquant from its origin, the isoquants in Figure 5 illustrate the instance of constant returns to scale. Each of the isoquants for outputs of 50, 100 and 150 units intersects any ray from the origin, such as OA, at equal distances (OD = DC = CB).

If the amounts of the inputs required to double the output is less

[19] Leftwich, The Price System and Resource Allocation, (1961), P. 140.

FIGURE 5

ISOQUANT MAP ILLUSTRATING CONSTANT RETURNS TO SCALE

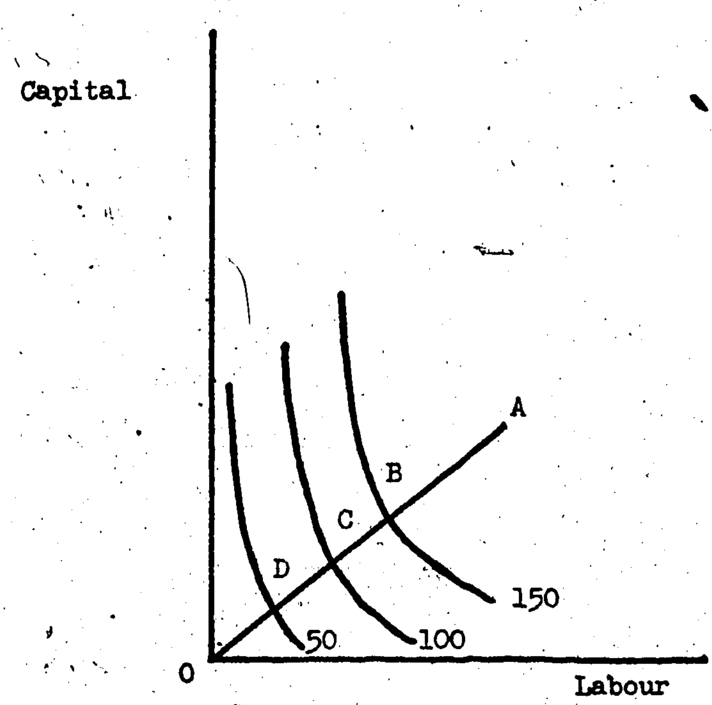
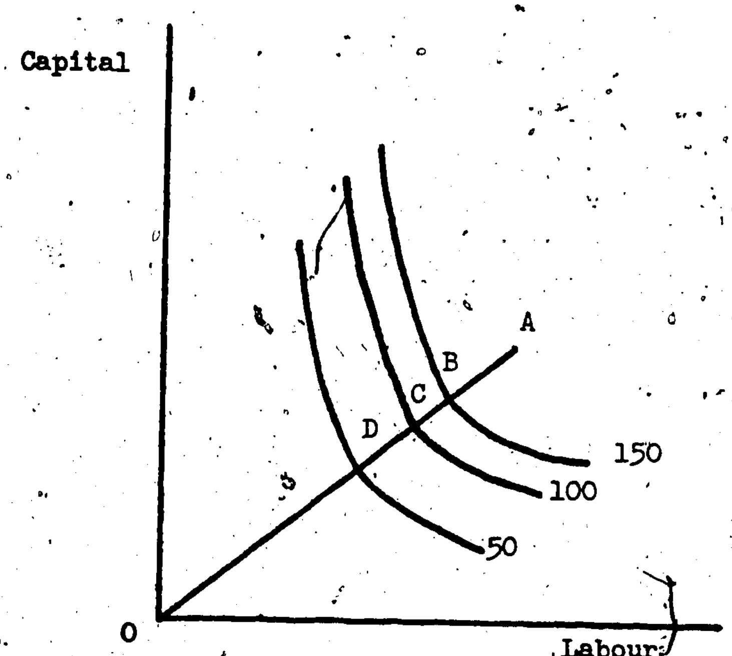


FIGURE 6

ISOQUANT MAP ILLUSTRATING INCREASING RETURNS TO SCALE



than double then 'increasing returns to scale' are present. Figure 6 depicts increasing returns to scale where successive isoquants become close together (OD>DC>CB). Finally, Figure 7 illustrates the case of 'decreasing returns to scale'. Decreasing returns are encountered if the amounts of both inputs must be more than doubled in order to double the output (OD<DC<CB).

Economies and Diseconomies of Scale

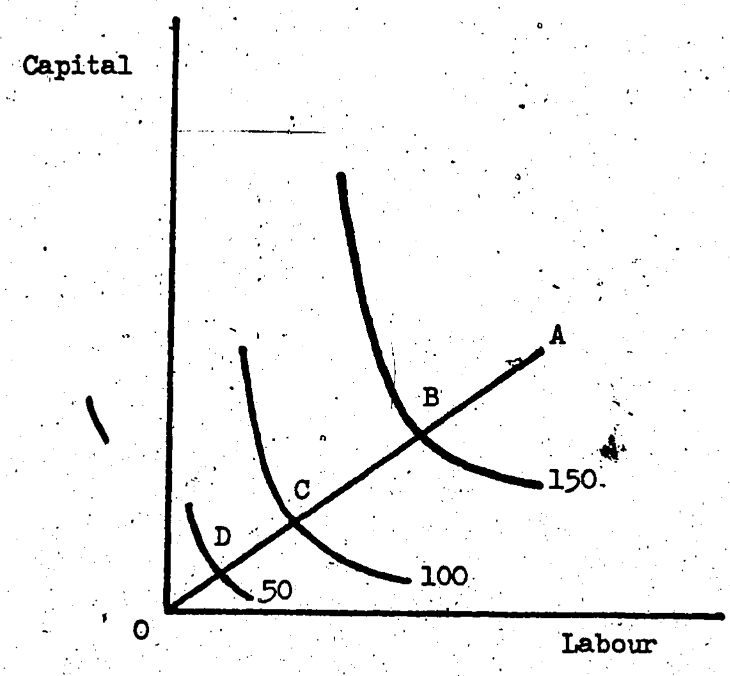
A production process generally encounters economies of scale in the long-run as it becomes larger. "That is, after adjusting all inputs optimally the unit cost of production can be reduced by increasing the size of the plant".[20] As scale is increased a number of efficiencies are encountered. There is increased specialization and division of labour and the productivity of labour increases accordingly. Machines which are interdependent in a process but which have different rates of output, may become more productive if increasing scale permits them to be arranged in optimal combinations. An increase in output may necessitate the acquisition of larger pieces of equipment which enable more output to be produced faster. Increased scale may also enable a production process to become more automated. Increasing scale enables more units of output to be produced from the inputs used.

The effects of economies of scale are operative over a range of plant sizes. The range varies between industries. At some point, however, in a production process inefficiencies begin to develop when

[20] Ferguson, Microeconomic Theory, P. 211.

FIGURE 7

ISOQUANT MAP ILLUSTRATING
DECREASING RETURNS TO SCALE



scale is increased. The inefficiencies result in diseconomies of scale. Initially, the effect of the diseconomies is to decrease the rate of increases in production. Eventually, diseconomies of scale cause output to actually decrease when the plant becomes very large. Communication breakdowns and increasing administrative problems are believed to be the cause of the diseconomies of scale often encountered in very large production processes.[21]

In an isoquant map showing the total long-run production process, it may be anticipated that near the origin the isoquants would get closer together as we moved out from the origin. Economies of scale would yield increasing returns, initially. Then farther out from the origin the intervals between the isoquant would become constant as the diseconomies equalled the economies of scale. Moving still farther out from the origin, increasing intervals between the isoquants would be encountered. Decreasing returns to scale would become predominant as the diseconomies became excessive.

---

[21] *Ibid.*, P. 212.

The Theory of Cost

The cost of production of any good or service includes more than just the total dollar outlay necessary to obtain the resources required as inputs for the production process.

The Nature of Cost

It has been suggested that there are two legitimate costs of production which, under ideal circumstances, are one and the same.[22] They are the 'social cost of production' and the 'private cost of production'.

The social cost of production includes those costs which are external to the production process and which are borne by society at large. It includes the cost of the resources utilized in the production process. Social costs include 'external' costs such as pollution, which may be by-products of production. The social cost also includes the alternative (or opportunity) cost of society's resources that are utilized in a particular production process. The opportunity cost of a production process is equal to the value of the goods or services which could be produced by using the same resources in the next

---

[22] Ibid., P. 185.

Ferguson suggests that the 'ideal circumstance' occurs when the individual firm or entrepreneur operates in such a manner that the economic choices made are those which society would also make. The individual firm or entrepreneur maximizes the net benefit accruing to society as a consequence of actions taken to soley maximize the net benefits for the firm or himself.

possible alternative production process.[23]

Society's supply of resources is finite and so society seeks to maximize the return it receives when its limited resources are expended. There are constraints which inhibit the ability of society to maximize its return. These are the constraints of imperfect knowledge regarding the future and also the variety of alternative uses to which the resources could be applied.

The private cost of production refers to the costs incurred by the individual entrepreneur in operating a production process.

The Cost Function

As with the production process which is best described by the production function, the cost of production can also be described by a function. The cost function is a schedule, table or mathematical equation which shows various relationships between the firm's costs and its rate of production. "The firm's production function and the prices it pays for inputs determine the firm's cost function."[24] The cost of production may be expressed mathematically by the general form:

$$C = f\,(Q, P_i) \qquad i = 1,2,3,\ldots n$$

where C represents the total cost of production, Q the firm's output and P_i is the price of the various inputs.

In using the cost function, the only costs considered are those presumed to be the payments that must be made to obtain the resource inputs necessary for the production process. The price of each input

---

[23]Lancaster, Introduction to Microeconomics, P. 94-95.

[24]Mansfield, Microeconomics, P. 159.

is assumed to be constant no matter what the quantity purchased. It is further assumed that in the long-run the inputs can be purchased as and when required in any quantity. These assumptions imply perfect competition exists in the input markets.[25]

It may be recalled that the production function assumes that any production process operates in the manner which is most efficient given the existing technology. From the cost point of view a related assumption is made. "If there are various ways of achieving the same outcome, but with different costs, all decision makers in the economy will always choose the least-cost method from among those known and available to them."[26] It should be noted that the most technically efficient manner of producing one unit of output in a given set of circumstances is not necessarily the least cost method. The rational firm or enterprise minimizes its costs at each level of production even if it employs a technology which is less than optimally efficient.

The Short-Run

When planning the short-run operation of a production process, most producers operate on cost curves similar to those depicted in Figures 8 and 9.

In the short-run the producer encounters fixed costs. Fixed costs represent the total unavoidable financial obligation of the producer during a given period. Fixed costs include such items as land, depreciation cost of the physical plant and,

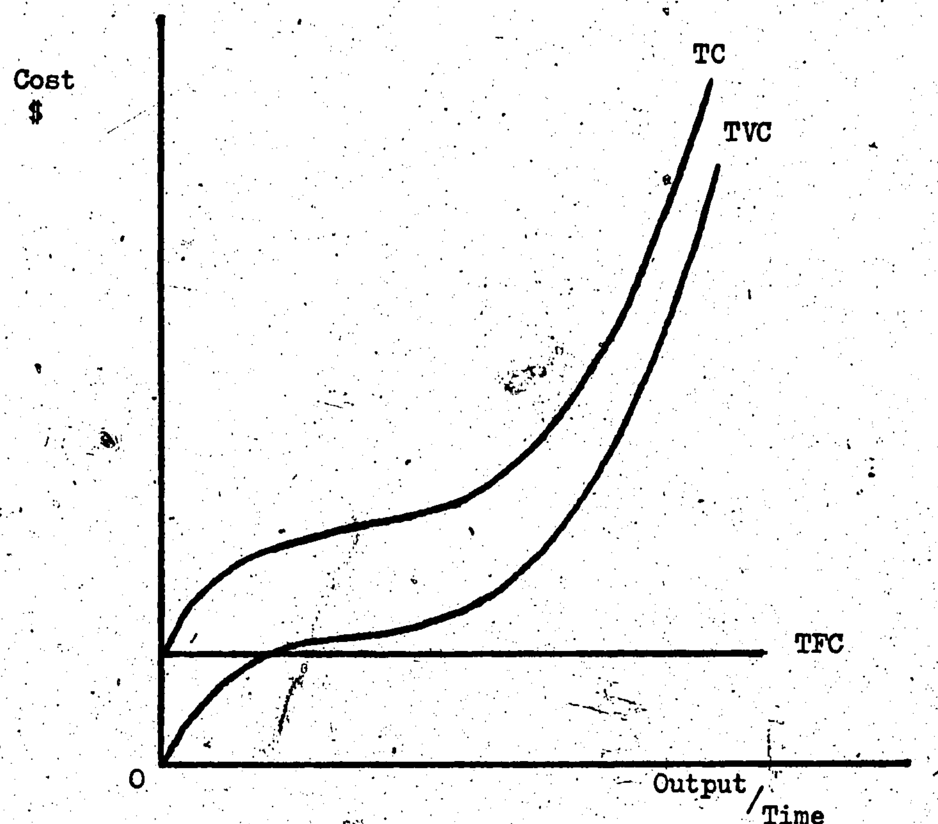
---

[25] Lancaster, Introduction to Microeconomics, P.95.

[26] Ibid, P.96.

FIGURE 8

SHORT-RUN TOTAL COST CURVES



frequently, the salaries of top management. In Figure 8, the total fixed cost curve is a straight line which is parallel to the horizontal axis. Fixed costs do not vary with the level of output within the short-run period.

The producer also encounters variable costs in the short-run. These costs increase as a producer increases his level of output because larger quantities of variable resources must be combined with the fixed resources to increase production. The shape of the total variable cost curve results from increasing and diminishing marginal returns of the variable resources. Initially, the amounts of the variable resources such as labour are very small in relation to the fixed resources and the efficiency of the production process is impaired. As more of the variable resources are added, efficiency increases and the rate of increase in variable cost lessens. Eventually, if it happens that very large amounts of variable resources are combined with the fixed resources, the law of diminishing marginal returns becomes operative. Variable costs begin to increase more rapidly as production increases. No matter what size a plant may be, production will reach full capacity at some point. Beyond this point output cannot be increased and the total variable cost curve begins to go straight up.

It is possible in a production process for the law of diminishing marginal returns to be in effect at every level of output. In such a case variable cost rises at an increasing rate because it is more costly to produce each subsequent unit of output. (i.e. Marginal cost is rising over the total range of output). The total variable cost curve in this situation rises continuously throughout its length and, eventually, it goes straight up when full capacity is reached.

If it should occur that there are constant marginal returns in a production process then variable cost will increase at a constant rate. (i.e. Marginal cost of production remains at one constant level). The total variable cost curve in this special instance will be a straight line which runs continuously upward and to the right from the origin.

Total cost in the short-run represents the sum of the fixed and variable costs for each level of output. The total cost curve and the total variable cost curve both have the same shape since an increase in output increases both variable cost and total cost by identical amounts.

Unit Cost Curves

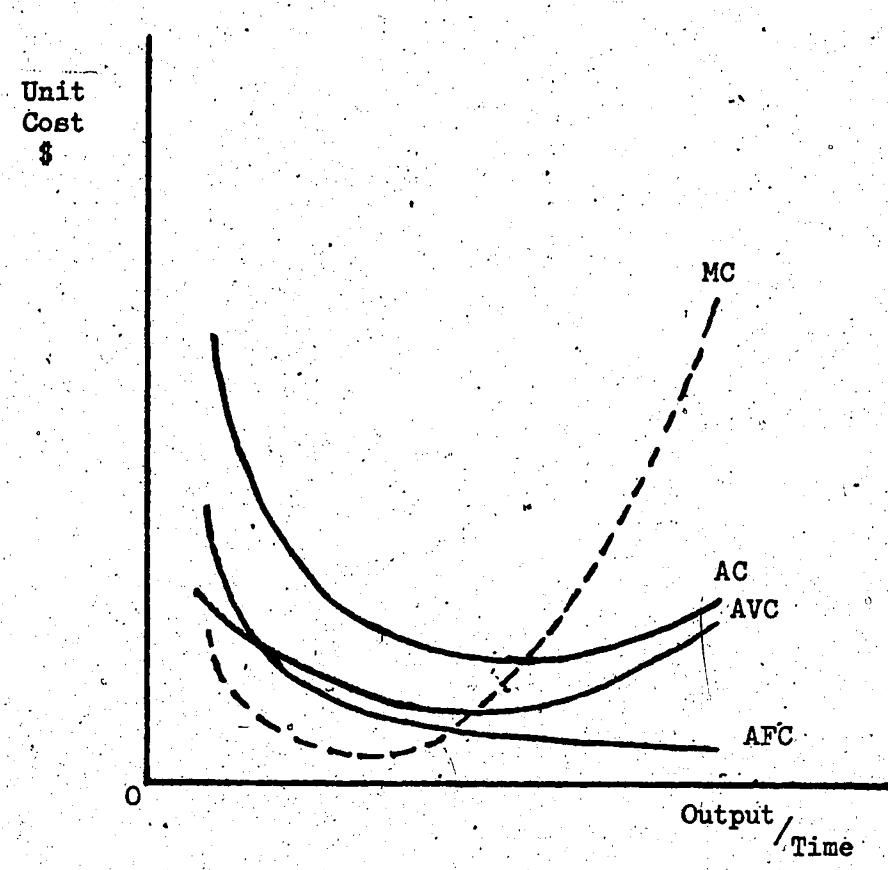
There are four types of unit cost curves which need to be considered in short-run planning. These curves are plotted in their general forms in Figure 9.

The average fixed cost for the various levels of output is obtained by dividing the total fixed cost by that output. The average fixed cost curve always declines as output increases because the total fixed cost is spread over more units of output. A producer may reduce his average fixed cost by producing more units of output.

The average variable cost is calculated by dividing the total variable cost for each level of output by that amount of output. Traditionally the average variable cost curve has a U-shape when plotted. This shape occurs because initially, when increasing amount of variable inputs are combined with the fixed inputs, increasing marginal returns

FIGURE 9

SHORT-RUN UNIT COST CURVES



are encountered and output increases faster than the rise in cost. As a result, the average variable cost curve declines initially as production rises. When diminishing marginal returns are encountered at higher levels of production and the cost increases begin to exceed the production increases more and more, the average variable cost ceases declining and begins to rise.

The average cost of production is calculated from the total cost in a manner similar to the way in which the average variable cost is calculated. Traditionally, the average cost curve is also depicted as being U-shaped for the same reasons that the average variable cost curve is U-shaped. The average variable cost is always less than the average cost because fixed costs are included in the average cost.

"Marginal cost is defined as the change in total cost resulting from a one-unit change in output".[27] The changes form a U-shaped curve when plotted, generally. The marginal cost curve bears a unique relationship to the average cost curve. MC is less than AC where AC is decreasing as output increases. MC is greater than AC where AC is increasing as output increases. Therefore, it follows that MC is equal to AC at that level of output where AC is a minimum.

The Optimum Rate of Output

"That level of output at which the short-run average cost is lowest is the output at which a given scale of plant is most efficient."[28] At the low point on the average cost curve the cost of

---

[27]Leftwich, The Price System and Resource Allocation, (1961), P. 148.

[28]Ibid., P. 152.

the resource inputs per unit of output is least and at this point the optimum rate of output occurs. There is a short-run optimal output level for every scale of plant wherever U-shaped average cost curves are present.

The shapes of the short-run cost curves reflect the efficiency with which resources are used by the producer.

The Long-Run

Isocost Curves

If Capital (K) and labour (L) are the two inputs in a production process, the various combinations of each which can be purchased may be represented by a straight line. The unit cost of each input and also the total dollar outlay allowed for the purchase of these inputs by the producer must be known.[29] The straight line is called an isocost curve. Four isocost curves are illustrated in Figure 10.

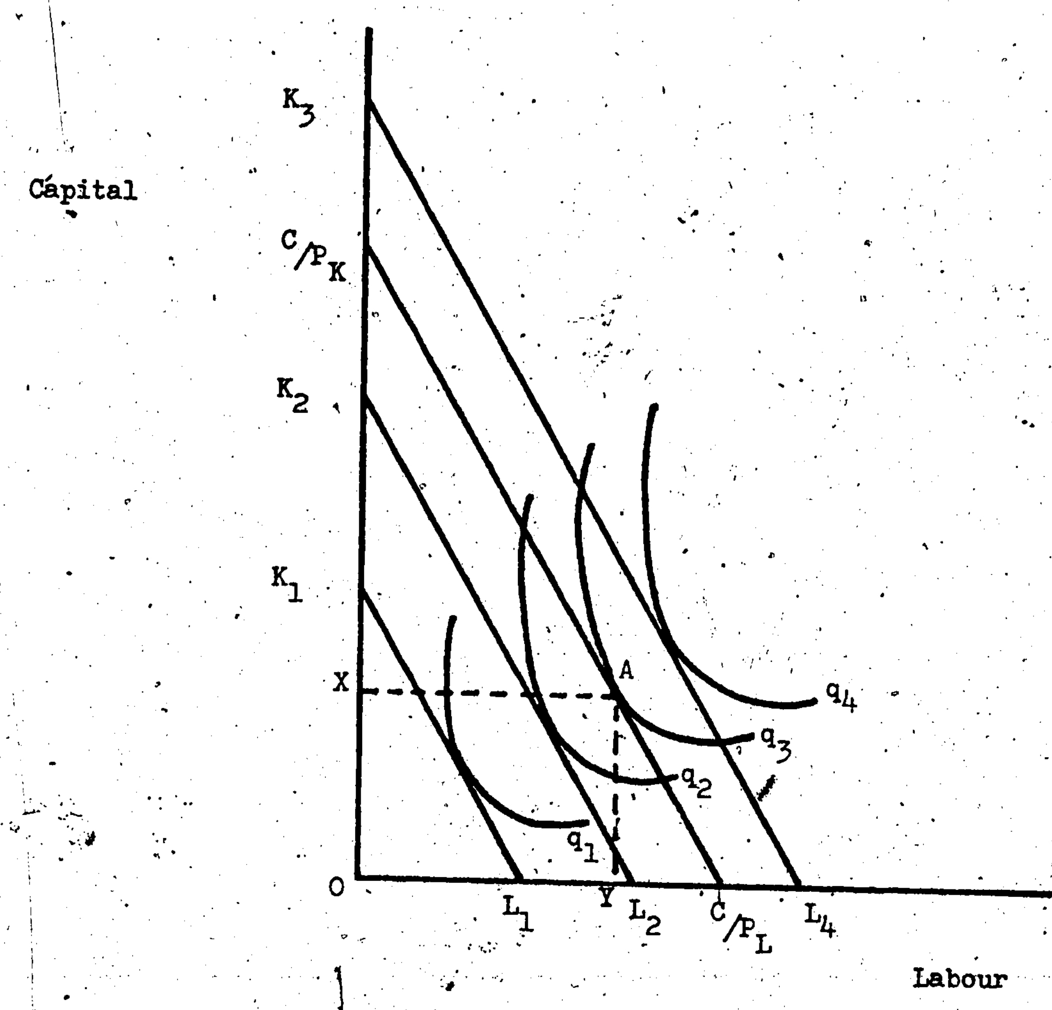
In Figure 10 for purpose of illustration it is assumed that a producer is prepared to spend only a finite amount on inputs. This amount may be represented by 'C'. If a producer were to only use capital in the production process, and P_K represents the unit cost of capital, he could purchase C/P_K units of capital. Similarly, if he only used labour he could only purchase a maximum of C/P_L units. Generally, the producer would use some combination of the two inputs and so his total expenditure could be expressed as follows:

$$C = P_L L + P_K K$$

[29] Ibid., P. 133.

FIGURE 10

COMBINED ISOQUANT AND ISOCOST MAP



The isocost curve which joins $^C/P_K$ and $^C/P_L$ shows all of the combinations of capital and labour which can be purchased for a total outlay of C dollars. K_1, L_1, K_2, L_2 and K_3, L_3 are isocost curves attainable at various other levels of total outlay.

The problem generally faced by a producer is that of getting the greatest amount of output from the total dollar outlay he is prepared to make for resources. In terms of isoquants and isocosts the producer desires to be on the highest possible isoquant which his isocost curve will permit. If $^C/P_K$, $^C/P_L$ is the maximum possible isocost curve, the producer will operate at Point A which represents the point where the isocost curve is tangent to the highest attainable isoquant. By using X units of capital and Y units of labour the producer maximizes his output from his total dollar outlay. To use any other combination would result in a decrease in total output because any movement along the maximum isocost curve away from point A will move the producer down to a lower isoquant. The producer operates so that the ratio of the marginal physical product of labour to the price of labour is equal to the ratio of the marginal physical product of capital to the price of capital.

$$\text{i.e. } \frac{MPP_L}{P_L} = \frac{MPP_K}{P_K}$$

This occurs at point A.

Under certain market conditions a producer may be restricted to producing only a certain amount of output. If a producer were restricted to being able to produce only q_3 units of output, he would attempt to operate at point A given the set of factor prices. For q_3 output point A represents the least cost combination of resources.

Long-Run Cost Curves

The long-run cost curves traditionally possess the general shapes as illustrated in Figure 11. Similarly, as with the production curves, the shapes of the long-run cost curves are determined by the effects of economies or diseconomies of scale. There are no fixed costs or average fixed costs in the long-run since all resources are considered to be variable. The long-run may be referred to as a 'planning horizon' because decisions which a producer makes for the long-run determine the short-run position that the production process will occupy in the future.[30]

Returns to Scale

Figure 12 depicts the shape and relationship of the cost curves of a production process which possesses constant returns to scale. Figure 13 illustrates the case of increasing returns to scale and Figure 14 the instance where decreasing returns to scale are in effect.

The Long-Run Average Cost Curve

The long-run average cost curve is commonly employed in analyses of the cost of production. The curve is traditionally considered to have a U-shape. It shows the minimum cost per unit of producing at each level of output where any desired scale of plant can be built.[31] The lowest point on the curve represents the lowest level of average

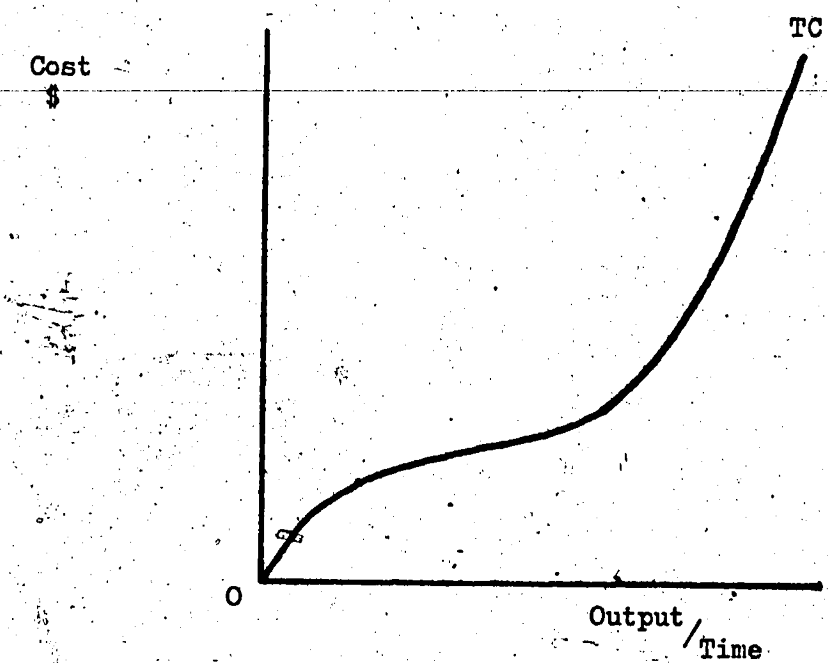
---

[30] Mansfield, Microeconomics, P. 173 and Ferguson, Microeconomic Theory, P. 198.

[31] Mansfield, Microeconomics, P. 173.

FIGURE 11

LONG-RUN COST CURVES



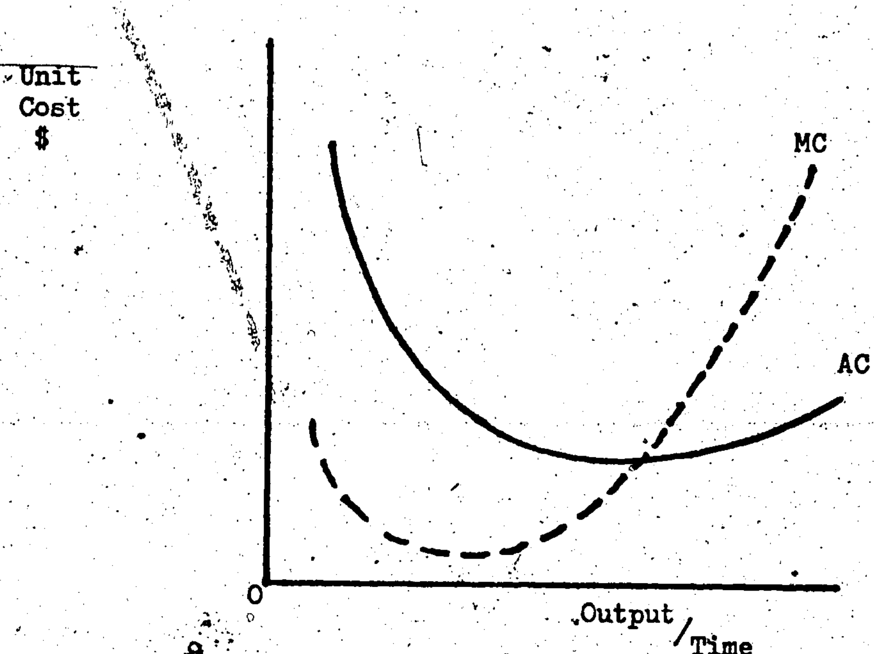
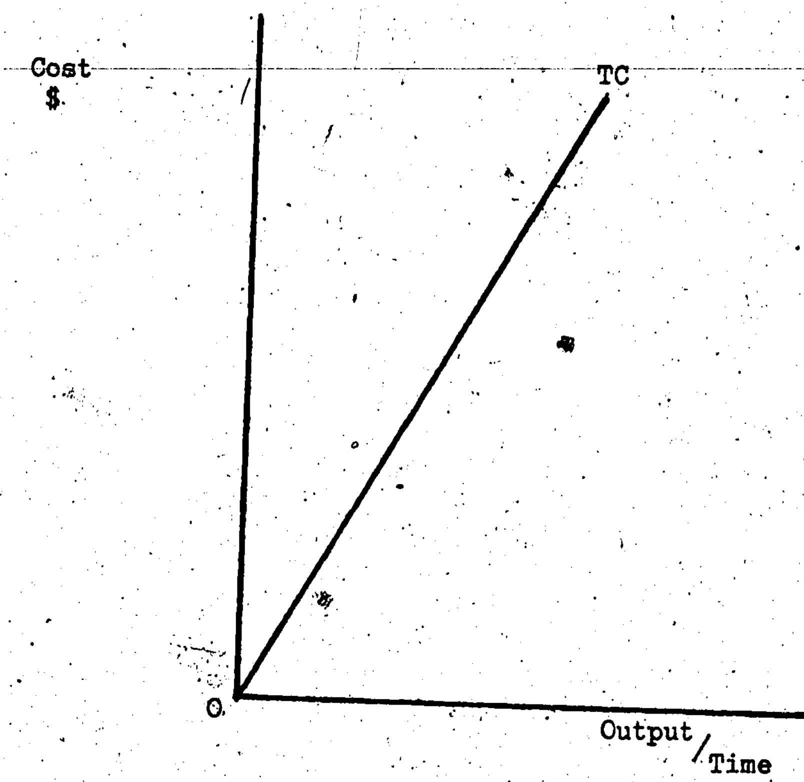


FIGURE 12

COST CURVES REFLECTING CONSTANT RETURNS TO SCALE



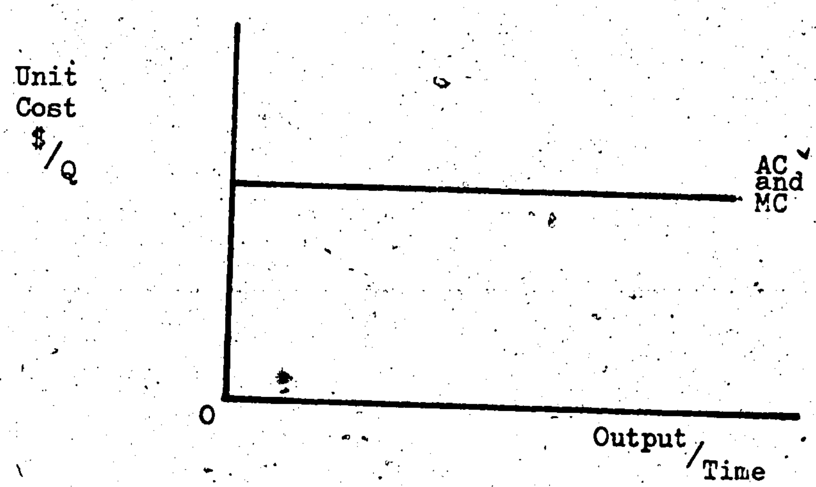


FIGURE 13

COST CURVES REFLECTING INCREASING RETURNS TO SCALE

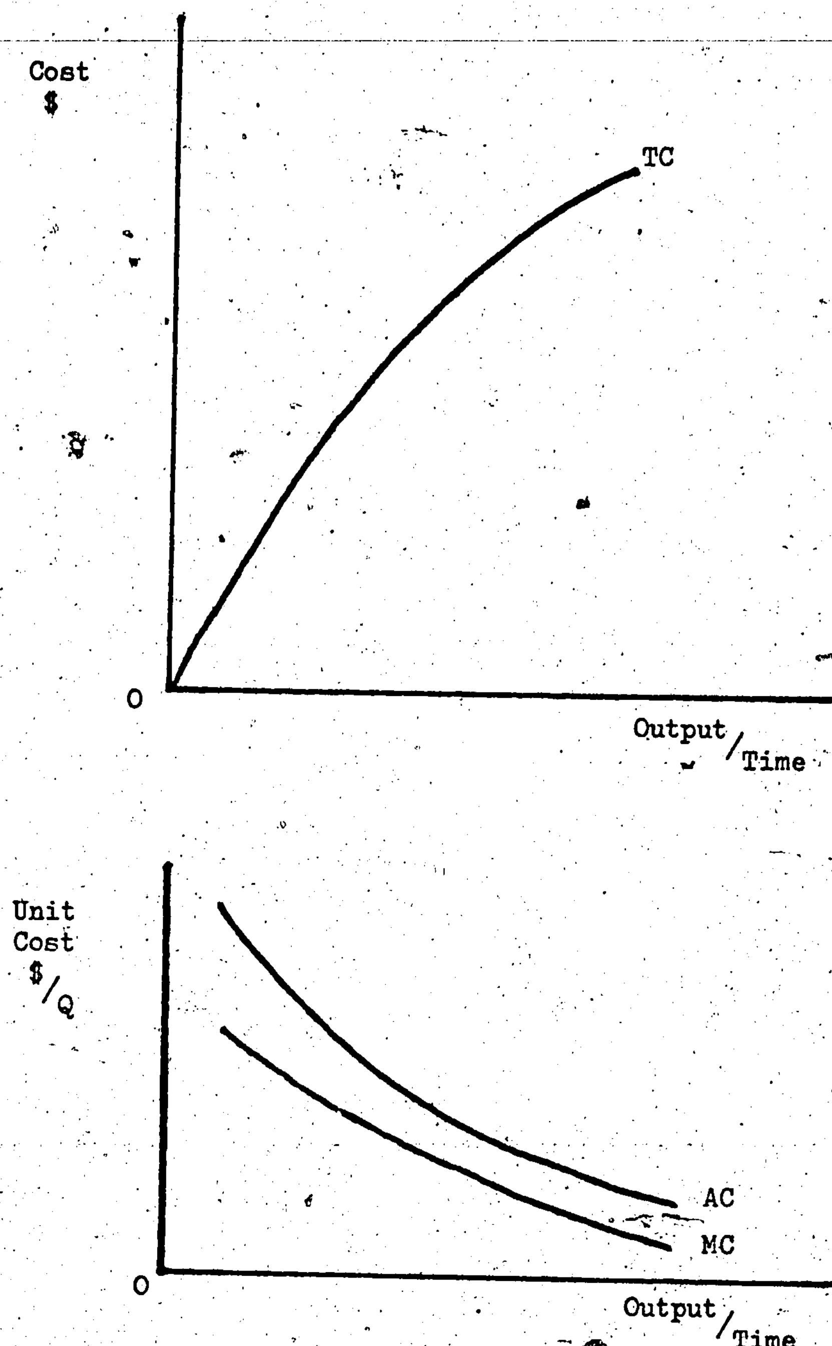
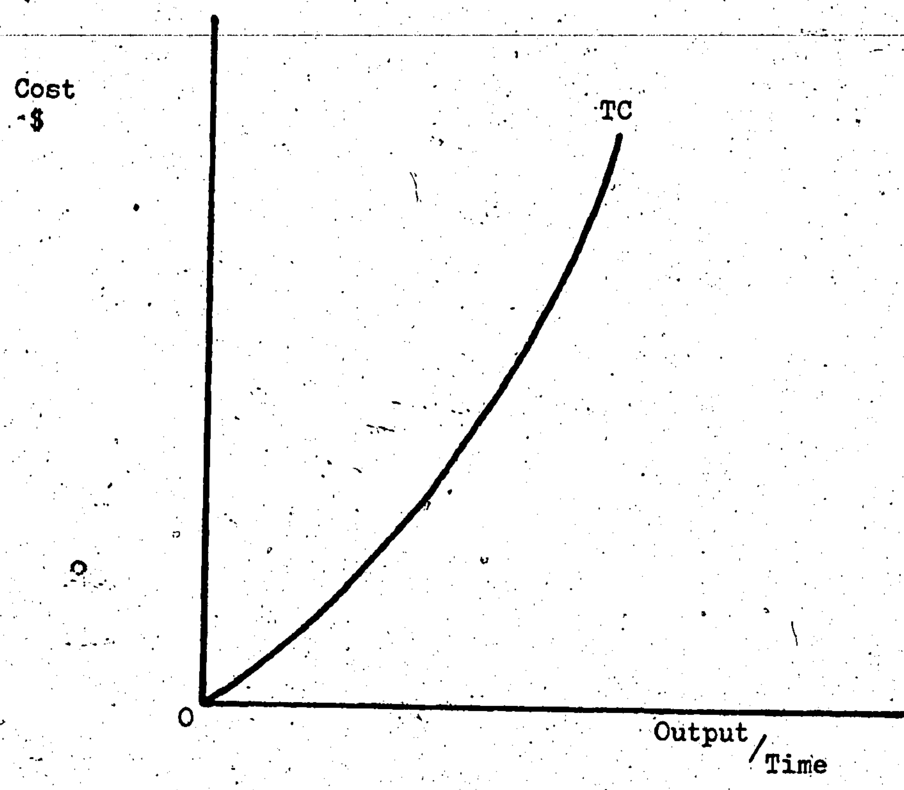
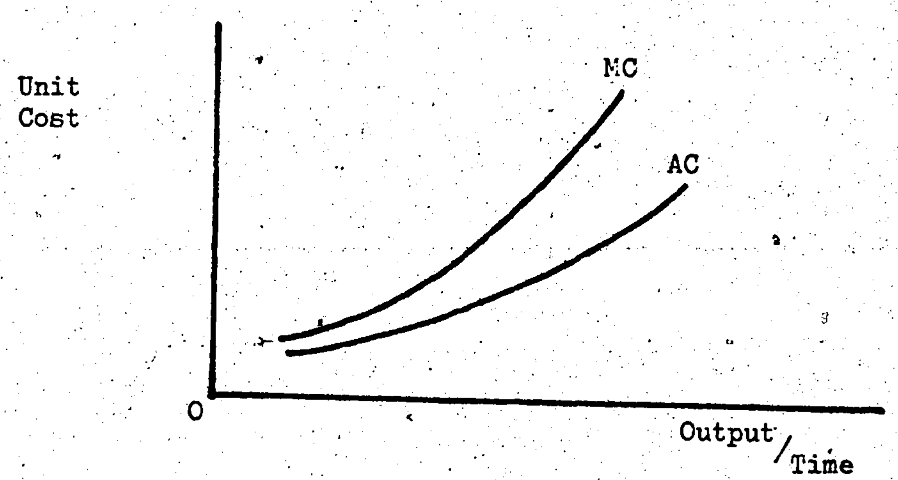


FIGURE 14

COST CURVES REFLECTING DECREASING RETURNS TO SCALE





cost attainable and the point at which the size of the plant is optimal.

It is possible to regard the long-run as a series of alternative short-run situations into any one of which the producer can move.[32] Figure 15 illustrates five such short-run situations. Each SAC curve is the short-run average cost curve for a given scale plant; the farther to the right, the larger the scale plant. The solid portions of the SAC curves form a long-run average cost curve. A producer would never operate on the broken portions of the SAC curves in the long-run because he can reduce his average cost by changing the scale of the plant.

In the long-run the scales of plant possible are unlimited. For each scale one slightly smaller or slightly larger plant may be built. Rather than only the five SAC curves of Figure 15 there is a multitude of SAC curves and their solid portions form the solid line of the long-run average cost curve. The long-run cost curve is considered to be just tangent to the short-run average cost curves of each of the many possible scales of plant.[33]

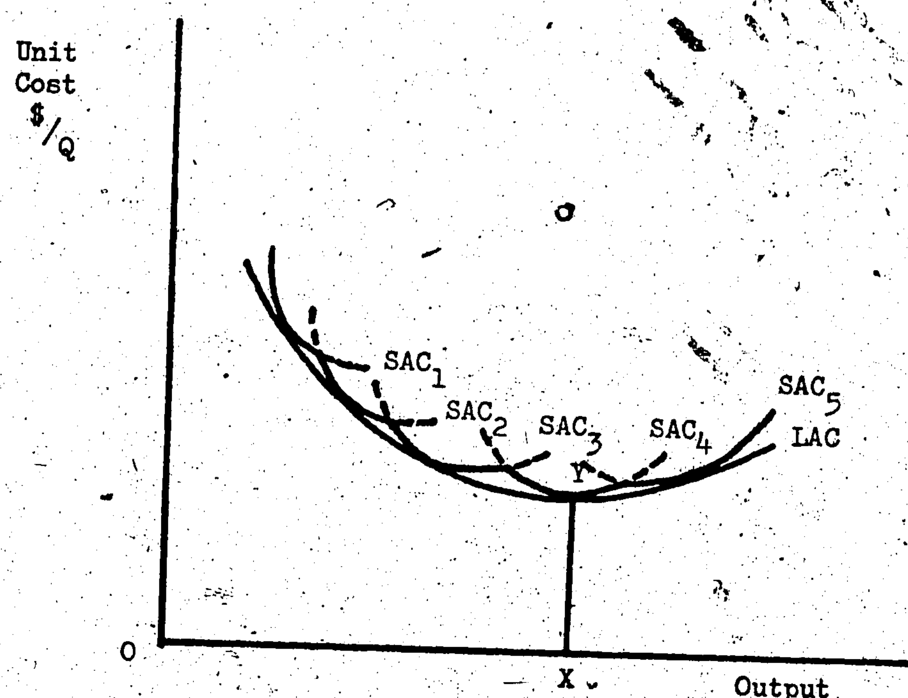
In the long-run increases in production are achieved by the construction of larger plants. As mentioned above, the forces causing the U-shape of the long-run average cost curve are those of economies and diseconomies of scale. When the curve decreases as production

[32]Leftwich, The Price System and Resource Allocation, (1961), P. 153.

[33]Ibid., P. 153.

FIGURE 15

LONG-RUN AVERAGE COST CURVE



increases, successively larger plants are more efficient than those smaller. The opposite is true when the long-run average cost curve is rising.

The Optimum Scale of Plant

The optimum scale of plant is the one which is of the most efficient scale possible. "The optimum scale of plant is the one whose short-run average cost curves forms the minimum point of the long-run average cost curve."[34] In Figure 15, SAC_4 represents the optimal scale of plant.

In cases of continuously increasing returns to scale (i.e. continuously decreasing LAC curve), the optimal scale of plant cannot be achieved because the production process cannot operate at a high enough level to reach the minimum point of the LAC curve. The producer has to operate a scale of plant which has higher short-run average costs than could be possible if the scale were larger.

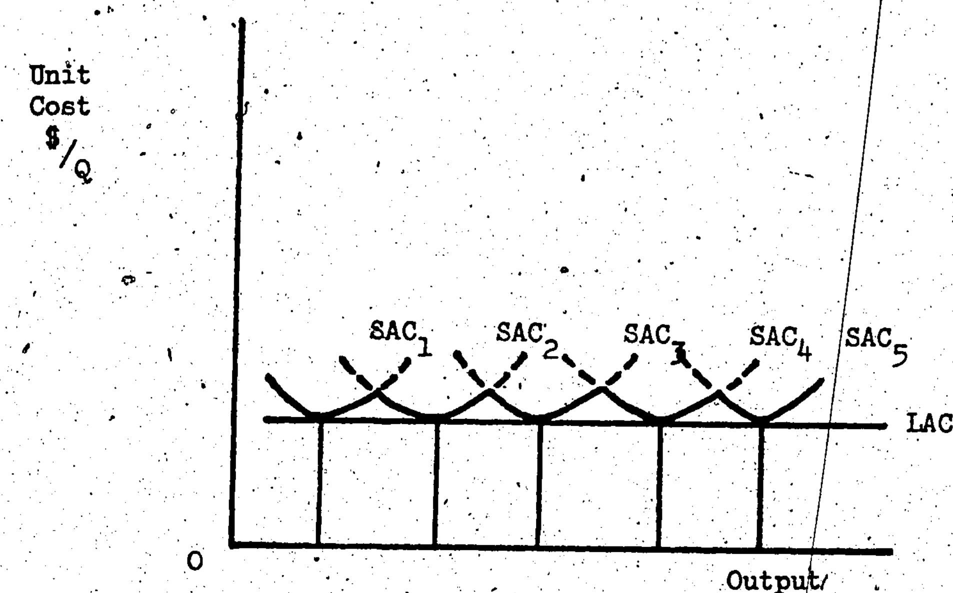
Where decreasing returns to scale are present (i.e. continuously increasing LAC curve) the optimal scale of plant also cannot be achieved. The level of production can be decreased towards zero output but output will be equal to zero before the optimal scale of plant is reached.

Figure 16 illustrates the instance of constant returns to scale (LAC curve parallel to the horizontal). In this case no one size of plant may be considered the most desirable because each SAC curve represents an optimal scale of plant.

[34] Ibid., P. 159.

FIGURE 16

LONG-RUN AVERAGE COST CURVE REFLECTING CONSTANT RETURNS TO SCALE



The Significance of the Long-Run Average Cost Curve

The shape of the long-run average cost curve is of great significance from the view-point of public policy making. If an industry is such that the long-run average cost curve decreases continuously (i.e. the curve is L-shaped) up to the level that represents all of the market demand for the output, then there is need for only one firm in the industry and competition is not desirable.[35] If there is more than one competing in such an industry, the costs of the product would be higher since each firm would not be able to produce an output large enough to enable it to reach the lowest point on the long-run average cost curve of the industry.

Traditional economic theory postulates that the optimally-sized production process is the one which operates at that level of production which coincides with the lowest point of the U-shaped long-run average cost curve. It has been observed in a number of empirical studies that in many industries diseconomies of scale do not occur within the range of observation currently obtainable. These industries appear to possess a long-run average cost curve which is either flat or L-shaped.[36] In industries such as these the production process of optimal size cannot be achieved.

Whether or not economies and diseconomies of scale occur in the production of hospital services is a most important question for those

---

[35]Mansfield, Microeconomics, P. 177.

[36]For a list and brief summary of these studies see either M.S. Feldstein, Economic Analysis for Health Service Efficiency, (Amsterdam: North Holland Pub. Co., 1967), P. 57-58 or Mansfield, Microeconomics, P. 180-182.

who formulate policies and plans regarding the delivery of hospital services. If the long-run average cost curve for hospitals possesses a U-shape then the optimal size for all hospitals would be known. Planners and policy makers could then aim to increase the efficiency of the hospital industry by constructing new hospitals of only the optimal size and adjusting the size of the previously existing ones to this size.

The question of what is the shape of hospital cost curves is the one with which the literature to be reviewed is concerned.

CHAPTER III

CRITICAL SURVEY OF THE LITERATURE

The purpose of this chapter is to survey the economic literature which deals with cost relationships in general acute hospitals. Chapter IV contains a critical examination of the literature.

The literature has been categorized for purposes of organization in this thesis by two characteristics: a) the nature of the cost specification: total or average costs; and b) the nature of the data: time-series or cross-sectional. A total of four categories are possible:

I) Total cost time-series;

II) Total cost cross-sectional;

III) Average cost cross-sectional; and

IV) Average cost time-series.

In Table III-1, the economic literature is arrayed by category in the order in which it will be discussed. A number of additional characteristics from each piece of literature are also included. Time periods in Table III-1 are reported as stated in each study where available. If not stated, the time period is inferred, firstly, from the method of analysis and, secondly, from the conclusions if the method was not indicative. In summary, Table III-1 is a tabular review of the literature.

Before continuing, the reader should note that, when a specification is reported hereafter, the value of the standard error appears

TABLE III-1

TABULAR REVIEW OF THE LITERATURE

Name	Nature of Cost Specification	Nature of Data	Time Period	Degree of Specification	Number of Hospitals	Hospital Size Range	Shape of AC Curve	Conclusions
P.Feldstein P.1-57	TC	T-S	SR	1	1	242	L	Hospital operated on decreasing part of short-run average cost curve.
P.Feldstein P.60-64	TC	X-S	LR	1	60	48-453	L	Decreasing average cost over sample range.
Cohen 1967 1)	TC	X-S	LR	2	23	NS	U	Minimum point between 290-295 beds.
2)	TC	X-S	LR	2	53	NS	U	Minimum point between 160-170 beds.
Cohen 1970	TC	X-S	LR	2	25	NS	U	Minimum point between 540-570 beds.
Carr & Feldstein 1)	TC	X-S	LR	2	3,147	NS	U	Minimum point: 190 beds.
2)	TC	X-S	SR	2	3,147	NS	U	Declining average cost in small hospitals but they likely rose in large hospitals.
Fraser	TC	X-S	SR	1&2	1,027	NS	L	Decreasing LR average costs.
Kushner 1)	TC	X-S	LR	3	95	NS	U	Minimum point: 359 beds.
2)	TC	X-S	LR	3	179	NS	U	Minimum point: 348 beds.

Ingbar & Taylor		AC	X-S	LR	2	72	31-331	weak inverted U	Maximum point: 150 beds. 150-200 bed hospitals provided more services than efficiency required.
Berry		AC	X-S	SR	1	794	NS	L	Economies of scale prevailed.
M.Feldstein	1)	AC	X-S	LR	2	177	72-1064	weak U	Minimum point: 310 beds.
	2)	AC	X-S	SR	2	177	72-1064	weak U	Minimum point between 250-350 beds.
Francisco	1)	AC	X-S	SR	1	1,328	NS	L	Small hospitals experienced decreasing average cost; large hospital constant returns to scale.
	2)	AC	X-S	SR	1	4,710	NS	L	As above.
	3)	AC	X-S	LR	1	1,328	NS	weak L	Slight continuously decreasing average costs.
Evans		AC	X-S	LR	2	185	NS	—	Constant average costs.
Ro		AC	T-S: X-S	LR	1	68	36-794	L	Economies of scale present over range of sample.
Lave & Lave	1)	AC	T-S	SR	2	74	NS	—	Results inconclusive.
AER	2)	AC	T-S: X-S	LR	2	74	NS	—	Nearly constant LR average cost.

Lave & Lave Inquiry	AC	T-S: X-S	LR	2	109	NS	—	Nearly constant LR average cost.

Abbreviations

TC: Total Cost
AC: Average Cost
SR: Short-Run
LR: Long-Run
X-S: Cross-Sectional
T-S: Time-Series
NS: Not Stated

in parentheses immediately beneath each coefficient. The value of the 't' statistic appears below the standard error with no parentheses. Occasionally, the standard error is omitted completely.

I Total Cost Time-Series Literature

Paul Feldstein[1]

Feldstein's work, published in 1961, represents the earliest piece of literature in which the technique of regression analysis was used to study hospital cost relationships. Other economic analyses performed at about the same time used relatively less sophisticated techniques.[2] Little of value appeared in the literature prior to 1961 regarding hospital cost relationships.

In the main body of his work, Feldstein estimated a short-run total cost function for one 242 bed hospital. He collected operating cost data on a monthly basis for over two years. In the linear specifications, which he estimated for some hospital services, he used the patient day as his measure for output. Feldstein acknowledged that it was only a surrogate, however. He summed the departmental cost estimates to obtain an estimate of the total monthly operating cost of the entire hospital. In each of the individual departmental estimates

---

[1] P.J. Feldstein, An Empirical Investigation of the Marginal Cost of Hospital Services, (Chicago: University of Chicago, 1961), P. 1-57.

[2] See H.E. Klarman, The Economics of Health, (New York: Columbia University Press, 1965), C.G. Skinner, "Hospitals and Allied Institutions", Hospital and Medical Economics, Vol.II, ed. W.J. McNerney et al., (Chicago: Hospital Research and Educational Trust, 1962), P. 857-868, and D. Saathoff and R. Kurtz, "Cost Per Day Comparison Don't Do The Job", Modern Hospital, Vol. 99, 1962, P. 14-16.

Feldstein adjusted the cost data to reflect price and wage-rate changes during the year. The following equation reflected the cost-output relationships:

$$\begin{aligned} TMOE = \; & 175{,}016.92 + 4.9[illegible]\,PD + 1.56\,MSPD + 3.06\,OBPD + 3.35\,OBD \\ & + 2.10\,LP + 2.35\,XP + 3.21\,EKG + .18\,PT + 4.78\,OP \\ & + .50\,S_{t-1} + .73\,FC_{t-1} + 78.20\,T \end{aligned} \qquad n = 288$$

TMOE = total monthly operating expense,

PD = patient days per month,

MSPD = medical and surgical patient days,

OBPD = obstetrical patient days,

OBD = number of deliveries,

LP = laboratory patients,

XP = radiology patients,

EKG = EKG patients,

PT = physical therapy patients,

OP = number of operations,

S_{t-1} = O.R. monthly supplies expense,

FC_{t-1} = monthly food costs, and

T = a continuous time variable

No value was reported for either R or R^2. Each of the coefficients reported above was significant at the 5 per cent level in the individual department estimates.

Feldstein observed that the marginal cost of a medical-surgical patient day was \$6.48 (PD + MSPD) and that of an obstetrical patient day was \$7.98 (PD + OBPD). During the particular year he studied, the average cost of a patient day was \$28.79. He, therefore, concluded that marginal cost was approximately 21 to 27 per cent of average operating cost. Since marginal cost was so much less than average cost, Feldstein concluded that the hospital was operating on the declining portion of the short-run average cost curve. The hospital was not experiencing diminishing marginal returns at its size of 242 beds. Feldstein felt that future studies of hospital costs should concentrate on establishing the shape of the average cost curve for each department in individual hospitals. This approach would, "--- provide administrators with the knowledge necessary to operate their hospitals as efficiently as possible".[3]

II Total Cost Cross-Sectional Literature

Paul Feldstein[4]

In an appendix to the study discussed above, Feldstein specified a linear relationship between total operating expense per year (TOE) and the number of patient days per year (PD). He used cost data gathered from sixty U.S. general hospitals during one year. When a regression line was fitted to the data he obtained the following result:

$$TOE = 267{,}692 + \underset{\substack{(1.28)\\ 17.9}}{22.86}\,(PD) \qquad R=.92,\; n=60$$

The standard error of the coefficient of PD indicated that the result was significant and the 't' statistic, (17.9), when calculated, confirmed

[3] P. J. Feldstein, An Empirical Investigation, P. 56.
[4] Ibid., P. 60-64.

that it indeed was highly significant at the 1 per cent level.

Feldstein's conclusion was that because there was a positive constant in the total cost equation, marginal cost was less than average cost, and therefore, the long-run average cost curve was decreasing over the range of hospitals included in the sample; i.e. 48 to 453 beds.

Harold Cohen.[5,6]

Two studies by Cohen are considered in this thesis.

In a 1967 study, Cohen discarded the patient day as an acceptable measure of output and developed a composite measure (S^k) as an output surrogate. In doing this, he extended the development of a concept initially described by Saathoff and Kurtz.[7] Cohen weighted thirteen selected hospital services by their relative average costs, assigning a value of '1' to the 'adult and pediatrics day'. He derived his average cost data from direct operating costs and quantity statistics for these selected services from twenty-three New York City hospitals which were members of the United Hospital Fund.

Cohen recognized that S^k was not an all inclusive measure of the

---

[5] H.A. Cohen, "Variations in Costs Among Hospitals of Different Sizes", Southern Economic Journal, Vol.33, No.3, Jan., 1967, P.335-356.

[6] H.A. Cohen, "Hospital Cost Curves With Emphasis on Measuring Patient Care Output", Empirical Studies in Health Economics, ed. H.E. Klarman, (Baltimore: Johns Hopkins Press, 1970), P. 279-293.

[7] D. Saathoff and R. Kurtz, "Cost Per Day Comparisons Don't Do The Job",

variety of hospital services; however, he used it to estimate two total cost functions. His purpose was to determine which size hospital provided 'ordinary patient care' most efficiently.

In the first specification, Cohen used total cost data from the twenty-three hospitals from which he computed S^k. Secondly, he estimated a similar function using cost data gathered from fifty-three hospitals located in six north-eastern states.[8] He again employed S^k as output. In the second estimate Cohen also attempted to adjust the total cost data for inter-hospital wage differentials because he felt that variations in salaries were induced exogenously to hospitals by factors other than size. He adjusted by eliminating cost differences attributable to variations in the starting salaries of each hospital.

The results of his two analyses were:

1) $$TC = 499{,}446.3 + \underset{\substack{(2.33257)\\ 7.5}}{17.16104\, S^k} + \underset{\substack{(.000005512)\\ 4.3}}{.00002395\,(S^k)^2} \qquad R^2 = .9925,\; n = 23$$

2) $$TC = 88{,}802.6 + \underset{\substack{(1.76727)\\ 11.2}}{19.09026\, S^k} + \underset{\substack{(.000003908)\\ 3.3}}{.000013066\,(S^k)^2} \qquad R^2 = .9\text{[illegible]},\; n = \text{[illegible]}$$

Cohen reported only the standard errors; however, when computed the 't' statistic demonstrated that both coefficients in each of the equations were significant at the 1 per cent level. It would appear that both of these quadratic relationships could have represented an L-shaped average cost curve because the $(S^k)^2$ coefficients were relatively small, however, Cohen's conclusion was that the curve had a U-

---

[8] Cohen sent out survey questionnaires to 399 hospitals in six states. Of the fifty-three that responded thirty-five were located in New York City.

shape. Using a factor to convert the S^k values he had calculated, Cohen found that the minimum point on the average cost curve fell in the 290 to 295 bed range for the first estimate and in the 160 to 170 bed range for the second.[9]

In 1970 Cohen published a second study estimating the relationship between total cost and S^k. He observed that his composite measure alone did not adequately reflect differences in output quality between hospitals since it implicitly assumed output to be homogenous in all hospitals. To control for quality differences he introduced a dummy variable into the specification which he made equal to '1' when a hospital was associated with a medical school and equal to '0' when it was not affiliated. The author used cost data from twenty-five New York City hospitals which all belonged to the United Hospital Fund. The total cost included only the direct costs associated with the services offered routinely by these hospitals. All indirect and non-patient care costs including capital costs were excluded.

Cohen's results were:

$$TC = 4{,}100{,}000 + \underset{\substack{(.000003)\\ 17.3}}{.000052}\,(S^k)^2 \qquad R^2 = .92,\ n = 25$$

$$TC = 3{,}700{,}000 + \underset{\substack{(.006)\\ 2.1}}{.013}\,(d) + \underset{\substack{(.000004)\\ 12.5}}{.000049}\,(S^k)^2 \qquad R^2 = .93,\ n = 25$$

d = 1 : affiliated

d = 0 : nonaffiliated

[9] Cohen used the conversion factor $1S^k = 500$ patient days. He neglected to provide it in his 1967 study but he provided it in the 1970 study. The conversion factor was without statistical support.

Although Cohen did not report the 't' statistics, when calculated they indicated that the values of the coefficients of $(S^k)^2$ in both estimates were highly significant at the 1 per cent level. The coefficient of the affiliation dummy was barely significant at the 5 per cent level. The value of R^2 was only slightly improved by the attempt to control for differences in quality.

After he converted S^k into terms of beds, Cohen reported that a U-shaped average cost curve was indicated by both estimates. It reached a minimum point between 560 and 570 beds when the affiliation dummy was not used and a minimum point between 540 and 555 beds when the dummy was included in the estimate. Again it was questionable that the average cost curve was not actually L-shaped considering that the $(S^k)^2$ coefficient was so small.

John Carr and Paul Feldstein[10]

Carr and Feldstein collaborated in an extension and an elaboration of Feldstein's work.[11] They used operating cost data reported to the American Hospital Association by 3,147 non-profit, U.S. general hospitals. No adjustment was made to control for depreciation. They did adjust their cost data for geographic wage differences by multiplying the number of full-time equivalent employees in each hospital by the average yearly wage rate paid by all of the hospitals in the sample. This amount they combined with non-payroll costs to give an adjusted total cost figure.

[10] W.J. Carr and P.J. Feldstein, "The Relationship of Cost to Hospital Size", Inquiry, Vol.4, No.2, June, 1967, P. 45-65.

[11] P.J. Feldstein, An Empirical Investigation, P. 60-64.

Carr and Feldstein performed two separate analyses of their data. In both they specified a total cost function in quadratic form and used patient days (PD) to represent size. The ultimate aim of their work was to estimate the net effect of hospital size upon the cost of in-hospital care.

In the first analysis Carr and Feldstein estimated an aggregate, long-run, total cost function for their sample. They included eight measures of service capability in the specification as dummy variables.

S - number of services and facilities available.

S.PD - the number of services and facilities times the number of patient days.

OPV - the number of outpatient visits.

NS - the presence of a professional school of nursing.

N - the number of student nurses.

IRP - the number of types of internship and residency programs offered.

IR - the number of interns and residents.

MS - affiliation with a medical school.

The estimate of the total cost function was:

$$
\begin{aligned}
TC = & -307{,}568 + \underset{\substack{(1.19)\\ 29.1}}{34.70}\,PD + \underset{\substack{(.0000029)\\ 12.1}}{.0000351}\,(PD)^2 - \underset{\substack{(.07)\\ 4.4}}{.31}\,(S.PD) + \underset{\substack{(31{,}593)\\ .73}}{23{,}188}\,NS \\
& + \underset{\substack{(617)\\ 8.2}}{5{,}034}\,IR + \underset{\substack{(3{,}619)\\ 9.3}}{33{,}827}\,S + \underset{\substack{(.34)\\ 14.1}}{4.81}\,OPV - \underset{\substack{(295)\\ 6.1}}{1{,}805}\,N + \underset{\substack{(5{,}480)\\ 10.1}}{55{,}347}\,IRP \\
& + \underset{\substack{(43{,}744)\\ 3.7}}{174{,}796}\,MS \qquad R^2 = .947 \quad n = 3{,}147
\end{aligned}
$$

Although Carr and Feldstein did not report any 't' statistics, when calculated they confirmed that all of the coefficients were significant at the 1 per cent level except NS the result for which was not significant even at the 5 per cent level. They attributed the unexpected negative constant to a high degree of correlation between the various independent variables. They calculated an estimate of the average cost function by dividing the coefficients of PD, PD^2 and the constant by the total number of patient days per year. The other independent variables were not included in this calculation. The average cost equation which resulted provided a U-shaped curve which reached its minimum point at 190 beds.

In a second analysis Carr and Feldstein stratified their sample into five service-capability groups according to the number of services offered by each hospital. In doing this, they expected to limit the effect which the number of services had on the cost-size relationship. For each of the five groups they estimated a total cost function using the same independent variables as in the first analysis omitting only S. The net result of the stratification was to produce five short-run situations.[12] Table III-2 reports their results for the constant term, PD and PD^2 in each group.

The coefficients of PD were all highly significant at the 1 per cent level except for the PD coefficient of Grou[illegible]ich was only significant at the 5 per cent level. The coefficient of PD^2 was significant at the 5 per cent level for Group 2 and at the 1 per cent level

[12] See the discussion in Chapter II regarding the long-run average cost curve and, in particular, Figure 15 which illustrates that the long-run average cost curve is composed of a series of short-run curves.

TABLE III-2

CARR AND FELDSTEIN'S ESTIMATES OF THE RELATIONSHIP BETWEEN TOTAL COST AND OUTPUT BY SERVICE CAPABILITY GROUPS

Group	Number of Services	Constant	PD	PD^2	R^2	n
1	0- 9	13,422	21.36	-.0000182	.844	680
			(3.06)	(.0000515)		
			7.0	.4		
2	10-12	55,003	9.58	.0000667	.861	693
			(4.50)	(.0000294)		
			2.1	2.3		
3	13-16	83,756	27.76	.0000146	.875	729
			(4.49)	(.0000165)		
			6.2	.8		
4	17-19	134,677	36.19	-.0000149	.850	490
			(6.85)	(.0000148)		
			5.3	1.0		
5	20-28	570,398	27.25	.0000370	.896	555
			(5.25)	(.0000069)		
			5.2	3.4		

for Group 5. The authors converted their results to average cost per patient day. They found evidence which indicated that diminishing marginal returns occurred only for the larger hospitals in Group 5; the highest service-capability group. It was concluded that the level of average cost was greater the more services a group contained.

After considering their results, Carr and Feldstein concluded that at the aggregate level, the average cost per day initially fell as size increased but probably rose in very large hospitals.

R.D. Fraser[13]

Fraser was given access to cost data gathered by the Dominion Bureau of Statistics from 1,266 'public general and allied special hospitals' in Canada. With this data he estimated 290 total cost functions, ninety of which were for public general hospitals alone. He specified both simple linear and quadratic relationships between the cost and the independent variables.

Fraser used three different measures of total cost in his estimates: 1) total operating cost plus the sum of the estimated depreciation on buildings, land, improvements and major equipment; 2) total operating cost plus one-tenth of the total value of plant assets; 3) total operating cost alone. In addition, he considered four different measures of output in his analyses: 1) patient-days; 2) the number of admissions; 3) the rated bed capacity; and 4) the composite measure

[13] R.D. Fraser, Canadian Hospital Costs and Efficiency, Special Study No. 13, Prepared for the Economic Council of Canada, (Ottawa: Information Canada, 1971). In his first estimate of total cost Fraser included depreciation of land. Under normal accounting practice land is normally not depreciated.

of output (S^k) developed by H.A. Cohen described above.[14] In some of his equations Fraser included dummy variables to represent the presence of training programs for nurses and interns as surrogates representing quality. He also included a capacity utilization dummy variable to indicate the presence of rigid patterns of patient movement through hospitals in some estimates.

Fraser estimated a number of long-run total cost relationships by province in which he used data reflecting the behaviour of all Canadian hospitals. He used various combinations of his three measures of total cost and his four measures of output. He found that the results obtained with the estimates incorporating the composite measure of output (S^k) were the most supportive of his hypothesis of decreasing average cost. He employed the composite measure in the majority of his estimates.

Fraser grouped the cost data for acute care general hospitals and then estimated ninety total cost relationships using only S^k for output. All but nine included a measure of capital cost. He stratified his sample into nine groups according to bed size and, in effect, analyzed nine short-run situations.[15] The groups were as follows: 1-9 beds, 10-24 beds, 25-49 beds, 50-99 beds, 100-199 beds, 200-299 beds, 300-499 beds, 500-999 beds and 1,000 beds or more.

---

[14] H.A. Cohen, "Variations in Costs Among Hospitals of Different Sizes".

[15] Appendix A contains the ten tables of results obtained by Fraser in his analysis of acute care general hospitals in Canada.

Fraser's results indicated that in simple linear estimates the values of the coefficients of S^k were generally large and highly significant at the 1 per cent level for every size hospital. Only the group of 1,000 beds or more had some coefficients which were not at all significant. The estimates of the quadratic relationship produced results which indicated that the coefficients of S^k were generally large and significant at the 5 per cent level except for the two groups of 300-499 beds and 500-999. The coefficients of $(S^k)^2$ were frequently negative but in almost every instance, whether positive or negative, they were not significant at the 5 per cent level. Fraser's results indicated that total operating costs were generally rising at a decreasing rate for almost every size general hospital in Canada. There was not any major difference between the results using operating cost and those which also included capital costs.

Fraser concluded that it would not be unlikely if decreasing long-run average cost curves generally existed for Canadian general hospitals. This conclusion was at odds with that of Cohen who concluded that the average cost curve had a U-shape and that an optimum level of output existed. Both Cohen and Fraser used S^k to represent output, both specified quadratic total cost relationships and for both their results were significant. Fraser did, however, have a larger sample which likely contained far more small and non-urban hospitals than did Cohen's sample. Fraser's sample also likely contained fewer large size hospitals. It could also be that large Canadian hospitals do not approach the size of their American counterparts particularly in a large urban center such as New York City. Fraser disaggregated his data into nine size categories whereas Cohen retained an aggregate approach. It is possible that their different results were related to some of the differences in data, samples or technique.

Joseph Kushner[16]

Kushner prepared two estimates of a total cost function, utilizing data which reflected the operations of acute care general hospitals in Ontario.

Kushner restricted his sample in his first estimate to include only the ninety-five accredited hospitals of Ontario. By doing this, he hoped to obtain a relatively more homogeneous sample with regard to quality of output. He attempted to further increase the homogeneity of his cost data by excluding amounts equal to the costs of providing medical and other treatment services and education from the total cost figure for each hospital. Kushner arbitrarily allocated hospital costs and found that treatment services represented 26 per cent of the total and education 4 per cent. The remaining 70 per cent he described as 'hotel costs' (THC). He adjusted these hotel costs to include an allowance for depreciation on plant and equipment. He anticipated that by using hotel costs in his analysis he could eliminate cost variations which were not directly related to the size of a hospital but rather to the type of patients, to their illnesses and to education programs within an institution.

16 J. Kushner, Economies of Scale in the General Hospital Industry, Unpublished Ph.D. Disertation, (London, Ontario: University of Western Ontario, 1969).

Kushner developed his own measure of output, the adjusted patient day (APD). It was equivalent to the sum of the patient days per year plus a value equal to seventy-five per cent of the empty beds maintained during a year by a hospital. He implicitly assumed that some output was always being produced in a hospital whether or not there were any patients. In his estimate of the relationship between total hotel costs and adjusted patient days, Kushner specified a polynomial relationship of the third degree. No other independent variables were included in the equation.

$$THC = \underset{19.79}{23.20\ APD} - \underset{-4.098}{.00007593\ APD^2} + \underset{4.594}{.0000000003025\ APD^3}$$

R^2 = .97[illegible]05
n = 95

Kushner reported the 't' values for each of the three APD terms in the estimate. Each estimated coefficient was significant at the 1 per cent level. His result supported his hypothesis that the long-run average cost curve for all hospitals was U-shaped. The curve reached its minimum point at 359 beds.

In the second analysis, Kushner again specified a total cost relationship of the third degree. He used data gathered from all 179 Ontario general hospitals but he did not adjust for any factor including the cost of capital. He used the term potential patient days (PPD) to represent size. He felt that PPD reflected the actual number of beds set up.

Initially, Kushner specified an equation of fifteen independent variables which represented output in eight service areas. The rest were dummies used to represent quality. Kushner estimated a relationship and then proceeded to alter it into the form he felt it should 'logically' take. His fourth and final estimate upon which he based his conclusions was:

$$
\begin{array}{l}
TC = -\underset{-.1674}{429.6}\ MED + \underset{1.46}{6{,}208}\ MAT - \underset{-.4165}{1{,}420}\ PAED - \underset{-.7592}{8{,}877}\ CONV - \underset{-4.284}{14{,}840}\ CHR \\
\quad + \underset{1.144}{2{,}504}\ OCC + \underset{6.062}{739{,}300}\ MS + \underset{1.037}{2{,}299}\ GN + \underset{20.76}{35.68}\ PPD - \underset{-1.773}{.00002395}\ PPD^2 \\
\quad + \underset{3.696}{.00000000009031}\ PPD^3
\end{array}
$$

$R^2 = .98889$
$n = 179$

where MED = ratio of medical patients to total patients,

MAT = ratio of maternity patients to total patients,

PAED = ratio of paediatric patients to total patients,

CONV = ratio of convelescent patients to total patients,

CHR = ratio of chronic patients to total patients,

OCC = occupancy rate,

MS = presence of a medical school; and

GN = ratio of graduate nurses to total nursing staff.

The coefficients of PPD and PPD^3 were both significant at the 1 per cent level; the PPD coefficient was highly significant. The PPD^2 coefficient was not significant at the 5 per cent level. Of the other independent variables only MS and CHR were significant. The shape of the total cost curve was similar to the first one estimated by Kushner. He concluded again that the long-run average cost curve was U-shaped and that in this case, the minimum level was 348 beds.

Kushner considered both of his results and concluded that the optimum hospital size was somewhere within the 300 to 500 bed range.

III Average Cost Cross-Sectional Literature

M. L. Ingbar and L. D. Taylor[17]

Ingbar and Taylor performed two separate aggregate analyses of hospital average cost. They had access to a wide range of cost information supplied by the accredited general hospitals of Massachusettes which ranged in size from thirty-one beds to 331. Initially, the authors hypothesized that over 100 variables might affect hospital costs. They reduced these variables to six factors which by themselves accounted for much of the variation in the original list of variables. The six factors were used as the independent variables in the specification. The six were:

Y = hospital service expense per patient day (PD),

S_{55} = medical and surgical physical expense per PD,

S_{77} = weighted operations per PD,

S_{88} = outpatient radiological weighted films per PD,

S_{90} = private patient days per PD,

X_{38} = beds; and

X_{66} = occupancy rate.

The data did not reflect capital costs.

In the first analysis Ingbar and Taylor specified a quadratic relationship. They used cost data gathered from seventy-two hospitals combined for the two years of 1958 and 1959. In a second estimate they specified a similar relationship and used combined costs for the years

17. M.L. Ingbar and L.D. Taylor, Hospital Costs in Massachusettes, Cambridge, Mass.: Harvard University Press, 1968).

1962 and 1963 from sixty-seven hospitals. The result for the first estimate was:

$$Y_1 = 29.86 + \underset{\substack{(2.35)\\ 4.35}}{10.19}\,(S_{55}) + \underset{\substack{(41.90)\\ 3.60}}{151.03}\,(S_{77}) + \underset{\substack{(2.3[illegible])\\ 5.98}}{13.[illegible]}\,(S_{88}) + \underset{\substack{(5.94)\\ 1.99}}{11.52}\,(S_{90})$$

$$+ \underset{\substack{(1.93)\\ 2.04}}{3.93}\,(X_{38}) - \underset{\substack{(.61)\\ 2.15}}{1.31}\,(X_{38})^2 - \underset{\substack{(4.59)\\ 3.34}}{15.33}\,(X_{66}) \qquad \begin{array}{l} R^2 = .568 \\ n = 72 \end{array}$$

The second estimate yielded the following specification:

$$Y_2 = 30.83 + \underset{\substack{(1.96)\\ 3.35}}{6.57}\,(S_{55}) + \underset{\substack{(50.76)\\ 2.41}}{122.38}\,(S_{77}) + \underset{\substack{(3.13)\\ 4.[illegible]}}{13.79}\,(S_{88}) + \underset{\substack{(7.66)\\ 2.19}}{16.77}\,(S_{90})$$

$$+ \underset{\substack{(3.37)\\ 1.15}}{3.87}\,(X_{38}) - \underset{\substack{(1.00)\\ 1.02}}{1.02}\,(X_{38})^2 - \underset{\substack{(7.20)\\ 1.02}}{7.35}\,(X_{66}) \qquad \begin{array}{l} R^2 = [illegible] \\ n = 67 \end{array}$$

The results indicated that the average cost curve had an inverted U-shape which peaked at 150 and 190 beds respectively when 1958-9 and 1962-3 data were used. The variables representing size (X_{38}) and $(X_{38})^2$ were both significant at the 5 per cent level in the first estimate but neither was significant in the second estimate. The inverted U-shape was contrary to the expectation of Ingbar and Taylor. They concluded, however, that since the inverted U was relatively shallow and since the coefficients in the second estimate were not significant, the results were of questionable value for purposes of policy formulation. They speculated that their results reflected the fact that hospitals in the 150 to 200 bed range might have been providing more services than was necessary for maximum efficiency.

Ralph Berry[18]

For his study Berry used cost data from 5,293 non-federal, short-term U.S. general hospitals. He observed that in investigating the existence of economies of scale in the hospital field the patient day should not be used as a measure of output. Since large hospitals generally offer more services than do small ones, he argued that to compare hospitals on the basis of their cost per patient day was to actually compare the cost of a variety of different 'products'. In his analysis Berry attempted to reduce the variety of products by stratifying his sample population.

Berry observed that an inter-relationship existed between hospital size, average cost and scope of services. Since his aim was to examine the effect of size on average costs, he attempted to eliminate the effect which scope of services had on average cost. He stratified the sample by grouping the hospitals according to which combination of twenty-eight facilities and services each offered. The services and facilities included items such as an operating room, a laboratory, a delivery room and a pharmacy. Berry's procedure assumed that in any one group all of the hospitals offered a like combination of services and facilities, and that each produced service in identical proportions.

Out of over 3,000 groups of hospitals Berry used only those which included ten or more hospitals. Although this procedure eliminated 85 per cent of his original sample population, it provided fourty

---

[18] R.E. Berry, "Returns to Scale in the Production of Hospital Services", Health Services Research, Vol.2, No.2., Spring 1967, P. 123-139.

groups of hospitals each homogeneous as to services and facilities. He specified a simple linear relationship between average cost per patient day (Y) and total patient days (PD). Since he stratified his population, he, in effect, estimated fourty short-run average variable cost functions. (His results are presented in Table III-3)

Berry found that in thirty-six of his fourty estimates, average cost and patient days were negatively correlated. In the same thirty-six estimates, the coefficient of patient days was negative. Using a 'rule of thumb' which required that a coefficient just exceed the value of its standard error, Berry concluded that twenty-six of the thirty-six negative relationships were 'significant'. Berry made no conclusions based on the significance of the 't' values. His conclusion was that since declining average costs were present in so many of his estimates that they could only have resulted if in fact economies of scale prevailed.

(When the 't' values are considered, one finds that one estimate with a positive PD coefficient is significant at the 1 per cent level, five with negative PD coefficients are significant at the 1 per cent level and one negative PD coefficient is significant at the 5 per cent level. This hardly constitutes overwhelming support for Berry's conclusion regarding the prevalance of economies of scale in the hospital field.)

TABLE III-3

BERRY'S ESTIMATES OF THE RELATIONSHIP BETWEEN AVERAGE COST AND OUTPUT

(Grouped by Like Combinations of Services and Facilities)

Group	Number of Services In Common	Estimate AC=	t	n
1	7	29.41-.0003 PD	1.30	92
2	6	35.72-.0008 PD	3.27 *	69
3	7	29.05-.0003 PD	.56	45
4	8	35.05-.0006 PD	1.31	33
5	8	31.83-.0004 PD	1.05	26
6	8	37.02-.0006 PD	.98	24
7	8	35.62-.0007 PD	1.08	21
8	8	35.21-.0007 PD	1.63	21
9	5	27.48-.0001 PD	.10	20
10	7	42.41-.0015 PD	[illegible]	[illegible]
11	7	36.71-.0012 PD	[illegible]	19
12	9	38.95-.0005 PD	[illegible]	18
13	8	35.92-.0004 PD	[illegible]	18
14	9	33.43-.0006 PD	1.7[illegible]	17
15	7	27.83-.0004 PD	1.51	16
16	26	58.63-.0001 PD	2.63	15
17	8	33.50-.0005 PD	.88	15
18	8	33.58-.0003 PD	.55	15
19	8	31.48-.0008 PD	1.17	15
20	8	37.86-.0008 PD	1.41	15
21	7	23.93+.0003 PD	.62	14
22	13	44.11-.0003 PD	1.37	13
23	7	22.96+.0003 PD	.73	13
24	11	44.35-.0010 PD	4.28 *	12
25	11	37.46-.0006 PD	1.73	12

26	10	33.30-.0003 PD	1.46	12
27	9	30.06-.0004 PD	.84	12
28	8	30.99-.0003 PD	.56	12
29	8	38.14-.0005 PD	1.01	12
30	5	43.96-.0020 PD	1.69	12
31	9	37.48-.0005 PD	1.35	11
32	9	34.80-.0004 PD	1.06	11
33	9	31.02-.0006 PD	.52	11
34	9	32.05-.0003 PD	1.15	11
35	9	39.45-.0013 PD	1.10	11
36	12	44.23-.0001 PD	.49	11
37	10	49.17-.0016 PD	2.46 *	10
38	9	32.34-.0004 PD	.89	10
39	8	18.46+.0016 PD	.04 *	10
40	22	34.90+.0001 PD	.52	10

* - significant at the 1 per cent level

z - significant at the 5 per cent level

Martin Feldstein[19]

In a comprehensive health study, Feldstein devoted one section to an analysis of the effects of scale on hospital average cost. He gathered his data from a sample of 177 acute non-teaching hospitals in England and Wales. The hospitals ranged in size from seventy-two to 1,064 beds with a mean size of 302.9 beds.

Martin Feldstein opined that the 'number of cases' was superior to any other alternative measure for hospital output because it included both live discharges and deaths. Subsequently, he specified average cost functions in both linear and quadratic forms and his dependent variable (C_i) represented the average cost per case. Capital costs were, however, omitted from the cost data. Feldstein used the number of beds (B_i) to represent scale as the independent variable. His estimates of the average cost function unadjusted for case-mix were:

$$Ci = 49.71 + \underset{\substack{(.55) \\ 2.92}}{1.62} (10^{-2})\, Bi \qquad R^2 = .048 \quad n = 177$$

$$Ci = 49.16 + \underset{\substack{(1.77) \\ 1.08}}{1.92} (10^{-2})\, Bi - \underset{\substack{(1.89) \\ .20}}{.38} (10^{-5}) Bi^2 \qquad R^2 = .048 \quad n = 177$$

Feldstein only considered the R^2 values, both of which were very small. He felt that in any analysis of hospital cost there must be an adjustment for casemix which would reduce the heterogeneous nature of the data. The results of his first two estimates confirmed his belief that estimates of cost functions were of little value if no allowances were

[19] M.S. Feldstein, Economic Analysis for Health Service Efficiency, (Amsterdam: North-Holland Publishing Company, 1967).

made for casemix. (It is interesting to note that the coefficient of scale was significant at the 1 per cent level in the linear estimate. Neither scale coefficient in the quadratic estimate was significant. The linear function indicated slightly rising average costs and the quadratic function a weak inverted U-shape curve.)

In a preliminary analysis, Feldstein had determined that he could group all types of cases into nine distinct categories. The nine included general medicine, paediatrics, general surgery, ear-nose-and throat, traumatic and orthopaedic surgery, other surgery, gynecology, obstetrics and miscellaneous. With these he developed an independent variable and introduced it into his estimates as an aggregated representation of casemix. It reflected the various proportions of total cost contributed by the nine categories. Feldstein re-estimated his first two specifications and included the adjustment for [illegible] ($Pi\delta$).

$$Ci = .295\,(10^{-2})\,Bi + Pi\delta i \qquad R^2 = .308$$
$$\quad (.567) \qquad\qquad\qquad\qquad\qquad n = 177$$
$$\quad .52$$

$$Ci = -.581\,(10^{-2})\,Bi + .934\,(10^{-5})\,Bi^2 + Pi\delta i \qquad R^2 = .310$$
$$\quad (1.728) \qquad\qquad (1.741) \qquad\qquad\qquad n = 177$$
$$\quad .34 \qquad\qquad\qquad .54$$

where $Pi\delta i = P_1\delta_1 + P_2\delta_2 + \ldots + P_9\delta_9$.

The R^2 values were considerably larger than for the estimates in which a casemix adjustment was omitted and this increase he attributed entirely to the influence of casemix. The linear estimate indicated a slightly rising average cost curve and the quadratic specification indicated that the average cost curve had a shallow U-shape which reached its minimum level at 310 beds. The scale coefficients in both the linear and quadratic specifications were not significant at the 5 per cent level.

Feldstein next stratified his data according to bed size and then estimated a separate average cost function for each size group. He initially divided the sample into two subgroups: seventy-two to 302 beds and 303 to 1064 beds. Then he divided it into four subgroups: seventy-two to 117 beds, 118 to 302 beds, 303 to 488 beds and 489 to 1064 beds. He specified a quadratic equation adjusted for casemix for each subgroup. He found that, "The more dissaggregated models do not provide a better model of the relation between size and average cost per case than the original single equation"[20]. The dissaggregated estimates indicated that the average cost curve had a slight U-shape with a minimum range of 230 to 350 beds. None of the coefficients for either Bi or Bi^2 was significant at the 5 per cent level in the functions estimated for the stratified subgroups.

Feldstein concluded that he could neither accept nor reject the hypothesis that diseconomies of scale prevailed among large hospitals based solely on the results of his analysis. He felt, however, that the shape of the average cost curve was slightly U-shaped reaching a minimum in the 250 to 350 bed range. He also felt that average cost likely rose slightly to about the 600 bed level and then flattened out. He concluded that it appeared that medium size hospitals were at least as efficient at providing general ward care as large hospitals.

[20] M.S. Feldstein, Economic Analysis, P.71.

Feldstein commented in his conclusion that the large hospitals in his study generally had longer lengths of stay than did the small hospitals. He felt that if the length of stay could be reduced in large hospitals, economies of scale could possibly occur over the entire size range of all hospitals. He believed that the managerial and sociological problems generally associated with large scale organizations which acted in large hospitals to prevent shorter lengths of stay and a lower cost per case.

Edgar Francisco[21]

In a recent study, Francisco performed three analyses using operating cost data which reflected the performance of 4,710 U.S., short-term, general hospitals. He used the patient day as his measure of output throughout.

In the first analysis, Francisco stratified his sample by grouping the hospitals which had identical combinations of services and facilities. He repeated the technique, initially developed by Berry, in his study discussed above.[22] He discarded any group which did not contain at least thirty hospitals and was left with twenty-five groups containing 1,328 hospitals. Francisco first estimated a linear total cost function for each group. Although he did not report his results, in his narrative he provided a summary. His total cost analysis revealed decreasing average cost in twenty-one of the groups, although, the

---

[21] E.W. Francisco, "Analysis of Cost Variations Among Short-Term General Hospitals", Empirical Studies in Health Economics, ed. H.E. Klarman, (Baltimore: Johns Hopkins Press, 1970), P. 321-332.

[22] R. Berry, "Returns to Scale".

results were significant at the 5 per cent level for only four.

Francisco then estimated linear average cost functions for each of the twenty-five groups. (see Table III-4). He found decreasing average cost for twenty-two of the twenty-five groups but in only seven of these was the patient day coefficient significant at the 5 per cent level. Each of the seven, however, happened to have an average size of less than fifty-seven beds. From the two analyses of the twenty-five groups Francisco concluded that economies of scale appeared likely in hospitals with relatively few services and facilities with fewer than 100 beds. He also concluded that larger hospitals with more services may well experience constant returns to scale. Francisco's results were very similar to those derived from Berry's reported results, once the 't' values were computed. Of Berry's fourty groups, six were significantly negative compared to Francisco's seven of twenty-five. Similar conclusions could be made from both sets of results.

Francisco subsequently stratified all 4,710 hospitals with regard to their total number of services and facilities. He obtained seventeen groups of hospitals and he then estimated a linear average cost relationship for each group. (See Table III-5) In fifteen of the groups decreasing average cost resulted but only eight of these had patient day coefficients significant at the 5 per cent level. Of these eight, seven were groups in which the average size was 135 beds or less. Only one group of a larger average size exhibited significantly decreasing average cost. These results supported his earlier conclusions.

From the result of an analysis he did not report, Francisco observed that output (patient days) accounted for little variation in

TABLE III-4

FRANCISCO'S ESTIMATES OF THE RELATIONSHIP BETWEEN AVERAGE COST AND OUTPUT

(Grouped by Like Combinations of Services)

Group	Number of Services In Common	Estimate AC=	n
1	1	33.51- .40X	125
2	2	48.26-1.63X *	40
3	2	36.03- .55X *	150
4	2	33.17- .28X	77
5	2	41.27- .63X	41
6	3	35.44- .65X	39
7	3	38.69- .74X *	31
8	3	36.97- .55X *	68
9	3	31.93- .08X	30
10	3	40.11- .46X *	66
11	3	41.39- .60X *	37
12	4	36.80- .91X *	34
13	4	40.16- .39X	69
14	5	45.42- .30X	32
15	6	55.64- .42X	41
16	6	50.32- .25X	53
17	7	52.46- .16X	59
18	8	43.40- .11X	32
19	8	57.56- .28X	35
20	9	48.91- .08X	59
21	10	41.76+ .02X	47
22	12	39.98+ .06X	49
23	13	49.51- .05X	30
24	14	41.71+ .05X	30
25	16	56.45- .02X	54

X = patient days

* = significant at 5 per cent level

TABLE III-5

FRANCISCO'S ESTIMATES OF THE RELATIONSHIP BETWEEN AVERAGE COST AND OUTPUT

(Grouped by Total Number of Services)

Total Number of Services	Estimate AC=	n
0	27.01+ .28X	28
1	32.36- .26X	167
2	37.28- .48X *	391
3	40.07- .51X *	457
4	41.54- .32X *	474
5	42.14- .19X *	464
6	44.47- .16X *	421
7	44.18- .09X *	375
8	46.75- .09X *	370
9	45.57- .05X	310
10	46.24- .03X	318
11	48.50- .05X	258
12	48.82- .04X	204
13	54.82- .07X *	202
14	53.46- .03X	136
15	49.05+ .01X	81
16	56.45- .02X	54

X = patient days

* = significant at 5 per cent level

average cost once he introduced dummy variables into the specification which represented services and facilities. Subsequently, he developed an unweighted index of services and facilities (EF) which he reported to be a satisfactory substitute for these dummy variables.

For his third analysis, Francisco used the twenty-five groups, containing 1,328 hospitals. He combined them into two groups; - one of hospitals offering five services or less and the other of hospitals offering six or more. In his linear estimates he included the unweighted index of services and facilities.

Small hospitals

$$AC = 31.21 - .37\,PD + 1.97\,EF \qquad R = .21$$
$$(-5.5) \qquad (4.8) \qquad n = 839$$

Large hospitals

$$AC = 39.03 - .01\,PD + .74\,EF \qquad R = .15$$
$$(1.4) \qquad (3.2) \qquad n = 489$$

All hospitals

$$AC = 29.04 - .3\,PD + 1.78\,EF \qquad R = .46$$
$$(-3.3) \qquad (15.8) \qquad n = 1,326$$

The results reflected decreasing average cost with increases in output for both the small and large hospitals. The rate of decrease was very small and not significant at the 5 per cent level for large hospitals. The coefficient for PD was larger for the smaller hospitals and it was significant at the 1 per cent level. When both large and small hospitals were combined, the negative PD coefficient was also significant at the 1 per cent level. Francisco concluded that only in small hospitals did it appear that output significantly influenced the behaviour of the average cost function.

Francisco concluded from his three analyses that the relationship between average cost and output was weak and negative which indicated a weak L-shaped average cost curve.

Robert Evans[23]

Evans' study represents the third Canadian study to be considered in this review. His sample consisted of 186 Ontario hospitals. Evans obtained access to this data in his capacity as a senior researcher with the Ontario Hospital Services Commission.

Evans replicated the technique of M. Feldstein[24] discussed above, using Ontario data. Some of the functions he estimated reflected the variation in the proportions of the different types of cases treated in Ontario. Initially, fourty-one diagnostic categories were represented in the proportions.[25] He eventually employed only thirty categories because he found that the estimated size coefficients yielded 'nonsense' values with fourty-one categories. Each diagnostic category was adjusted to reflect the age and sex composition of the patient load. The categories were converted to proportions which reflected the percentage of total days of care necessitated by each type of case.

Evans estimated average cost functions for both average cost per

---

[23] R. Evans, 'Behavioural' Cost Functions for Hospitals, Discussion Paper No. 38, Department of Economics, (Vancouver: University of British Columbia, 1970).

[24] M. Feldstein, Economic Analysis.

[25] A list of the fourty-one categories appears in Footnote No. 11, R. Evans, 'Behavioural' Cost Functions, (no page number shown). These same fourty-one categories are used by O.H.S.C. to prepare the Relative Stay Index for Ontario hospitals by which hospital performance is evaluated.

case (Casex) and average cost per day (Dayex) focusing on inpatient care only. Evans calculated inpatient costs by subtracting the direct costs of all non-inpatient activities reported by each hospital from its total expenditure. These included education, research and out-patient care costs. He did not allocate joint costs generated partly by these activities such as the costs of housekeeping, maintenance services, senior administrative salaries and other 'hotel' and administrative services. (This likely caused overstatement of inpatient costs.) Evans also excluded depreciation, interest and other capital costs reported by the hospitals from the total cost figure.

Evans selected the total number of rated beds to represent scale. He estimated equations having the two forms:

1) (Casex) $= a + B_1X + B_2X^2 + B_3Y + B_4Z$; and

2) (Dayex) $= a + B_1X + B_2X^2 + B_3Y + B_4Z$ where

Casex = average cost per case,

Dayex = average cost per day,

X = number of rated beds in each hospital,

Y = total occupancy rate,

Z = acute care average length of stay.

When Evans employed Casex as the independent variable, he obtained the following results. These equations marked * were adjusted for casemix.

Equation	a	X	X^2	[illegible]	Z	R^2	n
*1						.805	185
*2						.796	185
3	321.6	.038 (2.3) .02				.024	185
*4		.158 (3.2) .05				.808	185
5	266.5	.327 (6.3) .05	$-.336 \cdot 10^{-4}$ (5.9) 0.0			.174	185
*6		.16 (1.7) .09	$-.253 \cdot 10^{-5}$ (.03) 0.0			.806	185
7	45.7	.173 (7.3) .02		-1.5 (3.1) .48	35.8 (24) 1.49	.799	185
*8		.101 (1.3) .08	$-.744 \cdot 10^{-6}$ (.01) 0.0	-1.3 (3.2) .42	30.1 (8.9) 3.4	.877	185

Equation 1 was adjusted for fourty-one diagnostic categories and equations 2, 4, 6 and 8 were adjusted for thirty categories. The slight difference between the values of R^2 for equations 1 and 2 demonstrated that little explanatory power was lost when thirty categories were used instead of fourty-one. The coefficients of X and X^2 were very small in every estimate and never significant at the 5 per cent level. They also did not add very much to the explanatory power of the specification. Evans concluded that flat long-run average cost curves appeared likely based upon his Casex results.

Evans estimated a similar set of equations for Dayex:

Equation	a	X	X^2	Y	Z	R^2	n
*1						.467	185
*2						.509	185
*3	30.3	.001 (2) 0.0				.017	185
*4		.008 (4.1) .02				.553	185
5	27.9	.0138 (9.6) .001	$-.148 \cdot 10^{-5}$ (9.3) 0.0			.329	185
*6		.0135 (11.0) .001		-.0975 (3.9) .02	-.0134 (1.7) .01	.397	185
*7		.0008 (.20) .004	$.622 \cdot 10^{-5}$ (1.8) 0.0	-.0886 (4) .02	.2119 (1.3) .16	.605	185

Of those estimates marked * equation 1 was adjusted for fourty-one diagnostic categories and the balance for thirty. The results were similar to those with Casex, although the degree of explanatory power was generally less overall. The coefficients of X and X^2 were again very small; their small standard errors forewarned that when 't' values were calculated they would not be significant. The influence of the number of beds on the average cost per day was virtually nil. Evans concluded that the average cost curve for Dayex was U-shaped reaching a minimum of about 300 to 400 beds and then rising rapidly after 1,000 beds.

It appeared from the analyses of both Casex and Dayex that the that the casemix was related in some manner to changes in average cost although the relationship was weaker with Dayex. The rate of occupancy and the average length of stay also were related to changes in average cost although these relationships were not often significant. A review of the very small values for all of the size coefficients and their universal lack of significance at any acceptable level begs the question as to whether Evans should have made any conclusions at all regarding the relationship between size and average cost. Evans results per case were similar to those obtained by M. Feldstein in size, direction and significance, however, Evans avoided replicating his lengthy speculations.

IV Average Cost Time Series Literature

Kung Ro[26]

In the introduction to the report of his analyses, Ro hypothesized that certain factors logically were related to the cost of hospital care. The six factors were: 1) size-volume, 2) capacity utilization, 3) scope of services, 4) technology, 5) exogenous variables, and 6) educational programs. In his analysis, he considered the effects of only the first four of these factors.

Ro's first step was to select surrogate measures to represent

[26] K.K. Ro, "Determinents of Hospital Costs", Yale Economic Essays, Vol. 8, Fall, 1968, P. 185-256.

each of the four factors. With regard to the size-volume factor, he speculated that four possible surrogates existed including patient days, number of beds, number of admissions and births. By means of step-wise regression analysis, he found that the number of admissions best represented the size-volume factor. He similarly developed surrogates for each of the other three factors.

Ro employed data which reflected the operations of sixty-eight western Pennsylvania general hospitals over a period of eleven years. The hospitals ranged in size from thirty-six to 794 beds. Ro used inpatient expenses only in calculating the operating cost, although, he did not explain his cost allocation method. (Presumably, Ro has estimated an average variable cost function, since he included no depreciation or capital costs. Also, his total cost calculation possibly over stated the costs associated with inpatient services since joint costs were not likely allocated fully.) Ro divided the total cost of inpatient services by two different measures of output to obtain his dependent variable. He combined the data from each hospital for each of the eleven years with the data from all of the other hospitals in the sample. He thus specified a 'linear additive form' for the relationship between average cost and the independent variables. The basic model Ro used had the following form:

$$Y_{ht} = a + b_1A_{ht} + b_2O_{ht} + b_3P_{ht} + b_4X_{ht} + U_{ht} \quad \text{where;}$$

$h = 1,2,\ldots H = 68$ (hospitals)

$t = 1,2,\ldots T = 11$ (years)

Y = average cost

a = constant

A = number of admissions (size-volume surrogate)

O = occupancy rate (capacity utilization surrogate)

P = ratio of total patient care expenses to inpatient operating expenses (scope of services surrogate)

X = patient days per personnel (technology surrogate)

U = random error

Ro reported the results of three estimates of the average cost function. The result of the first estimate in which Y_1 equals the average cost per patient day was:

$$Y_1 = 29.64 - \underset{\substack{(.0094)\\1.54}}{.0145\,A} - \underset{\substack{(.0096)\\7.53}}{.0721\,O} + \underset{\substack{(.0241)\\5.36}}{.1291\,P} - \underset{\substack{(.0032)\\11.03}}{.0356\,X} \qquad R^2 = .892,\; n = 68$$

The relationship between Y_1 and the size-volume factor was negative, relatively small and not significant at the 5 per cent level. The three remaining coefficients were significant at the 1 per cent level.

In his second analysis of the same data, Ro introduced two dummy variables which altered the results somewhat. One variable reflected the proportion of the population served which was urban (U) and the second indicated whether or not a school of nursing was associated with the hospital (N).

$$Y_1 = 22.36 - \underset{\substack{(.0092)\\2.13}}{.0196\,A} - \underset{\substack{(.0094)\\7.37}}{.069\,O} + \underset{\substack{(.0238)\\6.34}}{.1512\,P} - \underset{\substack{(.0032)\\10.96}}{.0346\,X} + \underset{\substack{(.034)\\2.24}}{.0763\,U}$$

$$+ \underset{\substack{(.4252)\\5.36}}{2.2814\,N} \qquad R^2 = .897,\; n = 68$$

Once again, the relationship between Y_1 and the size-volume factor was negative and relatively small, however, it was significant at the 5 per cent level in this estimate. All of the remaining five coefficients were also significant. The addition of the urbanization and nursing variables only slightly improved the explanatory power of the equation.

The third estimate of the average cost function was specified in a slightly different form than the first two. In this estimate, Ro used the inpatient expense per admission as the dependent variable Y_2. He also dropped occupancy rate (O) as an independent variable representing capacity-utilization and replaced it with turn-over rate (T).

$$\begin{array}{llllll} Y_2 = 212.48 - & .2982\,[illegible] & - 1.813\,T & + 1.8265\,P & - .1985\,X & R^2 = .918 \\ & (.1051) & (.1933) & (.2616) & (.034) & n = 68 \\ & 2.84 & 9.38 & 6.98 & 5.84 & \end{array}$$

The negative relationship between Y_2 and the size-volume factor was significant at the 1 per cent level; the other coefficients were similarly significant. (The interpretation of this relationship is that the average cost of one admission is equal to its marginal cost plus the three other variables. It is no wonder the value for R^2 is so elevated and the relationship so highly significant.)

The linear relationship between average cost and the size-volume factor indicated that the average cost curve was L-shaped and that average costs were declining. Ro concluded that the results strongly indicated that economies of scale were present in his sample of hospitals and that hospital size was related positively to efficiency. The greater efficiency of larger hospitals Ro accredited to a greater degree of specialization and the employment of better levels of technology. Ro ignored the fact that his linear estimate reflected only the strongest relationship between average cost and size-volume and that it could have concealed a positive relationship between these two variables among large hospitals. His result did not in any way show that U-shaped average cost curves did not exist for the sample he used in the long-run.

Judith and Lester Lave[27,28]

In June, 1970, Lave and Lave reported the results of their analysis of average cost in the hospitals of Pennsylvania in two separate journals. The first and main part of their work was reported in the American Economic Review (AER) while that reported in Inquiry replicated the analysis using different data.

The primary aim of Lave and Lave was to develop a method of estimating the hospital average cost function which would overcome the problem associated with the multi-product nature of hospital output. They wanted to avoid the difficulty of being unable to estimate a 'true' specification. They did not want to use an ambiguous surrogate output measure which did not reflect the quality of the various hospitals' outputs. The Lave's technique was to finess the problem associated with multi-product output by making the assumption that over a short period of time the mix of hospital outputs would remain constant. They concluded that if the variety and quality of the output was constant in each hospital they could estimate the short-run average cost function omitting 'product-mix' altogether from the specification. Finally,

27 J.R. Lave and L.B. Lave, "Hospital Cost Functions", American Economic Review, Vol. 60, No.3, June, 1970, P. 379-395.

28 J.R. Lave and L.B. Lave, "Estimated Cost Functions for Pennsylvania Hospitals", Inquiry, Vol. 7, No.2, June, 1970, P. 3-14.

they assumed that hospitals could be characterized by a single cost function and they pooled their data to estimate a cross-section, time series model.

In the American Economic Review, Lave and Lave applied their technique to data gathered from seventy-four Western Pennsylvania hospitals over a period of seven years. Fourteen semi-annual 'observations' were made of the costs reported to Blue Cross of Western Pennsylvania by the individual hospitals. It was unclear in their description of the cost data whether or not it represented inpatient costs only or all hospital generated costs including depreciation or cost of capital.

Of the analyses reported in AER, two were relevant to this thesis. In the first, a time series analysis, the Laves specified an equation of the form:

$$\log AC_{it} = a_{it} + a_{i1} + a_{i2} \log U_{it} + a_{i3} \log S_{it} + e_{it}$$

where: i = the i th hospital,

t = aggregate time dummy,

U = utilization rate,

S = size expressed in terms of beds, and

e = random error.

They specified the function in this form to be consistent with the shape that they expected the short-run average cost curve to exhibit. i.e. L-shape. The Lave's estimated an average cost function for each of the seventy-four hospitals. They did not display these results but discussed them briefly in a narrative. They reported that in not one estimate was the value of R^2 less than .9. The coefficient of size ranged from -13.4 to 8.5 with a mean of -.333 and a standard deviation of 2.418. In almost none of the seventy-four estimates did they find the size coefficient to be significant. They concluded that they could say very little about the shape of the short-run average cost curve from these results.[29] (They did, however, persist in making a conclusion in this regard in their final conclusions. This is noted below.)

Lave and Lave next reorganized their initial model. They amalgamated the time-series data from all of the hospitals and they transformed AG, U and S into the index number I, J and K respectively. Their basic specification was:

$$\log I_{it} = a_{io} + a_{i2}t + a_{i2} \log J_{it} + a_{i3} \log K_{it} + e_{it}.$$

Upon this model, they built two expanded specifications. In one, the linear time variable was replaced by six annual dummy variables. In the second expansion, they also added dummy variables to represent hospital teaching status, location and the initial values observed for size, utilization and average cost. These variables were:

AT = advanced teaching hospital,

T = teaching hospital,

P = central city location,

M = general metropolitan location,

AC = average cost,

U = utilization rate,

[29] In a secondary analysis, the Lave's determined that size was positively and significantly related to the rate of cost increases in hospitals.

S = size in terms of beds.

~~The results for these three specifications were:~~

1) $$\log I = 6.67 - \underset{-11.41}{.357} \log J - \underset{-.22}{.006} \log K + \underset{62.19}{.031}t + \underset{14.81}{.028}F \qquad R^2 = .855,\ n = 1036$$

2) $$\log I = 6.503 - \underset{-11.74}{.317} \log J - \underset{-.15}{.003} \log K + \underset{8.05}{.021}t_1 + \underset{16.18}{.042}t_2$$
$$+ \underset{24.98}{.067}t_3 + \underset{34.03}{.096}t_4 + \underset{45.36}{.130}t_5 + \underset{66.56}{.195}t_6 + \underset{18.03}{.029}F \qquad R^2 = .894,\ n = 1036$$

3) $$\log I = 6.638 - \underset{-12.42}{.334} \log J - \underset{-.33}{.007} \log K + \underset{3.18}{.045}t_1 + \underset{3.25}{.091}t_2$$
$$+ \underset{3.33}{.139}t_3 + \underset{3.45}{.192}t_4 + \underset{3.6}{.25}t_5 + \underset{4.07}{.34}t_6 - \underset{-2.63}{.008}t(\log AC)$$
$$+ \underset{.33}{.001}t(\log U) + \underset{.32}{.002}t(\log S) + \underset{2.45}{.007}t(AT) + \underset{3.14}{.002}t(T)$$
$$- \underset{-5.95}{.041}AT(\log S) - \underset{-3.6}{.006}T(\log S) + \underset{1.56}{.008}AT(P) + \underset{2.08}{.013}AT(M)$$
$$+ \underset{2.27}{.009}T(P) + \underset{2.43}{.008}T(M) - \underset{-1.7}{.001}t(P) - \underset{3.64}{.002}t(M) + \underset{2.24}{.025}F$$
$$R^2 = .9,\ n = 1036$$

The high value of R^2 in equation 1) indicated that the specification fitted the pooled time-series, cross-sectional data well and the significant F-value indicated that no explanatory power was lost when the data was pooled. In each of the three specifications, the size variable (K) was very small, negative and not significant. K was almost totally unaffected by the addition of the different variables in equations 2 and 3.

The main purpose of these analyses was to demonstrate that the technique of overcoming the problem of the multi-product output was valid. The Lave's showed that little explanatory power was lost when the time-series data was combined to provide a cross-sectioned study. A secondary 'spin-off' of their work related to the shape of the average cost curve. From results of their time-series analyses of the seventy-four hospitals, the Lave's concluded that in the short-run the average cost curve was L-shaped. They concluded from their analyses of the pooled data that economies of scale were not very strong in hospitals over the long-run if they existed at all. i.e. the long-run average cost curve was nearly flat.

In Inquiry, Lave and Lave again employed the seventy-four hospitals from western Pennsylvania they had used in AER, but they also included data from thirty-five hospitals in eastern Pennsylvania. For both samples, they obtained annual observations over an identical seven year period. The purpose of the work reported in Inquiry was to again test their technique of analyzing cost when output was assumed to be constant. They estimated additional functions for each of the three models introduced in AER which used pooled, cross-section, time-series data.

Lave and Lave performed regression analyses, initially using cost data from western Pennsylvania hospitals only. With each of the three specifications the coefficient of size was very small and never significant. They repeated these analyses using only the eastern Pennsylvania data and they found that the K coefficient was larger, negative and significant at the 1 per cent level for each form of the expression. Finally, the Lave's combined the cost data from the two portions of Pennsylvania and then re-estimated each of the three specifications. (Their results for these estimates are shown below but only the values for the constant utilization and size terms are reported here.)

1) $$\log I = 6.782 - \underset{-9.12}{.335} \log J - \underset{-1.59}{.047} \log K \qquad R^2 = .862,\ n = 763$$

2) $$\log I = 6.578 - \underset{-10.24}{.334} \log J - \underset{-4.06}{.119} \log K_1 + \underset{-.04}{.001} \log K_2 \qquad R^2 = .896,\ n = 763$$

3) $$\log I = 6.692 - \underset{-10.98}{.346} \log J - \underset{-3.7}{.179} \log K_1 - \underset{-.18}{.005} \log K_2 \qquad R^2 = .909,\ n = 763$$

where K_1 = eastern Pennsylvania hospitals, and

K_2 = western Pennsylvania hospitals.

In the basic specification (#1), the value of the K coefficient was not significant where the data was combined completely. In the second and third expressions, the affect of size in the two parts of Pennsylvania were reported as separate independent variables. In both specifications, the two size coefficients were negative but only K_1, representing eastern Pennsylvania hospitals, was significant, (at the 1 per cent level).

Lave and Lave concluded that their technique permitted the study of various aspects of cost inflation in hospitals and that it had universal application to other hospitals. Specifically, with regard to the relationship between size and average cost, the Lave's concluded that the short-run average cost curve was L-shaped and that in the long-run, if economies of scale do exist, they were not very strong. These two conclusions were identical to those they reached in the American Economic Review.

The Lave's and Ro both employed cost data from hospitals in western Pennsylvania. Both specified an average cost function and both combined time-series and cross-sectional data in their estimates. Their long-run conclusions with regard to the shape of the average cost, however, were different. Lave and Lave concluded the curve was nearly flat, whereas Ro concluded it was L-shaped. It should be observed that these two analyses utilized different methods in initially specifying their respective equations. Ro performed covariance analysis to obtain the factors he used as independent variables. The Lave's simply specified variables which they assumed to be valid. Finally, Ro's specification was linear while Lave and Lave's was second degree. The different results obtained in these two studies appear to be related in their different specifications.

CHAPTER IV

COMMENTARY ON THE LITERATURE[1]

The literature which was reviewed in Chapter III will, initially, be criticized and compared in this chapter. Subsequently, a general discussion will be presented along with some comments and observations.

Criticism of the Literature[2]

As Chapter III demonstrated, the economic literature related to the hospital field is varied with regard to both methods and results. Such variety may possibly imply that the use of economic techniques of analysis with hospital data is in an early development stage. It might also indicate that problems exist in adequately defining parameters representative of hospital activities. (It is possible that the presently existing techniques of analysis could adequately reflect hospital behaviour if satisfactory parameters were available.) In spite of the variety, it should be noted that four studies developed out of previous

---

[1]Recently an unpublished Ph.D. disertation by Dr. D.M. Baker was brought to the author's attention. This thesis dealt with disaggregate hospital cost analysis. A brief discussion of this work is contained in Appendix II. See D.M. Baker, Hospital Cost Control and Product Pricing with an Empirically Derived Factor Analytic Model of Inpatient Output, unpublished Ph.D. disertation, (Los Angeles, University of California, 1973).

[2]Throughout Chapter IV reference will be made to the various pieces of literature reviewed in Chapter III. Only where the reference is to a specific part or to a direct quotation will it be annoted.

work either wholly or in part. Carr and Feldstein conducted a more extensive analysis using the aggregate, cross-sectional approach Paul Feldstein initially had introduced. Fraser repeated Cohen's analysis with the composite measure of output S^k but he substituted Canadian for New York City data. Francisco stratified his sample by like combinations of services and facilities in the same manner as Berry; and Evans replicated Martin Feldstein's study of hospital cost per case including an adjustment for casemix.

Criticism of the literature will relate to the sample data, specifications and parameters. Specific references will be made to the literature where it illustrates a point.

Sample Data

Degree of Aggregation

Throughout the literature aggregate data was most frequently employed. Carr and Feldstein, Berry and Francisco each used the entire population of U.S. general hospitals. Fraser considered one sample which included all types of Canadian hospitals and another representing all of the acute general hospitals in Canada. Some studies were less aggregate because the geographic origin of the samples was smaller. For example, Lave and Lave, Ro, Evans, Ingbar and Taylor, Kushner and Cohen sampled hospitals in specific states or provinces. The literature generally appears to have relied almost entirely upon existing sources of hospital cost data. These sources were almost always third party payment agencies which collected accounting data for the purpose of reimbursing hospital costs. Such data did not necessarily reflect hospital operations from the economic viewpoint.

Certain analysts attempted to reduce somewhat the aggregate nature of their data. Martin Feldstein, who studied British hospitals, included only non-teaching institutions in his sample. Kushner considered only the accredited hospitals of Ontario in one of his analyses and Ingbar and Taylor the accredited Massachusettes hospitals only. Both Berry and Francisco reduced the size of their samples by considering only those hospitals with specific combinations of services and facilities. Carr and Feldstein, Fraser and Francisco stratified their samples according to bed size. Only two disaggregate analyses were to be found in the literature.[3] One was Paul Feldstein's time-series analysis of one hospital by department. He estimated a total cost function for each department and then totalled them to obtain an approximate total cost function for the overall hospital.

Nature of the Data

Often in the literature, the scale range of the hospitals included in a sample was omitted. Reference to Table III-1 will show where in the literature range was not reported. While it may be argued that the 'number of beds' is questionable as an effective measure of scale, if the results and conclusions are to be evaluated and given perspective, then one needs to know the scale range of the sample. To illustrate the significance of this point, it might be observed that Ingbar and Taylor and Ro, who both reported the ranges of their respective samples, failed to observe the limitation inherent in a range. In both studies it was concluded that economies of scale are continuous over the entire

[3] The second disaggregate study was that of Dr. D.M. Baker which is discussed in Appendix II.

range of possible hospital sizes. The largest hospital in Ro's sample was 794 beds and in the other 330 beds. Both Ingbar and Taylor and Ro should have qualified their conclusions. The reported economies were continuous only within the range of hospitals represented by their samples. In his cross-sectional analysis of sixty hospitals, Paul Feldstein recognized that he could not make a valid conclusion regarding cost behaviour in hospitals larger than 453 beds as this was the size of the largest hospital in his sample.

Geography is another factor which should temper general acceptance of results and conclusions. Martin Feldstein used British hospital data. Fraser, Evans and Kushner employed Canadian samples and the remaining literature was of United States origin. Some of these used samples from specific states or provinces. The results of these studies reflected various forces present within certain nations and which likely were not shared identically with any other country. Such forces could have included the historical development of a nation's hospitals, the political attitudes within a country, the means by which hospitals were financed and the various alternative forms of health care emphasized within each country. Aggregate, cross-sectional studies of hospitals in a particular nation could well produce results not at all related to situations in any other country. This comment merits consideration before the results of foreign analyses are applied to the Canadian hospital field.

Specifications

Form

In each piece of literature a relationship has been specified between cost and some measure of hospital scale. The specifications have taken three forms. Reference to Table III-1 will show that six analysts specified linear functions. (Fraser also specified a curvilinear form of the relationship for some of his estimates.) The balance of the literature was curvilinear with all but Kushner specifying equations of the second degree. Kushner utilized a specification of the third degree.

A linear specification implicitly assumes that average cost will be constant or that it will increase or decrease at a constant rate as scale increases. The resulting average cost curve in every linear analysis in the literature sloped downward to the right indicating a negative relationship. Continuous economies of scale were either concluded to exist or else inferred. It is questionable, however, that the linear form adequately reflected the true cost-scale relationship. A linear specification is capable of representing general cost behaviour and may well conceal other behaviour. It is probably that the majority of hospitals are of small or medium sizes and it is possible that economies of scale are present in hospitals of these sizes. Large hospitals are fewer in number but diseconomies of scale could exist in their activities. In aggregate, cross-sectional analysis, the influence of the small and medium-sized hospitals, if they are more numerous, may overwhelm any diseconomies in large hospitals in a linear specification.

Specifications of the second degree permit examination of the possibility that the relationship between cost and scale varies with changes in scale. In addition to the three results possible using a linear specification, a second degree estimate also permits the influence of economies and diseconomies to be considered as separate forces.

The results reported in the literature where second degree specifications were employed were varied. Carr and Feldstein, Cohen and Martin Feldstein found the average cost curve to be U-shaped, although Feldstein reported that the U-shape was weak in his analysis. Fraser found the average cost curve was generally L-shaped for most of his second degree estimates. Evans concluded the curve was flat and Lave and Lave surmised that the average cost curve had a weak L-shape. Ingbar and Taylor reported the long-run average cost curve had the shape of an inverted-U. As the literature demonstrated, equations of the second degree can yield a convex or concave curve or possibly a straight line. Third degree estimates, such as Kushner's, permitted the analyst to observe whether or not there were two trends within the overall relationship between scale and cost. Despite estimating this more complex form of the relationship, Kushner concluded that the long-run average cost curve was L-shaped. This result might be a preliminary indication that equations of the second degree are adequate for the study of the relationship between hospital cost and scale.

Reflecting Quality

The quality of service and patient care likely varies between individual hospitals, between different types of hospitals and between hospitals of different sizes. Quality appears to have a somewhat ambiguous relationship to cost in hospitals. Much of the literature attempted to adjust for quality variations with limited results.

Ingbar and Taylor and also Kushner, in his first study, included only accredited hospitals in their samples on the assumption that accreditation was a guarantee of quality. Accreditation, however, is only a guarantee that certain minimum standards have been met. It may be that the quality of care in accredited hospitals is generally superior to that in non-accredited hospitals, however, it is not likely that quality is necessarily completely absent in hospitals which lack accreditation. Within a group of accredited hospitals, quality may vary widely between individual hospitals because accreditation is only a guarantee of a minimum standard.

Another approach observed in the literature was to make the yet unproven assumption that a hospital associated with an educational program necessarily provided a better quality of care than those which were not. (The question of whether or not the presence of students in a hospital acts as a catalyst to improve quality remains unanswered. It is possible that their affect is to reduce quality.) The next step was to introduce dummy variables into the specification to represent 'quality'. Carr and Feldstein and Cohen both used 'affiliation with a medical school' and Carr and Feldstein, along with Fraser, included the sponsorship of a school of nursing and an intern program. Kushner inserted a dummy to represent a school of nursing in his specification. He also used the ratio of graduate nurses to the total nursing staff in order to represent that the hospitals with more trained nurses on staff provided better quality care. Carr and Feldstein also represented quality by including dummies reflecting the absolute number of interns and residents in a hospital and the number of types of programs for interns and residents. Ro utilized a dummy representing a nursing school to indicate quality. Lave and Lave avoided the question of quality altogether through the technique of omitting output from their specification. This step followed from their assumption that output does not vary in any way in a short term period.

Ro included a dummy variable in his specification which represented the ratio of patient days to total hospital personnel. Presumably, Ro assumed that care was improved when the total number of staff grew faster than the number of patient days, irregardless of occupation. In three pieces of literature, the question of quality was not considered in the specifications at all; Paul Feldstein, Berry and Francisco. In the work of Berry and Francisco, where they stratified their respective samples by the offered combinations of services and facilities, they might have assumed that those hospitals which offered more service, therefore, offered better quality but this is a questionable assumption.

The approach of Martin Feldstein and also Evans was to assume that quality was reflected by the nature of the casemix. Those hospitals with a more difficult casemix appeared to be the ones with the more sophisticated equipment and more highly trained staff necessary to handle it. These hospitals presumably produced a better quality of care. Dummy variable(s) introduced into the specifications represented the casemix and adjusted output for quality.

Cost would appear to be some function of quality. Ignoring quality or simply equating it with educational programs does not provide an adequate representation of the relationship between cost and size. Of the specifications considered, the ones in which casemix was part of the specification are those in which quality appeared to be reflected most satisfactorily.

Parameters

The Nature of Cost

There were problems observed in the literature related to the nature of cost; specificly in two areas: 1) operating cost, and 2) capital costs.

i Operating Cost - Many hospitals offer more services and facilities than simply those required to treat inpatients only. Some services and facilities are shared by inpatients with patients who may be otherwise classified. These alternative forms of care could include outpatients and ambulatory patients; educational, day care, home care or self-care programs; mobile treatment units; and even hospital-based community health centers. The problem which arose in the literature was that costs generated by overall hospital operations were used in specifications in which the measure of scale reflected inpatient activity only. e.g. patient day, beds or admissions. It is possible that a biased result occurred where costs were not generated by inpatient services and facilities alone.

If larger hospitals offer proportionately more non-inpatient services than smaller ones, a biased estimate will occur where the cost used reflects more than simply inpatient services. It should be noted that Carr and Feldstein, Fraser, Cohen, Ro and Kushner did attempt to reduce this bias. They inserted various independent variables into their specifications including the number of outpatient visits, the presence of a nursing school or intern program or the number of services provided. They adjusted their independent variables and left cost, the dependent variable, intact. These adjustments were inadequate. Throughout the literature, cost generally did not appear to be related in the various specifications to the activities which actually generated the cost. In summary, it appears that the existing body of hospital economic literature underestimated the economies of scale in large hospitals.

ii Capital Costs - Fixed assets such as a physical plant and items of capital equipment depreciate in value with time and use. The cost of capital is part of the overall economic cost of the provision of hospital services. In the literature, capital costs were generally overlooked, perhaps because such data was not readily available. Throughout almost all of the literature, the cost data was obtained from a third-party payment agency which recorded only accounting costs for purposes of reimbursing operating costs to hospitals. Only Fraser and Kushner (in his first analysis using hotel costs) included the cost of capital in their estimates. Neither provided the detail of how they evaluated capital goods nor did they explain their respective methods for allocating annual depreciation, except in one of the methods Fraser employed. (He added one-tenth of the value of the fixed assets onto operating cost.) Fraser's estimates of the average cost function were similar whether capital costs were included or not.

Another general criticism of the literature is also related to the omission of capital cost. With two exceptions, the literature estimated variable cost functions; whether average or total cost; long-run or short-run. Only Fraser and Kushner included capital cost and, therefore, only they estimated 'true' average and total cost functions respectively. The results and conclusions of the remaining literature ought to be considered as estimates of variable cost functions rather than as the total and average cost functions they purported to be. It was a difficult task, however, to judge the amounts of capital by which these cost functions were understated without some detail of the capital costs. The large size constants in the various specifications indicated, in part, that these amounts could well have been substantial.

Omitting capital cost may also produce further bias in the results unless both large and small hospitals have identical ratios of output to capital. If this ratio is greater in larger hospitals than in smaller ones and the cost of capital is omitted from the estimated function, diseconomies of scale in larger hospitals may be understated. The reverse is true if the ratio is greater in smaller hospitals. In present day health care, the assumption may well be true that larger hospitals generally have bigger physical plants than small hospitals and also that they tend to have more items of equipment of a complex and costly nature. If this assumption is true then quite likely some diseconomies of scale associated with large hospitals are understated in the literature.

The Measurement of Scale

Every hospital produces a variety of services. The problem in specifying a cost function is to accurately reflect the relationship between cost and scale. Throughout the literature, scale has been represented in two different fashions. It has either been reported in terms of output or of capacity. The measures of capacity were employed both adjusted and unadjusted for output. The following lists show how the literature represented scale.

Output	Capacity	Both
Paul Feldstein	Ingbar and Taylor	Fraser
Cohen	Martin Feldstein	Kushner
Carr and Feldstein	Evans	
Berry	Lave and Lave	
Francisco		

i <u>Output</u> - The most commonly encountered measure of output in the literature was the 'patient day', although the 'number of admissions' also appeared. These two measures were used as gross representations of the total volume of hospital services produced. Their use indicated that an assumption of questionable validity had been made. Such output surrogates implied that they were standard units of output each representing a certain quantity and quality of hospital production and that these units were comparable between hospitals. Neither of the surrogates reflected the various hospital services nor the effort directed at patients who were not in the inpatient classification. Neither did they permit recognition of variation in the combinations and quantities of inpatient services which could be provided by different hospitals. For example, the services embodied in one patient day unit at a Red Cross outpost hospital are not identical to that of a large metropolitan teaching facility. The staff and equipment in the urban hospital would be more extensive and more sophisticated than at the outpost. Similarly, although less dramatic, variations occur between all hospitals because no two hospitals are exactly alike.

Hospitals may be defined as multiproduct firms. The services they produce are numerous and include education, research and community services, in addition to various types of patient care programs and services. Hospital output is heterogeneous and it varies over time within an individual hospital and it varies between hospitals. One assumption which underlies the theories of production and cost is that all output can adequately be measured in terms of one standard unit. This implies that a firm only produces one good or service. The output of the hospital field has been described as, "Unmeasurable and even to some extent undefineable".[4] Neither the patient day nor the number of admissions is an adequate measure of the heterogeneous output of hospitals unless the degree of heterogeniety in the sample is virtually nil.

Two items in the literature undertook to develop an improved output measure for scale. In his analysis using total 'hotel' costs, Kushner measured output in terms of patient days, however, he made an adjustment which was designed to reflect the effort a hospital made to maintain empty beds. Such effort, he claimed, was part of the total

[4] J.R. Lave, "Review of Methods Used to Study Hospital Costs", Inquiry, Vol. 3, No.2, May, 1966, P. 59.

hospital effort. Kushner stated that the cost of an empty bed was seventy-five per cent of that of an occupied bed irregardless of the level of hospital operation.[5] He adjusted his output surrogate to reflect the empty beds by adding an amount equal to seventy-five per cent of the total empty bed-days to the total number of patient days. Kushner's adjustment implied that a hospital would produce output whether or not there were any patients in its beds. Stated in other terms: an empty hospital would produce output at a rate equal to seventy-five per cent of the amount it would produce if it was full. Kushner's adjusted output was specified in a function with hotel costs only. It reflected his arbitrary assumption that treatment and education costs were only twenty-five per cent of the cost of a patient's daily stay. The usefulness of Kushner's adjustment is doubtful, because the purpose of a hospital, hopefully, is to provide service to its patients. It is difficult to accept a concept in which an empty hospital is considered to be productive.

Cohen approached the problem of hospital output another way. He computed a composite measure for output, S^k, which was equal to the sum of the indexed weights of thirteen services multiplied by the volume produced of each service. A weight of 'one' was assigned to the average cost of a patient day and the other twelve services were assigned index values according to the relative size of their average costs. (The concept of a weighted measure which would reflect the relative volume of

---

[5] Joseph Kushner, Economies of Scale in the General Hospital Industry, unpublished Ph.D. disertation, (London, Ontario: University of Western Ontario, 1969), P. 35.

output in each service deserves further research.) Cohen's method was unfortunate, however, because he has weighted each service by its average cost rather than the actual amount of effort required to produce one unit. He specified a relationship between total cost and an output surrogate calculated directly from cost rather than output. Naturally, the R^2 values for his functions were high reflecting the close movement of changes in total cost and S^k. The average dollar costs did not appear to necessarily reflect 'logical' values in terms of either resource inputs[6] or patient benefits. Cohen, for instance, assigned a weight of .9 to a blood transfusion but laboratory and radiology examinations were .07 and .16 respectively and an outpatient visit had a weight of .28. Cohen's weights were questionable in themselves aside from the conceptual problem of estimating total cost from units derived from average cost.

Similar criticism and comment may be made with regard to the functions that Fraser estimated in which he employed the composite measure of output S^k developed by Cohen. Fraser adopted Cohen's weights without question and applied them to his Canadian data. He found that functions in which S^k was used yielded the highest R^2 values of any of his four scale measures and, as a result, he employed S^k most frequently in his specifications. Fraser not only employed a technique which was faulty in itself but he implicitly accepted the weights used to derive it. These weights were derived under a specific set of operational conditions in a select group of New York City hospitals. It was highly unlikely that the weights reflected Canadian values for these

[6] Baker, Hospital Cost Control, As noted in Appendix II this analyst attempted to develop an output measure based on the value of resource inputs utilized by disease groupings.

services.

ii <u>Capacity</u> - When output is used for scale, economic theory assumes implicitly that it has been produced at the minimum cost level. This is not true necessarily in a hospital. If capacity is used for scale, this assumption is not required. In each instance where a measure of capacity was employed in the literature the surrogate 'number of beds' appeared in some form. Although only Martin Feldstein specifically stated that he used capacity as a surrogate for scale in order to avoid the 'regression fallacy', this benefit accrued to each capacity study.[7]

In regression analyses of cost in which scale is represented by output and no adjustment is made for the depreciation of capital items, the influence of the regression fallacy is likely. It is probably an influence in all of the literature discussed above under i <u>Output</u> except for Fraser and Kushner. Hospitals operating at a level of production in excess of the optimal level for their scale of plant tend to consume their capital assets at a faster rate than hospitals which operate at their optimal level or below it. When depreciation is omitted from the calculation of total cost, hospitals which operate beyond their optimal level may appear to have lower average cost per unit of output than those operating at or below the optimal level. Economies of scale may be reported for such hospitals which in fact are not real because of the excessive consumption of capital.

M. Feldstein and Evans both reasoned that, while cost might well

---

[7] M. Feldstein, <u>Economic Analysis for Health Service Efficiency</u>, (Amsterdam: North-Holland Publishing Co., 1968), P. 60.

be related to the number of beds, it was also influenced by the nature and number of cases treated. They each inserted variables into their specifications which represented casemix and volume. Their results were similar. Each found that economies of scale were weak in hospital operations.

In his second analysis in which he used total operating costs, Kushner's capacity variable was 'potential patient days'. This measure was equal to the total number of beds multiplied by the number of days in one year. Such a capacity measure implied that hospital cost behaviour was neither influenced by the presence or absence of patients nor by the nature of their ailments. A similar observation was in order with regard to the work of Fraser and also of Ingbar and Taylor, although 'number of beds' alone represented scale in their studies. Studies of hospital cost behaviour in which a physical capacity surrogate was the only measure of scale ignored the influence of the patient on cost and their results were accordingly suspect. Lave and Lave attempted to avoid this difficulty by assuming that in the short-run the output-mix was fixed and need not be accounted for in a specification. In order to represent scale they turned to capacity but they also included utilization (output as a percent of capacity) in their analysis.

The capacity variable 'number of beds' possibly introduces a bias into the functions in which it is used by itself. The number of beds is an inpatient surrogate which does not reflect the variety of other services and facilities a hospital may offer. A hospital's capacity to treat patients is simply not a function of the number of its beds alone. The more services a hospital provides to patients other than inpatients, the greater the possibility when capacity is the scale surrogate that its average cost will be over-estimated and its economies of scale left unobserved.

Overall Criticism

The economic literature in the hospital field consisted primarily of studies of specific groups of hospitals at or during specific times. The techniques of analysis were generally too crude to adequately define output. Often functions were estimated only in the first or second degree with the result that possible cost behaviour patterns could have been overlooked. The majority of the results in the literature was not statisticly significant. Altogether, the limitations of the literature virtually eliminated any universal application to the practical problems encountered by hospital planners and administrators. The literature did provide some useful approaches and directions for future research. In particular, Paul Feldstein's disaggregate study of cost function of the departments of one hospital provided one such direction for future research. In only this one study did a writer avoid the assumption that the behaviour of every hospital could be characterized in one overall aggregate cost function.

Discussion

The hospital industry is a sub-sector in the health sector of the nation's economy. In the production of any commodity, including hospital services, there is an economic problem. The problem involves: a) determining the amount of society's limited resources that is to be used to produce the commodity; b) utilizing these resources to produce the kind of commodity the consumers desire; and c) efficiently organizing the resources to ensure that the amount of the commodity produced is the maximum obtainable for the resources expended.[8] If the optimal size for a hospital could be determined, then at least part c) of the economic problem, as it relates to the hospital field, could be solved. Presumably, the major purpose behind much of the literature was to determine the level at which hospital output was maximized with the available resources. (Parts a) and b) of the economic problem are beyond the scope of this thesis.)

---

[8] Report of the Ontario Committee on the Healing Arts, (Toronto: Queen's Printer, 1970), Vol. 1, P. 109.

The Bed

The Number of Beds as a Representation of Hospital Size

The number of beds in a hospital is commonly regarded as a measurement of hospital size. The 'bed' is a convenient and readily available measure. It can be easily related to the frequently used proxy measurement of hospital output: the patient day. The relationship between the supply of hospital beds and total hospital operating cost has attracted study. The total cost of hospital operations has been found to vary directly with the number of beds available to consumers.[9] This relationship has led to the conclusion that once a hospital bed is made available, it will be used and that the more beds made available the higher will be the cost in terms of the nation's resources. Hospital cost and the number of beds are closely related because staff and facilities are maintained, and are available at all times, in enough numbers to fully service most of the beds even when the occupancy rate is very low. In the short-run, variations in the occupancy rate have only minor effects on hospital operating costs.

It has been suggested that the ultimate control of hospital costs lies in limiting of the number of beds available for the hospitalization of patients.[10] This method of minimizing or restricting the total amount of expenditure for hospital services does not ensure that the resources

---

[9] M. Shain and M.I. Roemer, "Hospital Costs Relate to the Supply of Beds", Modern Hospital, Vol. 92, No.4, April, 1959, P. 71 and Committee on the Healing Arts, Vol. 3, P. 117.

[10] Shain and Roemer, "Hospital Costs Relate to the Supply of Beds".

expended are utilized in an efficient manner nor does it ensure the actual services produced are the maximum return to be expected from such an expenditure. If there was a known optimal-size for a hospital, and all hospitals were of this size, then the return on expenditures for hospital services would be at the optimal level. The amount the nation is prepared to spend to secure hospital services will always be limited. It is the job of hospital planners and administrators to maximize the return to the nation from its expenditure on hospital services.

Limitations of the Bed as a Size Measurement

When the 'bed' is used to represent the size of a hospital, the assumption is implicit that only inpatients are treated there. Hospitals do, however, offer a variety of services to patients who never occupy a bed. These services and facilities may occupy space, have access to hospital treatment and diagnostic departments and they incur costs which are unrelated to the number of beds.

It has been reported that the average number of square feet of floor space per bed in hospitals built in recent years is more than thirty per cent higher than those built before 1950.[11] The same source also found that larger hospitals have many more square feet available per bed than do smaller ones. Variations between hospitals in the number of square feet available per bed reveal that there is a wide variation in the value of the 'bed' measure between hospitals of different ages and sizes. The value of inter-hospital comparisons based on the

[11] *Hospitals*, Vol. 41, No. 2, February, 1967, P. 150-1.

'bed' is, therefore, limited. A hospital which services a smaller proportion of inpatients than another hospital in relation to their respective patient loads appears to be less efficient when cost per bed statistics are prepared. The scale of its operation is under-represented by this measure.

Hospital Output

One of the basic assumptions underlying the Theories of Production and Cost is that the output of the production process can be precisely measured in unit terms. The theories also assume that only one commodity is produced. The application of the theories of production and cost to hospitals is a more complicated task than in an industry having one easily identified and unitized output. The difficulty arises because the quality of service offered by hospitals is not homogenous over time nor between hospitals. The identification and treatment of quality variations is one of the most vexing conceptual difficulties facing economists performing empirical studies of hospital production. In 1970, the Ontario Committee on the Healing Arts bemoaned that: "Were 'units of hospital service' standardized, this difficulty, quality measurement, would be eliminated. But at present such a standardization is not available".[12]

The Nebulous Nature of Hospital Output

Hospital output is sometimes expressed in general terms such as

[12] Committee on the Healing Arts, Vol. 3, P. 116.

the restoration of health and alleviation of pain.[13] Such nebulous terms are of limited use as operational definitions of output. Their imprecision makes them meaningless for the purpose of developing an empirical cost function. Restoration of health and alleviation of pain imply that the 'final' hospital product is the utility which the patient receives from consuming the various types of health related services. It could be argued, however, that a hospital's immediate output (i.e. short-run output) is patient dissatisfaction as opposed to satisfaction. Many patients experience pain, discomfort or anxiety related to their stay in hospital. Patient satisfaction derived from hospital services may accrue for many years after an operation or treatment, particularly if the patient's life was saved or a fuller measure of health was restored along with a lessening of pain and disability. The patient derives utility from the consumption of hospital services over the remainder of his life. This is a form of 'psychic' output.

It has been observed that to describe hospital output as the psychic benefit which the patient receives is much the same as saying that the output of a bank, an automobile manufacturer or a department store is "customer satisfaction".[14] Each produces goods and/or services and customer satisfaction is a consequence of the consumption of these goods or services. Similarly, the restoration of health and

[13] J.R. Lave and L.B. Lave, "Hospital Cost Functions", American Economic Review, Vol. 60, No. 3, June, 1970, P. 380.

[14] M. Brown Jr., "An Economic Analysis of Hospital Operations," Hospital Administration, Vol. 15, No. 2, Spring, 1970, P. 67.

the alleviation of pain are the benefits or utilities which an individual, or the community as a whole, derive through the consumption of hospital services. Psychic benefits do not provide operational definitions of hospital output because they are impossible to measure satisfactorily. For this reason, surrogate outputs have been used in order to study the hospital's cost function.

The Hospital as a Multiproduct Firm

The hospital may be designated as a 'multiproduct firm'; unfortunately, there is no general agreement on the composition of a hospital's total output aside from the catch-all term of 'service'. Hospitals produce various amounts of education, research and community service as well as care for a variety of patients. Measuring and quantifying these outputs is difficult because each type of service is not always easy to distinguish from the others. They are joint products. For example, in a research project where a patient receives treatments as part of the experiment, the research data and treatments are joint products. A precise aggregate measurement which adequately encomposes all forms of output appears unlikely in the hospital field. A problem of similar magnitude must be overcome with disaggregate measures if each of the 'n' different hospital services is to be represented in an analysis. It would be difficult to identify all of the individual outputs as separate entities. Joint products would limit the effectiveness of this approach.

Since the boundaries between the various service components is difficult to distinguish, it is, therefore, not a simple matter to assign production costs to specific outputs. Large amounts of joint costs exist in the operation of a hospital and complicate empirical analysis of the cost function. Joint costs are generated by products which are not separable. i.e. If a hospital operates an internship program it is virtually impossible to completely allocate this cost between the treatment and diagnostic services provided to the patients and the educational experience of the interns.

The disaggregate analysis of the production and cost functions of specific services within a hospital is one which deserves further investigation inspite of the difficulty of joint-costs and products. Paul Feldstein examined the total cost functions for a variety of services in one hospital in one disaggregate approach.[15] He included such units of measurement as the number of lab tests, radiology examinations and operations. Even these units reflected some degree of aggregation. More effort might be directed towards identifying and quantifying every service or task performed within a hospital in order to develop an accurate disaggregate approach. The question which arises with regard to a disaggregate method is whether or not the specific outputs of each hospital area, such as nursing or housekeeping are actually measures of overall hospital output. As P. Feldstein's work illustrated, it remains to be shown that the sum total of all individual hospital services is equivalent to the 'total output' of a

---

[15] P. Feldstein, An Empirical Investigation of the Marginal Costs of Hospital Services, (Chicago: University of Chicago, 1961). Baker, Hospital Cost Control. This other disaggregate study is reported in Appendix II.

hospital, whatever that may be.

Output Surrogates

The problem of defining hospital output has generally caused difficulty in the literature. In economic theory it is a requirement that there be an operational definition of the output of a production process before the study of the production and cost functions may be undertaken. In studies of the hospital industry, this problem has generally been circumvented through the use of a surrogate for output.

A surrogate is a substitute. The use of a surrogate enables a researcher to study other problems related to hospital production and cost without first having to precisely define output. In the literature the surrogate for output used most frequently was the 'patient day', although the 'number of admissions' was also used. A surrogate, such as the patient day, can be easily derived from hospital statistical reports. Once the number of patient days is calculated for a year, the average unit cost of output can be calculated by dividing the total annual expenditure by the total number of patient days recorded during the same year.

The patient day overcomes the problem of data heterogeniety but such an unidimensional measure of hospital output is not without flaws. It implicitly assumes that all patient days are alike; that within each patient day is contained an identical combination of services. It also assumes that the resource inputs embodied in a patient day are identical in every hospital. Output surrogates, including the patient day, have primarily been used in a variety of macro analyses.[16] They are of

[16]Refer to Table III-1.

little value in answering micro questions such as explaining the difference between the cost of a medical and a surgical patient day.

Another limitation of aggregate surrogates including the patient day originates in their inability to reflect variations in quality.[17] The present body of literature generally failed to adequately adjust for quality differences. It has been observed that: "Unfortunately, there is no feasible way to quantify quality for hospital services and its influence on costs cannot be estimated directly."[18] In fact, 'quality' in the hospital field is a term which conceals two effects. These are 1) changes in technology and 2) changes in psychic value. Costs tend to increase with changes in technology but may or may not with psychic changes. It cannot, therefore, be concluded that costs will always increase with increased quality. Both quality and cost might increase per radiological examination if a hospital replaces an x-ray machine obtained in 1940 with a more recent model. Both machines take x-rays but the technological superiority of the new piece of equipment would likely be greater. Quality may also increase if a nurse with no special training in social interaction is replaced by a trained nurse. Both would be paid at the same rate but the patient experiences, hopefully, a psychic improvement in the quality of the care he receives.

A major problem in the study of hospital costs is to distinguish the costs which are associated with improvements in technology from those associated with increases in the scale of hospital output. If no adjustment is made for technological differences when a surrogate out-

---

[17] R. Berry, "Returns to Scale in the Production of Hospital Services", Health Services Research, Vol.2, ~~No.~~ 2, Summer, 1967, P. 126.

[18] Ibid., P. 127.

put is used, then small hospitals or ones producing less technically efficient services may well appear to be operating more efficiently than larger hospitals or ones producing services, which contain a higher technological component. (As a point of interest, it should be observed that, generally, higher technology levels appear to be associated with larger hospitals than small ones. The larger hospital tends to offer a broader range of facilities and services and a better qualified and more diversified staff.) There appears to be some relationship between hospital costs and the level of technology which the literature has failed to adequately report. Future research and analysis in the field of hospital economics could usefully develop a technique to more adequately reflect this form of quality differences. Such a technique would be useful because no two hospitals are likely to produce outputs identical with regard to either volume or quality because no two hospitals possess identical equipment, offer identical services, employ identical staff or operate identical physical plants.

Hospital Efficiency

The Ontario Committee on the Healing Arts has observed that, "Increasing quantities of labour and other inputs.... cannot be taken as evidence of falling productivity in hospital operations".[19] In fact, the Committee felt that productivity actually increased with the addition of such inputs. As medicine has developed new techniques and

---

[19] Committee on the Healing Arts, Vol.3, P. 121.

procedures have resulted which in turn have led to new kinds of equipment and new job classifications. The new equipment could be designed to perform a function never before performed or to improve on the performance of older equipment. In either case, the new equipment represents an improvement in technology. Often the new equipment requires a new and more highly trained type of employee be added to the staff in order to operate it. In the hospital field improvements in technology frequently do not yield reductions in manpower, as is generally expected in economics, but rather staff must be increased. Much of the benefit of new equipment and jobs appears to be psychic, although patient volume may increase in many instances. Hospitals may add staff and equipment and, as a result, raise the total cost of operating without treating one additional patient. However, it cannot be said that productivity has dropped because the psychic benefits which the patients receive may well have been increased substantially. Inspite of this argument, it is doubtful that most hospitals operate at an optimal level.

In economics the production function traditionally represents in mathematical terms a firm operating in the most efficient manner possible at a given level of technology.[20] The production function is related to the premise that the goal of a firm is profit maximization. A firm which maximizes its profits operates on the edge of its production frontier. Any firm on the edge of its frontier is combining

---

[20] M. Brown Jr., "An Economic Analysis", P. 61.

the inputs in its production process in the most efficient manner possible to produce the maximum level of output possible with the current technological resources. The production function specifies the output of a firm on its production frontier.

In applying the theories of production and cost to the hospital industry, it must be recalled that it is a non-profit industry and as such does not maximize profits or minimize costs. It appears that hospitals generally operate in a less than perfectly efficient manner, and therefore, operate within the edge of their production frontier.[21] The usefulness of specifying hospital cost functions is primarily in the creation of 'bench marks' of hospital cost behaviour which provide guidelines and indicators for hospital planner and administrators. Given a level of production, which is likely not optimal, cost specifications, such as were contained in the literature, reflect the behaviour of costs and the level at which they are lowest for specific samples of hospitals.

---

[21]This behaviour occurs in part because there is little incentive for the administrator of a hospital to try to maximize efficiency. In order to ensure efficient operations, he must expend great effort and become involved in conflicts with the medical staff. Since hospitals in Canada have traditionally been reimbursed for the costs incurred, all savings accrue to the government paying agency. An administrator maximizes his utility by expanding his hospital in size and staff. This action increases his status and, generally, his salary and virtually eliminates conflict with the medical staff. This behaviour ensures that the net social benefit is not maximized.

CHAPTER V

SUMMARY, OBSERVATIONS AND RECOMMENDATIONS

Summary

Chapter I of this thesis examined the behaviour of costs in the health field in both Canada and the Province of Alberta during recent years with particular emphasis on hospital-generated costs. During the period considered, 1957 to 1970, costs were observed to escalate at a rapid rate in the health services and, especially, in the hospital field.

In light of such rapidly increasing costs, it was thought that a thorough review of the economic literature concerning hospitals would be a useful undertaking. It would reveal the nature of the current available literature:- its directions, results and conclusions; the concepts, methods and approaches meriting further research and investigation; and an indication of which size hospital is optimal.

A number of points emerged from the literature review:

1. Throughout much of the literature, the technique of analysis consisted of the cross-sectional approach. Of three time-series analyses in two cases, the data was combined after the time-series analyses were complete to provide a pooling of data.

2. Much of the literature consisted of aggregate analysis of large groups of hospitals. Two writers performed a type of disaggregate analysis. Most of the conclusions and results were, therefore, of a 'macro' nature and as such they are of little use in explaining the cost behaviour of individual hospitals or in decision making at the 'micro' level of hospital activity.
3. In most of the literature aggregated costs were employed and reflected the overall costs generated by all hospital activities. In all but one instance such costs were specified in relationships with independent variables representing only inpatient activity. In the literature, the specifications generally appeared to be inadequate because the actual costs were greater than those which the independent variables could explain.
4. Hospital scale has been represented in terms of output and capacity both of which have limitations. Output was generally measured in terms of a surrogate reflecting only inpatient activity. In most of the cross-sectional studies noted in Table III-1 in which output was employed, the influence of the 'regression fallacy' was operative because no adjustment was made for the cost of capital. Capacity was most often measured in terms of 'ability to serve' inpatients.
5. Capital costs were omitted from the cost calculation in all but two pieces of literature. The consequence was that, depending

on the specification, either total variable or average variable cost functions were specified rather than complete total or average cost functions.

6. Economies of scale for larger hospitals appeared to have been generally underestimated in the literature because neither operating nor capital costs were adequately reflected throughout the various specifications.

7. There was generally no discussion of nor attempt to adequately reflect the influence quality has on hospital costs. The technological and psychic differentiations of quality were not dealt with. This aversion to the problem of representing variations in quality appeared to reflect the inability of the health care industry to define standards.

8. Hotel costs were relatively 'fixed' in the short-run.

9. Many of the results in the literature were not statisticly significant.

10. The nature of the studies reported was such that their results have limited application to any population except for the one used in each particular study.

Concluding Observations

The pressures of the market place appear to be all but inoperative in the provision of hospital services in Canada and, as a result, growth and development have been less than optimal. The planning of the size and distribution of hospital facilities by external authority has developed since the lack of a price mechanism has caused cost to increase beyond a level which the public could tolerate. In Canada, the external authorities are primarily the Provincial Government, although, because of heavy financial involvement, the Federal Government also is influential.

Presently in Canada, two general policy approaches appear to be emerging in relation to hospital services. Neither appear to be directly concerned with whether or not hospitals operate efficiently. On the one hand, hospitals are characterized as wasteful institutions which must be regulated[1] and on the other hand, 'alternative forms of care are proposed.[2]

Hospital expenditures are to be held in check by such measures as restricting government aid for new or expanded active-treatment facilities, applying ceilings to the rates of increase in hospital operating costs and ordering hospitals to close certain wards or specific numbers of beds. The politically acceptable belief is evolving that the number of active-treatment beds per thousand population must be reduced to a uniform ratio across an entire province. (e.g. The expressed aim of the Government of Ontario, "Is to trim down the number of beds from the present 5.3 beds per thousand population to 4 beds per thousand."[3] Limiting the number of beds may slow the rate of cost increases but as was observed in Chapter IV, such action does not guarantee that the resources which will continue to be expended will produce an optimal return.

Recent reports eminating from government study groups have called for a complete re-organization of the entire health care field and of the mechanisms by which it is financed.[4] Hospitals have been recognized

---

[1] Canadian Hospital, Vol.50, No.1, January, 1973, P.4 and P.19; and Hospital Administration in Canada, Vol.15, No.1, January, 1973, P.4.

[2] The Community Health Centre in Canada, Report of the Community Health Centre Project to the Conference of Health Ministers, (Ottawa, Department of National Health and Welfare, 1972).

[3] Canadian Hospital, Vol.50, no. 1, January, 1973, P.19.

[4] The Community Health Centre in Canada; Report of the Ontario Committee on the Healing Arts, Vol.1 and 3, (Toronto: Queen's Printer, 1970); and Report of the Commission of Inquiry on Health and Social Welfare, Part 2, Vol.IV, Tome I, (Quebec: Quebec Official Publisher, 1970).

as important features of the health system in these reports but the recommendations generally have called for limitations on hospital functions and future development. Alternative forms of health care have been put forward as mechanisms which would increase the amount of health care provided, deliver it in a more efficient manner and, simultaneously, reduce the rate at which health costs are rising. The 'community health centre' concept, in particular, ~~has~~ been expounded as such an alternative along with extended care, home care and various forms of 'day' care both inside and outside the hospital setting. Expectations exist that, with such alternatives, hospital expansion can be avoided, or at least postponed, and that health care would come to be provided more efficiently and effectively.

It is possible that restraints imposed on hospitals and the development of alternative forms of health care may achieve many of the purposes for which they were conceived. Two items for thought have been germinated by this thesis regarding hospital optimal size which merit further consideration by those who support restraints and alternatives. First, an overview of all of the results in the hospital literature revealed that, while the results are varied, the majority of the analyses indicated that hospital average costs were decreasing or constant over the entire range of presently existing hospitals. This statement is also supported by the observation that costs and variables were often erroneously specified which likely yielded an underestimate of economies of scale in the literature. Although the evidence is far from conclusive, at the aggregate level it does appear that larger hospitals could possibly operate more efficiently than smaller ones and that the optimal hospital size may be very large. Additional research needs to be directed toward determining the optimal size at the aggregate level before the concept of hospital expansion is eliminated as an alternative in developing the health care system.

Aggregate economic studies of hospitals have not necessarily reflected what is desirable for one hospital as a unit nor as part of an integrated system. The hospital of optimal size at the aggregate level may be unsuitable for many communities. This is the second item which merits consideration. More study is required at the very disaggregate level of the individual hospital, its departments and of the community agencies which complement and supplement each particular hospital. Paul Feldstein provided a technique upon which further research could evolve.[5] The goal of such research would be to provide information which would enable each individual hospital to know when its size would be optimal, given its particular community, and what organizational arrangements would maximize operational efficiency, given a desirable size.

An extensive development of hospitals has taken place in Canada during the past twenty years. Currently, hospitals represent the focus of health care in many communities. Given the present patterns of medical practice and existing public attitudes, it is doubtful that the place of the hospital in health care delivery will be substantially

---

[5]Paul Feldstein, An Empirical Investigation of the Marginal Costs of Hospital Services, (Chicago: University of Chicago, 1961), P. 1-60.

reduced. Hospitals will continue to exist and each needs to be managed efficiently both as a unit and as a component of a community's health resources. Restrictions on size and patient volume may prevent a hospital from achieving its least cost level of production and keep the overall health costs for a community higher than if the hospital was left unrestrained. Restrictions may limit the volume of hospital service available but costs are not necessarily limited proportionately. If hospital costs are not reduced when the alternate forms of care are introduced, but hospital services are cut-back, the total cost of health care overall may be increased but not the volume of health care available.

Recommendations

This thesis has examined the existing economic literature related to hospitals which presently is limited. The field of hospital economics, however, merits further research at both the aggregate and, particularly, disaggregate levels.

Specific problems have been isolated, each of which is an opportunity for additional study. Cost specifications which encompass both capital and operating costs need to be developed. If meaningful inter-hospital comparisons are to be made, a more precise measure of size is needed than that of the 'number of beds'.[6] The influence of 'quality'-

---

[6]The problem occurs because the number of beds in no way reflects outpatient types of activity.

both technical and psychic - on costs is one which needs recognition and study. Better specifications are needed if results are to be statistically significant more often. Further research should be directed toward defining the product of a hospital in precise terms.

In the literature there was no consensus as to which hospital size was optimal. Continuously decreasing average costs were found frequently in the hospital industry. With only one exception, there was a definite concensus that small sized hospitals have higher average costs than medium and large hospitals despite the fact that the larger hospitals may offer a greater variety of services. Considerable research is required if the question of optimality is to be settled satisfactorily.

The most significant point raised in this thesis concerns public policy and the need for 'micro' studies at the level of the individual hospital. They are required for each hospital in order to ensure that it operates efficiently and at the same time that it effectively meets the unique health requirements of its community. With a disaggregate analysis of the hospital, the local planners could then allocate resources in such a fashion that an optimal return could be achieved within the overall 'macro' guidelines of provincial policy.

SELECTED BIBLIOGRAPHY

Baker, Dennis M. Hospital Cost Control and Product Pricing with an Empirically Derived Factor Analytic Model of Inpatient Output. Unpublished Ph.D. disertation, University of California, 1973.

Berry, Ralph. "Returns to Scale in the Production of Hospital Services". Health Services Research. (Spring, 1967), 123-139.

________. "Product Heterogeneity and Hospital Cost Analysis." Inquiry. (March, 1970), 67-77.

Brown, Max Jr. "An Economic Analysis of Hospital Operations." Hospital Administration. (Spring, 1970), 60-74.

Carr, W. John, and Feldstein, Paul J. "The Relationship of Cost to Hospital Size." Inquiry. (June, 1967), 45-65.

Cohen, Harold A. "Variations in Costs Among Hospitals of Different Sizes." Southern Economic Journal. (January, 1967), 355-366.

________. "Hospital Cost Curves With an Emphasis On Measuring Patient Care Output." Empirical Studies in Health Economics. Edited by H.E. Klarman. Baltimore: Johns Hopkins Press, 1970.

Evans, Robert G. 'Behavioural' Cost Functions for Hospitals. Discussion Paper #38. Vancouver: Department of Economics, University of British Columbia, 1970.

________. and Walker, H.D. Public Policy Problems in the Canadian Health Services Industry. Discussion Paper #41. Vancouver: Department of Economics, University of British Columbia, 1970.

Feldstein, Paul J. "Applying Economic Concepts to Hospital Costs." Hospital Administration. (Winter, 1968), 68-89.

________. An Empirical Investigation of the Marginal Costs of Hospital Services. Chicago: University of Chicago, 1961.

Feldstein, Martin S. Economic Analysis for Health Service Efficiency. Amsterdam: North-Holland Pub. Co., 1968.

Ferguson, C.E. Microeconomic Theory. Homewood, Illinois: Richard D. Irwin, Inc., 1969.

Francisco, Edgar J. "Analysis of Cost Variations Among Short Term General Hospitals", Empirical Studies in Health Economics. Edited by H.E. Klarman. Baltimore: Johns Hopkins Press, 1970.

Fraser, R.D. Canadian Hospital Costs and Efficiency. Special Study No. 13. Prepared for the Economic Council of Canada. Ottawa: Information Canada, 1971.

Hefty, T.R. "Returns to Scale in Hospitals: A Critical Review of Recent Research." Health Services Research. (Winter, 1969), 267-280.

Ingbar, M., and Taylor, L. Hospital Costs in Massachusetts. Cambridge, Mass.: Harvard University Press, 1968.

Lancaster, K. Introduction to Modern Microeconomics. Chicago: Rand, McNally and Co., 1969.

Leftwich, R.H. The Price System and Resource Allocation. Hinsdale, Illinois: The Dryden Press Inc., 1970 edition; also New York: Holt, Rinehart and Winston, 1961 edition.

Klarman, H.E. The Economics of Health. New York: Columbia University Press, 1965.

Kushner, Joseph. Economies of Scale in the General Hospital Industry. Unpublished Ph.D. dissertation, University of Western Ontario, 1969.

________. "Returns to Scale in Hospitals." Health Services Research. (Winter, 1970), 370-372.

Lave, Judith R. "Review of Methods Used to Study Hospital Costs." Inquiry. (May, 1966), 57-81.

________, and Lave, L.B. "Estimated Cost Functions for Pennsylvania Hospitals." Inquiry. (June, 1970), 3-14.

________. "Hospital Cost Functions." American Economic Review. (June, 1970), 379-395.

________. "Economic Analysis for Health Services Efficiency: A Review Article." Applied Economics. (1970), 293-305.

Long, M.F., and Feldstein, Paul J. "Economics of Hospital Systems: Peak Loads and Regional Co-ordination." American Economic Review. (May, 1967), P.119-129.

Mansfield, E. Microeconomics: Theory and Applications. New York: W.W. Norton and Co. Inc., 1970.

Plain, R.H.M., Resource Allocation in the Medical Service Markets in Alberta. SectionI. (Report submitted to the Alberta Hospital Services Commission, April, 1971.)

Richmond, S.B. Statistical Analysis. New York: Ronald Press Co., 1964.

Ro, K.K. "Determinates of Hospital Costs." Yale Economic Essays. (Fall, 1968), 185-256.

Mann, J., and Yett, D. "The Analysis of Hospital Costs: A Review Article." Journal of Business. (April, 1968), 191-202.

Saathoff, D., and Kurtz, R. "Cost Per Day Comparisons Don't Do the Job." Modern Hospital. (June ,1962), 14-16.

Skinner, C.G. "Hospitals and Allied Institutions," Hospital and Medical Economics, Vol. II. Edited by W.J. McNerney et al. Chicago: Hospital Research and Education Trust, 1962.

Shain, M., and Roemer, M.I. "Hospital Costs Relate to the Supply of Beds." Modern Hospital. (April, 1959), 71-73

Public Documents

Alberta. Annual Report of the Alberta Hospitalization Benefits Plan - 1969. Department of Health, 1971.

Canada. The Community Health Centre in Canada. Department of National Health and Welfare, 1972.

Canada. Economic Council of Canada. Patterns of Growth. Ottawa: Queen's Printer, 1970.

Canada. Research and Statistics Memo. Expenditures on Personal Health Care in Canada, 1960-1970. Department of National Health and Welfare, 1971.

________. Expenditures of Personal Health Care in Canada, 1957-1969. Department of National Health and Welfare, 1970.

________. Expenditures of Personal Health Care in the Provinces of Canada, 1957-1969. Department of National Health and Welfare, 1970.

________. Health Care Price Movements in Canada, April, 1961 to April, 1970. Department of National Health and Welfare, 1970.

Canada. Task Force Reports on the Cost of Health Services in Canada. Ottawa: Queen's Printer, 1969.

Ontario. The Relative Stay Index Report, 1971 Edition. Toronto: Ontario Hospital Services Commission, 1972.

Ontario. Report of the Ontario Committee on the Healing Arts, Vols. I and III. Toronto: Queen's Printer, 1970.

Quebec. Report of the Commission of Inquiry on Health and Social Welfare, Part 2, Vol. IV: Health. Quebec: Quebec Official Publisher, 1970.

APPENDIX I

ESTIMATED TOTAL COST FUNCTIONS

FOR CANADIAN GENERAL HOSPITALS

AS PER R.D. FRASER[1]

[1] R.D. Fraser, Canadian Hospital Costs and Efficiency, Special Study No. 13, Prepared for the Economic Council of Canada, (Ottawa: Information Canada, 1971), P. 115-124.

Table B-20

ESTIMATED TOTAL COST FUNCTIONS (OPERATING COST ONLY) IN RELATION TO COMPOSITE OUTPUT AND TWO OTHER EXPLANATORY VARIABLES FOR HOSPITALS BY BED SIZE, CANADA, 1966

Bed Size of Hospital	Number of Hospitals	Intercept (S.E.E.)	Composite Output (T-Value)	Composite Output² (T-Value)	Capacity Utilization (T-Value)	R^2 (F-Value)
1-9	32	2118.861 (838.467)	1774.595 (2.713)	-129.533 (-1.470)	-3.798 (-.502)	.529 (12.613)
10-24	200	4672.227 (2230.258)	1527.200 (6.937)	-4.916 (-.397)	-39.622 (-3.826)	.696 (153.118)
25-49	190	4293.695 (4787.293)	2389.472 (6.634)	-18.616 (-1.720)	-120.261 (-4.065)	.698 (146.325)
50-99	146	-4998.137 (9541.219)	2657.668 (5.187)	-9.710 (-1.322)	-130.531 (-1.848)	.761 (154.956)
100-199	127	18696.813 (25848.195)	1941.799 (2.842)	-.282 (-.060)	-142.404 (-.588)	.672 (86.994)
200-299	62	-55416.375 (43428.977)	4539.129 (2.612)	-8.862 (-1.339)	-1290.758 (-2.095)	.609 (32.635)
300-499	54	179082.500 (76577.875)	1088.561 (.888)	2.814 (1.015)	-1254.910 (-1.077)	.691 (40.443)
500-999	29	1465702.000 (172761.125)	-402.590 (-.170)	2.936 (1.014)	-10374.832 (-2.368)	.580 (13.909)
1000+	8	-6227196.000 (169674.500)	13803.828 (2.857)	-7.173 (-2.595)	19893.363 (1.409)	.755 (8.208)

Table B-21

ESTIMATED TOTAL COST FUNCTIONS (INCLUDING ASSET VALUE) IN RELATION TO COMPOSITE OUTPUT AND TWO OTHER EXPLANATORY VARIABLES FOR HOSPITALS BY BED SIZE, CANADA, 1966

Bed Size of Hospital	Number of Hospitals	Intercept (S.E.E.)	Composite Output (F-Value)	Composite Output[2] (F-Value)	Capacity Utilization (F-Value)	R^2 (F-Value)
1-9	32	2148.543 (905.178)	2197.495 (3.112)	-197.662 (-2.078)	-3.832 (-.469)	.495 (11.145)
10-24	200	4794.070 (3348.503)	2108.129 (6.378)	-27.284 (-1.468)	-52.084 (-3.346)	.569 (88.492)
25-49	190	6356.574 (6593.813)	2874.004 (5.793)	-21.814 (-1.464)	-166.282 (-4.081)	.631 (108.716)
50-99	146	3490.367 (12660.117)	3214.026 (4.727)	-10.512 (-1.079)	-313.353 (-3.343)	.718 ([illegible].812)
100-199	127	54667.000 (31784.387)	2338.709 (2.764)	-.254 (-.044)	-645.696 (-2.167)	.6[illegible] ([illegible])
200-299	62	-36545.438 (57039.563)	4820.684 (2.112)	-9.305 (-1.070)	-1347.908 (-1.666)	[illegible] (21.[illegible])
300-499	54	372484.313 (95917.875)	474.249 (.310)	5.210 (1.503)	-2519.939 (-1.730)	[illegible]6 ([illegible].676)
500-999	29	2346514.000 (231463.750)	-432.653 (-.137)	3.351 (.864)	-19772.000 (-3.368)	.553 (12.539)
1000+	8	-7706258.000 (305685.500)	15970.441 (1.834)	-8.522 (-1.771)	30081.520 (1.183)	.454 (2.943)

Table B-22

ESTIMATED TOTAL COST FUNCTIONS (INCLUDING DEPRECIATION) IN RELATION TO COMPOSITE OUTPUT AND TWO OTHER EXPLANATORY VARIABLES FOR HOSPITALS BY BED SIZE, CANADA, 1966

Bed Size of Hospital	Number of Hospitals	Intercept (S.E.E.)	Composite Output (T-Value)	Composite Output² (T-Value)	Capacity Utilization (T-Value)	R^2 (F-Value)
1-9	32	2302.512 (928.181)	1856.214 (2.564)	-152.421 (-1.563)	-2.303 (-.275)	.461 (9.855)
10-24	200	4847.953 (2449.945)	1721.902 (7.120)	-12.150 (-.893)	-43.181 (-3.796)	.677 (140.296)
25-49	190	5376.344 (5026.414)	2601.207 (6.878)	-23.137 (-2.037)	-138.603 (-4.462)	.687 (139.581)
50-99	146	-3391.969 (9866.438)	2790.269 (5.266)	-10.094 (-1.329)	-154.340 (-2.113)	.765 (158.417)
100-199	127	22692.625 (26208.449)	2165.858 (3.127)	-1.228 (-.257)	-237.644 (-.967)	.681 (90.842)
200-299	62	-50139.813 (44772.750)	4678.184 (2.611)	-9.140 (-1.339)	-1344.899 (-2.118)	.608 (32.531)
300-499	54	179933.500 (77070.625)	1463.163 (1.186)	2.137 (.766)	-1579.182 (-1.347)	.700 (42.222)
500-999	29	1472917.000 (174255.063)	83.606 (.035)	2.425 (.835)	-11294.164 (-2.556)	.595 (14.690)
1000+	8	-6440534.000 (157731.000)	14120.984 (3.143)	-7.265 (-2.828)	20798.121 (1.585)	.812 (11.050)

Table B-23

ESTIMATED TOTAL COST FUNCTIONS (INCLUDING DEPRECIATION) IN RELATION TO COMPOSITE OUTPUT AND TWO OTHER EXPLANATORY VARIABLES FOR HOSPITALS BY BED SIZE,* CANADA, 1966

Bed Size of Hospital	Number of Hospitals	Intercept (*S.E.E.*)	Composite Output (*T*-Value)	Composite Output² (*T*-Value)	Capacity Utilization (*T*-Value)	$\bar{R}^2$ (*F*-Value)
1-9	32	2302.512 (928.181)	1856.214 (2.564)	-152.421 (-1.563)	-2.303 (-.275)	.461 (9.855)
10-24	200	4533.727 (2712.735)	1627.448 (6.261)	-9.429 (-.643)	-34.127 (-2.756)	.629 (115.607)
25-49	190	3661.461 (5035.691)	2776.159 (7.568)	-28.683 (-2.620)	-132.642 (-4.305)	.697 (146.944)
50-99	146	-3391.969 (9866.438)	2790.269 (5.266)	-10.094 (-1.329)	-154.340 (-2.133)	.765 (158.427)
100-199	127	22692.449 (26208.449)	2165.858 (3.127)	-1.228 (-.257)	-237.644 (-.967)	.681 (90.842)
200-299	62	-50139.813 (44772.750)	4678.184 (2.611)	-9.140 (-1.339)	-1344.899 (-2.118)	.608 (32.531)
300-499	54	179933.600 (77070.625)	1463.163 (1.186)	2.137 (.766)	-1579.182 (-1.347)	.700 (42.222)
500-999	29	1556763.000 (171736.313)	-290.553 (-.130)	2.836 (1.026)	-11316.090 (-2.598)	.595 (15.213)
1000+	8	-6440534.000 (157731.000)	24120.904 (3.143)	-7.265 (-2.828)	20798.121 (1.584)	.81[illegible] (11.050)

* Those hospitals having a zero value for one or more variables are included in these regressions.

Table B-24

ESTIMATED TOTAL COST FUNCTIONS (INCLUDING DEPRECIATION) IN RELATION TO COMPOSITE OUTPUT AND FOUR OTHER EXPLANATORY VARIABLES FOR HOSPITALS BY BED SIZE, CANADA, 1966

Bed Size of Hospital	Number of Hospitals	Intercept (S.E.E.)	Composite Output (T-Value)	Composite Output² (T-Value)	Capacity Utilization (T-Value)	Intern Program (T-Value)	Nurses' Training Program (T-Value)	R^2 (F-Value)
1-9	32							
10-24	200							
25-49	190							
50-99	146	-3707.629 (9890.809)	2274.497 (5.188)	-9.868 (-1.287)	-148.476 (-2.019)	11418.328 (1.146)	374.918 (.097)	.764 (94.843)
100-199	127	11628.938 (24774.508)	2282.457 (3.473)	-3.836 (-.840)	-111.842 (-.477)	39659.445 (3.031)	13417.055 (2.501)	.715 (64.327)
200-299	62	-39508.188 (45481.496)	4480.898 (2.394)	-8.345 (-1.156)	-1293.685 (-1.900)	-5806.059 (-.380)	-2695.170 (-.199)	.595 (18.956)
300-499	54	167290.625 (73530.500)	1192.812 (1.008)	2.118 (.793)	-1269.435 (-1.113)	52687.773 (2.295)	21573.254 (.925)	.727 (29.217)
500-999	29	-1244.359 (21091.629)	816.809 (1.888)	36.282 (7.432)	44.939 (.320)	16242.047 (1.120)	-12564.766 (-.787)	[illegible] (205.535)
1000+	8							

Table B-25

ESTIMATED TOTAL COST FUNCTIONS (INCLUDING DEPRECIATION) IN RELATION TO COMPOSITE OUTPUT FOR HOSPITALS BY BED SIZE, CANADA, 1966

Bed Size of Hospital	Number of Hospitals	Intercept** (S.E.)	Composite Output (T-Value)	R^2 (F-Value)
1-9	32	3717.021 (935.005)	807.656 (5.170)	.454 (267.728)
10-24	200	3488.844 (2525.538)	1332.222 (19.558)	.657 (382.511)
25-49	190	4294.246 (5270.836)	1561.585 (19.025)	.656 (361.957)
50-99	146	-769.266 (9987.754)	1966.486 (21.409)	.759 (458.348)
100-199	127	14008.250 (26096.230)	1926.029 (16.551)	.684 (273.935)
200-299	62	12139.000 (46047.926)	2115.719 (9.332)	.585 (87.095)
300-499	54	-4069.813 (78520.750)	2203.989 (10.872)	.689 (118.200)
500-999	23	178374.438 (193439.063)	1962.023 (5.390)	.500 (29.050)
1000+	8	395818.000 (241450.125)	1728.252 (3.139)	.558 (9.855)

Table B-26

ESTIMATED TOTAL COST FUNCTIONS (INCLUDING DEPRECIATION) IN RELATION TO COMPOSITE OUTPUT AND ONE OTHER EXPLANATORY VARIABLE FOR HOSPITALS BY BED SIZE, CANADA, 1966

Bed Size of Hospital	Number of Hospitals	Intercept (S.E.E.)	Composite Output (T-Value)	Capacity Utilization (T-Value)	$\bar{R}^2$ (F-Value)
1-9	32	3715.373 (950.974)	801.196 (2.987)	.251 (.030)	.435 (12.920)
10-24	200	5376.027 (2448.695)	1519.088 (18.254)	-40.834 (-3.691)	.678 (210.259)
25-49	190	9382.000 (5068.539)	1862.950 (17.150)	-122.104 (-4.038)	.682 (203.866)
50-99	146	6275.707 (9892.926)	2102.666 (18.307)	-140.941 (-1.943)	.764 (235.476)
100-199	127	27118.625 (26109.102)	1991.556 (14.661)	-222.113 (-.936)	.684 (137.271)
200-299	62	85947.313 (45071.918)	2300.601 (9.498)	-1199.559 (-1.904)	.603 (47.267)
300-499	54	115945.563 (76754.688)	2392.205 (10.739)	-1958.291 (-1.850)	.702 (63.561)
500-999	29	1157944.000 (173209.125)	2043.548 (6.244)	-11964.305 (-2.770)	.599 (21.953)
1000+	8	-992977.000 (244309.875)	1469.774 (2.360)	18830.754 (.928)	.548 (5.243)

Table B-28

ESTIMATED TOTAL COST FUNCTIONS (INCLUDING DEPRECIATION)
IN RELATION TO COMPOSITE OUTPUT AND THREE OTHER EXPLANATORY VARIABLES
FOR HOSPITALS BY BED SIZE, CANADA, 1966

Bed Size of Hospital	Number of Hospitals	Intercept (S.E.E.)	Composite Output (T-Value)	Capacity Utilization (T-Value)	Intern Program (T-Value)	Nurses' Training Program (T-Value)	$\bar{R}^2$ (F-Value)
1-9	32						
10-24	200						
25-49	190						
50-99	146	5754.313 (9913.316)	2102.264 (18.266)	-135.775 (-1.860)	11643.840 (1.166)	953.605 (.248)	.763 (117.607)
100-199	127	25532.750 (24744.738)	1744.456 (11.971)	-67.942 (-.298)	38313.945 (2.953)	13129.180 (2.456)	.716 (80.426)
200-299	82	84085.063 (45625.340)	2316.944 (9.338)	-1141.078 (-1.702)	-9707.996 (-.650)	-4109.734 (-.303)	.593 (23.208)
[illegible]	[illegible]	103972.438 (73249.125)	2111.618 (8.836)	-1630.127 (-1.565)	53471.277 (2.340)	19953.320 (.768)	.72[illegible] (36.645)
500-999	29	1244.793 (30614.563)	3778.193 (15.440)	-372.446 (-1.992)	-24164.414 (-1.238)	31627.277 (1.471)	.896 (115.389)
1000+	8						

Table B-29

ESTIMATED TOTAL COST FUNCTIONS (INCLUDING DEPRECIATION) IN RELATION TO COMPOSITE OUTPUT AND FOUR OTHER EXPLANATORY VARIABLES FOR HOSPITALS BY BED SIZE, CANADA, 1966

Bed Size of Hospital	Number of Hospitals	Intercept (S.E.E.)	Composite Output (T-Value)	Capacity Utilization (T-Value)	Intern Program (T-Value)	Nurses' Training Program (T-Value)	Newborn & Post-Operative Infections (T-Value)	R^2 (F-Value)
1-9	32							
10-24	200							
25-49	190							
50-99	146	6976.863 (9856.082)	2145.743 (18.260)	-155.502 (-2.113)	10621.695 (1.067)	1981.166 (.511)	-290.073 (-1.625)	.766 (95.710)
100-199	127	25926.438 (24835.773)	1759.482 (11.477)	-76.699 (-.333)	38523.121 (2.955)	13053.461 (2.430)	-55.923 (-.327)	.714 (63.891)
200-299	62	89141.563 (45902.066)	2238.233 (7.818)	-1103.369 (-1.628	-11072.820 (-.727)	-5975.305 (-.426)	125.816 (.561)	.588 (18.406)
300-499	54	102621.438 (69252.000)	2227.794 (9.674)	-1596.087 (-1.621)	65927.438 (2.980)	17062.344 (.694)	-625.073 (-2.611)	.758 (34.162)
500-999	29							
1000+	8							

AN EXAMINATION OF A RECENT DISAGGREGATE STUDY OF HOSPITAL COSTS BY D.M. BAKER[2]

[2]D.M. Baker, Hospital Cost Control and Product Pricing With an Empirically Derived Factor Analytic Model of Inpatient Output, Unpublished Ph.D. disertation, (Los Angeles, University of California, 1973).

It has been noted previously in the main body of this thesis that satisfactory measures of hospital output were not evident in the existing literature. One of the main observations of the concluding chapter was that further research should be conducted into the development of such a measure. Subsequent to the final submission of this thesis a recent piece of literature which dealt specifically with the 'output issue' was brought to the writer's attention. This thesis would not be complete without a discussion of the work of Dennis M. Baker.[3]

Baker had two objectives in his work. The first was to define and measure, in a 'meaningful' way, hospital inpatient output. The second objective was to test the usefulness of the inpatient product model with regard to cost control and product pricing. Baker restricted his study to inpatients and selected the case (i.e. discharge) as his unit of output. He was critical of the assumption underlying common surrogate measures of output such as the patient day with which all types of patients are presumed to consume resources in equal quantities. In Baker's view "Products should be defined and measured in terms of the kinds and amounts of resources they use".[4] Recognizing that to analyze the costs associated with each case would be a monumental task, he proceeded utilizing the

---

[3] D.M. Baker, Hospital Cost Control and Product Pricing With an Empirically Derived Factor-Analytic Model of Inpatient Costs, Unpublished Ph.D. disertation, (Los Angeles, California, 1973).

[4] Ibid., P.12.

assumption that some diseases are related to others in a systematic manner, in terms of the kind and amounts of resources used for each. In his analytical work, Baker operationally defined the product of inpatient care in terms of disease groups.

Baker first distinguished the disease groups by means of multivariate factor analysis. His data consisted of a 20 percent random sample of all discharges at one hospital during the first six months of a particular year. From his sample he grouped diagnoses alike in resource utilization establishing major utilization patterns and then assigned the remaining unused diagnoses from the sample to appropriate groups.

i.e. $Qj = Tj + Aj$

where j = a particular product (diagnostic) group

Q = diagnoses composing a product

T = the diagnoses initially defining an output

A = other diagnoses most similar to T

The final form of Baker's product model cost function was of the general form:

$$Y = \sum_{j=1}^{k} b_j X_j$$

where Y = total inpatient direct cost

k = the number of products

j = one product

b = cost of product j

X = number of units of product j

Baker began with a very disaggregate sample but by means of multivariate analysis he aggregated his data and then developed a cost function for the particular hospital from which he drew his data. As Baker noted, one of the limitations of his analysis was the definition of the iso-resource diagnostic groups in terms of standard charges for inpatient services. He opted for this approach because he felt that an attempt to define output in terms of an appropriately limited number of physical amounts was not operationally feasible. As an example he stated that the unit of service produced in Radiology:- the X-ray examination;- has not the same value of a Nuclear Medicine Department:- the scan. Such measures did not reflect the relative "costliness"[5] of the service. Since, the hospital charges constituted a standard rate schedule that reflected relative costliness of different services, it was assumed that the charges were acceptable surrogates for defining relative quantities of inputs and, hence, iso-resource products or outputs.

Baker's multivariate factor analysis yielded three distinct factors representing disease categories defined in terms of resource utilization. Those diagnoses not included in the multivariate analysis were subsequently allocated to one of the three categories. Having defined output, Baker proceeded to estimate a linear, time-series total inpatient cost function for the particular hospital. He included only inpatient direct costs thus omitting from his cost estimate overhead, depreciation and amortization costs. Additionally, no attempt to reflect changes in technology was incorporated. The total inpatient direct cost function took the following form:

$$Y_i = \underset{\substack{(268.09)\\2.16}}{578.37}\, X_{i1} + \underset{\substack{(266.27)\\0.95}}{252.36}\, X_{i2} + \underset{\substack{(529.22)\\1.80}}{954.62}\, X_{i3} + \underset{\substack{(2.65)\\5.26}}{13.93}\, t_i \qquad R^2 = .7226$$

[5]*Ibid.*, p. 128.

where t = linear time variable, and

$i = 1 \ldots 16$ (months).

Baker interpreted his result as a long-run variable cost function. The value of R^2 reflected the high explanatory power of the model. Only the coefficient of X_{12} is not significantly different from zero whereas the others are significant at the 10 percent level. The long-run average variable cost curve revealed by Baker's analysis was one reflecting continuously increasing average costs or continuous diseconomies of scale. Naturally, his linear analysis is subject to all of the limitations associated with such analyses mentioned in the fourth chapter of this thesis.

Baker's study provided an interesting alternative approach to the definition of hospital output. No conclusion with regard to the superiority of his measure is possible based solely on the limited nature of the cost function which he estimated. His measure, based on diagnostic groupings according to resource utilization, does raise questions, some of which are noted above. Further, one might also question on an intuitive basis whether it is indeed possible to represent all inpatient outputs by means of only three diagnostic categories. In any case, Baker's methodology is one deserving replication in future studies of other hospitals.