THE DERIVATION OF A MEAL:WHEY PRODUCTION FUNCTION FOR PIGS.

A Thesis presented at
Massey University College of Manawatu
in part fulfilment of the requirements
for the Degree of Master of Agricultural Science
in Victoria University of Wellington.

R. J. TOWNSLEY

OCTOBER 1963.

ACKNOWLEDGMENTS.

I am indebted to Professor W. V. Candler for his encouragement, advise and constructive comment throughout this study.

I also wish to thank Mr. A. C. Dunkin for making available data from the 1960 Meal:Whey levels pig feeding trial. His continued interest and help was also greatly appreciated. The assistance of Dr. E. W. Jones and staff of the Applied Mathematics Laboratory, D.S.I.R., Wellington in the preparation of data and use of Treasury's augmented type IBM 650 computer for regression analyses presented in this thesis is gratefully acknowledged. I wish to thank Mr. J. N. Hodgson for helpfull discussion, especially with Management aspects of this study.

An assistantship in the Farm Management and Economics Department, together with Scholarships from the New Zealand Pig Producers' Council and the Manawatu Co-operative Dairy Company enabled the author to carry out this study and are gratefully acknowledged.

CONTENTS.

			Page
CHAPTER	I	Introduction	1
	II	Nature of the Production Function.	25
	III	Problems in the Estimation of Production Functions.	56
	IV	A Meal:Whey Production Function for Pigs - Estimation Procedure.	76
	V	Results of Regression Analyses.	96
	VI	Production Economics Analysis.	120
APPENDIX	I	Report on Meal and Whey Levels Trial 1960/61.	137
	II	Variables Used and Their Notation.	141
	IIIA	Meal and Whey Consumption Data After 11 Weeks For 27 Pigs.	143
	IIIB	Occurrence of Scouring Over Trial Period.	144
	IVA	Summary of TRAP - Multiple Regression Analysis Programme.	145
	IVB	Experimental Data - Availability.	147
	VA	Cumulative Production Model - Log Function.	150
	VB	Cumulative Production Model - Square Root Function.	151
	VI	Fortnightly Incremental Production Models.	152
	VII	Maximisation of a Quadratic Function Subject to Linear Inequalities.	155
	VIII	A Review of Literature on the Selection of Experimental Designs for Response Surface Estimation.	160

CHAPTER I

INTRODUCTION

This chapter discusses in some detail, the usefullness of knowledge about technological relationships, in the form of a production function, in management processes associated with pigmeat production in New Zealand.

1.1 The Management Process

In 1939 T.W. Schultz wrote a fundamental article[1] pointing out that the farm firm exists in a dynamic economy where production may be adjusted and co-ordinated in response to changing conditions. The motivation for change at the farm level is generally the expectation of progress in the attainment of a set of objectives held by the entrepreneur. The whole process of making adjustments and changes within the framework of the firm has become known as the "Management Process".

In the article referred to, Schultz pointed to the two main interests of Farm Management workers and Agricultural Economists, namely:

(1) a desire to provide a basis for guiding entrepreneurial decisions under dynamic conditions; or, in more up-to-date terminology, to assist farmers in carrying out the management

(1) "Theory of the Firm and Farm Management Research", T.W. Schultz, J.Farm Econ., Vol.21, 1939, p.570.

process with the aim of maximising their objective functions, and

(2) to provide results of use to policy makers in understanding the relationship between micro and macro adjustments in agriculture.

Schultz felt that Farm Management research was failing to further both of these interests and pointed to the reason as a lack of understanding of the dynamic nature of the managerial process employed by farmers. He therefore reviewed analytical tools suitable for understanding actions of the firm.

In a reply [(2)] J.D. Black noted that current research at that time was aimed at exploring the general shape of response surfaces of interest, and that such technological information would surely help farmers in carrying out their managerial processes.

Careful consideration of the views expressed in these two papers is useful. Schultz says that Farm Management researc with the aim of helping farmers attain their objectives, cannot be fully directed towards this aim until the management process is understood. The full understanding of micro and macro adjustments in agriculture, of use to policy makers, is also dependent on knowledge of the management process.

Black, on the other hand, pointed out that if knowledge on technological relationships is necessary for the efficient functioning of the management process, then it is worthwhile doing research on this aspect of the process, even though full understanding of other factors necessary in the process is not available.

---

(2) "Dr. Schultz on Farm Management Research," John D. Black, J. Farm Econ., Vol.22, 1940, p.570.

Both authors recognise the dynamic nature of the information required to assist farmers in the management process. The process itself is dynamic in nature, and its character will change with type of production, farmer education, time, etc. The viewpoints of Schultz and Black are complementary, and explai the development of research into the management process on one hand, and research into facets of agricultural production aimed directly at improving efficiency of the management process, on the other. Many examples of the complementary nature of these fields of knowledge exist. Experiments run by Heady, Catron, et al, on pork production (3) have explored the marginal rate of substitution between carbohydrates and protein with and without the use of antibiotics. Resulting knowledge of the pork production function has allowed production economics principles to be applied to the problem of computing least-cost protein-carbohydrate rations for pigs, as well as making a choice between different marketing weights under differing price situations for feeds and product.

These recommendations maximise revenue from a given litter of pigs, through the well known principle of equating marginal product value with marginal input cost. A consideration of pork production, however, leads to the conclusion that where pigs are produced on a continuous basis certain "length-of-run problems" arise. Where profit is the objective in the

(3) "New Procedures in Estimating Feed Substitution Rates and in Determining Economic Efficiency in Pork Production", by E.O. Heady, D.V. Catron, D.E. McKee, G.C. Ashton, and V.C. Speer, Research Bulletin 462, Nov. 1958, Iowa State College.

management process, the aim will be to maximise profit over time rather than maximise profit per litter of pigs. In general these aims will not lead to the same production plans.(4)

Thus, as new aspects of the management process become clear, research can be carried out to provide relevant information, and hypotheses can be formulated and tested.(5)

The development of Farm Management (of which the Management Process is the essential feature) as a discipline has been reviewed by Glenn L. Johnson.(6)

Early in the history of Farm Management it was felt that Agricultural Economics had a considerable contribution to make to Farm Management, which until then placed emphasis mainly on technical agricultural sciences. Other disciplines from both the sciences and the humanities are making much needed and important contributions to Farm Management. These include statistics, logic, sociology, home economics, psychology, philosophic value theory, as well as the physical and biological sciences.

---

(4) "Results from Production Economic Analysis", Glenn L. Johnson, J. Farm Econ., Vol.37, 1955, p.206.

(5) A notable example of a major research project which made contributions both to the theory of management and empirical testing in agriculture is the Interstate Managerial Survey. The results and description of this research have been reported in various articles in the J. Farm Econ. i.e., "Progress and Problems in Decision Making Studies", Vol.37, 1955, p.1097, H.R. Jensen, C.B. Haver, et al., and in bulletins of the seven experimental stations involved, i.e., "Information Needs in Farm Management", D.W. Thomas and R.J. Amick, Purdue University Research Bulletin No.705, 1960. A book has been written summarising the methodology and major findings - "A Study of Managerial Processes of Midwestern Farmers", G.L. Johnson, et al. (Eds), Iowa State Univ. Press, Ames, Iowa, 1961.

(6) "Agricultural Economics, Production Economics and the Field of Farm Management", Glenn L. Johnson, J. Farm Econ., Vol.39, 1957, p.441.

The conclusion reached after study of the literature pertaining to the field of management process, is that before doing research aimed at helping farmers attain their objectives, some knowledge of the relevant management process is necessary. As more is learnt about management processes used in the field of interest, research can be directed to areas of need. The relationship is dynamic and will continue to be so while we are faced with a dynamic economy.

With this brief background it is proposed to consider Farm Management research into pig-meat production in New Zealand, where the aim of such research is to assist farmers maximise their objective functions. The characteristics of pig-meat production in New Zealand are discussed and this leads to the broad description of features of likely management processes that exist in the industry. Avenues of Farm Management research will then be discussed with respect to these features.

1.2 Characteristics of Pigmeat Production in New Zealand

Pigmeat production in New Zealand is carried out under widely differing conditions, with a variety of breeds in use(7) and with different marketing opportunities - both during the season and between districts. The conditions of pigmeat production in New Zealand differ from those in other parts of the world because of the bulky nature of the main foodstuffs - skim milk (9% D.M.), whey (6.5% D.M.). The supply of these foods varies from almost zero in the winter months, to a peak in November, and then falls again as the dairying season advances.

(7) "Pig Production in New Zealand: History and Breeds", I.H. Owtram, N.Z.Jnl.of Agriculture, Vol.106,No.4, 15 April 1963, p.291.

The productive performance of the pig can be closely controlled (if desired) by the pig producer. The farrowing dates of sows can be controlled by hand mating. The subsequent litter may be weaned at 7-10 days with appropriate management[8] or at any time subsequent to this, commonly six or eight weeks. By mating directly after weaning sows may be farrowed twice yearly. The growth rate of a litter after weaning is determined to a large extent by both the rate of feeding and nature of the foodstuff (disease factors, breeding, housing, etc., will also play a part in growth rate of course). Pigs may be sold at almost any weight subsequent to weaning (where weaning occurs at 35-40 lb live weight). Pigs from 60-109 lb carcass weight are known as "Porkers", and from 110-140 lb carcass weight as "Baconers". Price of pigmeat varies during the year, general levels falling towards the end of the dairy season when farmers are forced to quit stock or over-winter them on supplementary foodstuffs. Some idea of the present relationship between pork and bacon prices are given in the following table[9]:-

<u>Fat Pigs</u>

Porkers (60-109 lb)	Av. 17d. per lb.
Baconers (110-140 lb)	Av. 18½d. per lb.
Baconers Special Grading Scheme	
Prime No.1	20½d.per lb.
Prime No.2	18½d.per lb.
Grade 2	16½d.per lb.

---

(8) "Early Weaning of Pigs," D.M. Smith, N.Z.J.Ag., 91,No.6 1955,p.594-599.

(9) Massey University College of Manawatu; Farm Management - Guide to 1962/63 Rural Costs and Prices; p.9.

Supplementary foodstuffs are most usually thought of as substitutes for liquid dairy by-products in times of milk scarcity. Whey, however, is low in protein and is most commonly supplemented with barley or meat meal to provide a more balanced diet. A variety of foodstuffs present themselves as supplements to liquid dairy by-products: skim milk powder and buttermilk powder, meat meal and barley-meal, leafy clover pasture, sugar beet and fodder beet and carrots are, perhaps, used most commonly.

The pigmeat producer in New Zealand thus has a high degree of flexibility available in deciding on rations for pigs, feeding, and hence subsequent fattening rates, and selling weights. As mentioned above, reasonable control can also be exercised over farrowing dates, and hence pig numbers during the season.

The remaining general aspect to consider in pigmeat production is that of uncertainty.(10) At the start of the dairy season pig producers will have information on present levels of pigmeat prices and expectations as to the movement of these prices during the year. Given an "average", "poor", or "good" dairy season, the majority of producers could make reasonable estimates, from past experience, as to the quantity of whey (or skim milk) they would be likely to have available in any

(10) Distinction between risk and uncertainty situations is made by E.O. Heady in chapter 15 of his book: "Economics of Agricultural Production and Resource Use", Prentice-Hall, Inc., N.J. 1952:- "Risk refers to variability or outcomes which are measurable in an empirical or quantitative manner, i.e. the statistical probability of a particular outcome is known with certainty. In contrast to pure risk, the probability of an outcome cannot be established in an empirical or quantitative sense for uncertainty". Although some of the things referred to in this section as uncertainties could be reduced to risks with adequate records, the term uncertainty will be used to refer to positions where lack of certainty about outcomes is likely to affect the management process.

month. Whether a "good", "bad" or "average" season is in store, however, will be uncertain. Uncertainty is present in the prediction of the number of piglets that will be farrowed (we might expect 7 or 8), or as to the number in a litter that will survive to weaning. Farmers are commonly uncertain as to the number and weights of pigs that will be on hand at some future date in time; naturally enough, this uncertainty will increase as the point of interest in the future is extended. Expected growth rates from a given feeding schedule may not be achieved.

Pigmeat production in New Zealand could therefore be said to be characterised by a high degree of flexibility in production possibilities. This degree of flexibility is complementary to uncertainty, also a characteristic of pigmeat production in New Zealand.

This then is a broad description of the general situation under which management processes concerned with pigmeat production are carried out in New Zealand. The objective now is to have a closer look at these management processes.

A pig production (or management) system on a farm is the result of a management process or processes, controlled by the farmer (the degree of control will vary from person to person). The endpoint of a pig production system is the sale of a number of pigs, at given weights, over a period of time. This end-point is the result of decisions involving combinations of production factors such as fattening and farrowing facilities, farrowing and fattening schedules, number of sows, labour supply, feed supplies, etc. Decision making is the focal point of the management process. The aim of Farm Management research that concerns us - to assist farmers in the management process - implies helping farmers make the right decisions in the context of maximising objective functions.

It may be possible to categorise decisions made in a particular production process. As we will see, this knowledge of decision categories is helpful in directing research and deriving possible management systems.

Decisions are based on expectations about technological production relationships. Thus we may be interested in elucidation of technological relationships already in use, and new technologies that might widen the field of production possibilities.

Producers may wish to compare alternative management systems in terms of economics and feasibility, as an aid to decision making.

Once the relevant decisions have been made (sometimes without reference to production relationships) and the management system has been adopted, we are interested in the success or failure of the system to fulfil the expected change in the producer's objective function.

Important avenues of Farm Management research, in line with this discussion, are then:

(1) Understanding and knowledge of the categories of decision making existing in the production process.

(2) Elucidation of new and existing technical production relationships.

(3) Development and examples of methods for economic comparison of management systems.

(4) Studies of success and failure will add to knowledge on technical production relationships, and give actual measures of success of management systems.

(5) A profit objective function may be assumed and a "best" management system calculated using results from the fields of research listed above. If the resulting system is not practiced, reasons should be ascertained, with a view to detecting voids and shortcomings in the knowledge associated with (1), (2) and (3). If the feasibility of such a system is simply not well known, extension methods may well result in the stimulation of management processes leading to adoption of such a system.

The widely differing conditions of pigmeat production, and the flexibilities and uncertainties facing the producer, suggest that the second and third avenues of Farm Management research, in conjunction with the first field, offer the greatest possibilities in assisting decision making in pig production in New Zealand. These aspects are discussed at greater length in the remainder of this chapter. Difficulties associated with the derivation of a "best" management system are also discussed.

The fourth avenue of research listed applies only to management systems at present in use within the industry. Neither time nor finance was available for a study of this breadth.

1.3 Decision Making In The Process of Pigmeat Production

Decision making in the process of pigmeat production is discussed under the following three categories:-

(a) Short term decisions.

(b) Intermediate term decisions.

(c) Long term decisions.

(a) Short term decisions: The existence of this category might be expected because of the degree of flexibility and uncertainty, already described, facing the pigmeat producer. These decisions will most likely relate to feeding and selling policies to be adopted for pigs on hand at the present time. These policies will be influenced by the number of pigs on hand, their weights, food supplies, prices of supplementary feeds, prices for various classes and grades of pigs, and expectations about these variables in the foreseeable future (i.e., might be one or two months).

(b) Intermediate term decisions: Relating to farrowing dates, fattening policies, provision of supplementary foodstuffs grown on the farm such as fodder beet, barley, etc. The contracting for future supplies of food, i.e., whey. It is possible for these decisions to be made on the basis of short term considerations, but if intermediate term planning is done then these are the likely fields of decision making.

(c) Long term decisions: Involving the general level and intensity of pig production in relation to labour and capital supply, size of farm, dairy cow numbers and possibility for expanding milk production; i.e., the place of the piggery in the general farm organisation.

The division of decision making as related to pigmeat production into three categories was entirely arbitrary, though not necessarily unrealistic. The existence of a greater number of levels will not affect the result of the following discussion.

Each of the levels or categories of decision making will be inter-related. Thus the feeding schedules and selling programme that maximises net revenue in the next month depends on the number and weights of pigs on hand at the beginning of the month, together with expected supplies of foodstuffs and increases in pig numbers in this period. The number and weights of pigs on hand will depend on farrowing dates and feeding schedules in the past. Food supply may be an uncontrolled variable, i.e., supply of whey or skim milk, or controlled in the case of barley meal that can be purchased in any quantity at any time of the year at a given price. In turn, the number of sows farrowed and the dates of farrowing will depend on farrowing facilities and date when last farrowed.

1.4 Provision of Technological Information

Technological information is always necessary in a decision making process. Information on the technical process of production is summarised in the form of a Production Function, which aims to predict the level of output, for a given number and levels of inputs on which output is dependent. Commonly the Production Function is specified in mathematical form:

$$y = f(x_1, x_2, \ldots, x_k)$$

where we say the output (pigmeat production) y, is a function of various inputs (meal, whey, disease factors, breeding, etc.) $x_1, x_2, \ldots, x_k$; where the level of output depends on the level of the various inputs. There are problems associated with the derivation and mathematical specification of production functions, that will be dealt with later.

1.5 Comparison of Alternatives

Alternative plans or courses of action may be compared, both economically and for feasibility, by methodological techniques that commonly use the technological information discussed in section 1.4 as part of their input. Methods for comparing production alternatives that suggest themselves in this case are -

(a) Partial budgeting,

(b) Linear programming,[11]

(c) Dynamic programming,[12]

(d) Simulation techniques.[13]

Partial budgeting and a linear programme have much in common. A series of partial budgets compare the profitability of alternative processes. Linear programming is a mathematical tool that enables the most profitable process, or combination of processes to be selected from a large number of alternatives. The concepts behind dynamic programming and simulation will become clear from the following discussion.

Having mentioned briefly the analytical techniques that we wish to use to assist in decision making, we now discuss their use with respect to the levels of decision hypothesized to exist in the pig production process.

---

(11) "Linear Programming Methods" Earl O. Heady and Wilfred Candler, Iowa State University Press, Ames, Iowa, 1958.

(12) "Dynamic Programming" R. Bellman, Princeton University Press, Princeton, N.J. 1957.

(13) "Scientific Programming in Business and Industry", A. Vazsonyi, J. Wiley and Sons, N.Y., 1958, Ch.13.

1.6 Short Term Decisions - Economic Analysis

The short term position of the pig producer is characterised by his resource situation, and prices in the immediate future for pigs and foodstuffs, and the fact that reasonable knowledge exists about these variables. Under these conditions we can expect decisions to be made. For example: towards the end of the dairy season a pig producer will have a certain number of pigs of given weights on hand, feed supplies from dairy by-products will be declining, as will be the price for pigmeat. The decisions to be made will generally revolve around the question: what is the most profitable course of action to follow with the pigs on hand? Should some pigs be sacrificed at pork weights so that sufficient feed will be available to take the remaining stock to bacon weights? What are the economics of wintering pigs on supplementary foodstuffs? Perhaps it would pay to mate some pigs as gilts to be sold in-pig during winter? The right answers to these questions (and others) will depend on the resource restrictions mentioned and the technological production relationships that exist. It might be sufficient to provide a farmer with knowledge as to the growth rate that can be expected when pigs of various weights are fed different amounts of foodstuff per week, to enable him to calculate his most profitable (or desirable) short run policy. In this situation partial budgets or linear programming may be used to assist in comparing alternatives. Both these analytical methods will make use of the technological production relationships given by the pigmeat production function. At other times of the year the short run decision process may involve greater or less opportunity for alternative courses of action.

It is known that some pig producers in the Manawatu are at present successfully exploiting the flexibilities that exist in pigmeat production by making mainly short term decisions, and roughly letting farrowing dates, etc. work themselves out, the pigs being dealt with in a manner appropriate to the particular occasion. A study of pig producers using this approach may indicate special features which are necessary for the success of this management system. However, without doing such a study it is possible to imagine one such feature of this management system. The provision of cheap sources of supplementary foodstuff, notably fodder beet where labour is plentiful, could result in pig production being less dependent on a variable supply of liquid dairy by-product such as whey. The removal of the necessity to ensure that pig numbers and appetite are fitted to available feed supplies over the year is probably an essential prerequisite to the success of the management system described. Rather then, feed supply is adjusted, via the use of relatively cheap and storable supplements, to pig supplies. A series of successful decisions can thus be made in the short run.

Of course, short run decisions will be made under other circumstances, the important thing to note is that short run decisions are made, that technological relationships are usually an essential input to the decision making process and that partial budgeting and linear programming could assist in comparing alternatives - both in terms of economics and physical feasibility.

1.7 Intermediate Term Decisions - Economic Analysis

Consider briefly the characteristics of short term decisions. In the short run, the right decisions will depend on the current resource situation, prices and expectations about future variables; the level of some being controlled by present decisions. The present resource situation is a result of past decisions. Past decisions therefore affect the profit it is possible to make in the immediate future; or, our range of present decisions is affected in part by past decisions. In the case where this situation exists (as it does in pigmeat production), the advantages that might exist in intermediate term (say 12 months) planning are obvious. This sort of economic reasoning forms the background for the analytical method of Dynamic Programming. The method applies to the production process where, what has been done in the past affects what can be done at present.(14) This is obviously the case in pigmeat production where past feeding rates will have affected the present weight of the pig, and the present weight together with feed supplies will determine the weight that can be attained at the end of the current period.

The general idea in dynamic programming is to select a period of time over which it is desired to maximise profit (or minimise cost, etc.). This period is divided up into a number of sub-periods, which might correspond to our short term decision periods. It is then possible to consider systematically the effect of making decisions in any following period. The decisions made in the last sub-period will not affect those made in prior sub-periods, we may therefore safely maximise profit for the last sub-period, subject of course to resource

(14) As well as referring to sequential decisions in time dynamic programming may refer to sequential decisions in space, or in general, to any n-dimensional problem that can be split into n one-dimensional problems.

levels at the beginning of the sub-period. These resource levels depend on decisions made the sub-period before (if they do not we have not got a "dynamic" situation), hence it is possible to make our decisions to maximise profit from both sub-periods. The process is continued backwards to the first sub-period and results in a plan that maximises profit for the whole production period. Suitable as dynamic programming sounds in theory for the planning of pig production over a period of say 12 months, severe limitations arise in practice. When there is more than one decision variable to consider in each of the sub-periods the size of the problem becomes rapidly unmanageable.[15]

As we have already discussed, pig production in New Zealand is characterised by some degree of uncertainty and a high degree of flexibility manifested in many alternative production possibilities. Where dairy by-products provide a major portion of pig food, intermediate term planning is necessary to organise pig numbers and appetite in such a way as to maximise profit from the variable feed supply. One aim, according to our definition of desirable Farm Management research, is to evolve methods of deriving plans that will

(15) Where the same parameters carry over from period to period, Candler has shown that the problem can be solved once and for all using Parametric Linear Programming: "Reflections on Dynamic Programming Models", W.V. Candler, J.Farm Econ.Vol.42 pp.920-926, Nov. 1960. However the author knows of no explicit discussion that reduces or simplifies the computational burden of dynamic programming where more than one decision variable is involved in each sub-period and where different decision variables may exist in different sub-periods.

maximise profit (or some other criterion) in any given resource/price situation. The theoretical possibility and practical problems associated with dynamic programming as such a method, have been discussed.

Linear Programming was described previously as a method of comparing the economics of alternative production processes with the aim of selecting the most profitable combination; and might therefore appear to be a suitable aid to intermediate term decision making (planning). However, the large number of production alternatives and restrictions that are, as we have noted, characteristic of pigmeat production, mean that the size of a linear programme to tackle this sort of planning becomes prohibitive. For example, consider the production situation where whey is the main source of pig food during the year. Let us divide the 12 month production period into 26 two-week periods; we may then estimate the expected quantity of whey available in each of these periods. Let us consider the unrealistic assumption that we will only produce pigs to pork weights, and that we wish to consider only three alternative rations that will enable us to do this. Sows may be farrowed, theoretically, in any of the 26 two-week periods during the year, thus each pork production ration results in 26 pork producing activities (or processes). Barley meal may be purchased in any period to supplement whey supply, and let us consider the case where we vary parametrically the total quantity of meal purchased from zero to some upper limit. Without adding any further restrictions such as fattening facilities, labour supply, etc., we have 27 restrictions, and 4 x 26 = 104 activities, (3 x 26 = 78 pork producing activities, and 26 buying meal activities). This gives an initial matrix of 27 x 105 = 2835 elements. Although a very considerable number of these elements will be zero in the initial tableau, they may not be after several iterations. A linear programme of this size is too big for convenient analysis by electronic computers in common use in New Zealand, such as the IBM 650 which has 2000 memory cells on drum storage. (Magnetic tape storage could be used but would be time consuming.)

Even if this problem could be solved on a larger computer, it can readily be appreciated that the formulation of production alternatives was unsatisfactory, and a solution would be of relatively little use in planning pigmeat production. The size of the problem could be reduced by either considering a shorter production period, or by dividing the 12-month period into fewer sub-periods, i.e., months. However, the introduction of further production alternatives and resource restrictions to give a more realistic description of the problem again result in the capacity of computers such as the IBM 650 being exceeded.

It appears, therefore, that the size of the problem prohibits the use of linear programming for intermediate term planning of pig production in New Zealand, at least in the case where a variable supply of dairy by-product provides the main source of foodstuff. If linear programming could be used, i.e. electronic computers with sufficient fast access storage space were available, an important aspect would be the specification of alternative rations for pig production to various weights. This information is derived from the pigmeat production function.

A third and less widely understood class of analytical methods could possibly be used for intermediate term planning for pigmeat production, where decision making may best be described as multistage. These are known as Simulation Techniques. Where distribution functions can be specified for the random variables that occur in the production process; i.e., the number of pigs per litter, the whey supply in a given month, the probability of getting a sow in pig at the first or second or third heat after weaning, etc., we have the possibility of simulating the production process. Physical production relationships will be given by the pigmeat production function. A pig production model for some period (i.e., 12 months) is set up with restrictions on labour supply during the year, fattening and farrowing facilities, number of sows, etc. Pigmeat prices at any time may be also specified in terms of a probability distribution.

The results of a simulation of the production process revolve around the answers to various questions that are asked during the run. The simulation may therefore proceed by asking how many sows are available for mating and how many of these should we mate at this time. The answers to these questions will then give subsequent farrowing dates and the numbers of piglets obtained and weaned will depend on two more questions and answers. A selected ration will result in predicted liveweight gains of the animals. Whether food supply is from whey or supplementary feed, or some combination, will depend on the expected quantity of whey available, and the competition from other animals for foodstuff, at that time; this information also being the result of questions and answers. Similarly with questions on selling weights - the answers to which may well be determined by expected future whey supply and prices. The method of "obtaining the answers" is given by random sampling techniques; resulting in the name of Monte-Carlo for this type of simulation. Thus the value of a particular variable is obtained by random sampling the relevant probability density function. A large number of simulations (carried out on an electronic computer) results in a large number of production plans. Some results occur more frequently and with higher profits than others. We thus obtain a large number of plans and some idea of their probability of occurring, and the expected profit from each can be calculated.

Whether or not sufficient information exists for the specification of the distribution functions of some of the important variables mentioned above is not known. The possibility of using simulation techniques for an analysis of the profitability of alternative methods of pigmeat production could be the subject matter of another study.(16) This method would, however, appear to offer definite possibilities as a method of attack in intermediate term planning of pigmeat production. Once again, it would be essential to be able to predict live-weight gains from various feeding rations as an integral part of the simulation technique.

We conclude, therefore, that the nature of the problem and the methods of analysis available make intermediate term

(16) For an introductory paper on the use of simulation as a research technique in agriculture see: "An Introduction To The Use Of Simulation In The Study Of Grazing Management Problems" P.L. Arcus. Proc.N.Z. Soc. of An. Prodn. Vol.23 pp. 159 - 68, 1963.

planning a relatively uncertain proposition. If suitable methods of analysis and adequate computing facilities were available, technological information from the pigmeat production function would be an essential part of such an analysis.

1.8 Long Term Decisions - Economic Analysis

Long term decisions will be concerned with the general level of pig production that is most profitable, or meets some other requirements. Pigmeat production competes with milk production for labour, capital, and in some cases (where supplementary crops are grown) for land. In making long term decisions as to the desirable level of pig production, Budgeting and Linear Programming are analytical methods that allow the relative profitability of various levels of pigmeat and milk production to be compared. However, the apparent profitability of any level of pigmeat production will depend on how well short and intermediate term decisions have been made.

1.9 Summary

We have discussed pigmeat production with the aim of doing Farm Management research to assist pig producers in maximising their objective functions. We realise that a production process is the result of decisions made by the farmer. The uncertainty and flexibilities associated with pigmeat production have been stressed. Three levels of decision making in pigmeat production were postulated, and Farm Management research was discussed in relation to each of these levels.

1.10 Conclusions

The inter-relation between levels of decision making is evident, and it is realistic to assume that this inter-relation and the relative importance of these levels will vary between farms. However, the existence of this situation should not deter us from doing Farm Management research into problems associated with these levels of decision making.

It would appear likely in this situation that elucidation of the technological relationships that exist in pigmeat production (the pigmeat production function) would be of value in making decisions, especially in the short term.

A study of methods of economic analysis, concentrating either on some form of Dynamic Programming, or Simulation Techniques, would be of greatest value for intermediate term planning. However, the pigmeat production function is also essential to economic analysis (crude though it may be) for intermediate term planning.

Successful long term decisions can best be made when based on optimum short and intermediate term planning of the piggery.(17)

Two main fields of Farm Management research are thus seen to be:

(1) Elucidation of new or existing technical production relationships.

(2) Development and experience of methods of economic analysis.

---

(17) Long term planning will often be done when these conditions are not fulfilled. This, however, could well result in less than optimum long term planning as the present profitability of pig production is sub-optimum with respect to resources in use.

In helping to decide in any particular instance which of these fields most warrants research, we should try to focus attention on the decision variables considered by farmers, and where the lack of knowledge exists that presumably hampers decisions. Time has not allowed an actual survey or case farm studies of farmers to determine these variables and knowledge voids with respect to pig production. However, it is felt that general knowledge and discussion with a few people in touch with pig production problems has been sufficient to describe fairly accurately the problem setting.

This discussion has lead to the conclusion that the correct estimation of the production function should be a primary consideration in research aimed at helping farmers make the right decisions in pigmeat production in New Zealand. The remainder of this thesis is therefore directed towards this aim.

CHAPTER II

NATURE OF THE PRODUCTION FUNCTION

The preceding chapter discussed the usefulness and place of technological knowledge, commonly expressed in mathematical form as a production function, in management processes associated with pigmeat production. This chapter considers in some detail what are desirable characteristics of a pigmeat production model.

For the moment then we hypothesize that the production function we are interested in can be conveniently expressed as a mathematical model, where the expected value of one variable (the yield) is defined as a function of the observed values of other variables (the inputs). In common notation we write:

$$\hat{y} = f(x_1, x_2, \ldots, x_k) \qquad (2.1)$$

where $\hat{y}$ is the predicted value of the yield, given the levels of the input variables $x_1, x_2, \ldots, x_k$. Regression analysis may be defined,[1] as the estimation or prediction of the value of one variable from the values of other given variables; and is therefore suited to the expression of production functions.

In carrying out a regression analysis there are three aspects to consider:

(1) The particular <u>variables</u> among which the relationship is hypothesized to exist.

---

(1) "Regression Analysis", E.J. Williams, J. Wiley and Sons Inc.,N.Y., 1959.

(2) The mathematical form of the model.

(3) The estimating procedure to be used.

These aspects should be considered with the aim of obtaining the most accurate description possible of the production process of interest. More specifically, this aim is usually interpreted as obtaining unbiased and minimum variance estimates of $\hat{y}$. The estimating procedure of Least Squares fulfils these requirements under certain assumptions and will be considered at a later stage.

The particular variables and the mathematical model used in the production function depends partly on our knowledge and understanding of the nature of the production process. Production economics principles may provide some basis for the selection of a suitable mathematical form.

Statistical tests may be used to check the effectiveness of different variables and mathematical models in describing the production process.

A combination of logic and statistical testing is therefore used in deciding upon variables and the mathematical form for the production function. The remainder of this chapter is concerned with the selection of these two characteristics for a pigmeat production function based on present knowledge of nutritional relationships in the nature of the production process.

2.1 Two Classes of Animal Products

Animal products can be divided into two broad classes on the basis of the biological processes involved: those that result from the growth of the animal, and those that are "manufactured" by the animal. Meats are the principle elements of the first class, and milk and eggs are examples of the second. We are interested in this thesis in the first type of biological process - pigmeat production.

2.2 The Pigmeat Production Function

We will be concerned with the problem of estimating liveweight gain for pigs over the liveweight range: weaning to bacon weights.

Meat production is unlike other types of production wherein a single level and combination of factors is used in the expectation of obtaining a single output. Instead, a series of factor combinations is employed over time as the animal is taken from the initial to the final weight.

The nutritional relationships that exist when feeding pigs from weaning to higher weights, with skim milk as the main feedstuff, have been explored by D.M. Smith, at Ruakura Animal Research Station.(2)

Fattening requirements for pigs are a function of efficiency of liveweight gain at different rates of feeding and at different liveweights. By efficiency of liveweight gain we mean: the rate at which the particular feed of interest is converted into liveweight gain; and is alternatively referred to as "feed conversion efficiency". The relationship between

---

(2) "Factors Affecting the Efficiency of Food Conversion by Pigs," D.M. Smith, Proc.N.Z. Soc.An.Prodn., Vol.11, 1951, p.89. "Feed Value of Meal When Used as a Supplement to skim-milk for Pig Feeding", D.M. Smith, N.Z.J.Sc.Tech., Vol.34, Sec.A, 1953, p.544. "Effect of Level of Feeding Upon the Efficiency of Utilization of Separated Milk by Pigs", D.M. Smith, N.Z. J. Sc. Tech., Vol.38, Sec.A, 1956, p.217.

liveweight gain from weaning (y) and feed consumption (x) is illustrated for a single pig in fig.1. This relationship may be expressed either as:

$$y = f(x), \quad \text{or} \quad x = f(y).$$

The derivative $\frac{dy}{dx}$ evaluated at any point along the feed axis gives the efficiency of feed conversion at that point in terms of: liveweight gain per unit of feed consumed. The derivative $\frac{dx}{dy}$ evaluated at any point on the weight gain axis is also a measure of efficiency of feed conversion, but in terms of: feed consumed per unit of liveweight gain.

We can easily prove that fattening requirements for pigs are a function of efficiency of liveweight gain. This may be done by writing the feed required (x') for any given liveweight gain (y') as follows:

$$x' = \int_{y=0}^{y'} \left(\frac{dx}{dy}\right).dy \qquad (2.2$$

This amount of feed required for the given liveweight gain can be a function of nothing else but the efficiency of liveweight gain measured by $\frac{dx}{dy}$.

Similarly, we note that liveweight gain achieved from a given quantity of feed depends implicitly on efficiency of liveweight gain thus:

$$y' = \int_{x=0}^{x'} \left(\frac{dy}{dx}\right).dx \qquad (2.3$$

This latter relationship gives us an exact production model for liveweight gain (y') from a given amount of feed consumed (x').

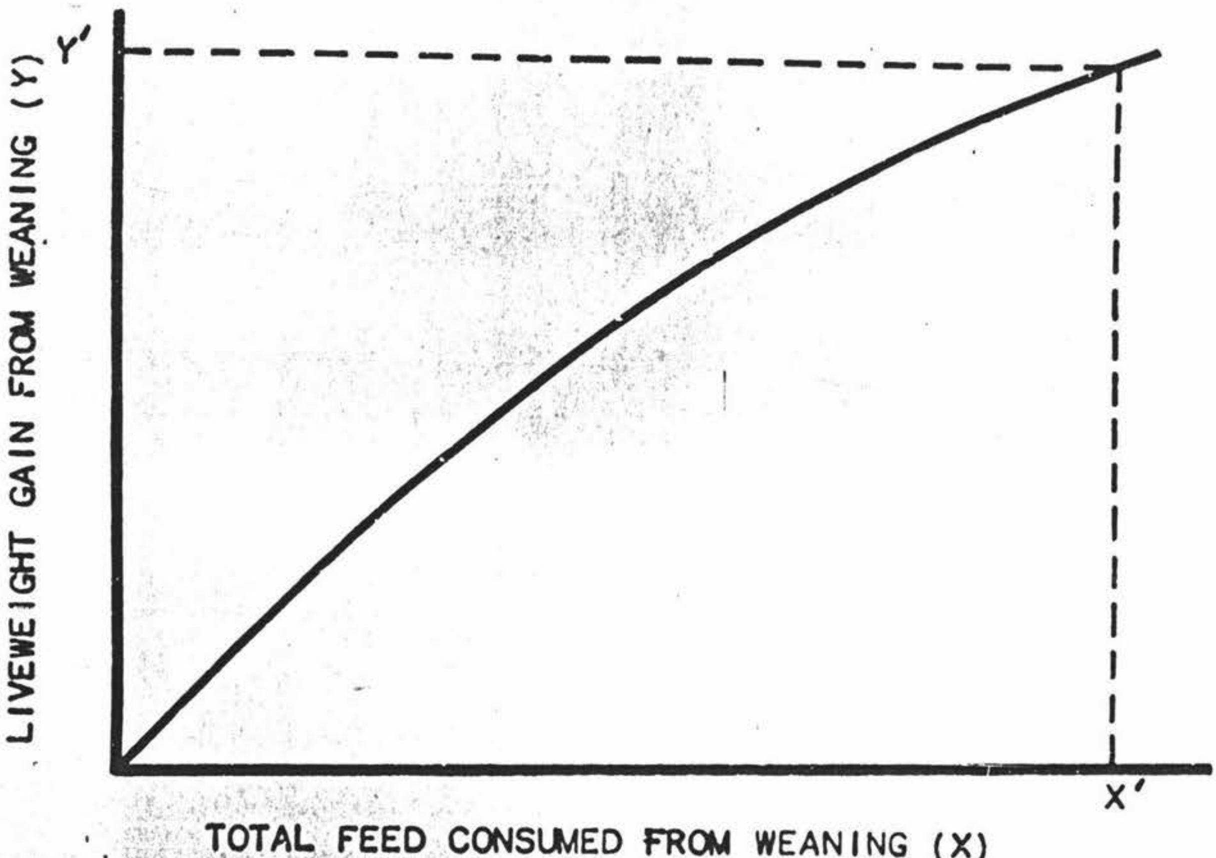


FIG.1 RELATIONSHIP BETWEEN FEED CONSUMED AND LIVEWEIGHT GAIN FOR A SINGLE PIG

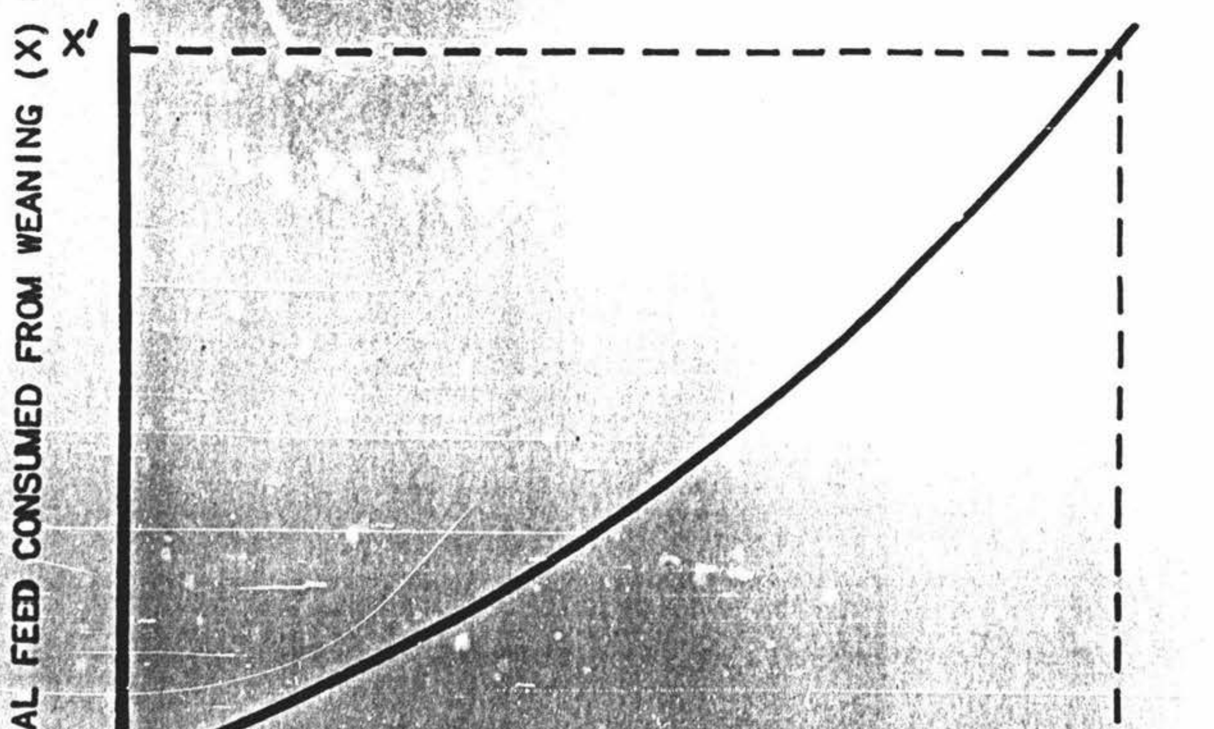


FIG.2 RELATIONSHIP BETWEEN FEED CONSUMED AND TIME FOR A SINGLE PIG

Similarly it can be seen that weight gain can be expressed exactly for a given time period by the function:

$$y' = \int_{T=0}^{T'} \left(\frac{dy}{dT}\right).dT \qquad (2.4)$$

where the derivative $\frac{dy}{dT}$ is commonly referred to as "rate of growth".

The production relationships given by equations (2.3) and (2.4) are exact in the sense that, given "efficiency of liveweight gain" for a quantity of feed, or "rate of growth" over a period of time, a definite value for weight gain, y, from this quantity of feed or in this period is implied. Where, however, "rate of growth" or "efficiency of liveweight gain" cannot be measured or estimated without error, weight gain, y, as given by (2.3) and (2.4) will be stochastic.

For a single pig the production process is summarized by the "overall" functions:

$$y = f(x) \quad ..... \quad (\text{fig.1})$$

and $x = f(T)$[3] ..... (fig.2)

Rate of growth $\frac{dy}{dT}$ is given at time T' by calculating $\frac{dy}{dx} \times \frac{dx}{dT}$, where $\frac{dx}{dT}$ is evaluated at T', and $\frac{dy}{dx}$ is evaluated at x' given by $x' = f(T=T')$.

We now show that in the more general case of more than one animal fed on the same ration, difficulties can arise if we attempt to describe the production process by the overall

(3) In this context it should be noted that x = f(T) refers to a single pig fed a specified ration, x, at a specified rate through time. In the case of restricted feeding through time this relationship can be specified exactly, i.e. it is non stochastic. For <u>ad lib.</u> feeding the relationship would be stochastic.

relationships sufficient for an individual animal. In this more general context it is found useful to consider factors affecting efficiency of feed conversion, measured by $\frac{dy}{dx}$. If a general relationship affecting $\frac{dy}{dx}$ could be found we should be able to (theoretically at least) evaluate the integral:

$$y' = \int_{x=0}^{x'} \left(\frac{dy}{dx}\right).dx \qquad (2.5)$$

over the range $x = 0 \longrightarrow x = x'$ of interest.

2.3 Two Pigs fed on the same Ration

Fig.3 illustrates the simplest case of this more general context referred to, where two pigs have consumed exactly the same quantities of ration but different weight gains have resulted. Unless we are willing to assume that the differences in weight gain are due to the inherent variability in feed conversion efficiency between pigs, it is meaningless to attempt to relate liveweight gain to feed consumed by the single relationship:

$$y = f(x)$$

where y is the expected total weight gain for any given quantity of feed consumed, x. These differences in total weight gain between pigs have arisen due to differences (including those due to inherent variability) in feed conversion efficiency, $\frac{dy}{dx}$. In describing the production process, as we wish to do, it is meaningful therefore to look for factors that affect feed conversion efficiency.

The nutrition studies carried out by Smith have confirmed that under New Zealand feeding conditions efficiency of liveweight gain, $\frac{dy}{dx}$, is affected by relationships between rate of

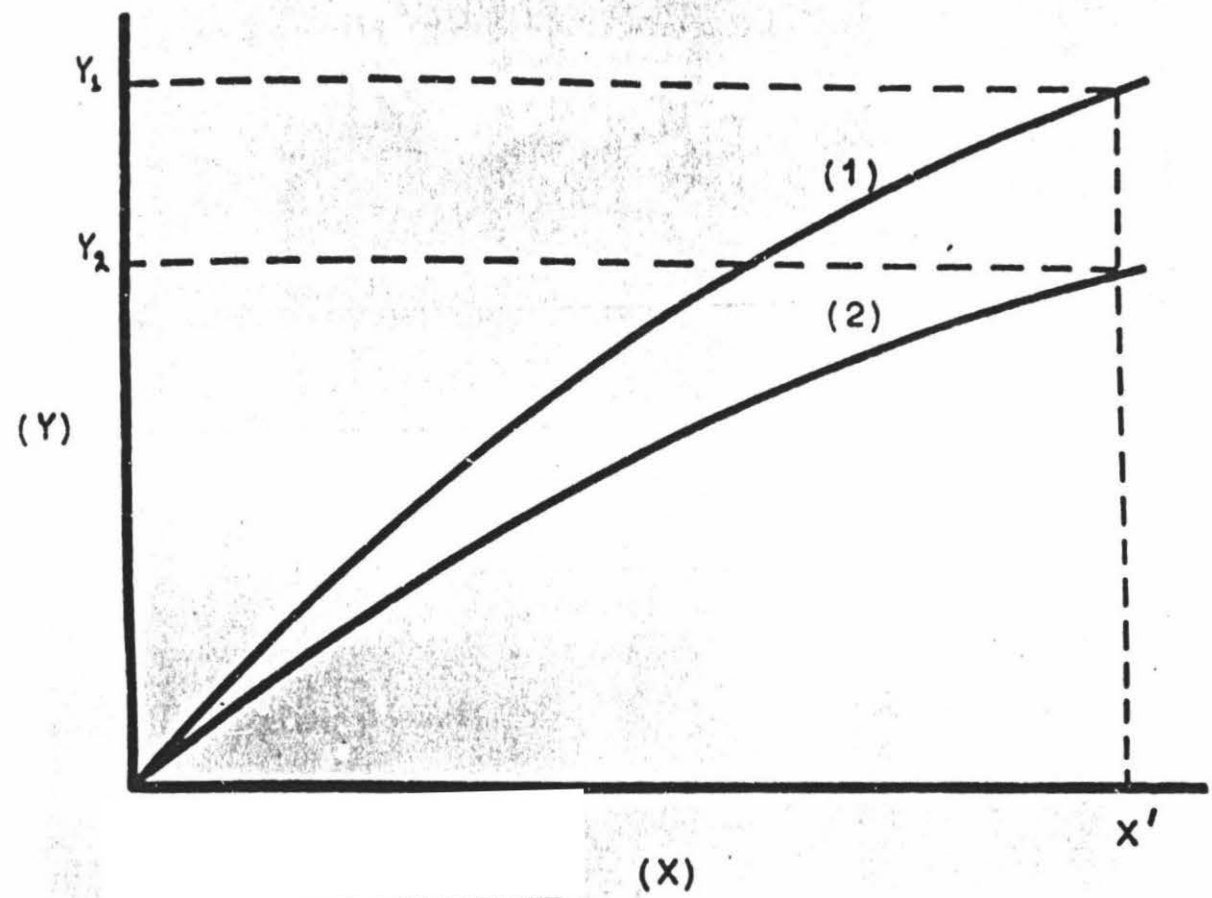


FIG.3 RELATIONSHIP BETWEEN (X) AND (Y) FOR TWO PIGS

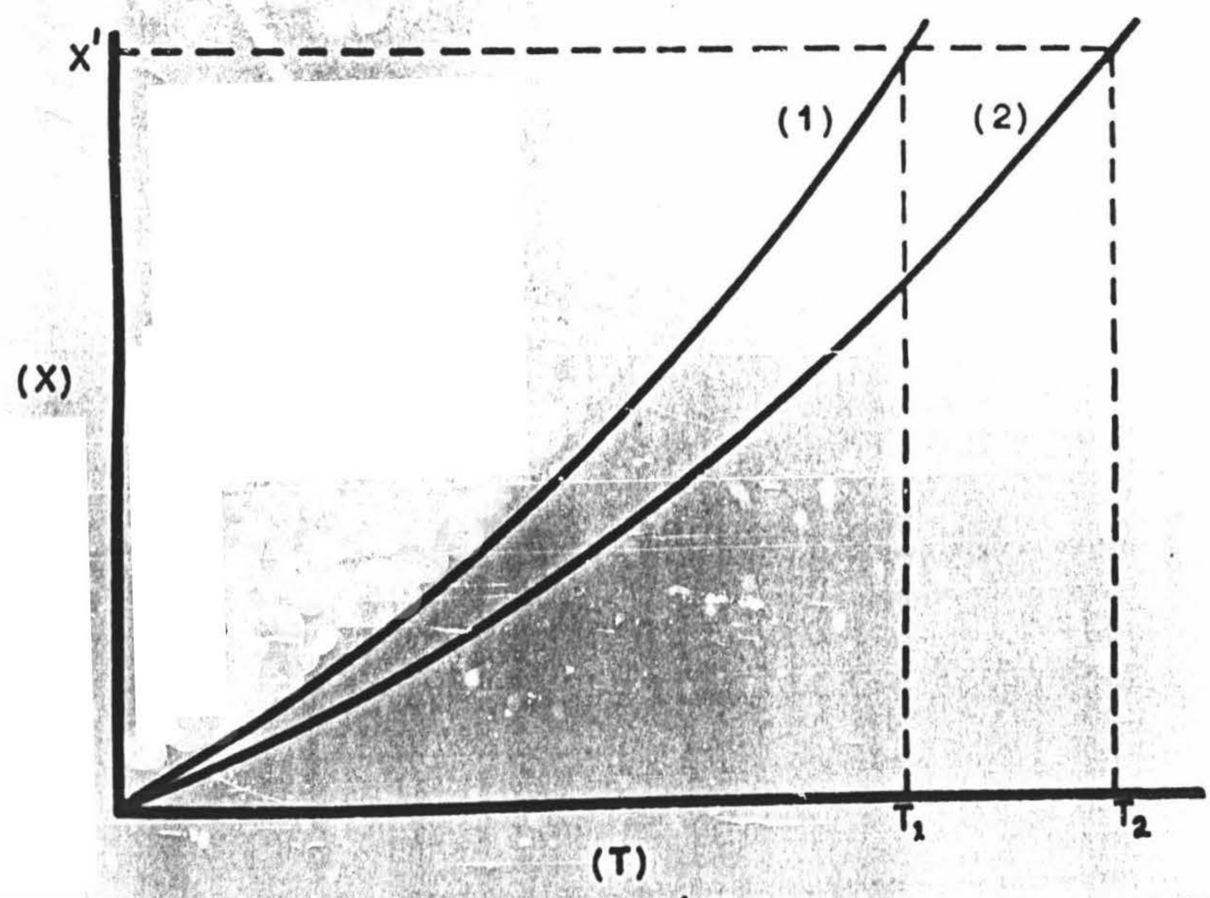


FIG.4 SAME QUANTITY OF FEED X' CONSUMED AT DIFFERENT RATES BY TWO PIGS

feeding and liveweight of the animal.[4] Animal nutrition specialists have divided total feed input into a maintenance and a production portion. The importance of time to the feed-output relationship can be seen because variations in time involve variations in total quantity of feed used for maintenance purposes. The more time, the more feed that must be used for maintenance. In fig.3, we noted that both pigs though consuming the same quantity of feed, reached different liveweights. If we relate total quantity of feed consumed to time, for these two pigs, the relationships shown in fig.4 may be found to exist. One of the reasons for the lower conversion rates realized by the second pig could well have been the higher total maintenance requirement due to the longer period of feeding. The nature of maintenance requirements suggests that they are related to animal liveweight in some way.

The situation as envisaged is represented in fig.5. At any liveweight we can assume a given maintenance requirement. The particular liveweight will also affect potential feed conversion efficiency and maximum rate of feeding that is possible. The actual rate of feeding, combined with maintenance requirements and liveweight thus affects the feed conversion efficiency achieved. This situation may be written in conceptual terms:

$$\frac{dy}{dx} = f(\text{liveweight}, \frac{dx}{dT}) \qquad (2.6)$$

which may be stated:

"feed conversion efficiency (efficiency of liveweight gain) is some function of liveweight of the animal and rate of feed intake".

(4) D.M. Smith; 1951, op.cit., and 1956, op.cit.

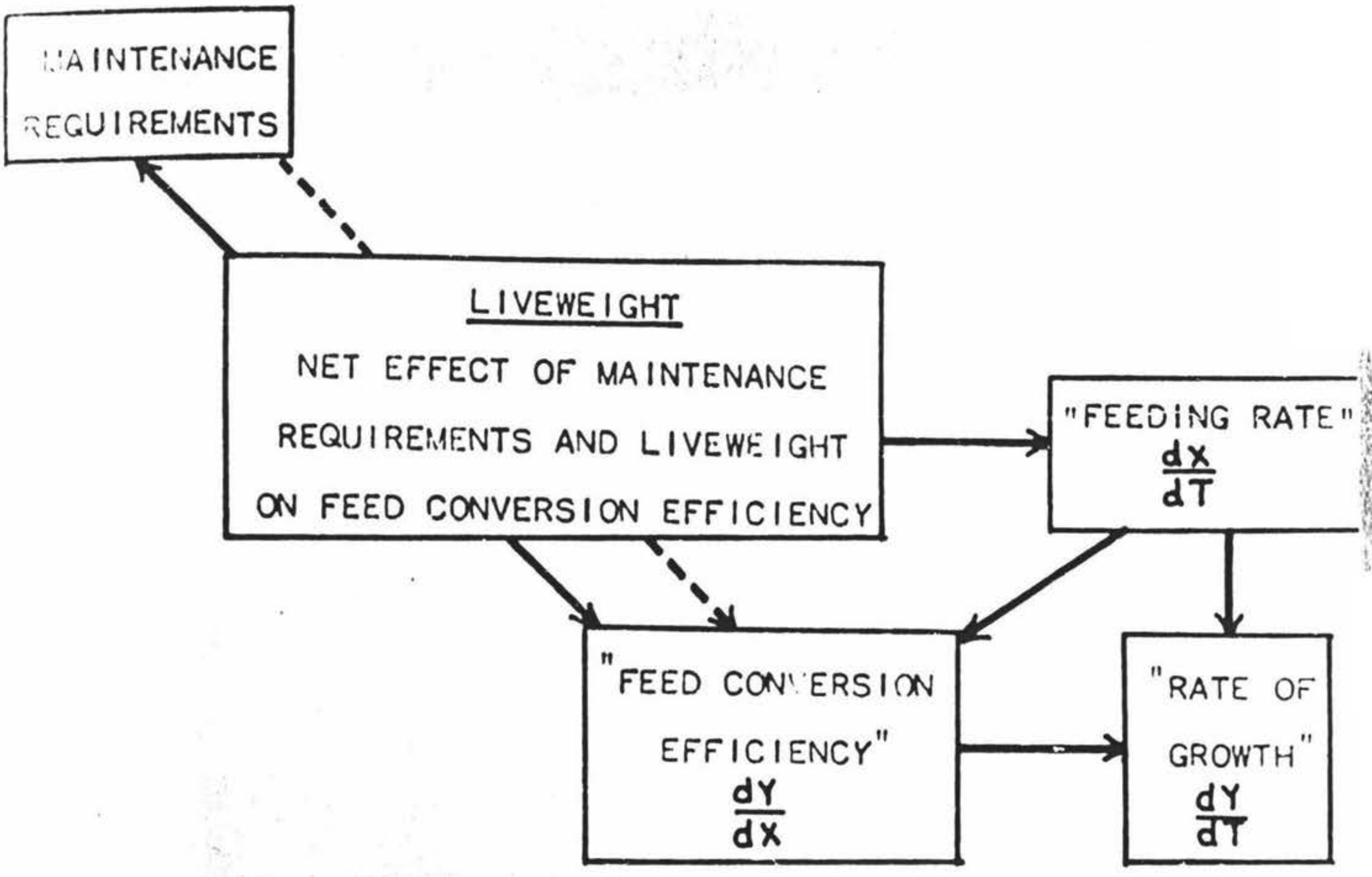


FIG.5 DIAGRAMATIC REPRESENTATION OF FEEDING FACTORS AFFECTING "FEED CONVERSION EFFICIENCY" AND "RATE OF GROWTH"

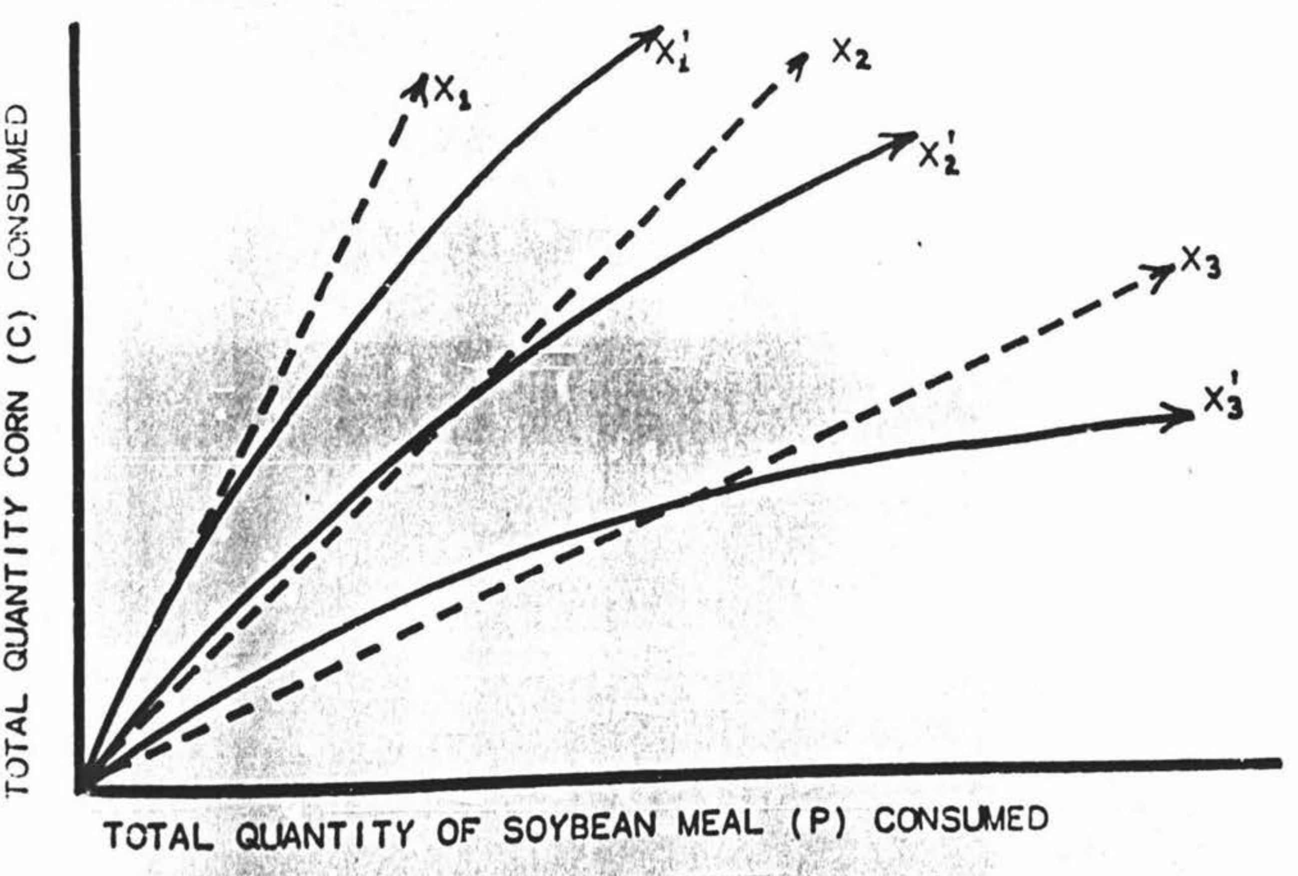


FIG.7 TYPES OF RATION LINE

Until now these relationships have been expressed in terms of first differentials, thus giving instantaneous rates of growth, feed conversion, and feeding rates. When estimating these relationships it is impossible to measure instantaneous rates and incremental rates must therefore be used, thus:

$\frac{\Delta x}{\Delta T}$ is the incremental rate of feeding, where Δx is the small increment of feed intake and ΔT is the small increment of time in which Δx was consumed. We therefore write:

$$\frac{\Delta y}{\Delta x} = f(\text{liveweight}, \frac{\Delta x}{\Delta T}). \qquad (2.7)$$

The term $\frac{\Delta y}{\Delta x}$ <u>estimates</u> incremental feed conversion efficiency and will be a random variable, i.e. equation (2.7) represents a stochastic relationship. This relationship is now defined for some time period ΔT, and it is assumed that the effect of liveweight and Δx on the ratio $\frac{\Delta y}{\Delta x}$ is unchanged over this period. Because we do not know *a priori* the mean liveweight in this period, liveweight at the beginning of the period could be used. The degree to which the above assumption is violated due to the convention of using liveweight at the beginning of the period is most likely to depend on the value of ΔT that is chosen, and Δy that occurs in this period.

This consideration might lead us to choose as small a value for ΔT as possible. Two factors may offset to some degree the desire to make ΔT small. These are:

(1) Errors of measurement. Where these are independent to some degree of say Δy, Δx, etc., larger values of ΔT will result in smaller percentage errors of measurement of Δy and Δx.

(2) There is likely to be some lag effect of rate of feeding in the last incremental period on the rate of growth in the next period. This effect would be reduced by making ΔT larger.

It should be possible to test statistically the effect of changing the incremental period ΔT over which the relationship is assumed to hold.[5]

For the above relationship, (2.7), Δy is the increment in growth occurring at a given liveweight when Δx is consumed. Both Δy and Δx refer to the time period ΔT. In this situation we have the relationship:

$$\frac{\Delta y}{\Delta x} \cdot \frac{\Delta x}{\Delta T} = \frac{\Delta y}{\Delta T}. \qquad (2.8)$$

Now equation (2.7) gives feed conversion efficiency:

$$\frac{\Delta y}{\Delta x} = f(\text{liveweight}, \frac{\Delta x}{\Delta T})$$

and for given Δx and ΔT we may write:

$$\frac{\Delta y}{\Delta T} = \frac{\Delta y}{\Delta x} \cdot \frac{\Delta x}{\Delta T} = \frac{\Delta x}{\Delta T} \cdot f(\text{liveweight}, \frac{\Delta x}{\Delta T})$$

or

$$\frac{\Delta y}{\Delta T} = f(\text{liveweight}, \frac{\Delta x}{\Delta T}), \qquad (2.9)$$

where we note that $f(\text{liveweight}, \frac{\Delta x}{\Delta T})$ is <u>not</u> used to represent any <u>particular</u> functional form.

This situation is intuitively obvious, but the position may be further illustrated graphically, as in fig.6. Fig.6(B) shows the relationship between feed conversion efficiency and rate of feeding at a given liveweight. If this relationship is known (we are hypothesising that it exists in some form) then we can calculate exactly how rate of growth changes with feeding

(5) This aspect is considered further in section 5.10.

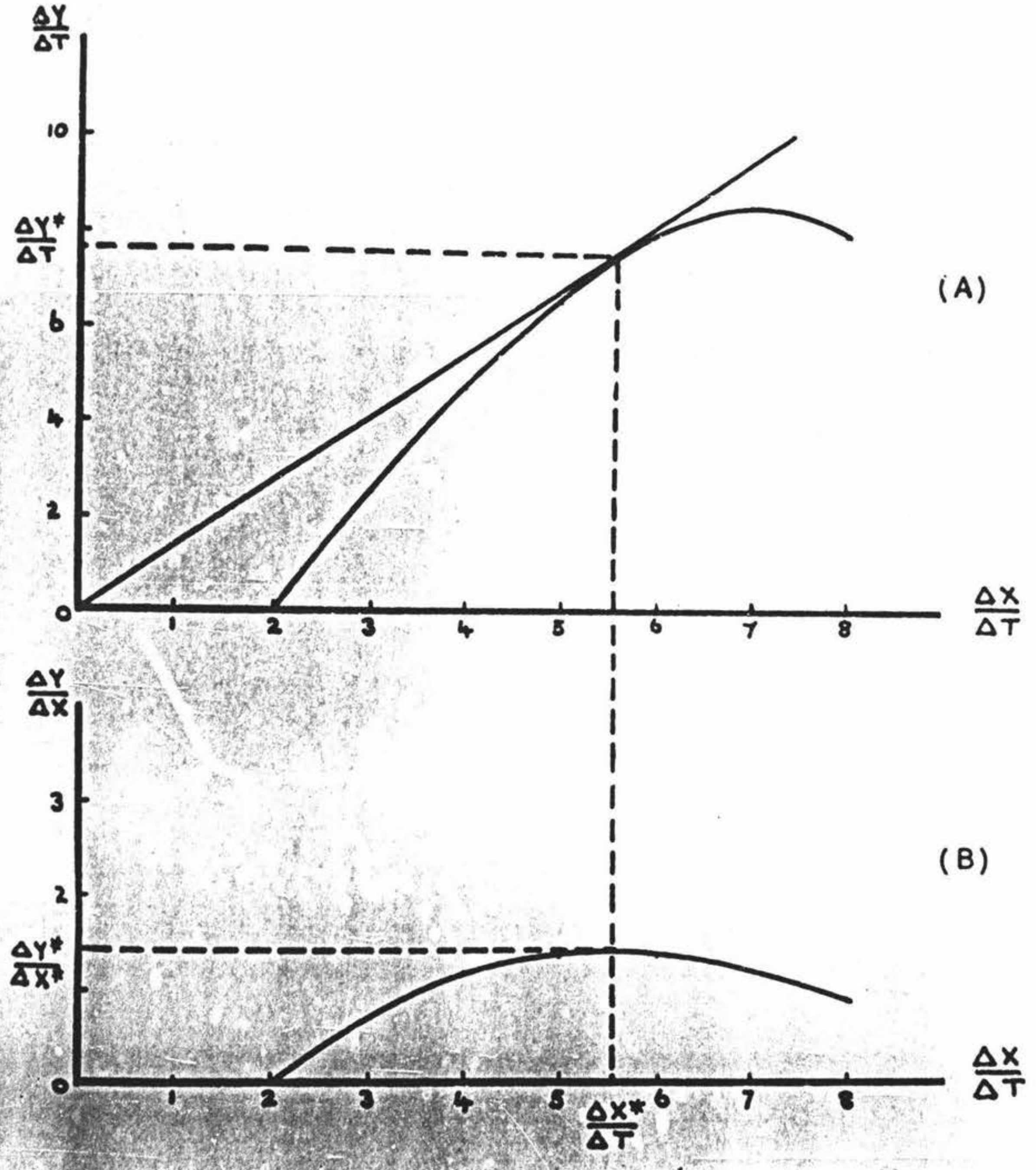


FIG.6 RELATIONSHIP BETWEEN FEED CONVERSION EFFICIENCY AND RATE OF WEIGHT GAIN AT A GIVEN LIVEWEIGHT

rate at the same liveweight. At any feeding rate $\frac{\Delta x^*}{\Delta T}$ we can read off a corresponding value for feed conversion efficiency $\frac{\Delta y^*}{\Delta x^*}$, and also calculate rate of growth $\frac{\Delta y^*}{\Delta T}$ from the identity (2.8) The resulting relationship drawn in fig.6(A) thus depends exactly on the relationship illustrated in fig.6(B). Changes in liveweight will result in different feed conversion efficiency curves (according to nutrition theory) and hence we will get correspondingly different rate of growth curves. With the exact relationship existing between $\frac{\Delta y}{\Delta T}$, $\frac{\Delta y}{\Delta x}$, and $\frac{\Delta x}{\Delta T}$ there is no reason why we should not estimate rate of growth $\frac{\Delta y}{\Delta T}$ directly from the factors affecting feed conversion efficiency, thus:

$$\frac{\Delta y}{\Delta T} = f(\text{liveweight}, \frac{\Delta x}{\Delta T})$$

We now discuss a major advantage of estimating the relationships involved in pigmeat production in this form.

2.4 More Than One Ration

A serious difficulty occurs when we wish to compare feed conversion efficiencies between rations. The ability to be able to do this necessitates generalising the relationship:

$$\frac{\Delta y}{\Delta x} = f(\text{liveweight}, \frac{\Delta x}{\Delta T})$$

to include more than one ration, as is the case illustrated in fig.7. The ratio $\frac{\Delta y}{\Delta x_j}$ refers to the feed conversion efficiency in terms of lbs. liveweight gain per unit of ration x_j.

This leads to the difficulty that there is no obvious way of defining a "unit of ration x_j". We might attempt to define a unit technically as: "a unit of ration x_j, is the amount needed to supply a unit of T.D.N." But we could equally define a unit of ration x_j as supplying a unit of S.E., or P.E., or of a given weight. Thus we see there are alternative definitions of a unit of ration x_j, and it follows that using any of these definitions, the relative "feed conversion efficiency" of the rations will depend, inter alia upon the definition used: an obviously unsatisfactory situation.

Equally the measure $\frac{\Delta y}{\Delta c + \Delta p}$ is unsatisfactory since this again involves the definition of Δc and Δp which will put them on a comprable basis, (where Δc, Δp refer to increments of corn and soybean meal respectively that make up the increment in ration Δx_j).

On the other hand $\frac{\Delta c_j}{\Delta T}, \frac{\Delta p_j}{\Delta T}$ are clearly defined, where $\frac{\Delta c_j}{\Delta T}$ $(\frac{\Delta p_j}{\Delta T})$ refers to the amount of corn (soybean meal) fed, using the jth ration, in the time interval ΔT.

Thus, in the more general case where it is desired to represent the production process for more than one ration we may estimate:

$$\frac{\Delta y}{\Delta T} = f(\text{liveweight}, \frac{\Delta c}{\Delta T}, \frac{\Delta p}{\Delta T})$$

or

$$\Delta y_i = f(\text{liveweight}, \Delta c_i, \Delta p_i)$$

where Δy_i = increase in liveweight in the ith incremental period (i.e., may be week) and Δp_i and Δc_i are the amounts of corn and soybean meal respectively consumed in this period. Furthermore, where these values, Δy_i, are estimated for n consecutive incremental periods, we may estimate total liveweight gain, y_n, over this period from:

$$y_n = \sum_{i=1}^{n} \Delta y_i$$

We have considered models based on nutritional relationships that will allow a meaningful description of the production process under three situations:

(1) A single pig fed on one ration,

(2) A number of pigs, with different feeding rate curves, fed on the same ration,

(3) A number of pigs, fed on different rations, where each ration may be fed at different rates.

With the aid of fig.7 we now distinguish between two types of rations and two methods of feeding. In fig.7 each of the ration lines x_1, x_2, x_3 is made up of a constant proportion of the two feeds, while for x_1', x_2', x_3' the proportion of the two feeds changes continuously. Where it is suspected that type of ration (in the sense just described) affects the nature of the production process we should be aware that knowing the results of using rations of the "constant proportion" type may tell us little about results we could expect from rations of the "continuously varying proportion" type. The type of ration line we are interested in depends to a large extent on convenience of feeding practices. It may be most convenient to mix up large quantities of corn and soybean meal in a given proportion. In the case of barleymeal and whey it may be convenient to feed a constant amount of meal per week, but an increasing quantity of whey per week, resulting in a ration line of the type x'.

We have discussed a theoretical model to allow for differences in liveweight that might occur as a result (at least in part) from feeding a ration at different rates. We now distinguish between <u>ad libitum</u> and restricted rates of feeding.

2.5 Production Models For ad libitum Feeding

Animals offered an amount of ration in excess of what they wish to consume, are said to feed ad lib. Animals offered an amount less than appetite are said to be fed a restricted diet. In America at least, it appears that ad lib. feeding is a common practice in meat production. Production functions based on ad lib. feeding studies at Iowa State University for pigs, broilers, turkeys and beef cattle have been reviewed by Heady and Dillon.(6) More detailed information on the derivation of pig production functions from experimental work at Iowa State University is contained in two research bulletins.(7) The essential features of the experiments used to generate the production data in these latter studies are:

(1) A number of rations with given percentages of corn and soybean meal were decided upon (cf. ration lines x_1, x_2, x_3 in fig.7).

(2) Groups of pigs were fed on each of these rations on an ad lib. basis.

(3) Pigs were weighed regularly and quantities of ration consumed noted.

Thus pigs fed ad lib. on ration x_2 in fig.7 consume progressively the quantities of corn and soybean meal traced out by this ration line. The important aspect of ad lib. feeding is that the total quantity of feed (x) consumed after total time (T) has elapsed depends implicitly on the pig concerned.

The situation for such a group of pigs may be represented by figs. 8 and 9, where all pigs are of the same initial

(6) "Agricultural Production Functions", E.O. Heady and J.L. Dillon, Iowa State University Press, Ames, Iowa, 1961.Chs.8 and 9,10,11,13.

(7) "New Procedures in Estimating Feed Substitution Rates and in Determining Economic Efficiency in Pork Production", Part I, by E.O. Heady, R. Woodworth, D.V. Catron and G.C. Ashton, Iowa Agr.Expt.Sta.Res.Bull.409,1954; Part II, by E.O. Heady, D.V. Catron, D.E. McKee, G.C. Ashton and V.C. Speer, Iowa Agr.Expt.Sta.Res.Bull.462,1958.

liveweight and age. Differences in time taken to consume a given quantity of feed will be due to differences in ad lib. feeding rates, as will be differences in feed consumed in a given time period. Variation in ad lib. feeding rates between pigs on the same ration is a result of the inherent variability of the animals involved and therefore considered as uncontrollable Depending on which of the variables is held at fixed levels while we observe values of the other variable, we may estimate:

$$\hat{x} = f(T) \quad \text{or} \quad \hat{T} = f(x)$$ (8)

where x is quantity of ration and T is consumption time. In the simple case of one ration the relationship estimated will depend on problem context, i.e., whether we are interested in predicting feed consumption in a given time period, or vice versa. Where we wish to generalise between rations, i.e., in the corn: soybean meal plane as in fig.7, we may write

$$\hat{T} = f(c,p)$$

or

$$(\hat{c},\hat{p}) = f(T).$$

The second function must be estimated either in the form

$$\hat{c} = f(T,p)$$

or

$$\hat{p} = f(T,c)$$

and as there is no a priori reason for estimating either one of the variables (c or p) alone as an expected value and not the other[9], it seems logical to estimate neither, but to use instead the function:

$$\hat{T} = f(c,p)$$

---

(8) We note that there is no real causal direction implied in either of these two functions, i.e. we may estimate either: "quantity of feed that will be consumed in a given time period", x, or "time taken to consume a given quantity of feed", T.

(9) For footnote (9) see page 44.

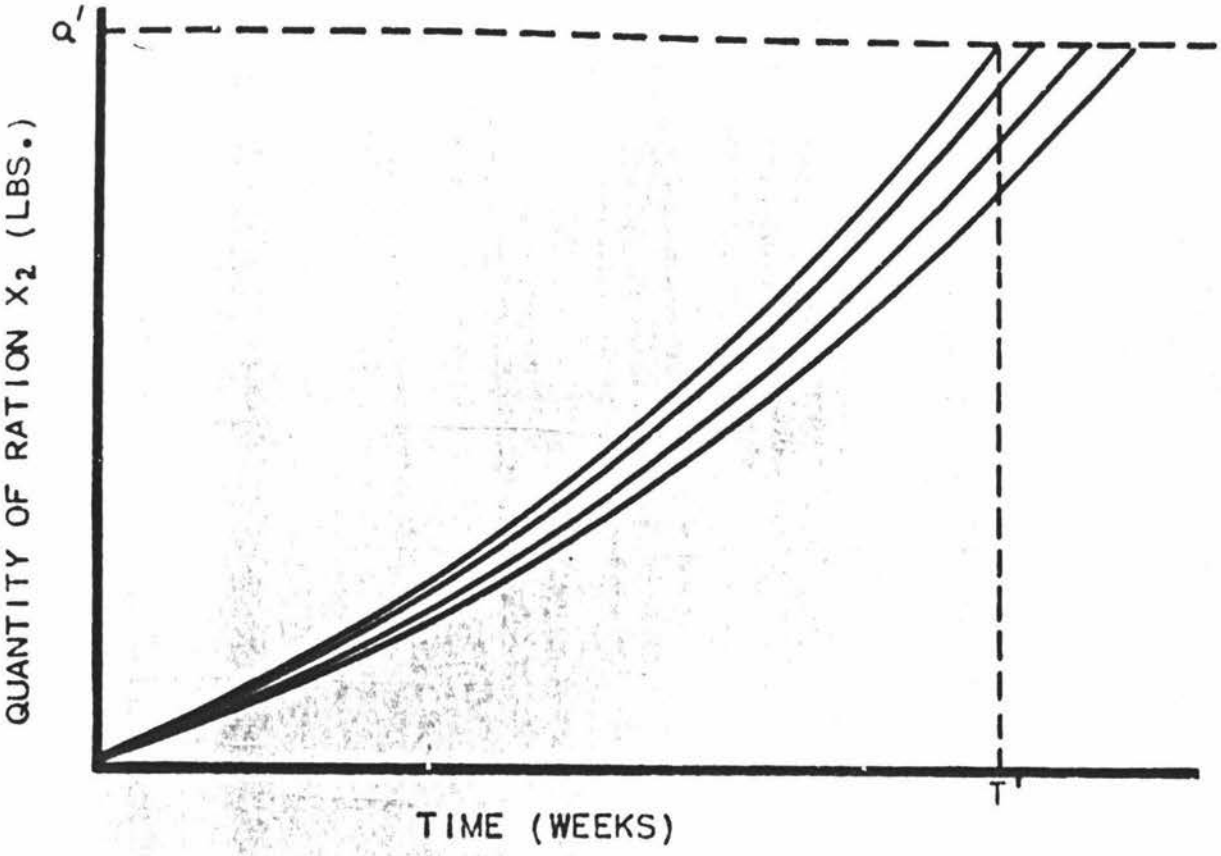


FIG.8 VARIATION IN AD LIBITUM RATES OF FEEDING - PIGS ON THE SAME RATION

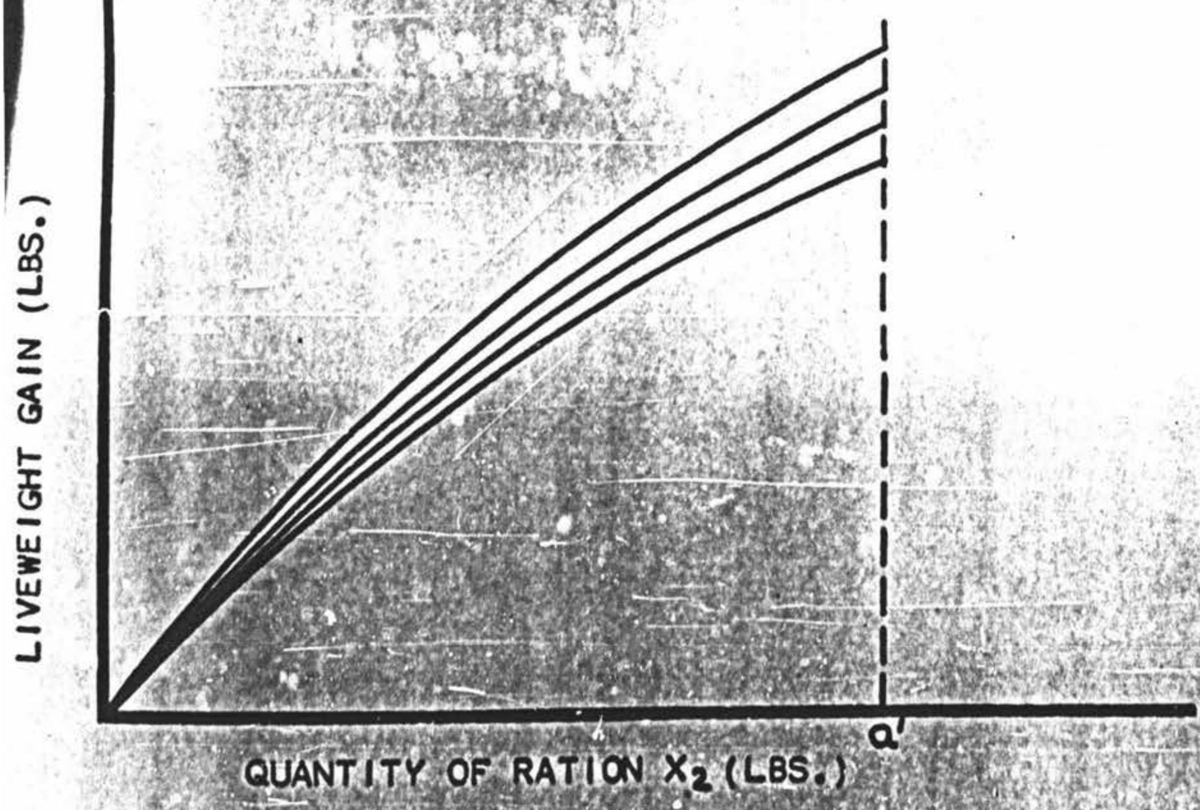


FIG.9 VARIATION IN LIVEWEIGHT GAIN - AD LIBITUM FEEDING RATION (X_2)

We may "explain" some of the variation in ad lib. feeding rates in terms of different genetic potential for growth or feed conversion efficiency between pigs. Thus, pigs having initially the same feeding rate may make different weight gains. Heavier animals may well have the capacity to consume greater quantities of feed, thus resulting in variation in feeding rates between pigs. If rate of feeding then affects rate of weight gain, these differences may well become accentuated under ad lib. feeding. However, although perhaps explainable to some degree, this variation under ad lib. feeding is uncontrollable.

Variations in liveweight gain made by pigs fed the same quantity of ration, fig.9, will be due to variations in feed conversion efficiency. Under ad lib. feeding it may reasonably be assumed that these variations (through the relationship: rate of feeding $\longrightarrow$ feed conversion efficiency) are uncontrollable and either of the two functions:

$$\hat{y} = f(x); \quad \hat{x} = f(y)$$

estimated. Where we wish to generalise the relationship between rations of the same type, the function:

$$\hat{y} = f(c,p)$$

has the same advantages as in estimating

$$\hat{T} = f(c,p).$$

Estimation of these two functions (as in the Iowa studies) allows us to plot in two dimensions, fig.10, liveweight gain and time

(9) However, in an article entitled: "A Method for Dealing with Time in Determining Optimum Factor Inputs", J.Farm Econ.,Vol.40,No.3,1958,p.666, Brown and Arscott state: "The rate of input of one factor should be expressed in terms of the input of the other factor and time, as in the following: $\hat{x}_1 = c_1T + c_2T^2 + c_3x_2$, where x_1 and x_2 are total feed inputs."
This article does not discuss the reasons for not estimating the functions : $\hat{x}_2 = f(T, x_1)$ or $\hat{T} = f(x_1, x_2)$.

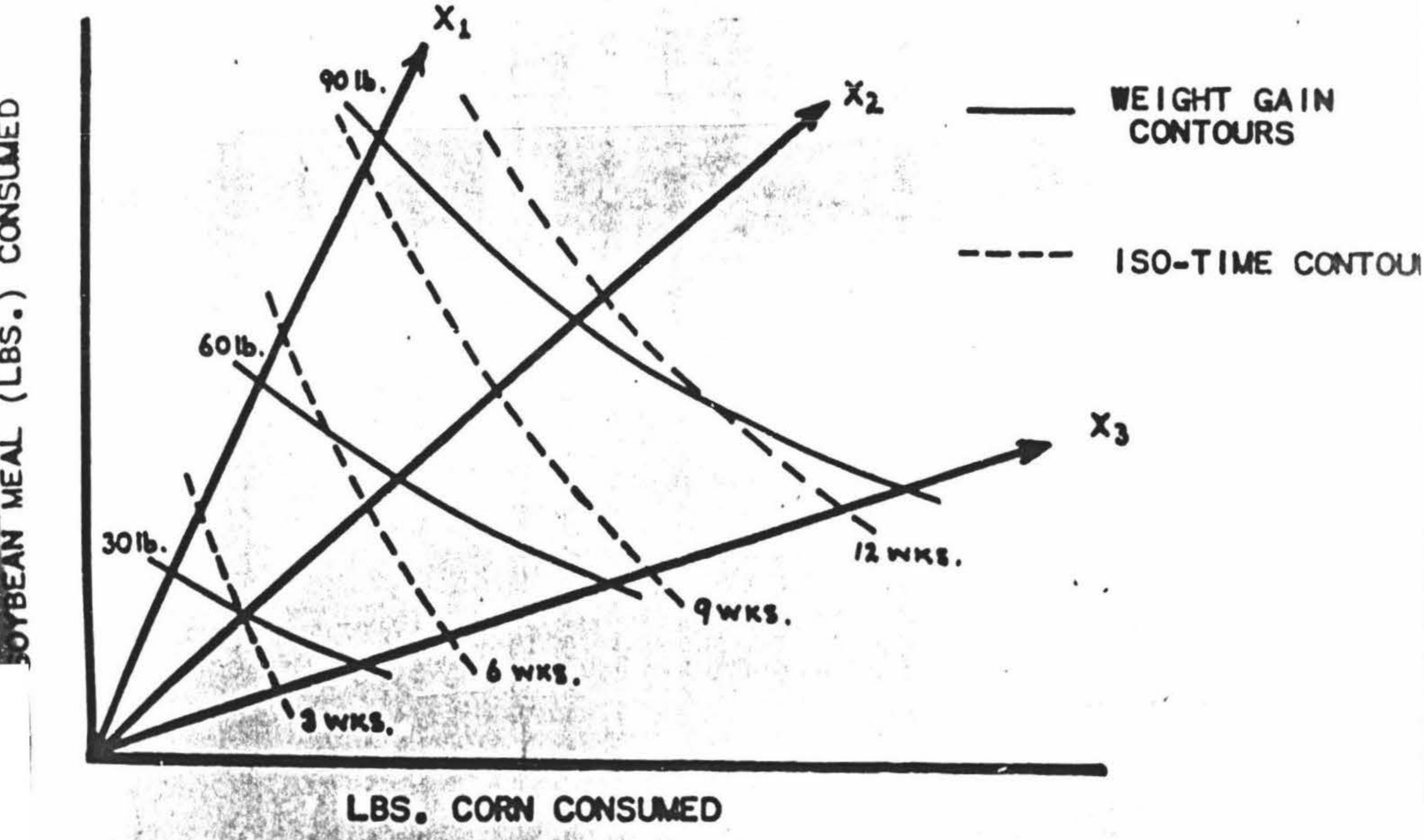
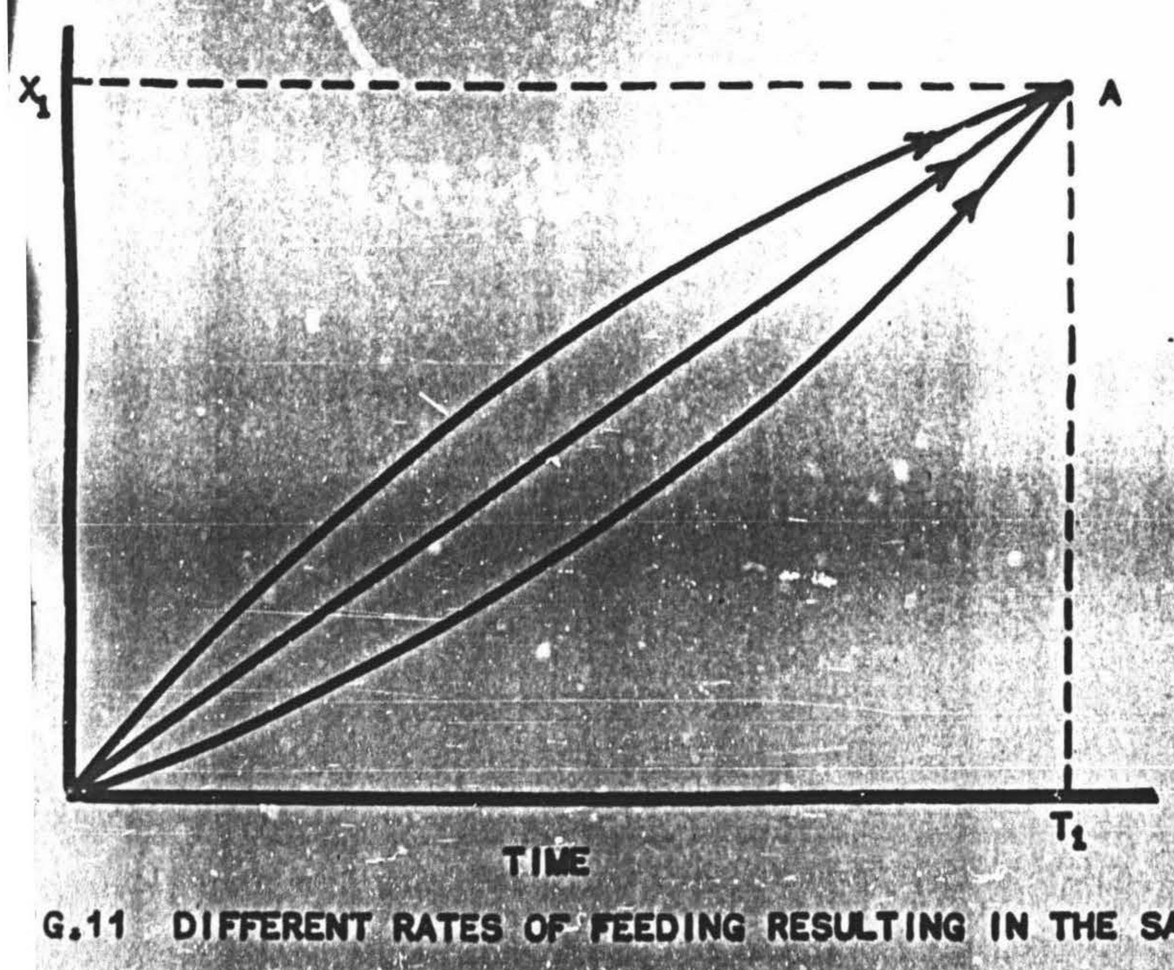


FIG.10 WEIGHT AND TIME CONTOURS FOR AD LIBITUM FEEDING



G.11 DIFFERENT RATES OF FEEDING RESULTING IN THE SAME TOTAL QUANTITY OF FEED (X_1) CONSUMED IN TIME T_1.

contours, for combinations of foodstuffs fed ad lib. Each corn and soybean meal combination within the region explored, gives us expected total weight gain and expected time to consume this feed combination. The variance of these estimates depends on the inherent variation in ad lib. consumption rates and feed conversion rates between pigs.

Because of the relationships outlined above, including the probable effect of liveweight on ad lib. feeding rates, we might expect estimates of weight gain and feed consumption time, to have increasing variances with increasing quantities of feed consumed. If this is the case then it is a characteristic of ad lib. feeding and under these conditions is uncontrollable - even if it is "explainable". Where this situation of heteroskedasticity exists, statistical methods may be used to place greater weight on the portion of the relationship that has lower variance. Estimation techniques which assume homoskedasticity are obviously suspect, and give "undue weight" to the portions of the production function with inherently large variances.

Therefore under conditions of ad lib. feeding, the estimation of the two functions:

$$\hat{y} = f(p,c)$$
$$\hat{T} = f(p,c)$$

would appear to provide a reasonable description of the production process.

2.6 Production Models for Restricted Rates of Feeding

The important difference between restricted and ad lib. feeding is that pigs, individually fed on the same ration, may be fed at exactly the same rate with restricted feeding. The same ration may also be fed at different controlled rates. We have already discussed one production model for different rations, each of which may be fed at different rates.

The situation for a single ration fed at different rates is illustrated in figs. 3 and 4. Referring to fig.3 we note that a given quantity of feed has been consumed in different time periods by the two pigs, and that weight gains have resulted according to the relationships shown in fig.4. A naive production model that suggests itself under these conditions is then:

$$\hat{y} = f(x,T)$$

where $\hat{y}$ is total weight gain over the period of time T and where x is the quantity of feed consumed in this period. Where rations are made up say of corn and soybean meal, this model may be generalised by writing:

$$\hat{y} = f(p,c,T).$$

At a particular feed:time combination however, past feeding history of the animal is recorded inadequately in terms of total quantities of corn and soybean meal fed and the time over which this has occurred. This inadequacy is illustrated with the aid of fig.11. Three pigs of the same initial liveweight would most likely make different liveweight gains as a result of reaching point A via the three feeding lines shown. According to nutritional relationships already described, these differences would be expected to occur from the way in which feed conversion efficiency is affected by liveweight and rate of feeding. However, the relationship:

$$\hat{y}_1 = f(x=x_1, T=T_1)$$

gives only one expected value for y_1, thus ignoring the possible effect of rates of feeding over this period.

Further, the point (x_1, T_1) in fig.11 allows us to calculate <u>only</u> the average rate of feeding: $\frac{x_1}{T_1}$. Over a reasonably long production period (say 10-15 weeks) differences in liveweight, occurring as a result of the way in which a given quantity of feed is consumed, might be considerable. Where we wish to describe the nature of the production process under feeding conditions where given feed:time combinations may be achieved by different rates of feeding within this time interval, the above function can have only limited meaning and should be replaced by the function:

$$\Delta y_i = f(\text{liveweight}, \Delta x_i)$$

as derived earlier.

Where given feed:time combinations are achieved from only one feeding line (i.e., no feeding lines intersect) the function:

$$y = f(x,T)$$

may be estimated. The usefulness of this function will depend on our ability to decide on a valid feeding line that will result in the predicted liveweight gain from a given feed:time combination. A basis that might prove reasonable for making such a decision would be to derive the hypothetical feeding line from some constant relationship between feeding lines from which the function was estimated. How reasonable this assumption might be could only be determined by testing the hypothesized ration and feeding rates.

Present knowlege of the nature of pigmeat production therefore favours the estimation of the relationships involved in the form:

$$\left.\begin{aligned} \Delta y_i &= f(\text{liveweight}, \Delta x_i) \\ \text{and} \qquad y_n &= \sum_{i=1}^{n} \Delta y_i \end{aligned}\right\} \qquad (2.10)$$

rather than: $y = f(x,T)$,

where we are interested in a number of rations, each of which may be fed at different rates.

We note that the estimation of pigmeat production relationships as given by (2.10) overcomes the "problem" of past rates of feeding by assuming that response to feed in a given subperiod depends on total liveweight at the beginning of this period, regardless of how this was attained. However, because of the undoubted importance of carryover effects from one rate of feeding to the next, this model may be very unsatisfactory if the Δ's refer to very short time periods. Also, estimating the relationships in the form given by (2.10) allows for:

$$y_n^* \neq y_n'$$

even though

$$\sum_{i=1}^{n} \Delta x_i^* = \sum_{i=1}^{n} \Delta x_i'$$

Statistical and production economics principles and estimational procedures may modify these conclusions based on present knowledge of nutritional relationships. Discussion has also been limited to the representation of feed variables as affecting meat production.

2.7 Other Factors Affecting the Pigmeat Production Process

Other factors affecting total weight gain may include: disease, breeding, environmental temperature, palatability of feedstuffs used, etc. It is often possible to make allowance for the effect of these factors on the production process in regressional analysis. However, where we cannot measure the level of occurrence of these factors, they often cannot be satisfactorily allowed for in regression analysis, and are therefore commonly considered as random effects that contribute to our inability to completely describe the production process. Thus, we may diagnose the presence of a particular disease, but to measure "how much" disease is present is usually out of the question. One symptom of nutritional upset in pigs is the occurrence of scouring. We may record for example the "number of days per week on which scouring occurred" for each pig, but the inadequacy of such measurements in reflecting digestive disorders are many. What is meant by "scouring" is hard to define in precise terms, the measure is subjective and may not therefore be repeatable. On the other hand, scouring may be closely associated with say efficiency of feed conversion, to the extent that even a crude measurement as suggested provides significant adjustment to production gains when used in regression analysis.

2.8 Characteristics of a Variable to be Measured

Increases in animal production may be measured on a liveweight or a carcass weight basis. Financial returns are most commonly based on carcass weight, hence in economic analysis it would be advantageous to express "efficiency of gain" or "rate of gain" on a carcass weight basis. In order to describe the production process, however, it is necessary to obtain these measures at regular intervals over the weight range of interest.

Under these conditions the expense of obtaining carcass weights becomes prohibitive. An equation expressing carcass weight as a function of liveweight has been derived from Ruakura slaughter data.[10] In an economic analysis, therefore, relationships expressed in terms of liveweight could easily be converted to a carcass weight basis. Alternatively the function could be used to express financial returns on a liveweight basis.

Commonly "type of feeds" used are considered as input variables, i.e., corn and soybean meal, barley meal and whey. When these combinations are common practice in feeding, the use of these as variables in the production function are justified. Where a greater degree of flexibility exists in choosing among feedstuffs to make up the ration, it might be better to use basic foodstuff constituents such as protein and carbohydrate as input variables in the production function. The production function so derived may then be used to calculate feed requirements for say a given liveweight gain, and linear programming used in turn to obtain a least cost feed mix that fulfils these requirements. The advantages of this approach and an example pertaining to broiler production are given in an article by Brown and Arscott.[11]

Where whey is used as a foodstuff for pigmeat production in New Zealand, it is commonly supplemented with barley meal, and there is little scope for the selection of least cost feed mixes.

---

(10) "Relationship Between Liveweight and Carcass weight Increments of Pigs", D.M. Smith, N.Z.J.Sc.and Tech., sec.A,Vol.38,No.8,Aug.1957,p.803.

(11) "Animal Production Functions and Optimum Ration Specifications", W.G.Brown and G.H. Arscott, J.Farm Econ.,Vol.42,No.1,1960,p.69.

gallons of whey and pounds of meal are intelligible units of measurement to farmers, whereas pounds of carbohydrate and protein require greater interpretation before becoming of practical use. In using carbohydrate and protein as variables we are also assuming that these are the properties of the foodstuff concerned that affect production. The seriousness of errors introduced in this assumption could be tested in practice by comparing models with "type of feed" variables on one hand, and "carbohydrate/protein" type variables on the other.

Similar comparison should allow us to decide whether or not increased accuracy of prediction of liveweight gains is obtained through measuring whey consumed in terms of "pounds dry matter" or "gallons of liquid".

This discussion illustrates that it is by no means obvious how we should measure the amount of whey and meal fed, i.e. gallons, or D.M., or T.D.N., or S.F., or P.E. All these measures may be highly correlated, but where there is a lack of correlation it might be better to replace a single measure of whey by the variables concerned, say: P.E.(whey) and S.E.(whey).

2.9 Mathematical Form of the Production Function

The nature of the pigmeat production process has allowed us to conceptualize forms of the production function for different feeding situations. The mathematical form of the function should allow for as many as possible of the biological features of the process that could exist.

An important factor that might affect the mathematical form chosen is the stages of growth we wish to describe. Pork production studies at Iowa State University[12] indicate that the marginal rate of substitution of high protein for high carbohydrate feeds is highest when the animal is young, and declines as the growing stage merges into the fattening stage. Nutrition studies such as those of Smith at Ruakura show that feed conversion efficiency usually declines as liveweight increases due to maintenance requirements and a direct liveweight effect.

The mathematical form of production models estimating total weight gain from weaning (approx. 40lb liveweight) as a function of feed input variables should therefore allow diminishing marginal productivity. The model should also allow the marginal rate of substitution between foodstuffs used to change with level of output.

In estimating the "incremental model", (for ration x_2 in fig.7):

$$\Delta y = f(\text{liveweight}, \Delta x_2)$$

we are assuming the existence of a family of curves as in fig.12.(13) There is no <u>a priori</u> reason for assuming a particular form of this relationship. Under these circumstances the mathematical model used should allow us to test whether or not the data concerned exhibits interaction between liveweight and shape of the response curve. (Fig.12 exhibits such interaction.)

In fig.12 ABC represents the production function:

$$\Delta y = f(\Delta x_2 \mid \text{liveweight} = 150\text{lbs.})$$

and where:

$\Delta x_2 = A$ is the number of units per week of ration x_2 necessary for maintenance of a 150lb. pig;

---

(13) For the production model: $\Delta y = f(\text{liveweight}, \Delta p, \Delta c)$ we assume the existence of a family of surfaces.

$\Delta x_2 = C$ is the maximum amount of ration x_2 that can be consumed per week by a 150lb. pig. We also note that at D and E, ad lib. feeding is hypothesized as resulting in irrational production (a greater weight gain could be achieved by a smaller input of ration x_2).

Actual mathematical forms that fulfil these requirements will not be discussed at this stage. The alternative forms that suggest themselves and have been used in the derivation of Total Product functions are adequately discussed elsewhere.(14)

---

(14) "Organisation Activities and Criteria in Obtaining and Fitting Technical Production Functions", E.O. Heady, J.Farm Econ.Vol.39,No.2, 1957,p.360. Chapter 3: Forms of Production Functions; in Agricultural Production Functions, by E.O. Heady and J.L. Dillon, Iowa State University Press, Ames,Iowa,1961.

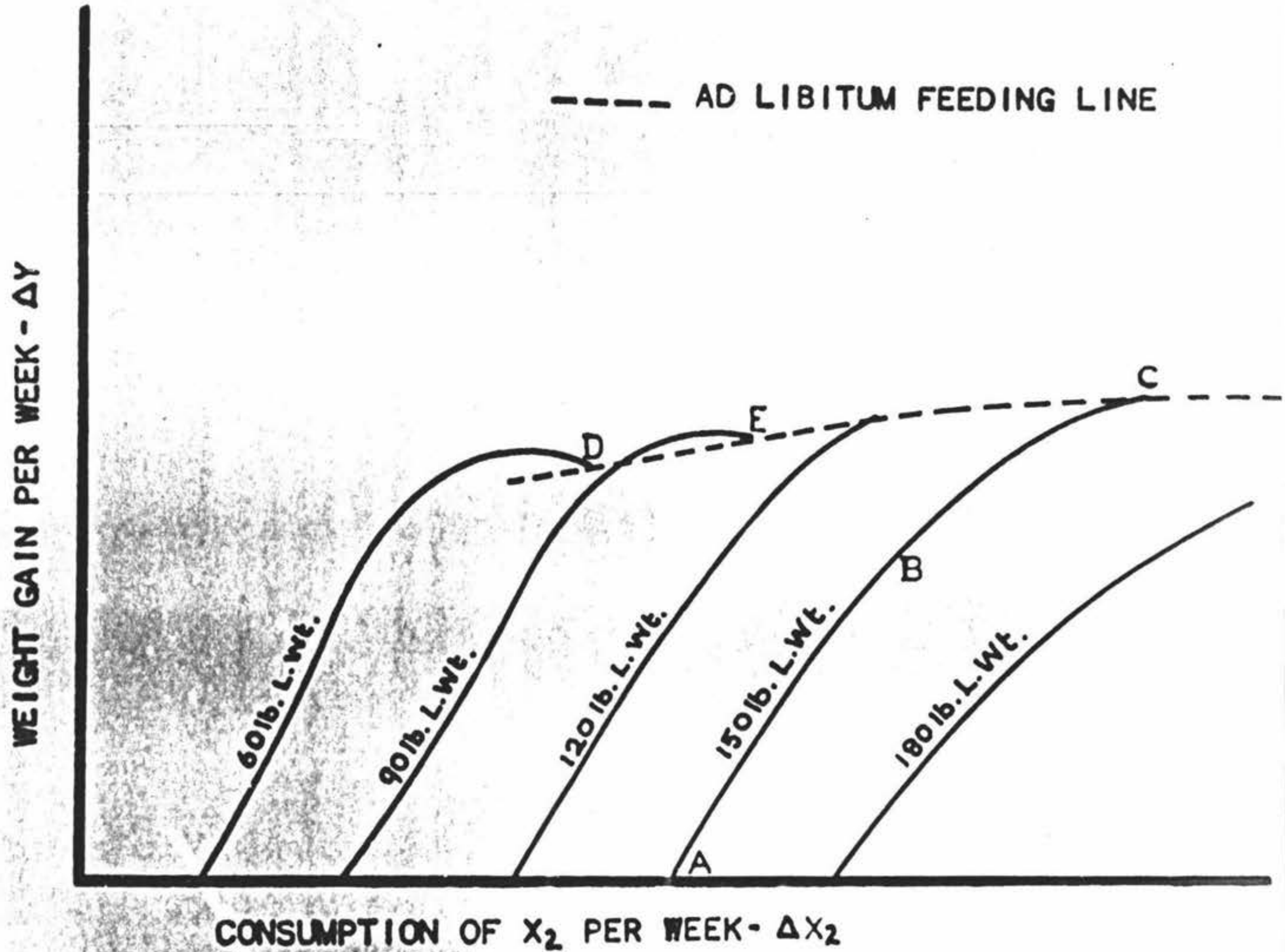


FIG.12 DIAGRAMATIC REPRESENTATION OF INCREMENTAL MODEL

CHAPTER III

PROBLEMS IN THE ESTIMATION OF PRODUCTION FUNCTIONS

The problems of estimating production functions will be discussed in the context of a pigmeat production function derived with respect to consumption of meal and whey.

To be of use in the management process, the estimated pigmeat production function should give accurate prediction of liveweight gain from given feed consumption. Where production economics principles are employed to select optimum rations, feeding rates, and marketing weights, marginal rates of substitution between feedstuffs must be known, and this necessitates the estimation of structural parameters.

Conceptual models based on knowledge of the biological nature of pigmeat production were derived for different feeding conditions in the preceding chapter. We are concerned with situations where more than one ration made up of two feedstuffs is employed. Estimates of the production function are based on liveweight and feed consumption data from a number of pigs. Successive observations on each pig are made at regular intervals over the weight range of interest. This data allows us to specify any of the single equation models developed in the previous chapter.

3.1 Multiple Regression

The discussion in this chapter is concerned with a statement made by Tintner[1]:

"The method of multiple regression is designed for the purpose of predicting Y (the dependent variable) from given levels of $X_1, X_2, \ldots, X_k$ (the independent variables). It is not, <u>in general</u>, appropriate for the estimation of the values of the constant b_0 and the regression coefficients $b_1, b_2, \ldots, b_k$ as they exist in the hypothetical infinite population corresponding to the sample."

In multiple regression we assume an underlying population relationship where the expected value of the dependent variable is given as a linear function of the fixed independent variables:

$$E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} \qquad (3.1)$$

The constants $\beta_0, \beta_1, \ldots, \beta_k$ are the population regression coefficients, sometimes called structural coefficients. $E(Y_i)$ is the expected value of the dependent variable for the ith. set of values of the independent variables. The observed value Y_i is a random real variable. Because we do not observe $E(Y_i)$ we say Y_i is subject to observational error ϵ_i given by:

$$Y_i - E(Y_i) = \epsilon_i$$

The observed values of the independent variables are not random variables and therefore not subject to observational errors. We may therefore rewrite (3.1) as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \epsilon_i \qquad (3.2)$$

where ϵ_i is a random real variable, commonly called the "error term" associated with the observed value Y_i.

We wish to estimate this relationship from n sets of

---

(1) "Econometrics", G. Tintner, J. Wiley and Sons, Inc. N.Y. 1952, p.85.

observed data, the sample, (where $n > k+1$) by:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki}$$

There are special situations in which the well known method of Least Squares yields best and unbiased prediction of $E(Y_i)$ and also best unbiased linear estimates of the regression coefficients. By unbiased we mean the expected values of the estimated coefficients are equal to the population values. By best linear estimates we mean those estimates that have the smallest variance among all linear unbiased estimates.

The assumptions on which the algebraic proof of the above properties of Least Squares estimates depends are:

(1) ϵ_i for every i has zero expected value, $E(\epsilon_i) = 0$

(2) The error associated with any one value of Y is independent of that associated with any other value of Y, $E(\epsilon_i \epsilon_j) = 0$, $i \neq j$.

(3) All independent variables are fixed and are measured without error.

(4) The variance of ϵ_i is constant for all i, $E(\epsilon_i^2) = \sigma^2$.

3.2 Evaluation of Regression Estimated by Least Squares

Under the above assumptions, Least Squares residuals, $Y_i - \hat{Y}_i = e_i$, provide unbiased estimates of the error term. Having estimated a relationship by Least Squares we may therefore check that the residuals conform to the assumptions made about the error term. If these assumptions do not apply to the residuals we cannot say the method of Least Squares has given us best linear unbiased estimates.

The method of Least Squares ensures that the first assumption about the error term is exhibited by the residuals as $\sum_i e_i = 0$.

Where assumption (4) does not hold we do not have best estimates, but the property of unbiasedness is not destroyed.

In order that estimates are unbiased the residual associated with any one value of Y must be independent of that associated with any other value of Y. (Where this assumption does not hold true the errors are said to be Autocorrelated.) Also the independent variables must be fixed and measured without error.

Thus, where assumptions (1) to (3) hold least squares estimates are unbiased, i.e. $E(\hat{Y}_i) = E(Y_i)$ and $E(b_j) = \beta_j$. Where in addition assumption (4) holds the estimates are minimum variance estimates.

Where the dependent variable is measured with error (u_i) in the sample, we check our assumptions against the net error term ($\epsilon_i + u_i$). If the assumptions hold true we have best linear unbiased estimates still. In this case the net error term is characteristic of the sample and population that we can measure or observe.

Under certain assumptions then, the method of Least Squares multiple regression applied to a sample of observations from the population, gives unbiased estimates of population regression coefficients and in addition these estimates have minimum variance of all estimates that could be derived from the particular sample. Where these properties of the estimates are considered to be a reasonable criterion for success, and where the assumptions discussed hold in the population and in the sample, the method of Least Squares may be described as a "reasonable" estimating procedure. However, other criterion of success, that may be thought equally reasonable, do exist. Thus the criterion that regression coefficients (b_0, b_1, ..., b_k) be selected so as to maximise the probability density of the particular sample of dependent variables that has occurred, leads to the method of Maximum Likelihood. In this instance the "reasonable" estimating procedure is that of Maximum Likelihood.[2]

The correct choice of estimating procedure is therefore largely dependent on what are considered desirable properties of the estimates.

In addition to knowing that estimates are unbiased, of minimum variance, etc., we may wish to obtain a measure of how closely the predicted series coincides with the observed values of the dependent variable.

The usual measure of "predictive success" is expressed as a ratio:

$$\frac{\text{Variance of observed values - Variance of Residuals}}{\text{Variance of observed values}}$$

commonly denoted as R^2, the coefficient of multiple determination. R^2 may also be obtained by calculating the empirical correlation between the observed and predicted values of the dependent variable. Least Squares minimises the variance of the residuals hence maximising R^2. On these grounds the magnitude of R^2 is often taken as an index of the predictive power of the estimated relationship. Where data includes replicated observations the method of Least Squares should be applied to the means of the

---

(2) In the special case where ϵ_i is $N(0,\sigma^2)$ the method of Maximum Likelihood leads to the method of Least Squares (Tintner, op.cit. p.87).

replicated observations, as these are estimates of $E(Y_i)$. Correspondingly R^2 should be evaluated in this case using these means as the "observed" values.

Implicit in the discussion so far has been the assumption of perfect knowledge about the variables among which the population relationship exists. Where this relationship may be represented by a single linear equation (3.1) and where the assumptions given hold true, the method of Least Squares gives unbiased and minimum variance estimates of the population regression coefficients and expected values of the dependent variable.

In actual practice we usually have incomplete knowledge about the variables which exert an influence on the dependent variable and the form which the underlying population relationship takes. Where we wish to estimate a relationship by Least Squares multiple regression we must think therefore about the following things:

(1) What are the variables amongst which we think the relationship of interest exists.

(2) Can the causal relationships be expressed as a single equation where the parameters to be estimated enter in a linear fashion.

(3) What mathematical form does this relationship take.

The dependent variable may be expressed in terms of the independent variables in almost an infinite variety of forms: i.e. a first, second, third, etc. order polynomial, a multiplicative relation that may be estimated linearly in terms of logs, a square root function, etc.

(4) Are the independent variables observed and measured without error.

If we obtain suitable answers to these questions Least Squares estimates may be derived. It is then necessary to think about the suitability of these estimates:

(1) Are they unbiased estimates of structural parameters?

(2) Does the estimated relationship have satisfactory predictive power?

The answers to these two questions will depend on:

(1) Characteristics of the data from which the relationship was estimated. Have we included all the variables which exert an influence on the dependent variable. And do in fact all the variables included exert an influence on the dependent variable.

(2) Have we chosen the correct mathematical form for estimating the relationship.

(3) The characteristics exhibited by the residuals from regression.

Unfortunately there is no real way of knowing whether or not all relevant variables have been included or whether the mathematical form does in fact coincide with the underlying "true" relationship. However, satistical tests of significance (based on the assumption that ϵ_i is $N(0,\sigma^2)$ for all i) and relative predictive success may be of use in deciding amongst theoretically acceptable variables and mathematical forms.

3.3 Statistical Tests of Significance - Variables

We assume for convenience that the relationship we wish to estimate may be represented by a polynomial equation. We are uncertain as to which of p independent variables should be included in the estimated relationship and whether or not a polynomial of first or second order adequately represents the form of the unknown function.

Where we have a first order regression equation in p variables we may test the significance of individual variables in two ways:(3)

(a) The reduction in Regression Sum of Squares that results from dropping the variable X_j from the regression equation represents the reduction in the sum of squares for "deviations from regression" that results when this variable is included in the model. Clearly this sum of squares depends on which specific variables remain in the model when X_j is dropped out, i.e. $X_1, X_2, \ldots, X_{j-1}$. This sum of squares is often called the sum of squares for X_j adjusted for $X_1, X_2, \ldots, X_{j-1}$. The resulting analysis of variance can be tabulated as follows:

Source	d.f.	S.S.	M.S.
Regression on X_1	1	S_1	S_1
Additional for X_2	1	S_2	S_2
. . .	.	.	.
. . .	.	.	.
. . .	.	.	.
. . .	.	.	.
Additional for X_p	1	S_p	S_p
Deviations	$n-(p+1)$	$(Y-\hat{Y})^2$	s^2
Total	$n-1$	$(Y-\bar{Y})^2$	

(3) "Simple and Multiple Regression Analysis", R.J. Hader and A.H.E. Grandage, p.109, in "Experimental Designs in Industry" Ed. by V.Chew. John Wiley and Sons, Inc. N.Y., 1958.

To test the significance of X_p adjusted for the other variables we use:

$$F = \frac{S_p}{s^2}$$

with 1 and n-(p+1) degrees of freedom.

(b) The significance testing procedure given above is limited in its usefulness by the particular way in which the variables were ordered in the regression estimation. If the variables X_1, X_2, X_3, etc. represent successive powers of a single variable X (i.e. the model is a polynomial in X), then a natural order is available.

However, the sum of squares for any variable X_j adjusted for all of the others is always available by taking $\frac{b_j^2}{c_{jj}}$.(4) On using this relationship to calculate the sums of squares of each variable we note that the sum of these does not in general add up to the Regression Sum of Squares. This is because in general non orthogonal data is used, i.e. the independent variables are correlated with one another.

Where these tests indicate that one coefficient is not significantly different from zero, it is not legitimate to replace the given estimate with a zero, for regardless of its magnitude, it is still the best estimate of the unknown coefficient. To replace this estimate by a zero would in effect

(4) "Statistical Methods", G.W. Snedecor, Iowa State College Press, Ames, Iowa. 1956. p.444. Note: This formula, $\frac{b_j^2}{c_{jj}}$, achieves the same as if we successively put each variable last in the ordering in the analysis of variance discussed in 3.3.(a).

be replacing an unbiased estimate by a biased one. However, if on the basis of these tests variables are dropped from the first order regression model, then the regression coefficients for the remaining variables must be recalculated.

Williams[5] has divided independent variables in regression analysis into "environmental" and "explanatory" variables. An explanatory variable is included as an estimator of the dependent variable, while an environmental variable is included as a correction to the dependent and other independent variables, so strengthening the relationship found between them. The present state of theory about the nature of the process we are interested in leads to the specification of explanatory variables. For these variables we are not so much interested in establishing the existence of the relation between them and the dependent variable as in estimating the regression coefficients. A particular sample of data may exhibit no significant relation between one of these explanatory variables and the dependent variable. This may be due to the particular sample of data rather than an invalid theory, and where possible the existence of the postulated relation would need to be checked further before the explanatory variable was dropped from the regression analysis.

On the other hand, we are only interested in the effects of environmental variables as they apply to the particular sample of data drawn. In this instant, nonsignificant variables would be dropped from the analysis.

---

(5) "Regression Analysis", E.J. Williams, p.118. John Wiley and Sons, Inc. N.Y. 1959.

3.4 Statistical Tests of Significance - Algebraic Form of the Relation

The algebraic form of the equation should allow theoretical characteristics of the function to be exhibited. We may compare forms on the basis of predictive power. Under conditions already noted the method of Least Squares gives best unbiased estimates of regression coefficients for a particular linear model. However, if the true underlying relation takes another form to the one estimated, the estimates will be biased due to an error of specification. We are unable to detect this bias however as we have no way of knowing the true underlying relation.

A polynomial equation of first or higher order is the most general algebraic equation form. We can obtain increasing predictive exactness, in the sense that the equation can be made to coincide with the mean of each value of dependent variable, by using polynomials of sequentially higher order. A polynomial therefore should only be replaced where a less general form, i.e. log model, is thought to possess special advantages. The characteristics of different algebraic forms used in estimating agricultural production functions has been adequately reviewed by Heady[6] and Heady and Dillon.[7] Where a polynomial equation is used it may be thought that production characteristics could best be represented by a second order model. Thus the fitted second order model for two independent variables would be:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_{11}X_1^2 + b_{22}X_2^2 + b_{12}X_1X_2$$

(6) "Organisation Activities and Criteria in Obtaining and Fitting Technical Production Functions", E.O.Heady, J.Farm Econ., Vol.39, No.2, 1957, p.360.

(7) "Agricultural Production Functions", Heady and Dillon,

It is possible however that for a particular sample of data a first order model would be representationally adequate. The test of significance in this instant is given in the following analysis of variance:

Source	d.f.
Mean	1
First order terms	2
Second order terms	3
Lack of fit (by subtraction)	n-k-6
Experimental Error	k
Total	n

The experimental error sums of squares is based on k replicated points. Where all experimental points are replicated the regression is fitted to the treatment means, lack of fit sums of squares is the failure of the regression to fit the means, and error sums of squares is obtained by remainder from the total sums of squares.

3.5 Properties of the Independent Variables

(i) Measurement errors: The most common source of such errors is lack of homogeneity in the data. Examples where we might expect this problem to occur in a pigmeat production model were given in the previous chapter, i.e. measurement of gallons of whey consumed, when dry matter content may differ from time to time. The measurement of liveweight almost certainly involves errors. Level of gutfill and whether or not an animal has recently excreted will affect the "liveweight" measured at any time.

Situations also exist, i.e. Demand and Supply relationships in Economics, where the independent variables are random variables and therefore subject to observational error.

In these situations Least Squares does not result in best linear unbiased estimates of the population regression coefficients. However, under the assumptions that there are no errors in the equation and that the errors associated with the independent variables are non-correlated, random and normally distributed, the method of weighted regression may be used.[8] The essential feature of weighted regression technique is that the variances and covariances of the observational errors are used as weights in estimating the regression coefficients. The variance and covariance estimates may be difficult to obtain, and the assumption of no errors in the equation is restrictive.

(ii) Autocorrelation in the Variables: Difficulties of analysis may arise from the fact that individual items of a time series are not independent - the variables concerned are said to be Autocorrelated. Several tests for the dependence between consecutive items in a series have been developed and are discussed in some detail by Tintner.[9]

Yule[10] has shown that the distribution of the correlation between two autocorrelated series tends to be U shaped with a majority of the correlations near ±1. This situation presents real difficulties in testing the validity of a relationship between two autocorrelated series.

---

(8) Tintner op.cit. p.121.

(9) Tintner op.cit.,p. 240.

(10) "Why do we Sometimes get Nonsense Correlations between Time-series?". G.U. Yule, Journal of the Royal Statistical Society, Vol. 89, 1926, p.1.

It is evident however that an autocorrelated series of a given length corresponds roughly to a pure random series of a shorter length. On the basis of an elaborate sampling study, Orcutt and James[11] have devised a method for testing the correlation between two autocorrelated series. The variance of the empirical correlation coefficient, V, is estimated, and where $V < 0.25$ the effective number of observations, n', is estimated from the formula: $V = \frac{1}{n' - 1}$. The number of effective observations n' is then used for entering the tables for tests of significance of the simple correlation coefficient for unrelated series.

(iii) Multicollinearity: Basically the problem is that so-called "independent" variables have moved together, so that it is impossible to tell which has been exerting a causative influence on the dependent variable, i.e. there exists more than one relationship between the variables being considered. This situation leads to unreliable estimates of the structural regression coefficients. If two independent variables are perfectly correlated then the value of their partial regression coefficients could be anywhere between $+\infty$ and $-\infty$, provided that an adjustment was made to the other coefficient.

It is of some importance to realise that multicollinearity does not reduce the predictive value of a regression equation, but where we are interested in functional relationships and hence structural coefficients, multicollinearity makes it impossible to obtain reliable estimates. The question: "at what level of correlation between two series of independent

---

(11) "Testing the Significance of Correlations between Time-series", G.H. Orcutt and S.F. James, Biometrika, Vol.35, 1948, p.397.

variables does multicollinearity become important?" appears to be largely a matter of general experience. It is felt that there can be few cases in which a correlation in excess of 0.8 between independent variables does not give unsatisfactory estimates. Where we decide multicollinearity is present in the data the question of which variable(s) to omit from the regression analysis depends on the nature of the model we are trying to estimate.

However, omitting one of two highly correlated variables only results in the effect of the omitted variable being "credited" to the remaining variable. Statistically, a more reliable estimate of joint effects is obtained by dropping one of two highly correlated variables, so long as they hold their old relationship.

3.6 Properties of Residuals

(i) Autocorrelation: One of the assumptions about the error on which the validity of the Least Squares method is based is:

"Successive errors are distributed independently of one another."

When this assumption is violated the Least Squares procedure breaks down at three points:[12]

(1) The estimates of the regression coefficients, though unbiased when the X's are fixed, need not have least variance.

---

(12) "Testing for Serial Autocorrelation in Least Squares Regression I", J. Durbin and G.S.Watson, Biometrika, Vol.37, 1950, p.409-28.

(2) The usual formula for the variance of an estimate is no longer applicable and is liable to give a serious underestimate of the true variance.

(3) The t and F distributions, used for making confidence statements, loose their validity.

Autocorrelation in the error term appears to be common in time series analyses. A considerable amount of research has been devoted to the problem of testing for the existence of correlation in the errors, but all too little on the more important problem of the best estimation procedure when correlations do exist. An excellent review of tests for the presence of autocorrelated errors, and estimating procedures that may be of use in this situation, and a detailed bibliography on the subject, has been given by Anderson.[13]

Cochrane and Orcutt[14] indicate three principle reasons that the errors in time series models tend to be positively autocorrelated:

(1) Faulty choice of the form of the regression model.

(2) Omission of important variables from the model.

(3) Use of incorrect variables or poor data.

Durbin and Watson[15] have presented upper and lower bounds on the significance levels of their "d-statistic"

---

(13) "The Problem of Autocorrelation in Regression Analysis", R.L. Anderson, Jnl.Amer.Stats.Assn., Vol.49, 1954, p.113.

(14) "A Sampling Study of the Merits of Autoregressive and Reduced Form Transformations in Regression Analysis" D.Cochrane and G.H. Orcutt, Jnl.Amer. Stats.Assn., Vol.44, 1949, p.356-72.

(15) "Testing for Serial Correlation in Least Squares Regression II", J.Durbin and G.S. Watson, Biometrika, Vol.38 (1951), p.159-78.

(based on a modification of the von Neumann ratio), which can be used for testing for the existence of autocorrelated errors in regression models, such as:

$$Y_i = \beta_0 + \sum_{j=1}^{r} \beta_j X_{ij} + \epsilon_i \quad i = 1, 2, \ldots, n.$$

where, the X's are fixed predictors, Y_i is the ith value of the dependent variable, and the ϵ's are normally distributed with equal variance. Where autocorrelation of the errors exists, the problem may be considered with respect to the likely reasons for its existence (given above).

An important point to note is that errors that are autocorrelated have arisen in a systematic way. To detect this, residuals must be tested in the order in which they arose. The methods for detecting autocorrelated errors referred to above apply when regression estimates are derived from a single time series. Successive growth and feed consumption observations on a single pig make up a time series. In estimating production relationships, growth data on a number of pigs fed different rations is commonly used. Systematic errors may arise in any one of these time series. The author has found no reference in the literature on dealing with problems of autocorrelation in regression analysis when the relationship is estimated from a number of concurrent time series as is the case in nutrition experiments. The existence of a problem is easily illustrated:

Consider the residuals from a regression fitted to time series data for a number of pigs <u>fed different rations</u>. The residuals $e_{1j}, e_{2j}, \ldots, e_{nj}$ for the jth pig, where e_{ij} is the ith such residual, may be autocorrelated, i.e., not independent, as may be the residuals $e_{1k}, e_{2k}, \ldots, e_{nk}$ for the kth pig. However, the two sets of residuals are almost certainly independent as they have arisen from different sources under different conditions - different pigs fed different rations. To test the series:

$$e_{1j}, e_{2j}, \ldots, e_{nj}, e_{1k}, e_{2k}, \ldots, e_{nk}$$

for autocorrelation is obviously meaningless. How then do we test for the presence of autocorrelated errors in a regression fitted to a number of time series? Is it necessary to test the residuals for each individual series? However, the regression is estimated from all the data and it is difficult to imagine the form adjustments could take if some series had autocorrelated errors while others did not!

(ii) Unequal Error Variance: One of the assumptions under which the Least Squares principle gives minimum variance estimates is that the errors have the same variance, i.e. $E(\epsilon_i^2) = \sigma^2$. If conditions of homogeneity of variance do not hold, each value of the dependent variable must be weighted inversely as its variance, to obtain efficient estimates of the regression coefficients. The unweighted regression coefficients will still be unbiased if the other assumptions hold true, but they will be less accurately determined than the weighted coefficients[16]. Unequal error variance may be detected by Bartlett's test[17]. A straightforward technique for handling

(16) Williams, E.J. op.cit. p.19.

(17) "Statistical Methods", G.W. Snedecor, Iowa State College Press, Ames, Iowa, 1956, p.285.

unequal error variances is given by Anderson and Bancroft.[18]

3.7 Summary

The major complications in Least Squares multiple regression analysis are due to

(1) The existence of simultaneous relationships between the variables.

(2) The presence of autocorrelated error terms.

(3) The presence of errors of observation in the independent variables.

These complications do not in general affect the predictive value of Least Squares estimates; at least where minimising the sum of squared residuals from regression is taken as the criterion for predictive success. For the purpose of estimating structural parameters it is necessary to find a method of dealing simultaneously with all three complications, or at least some indication of their relative importances.

We are now in a position to agree with Tintner[19]:

> "The method of multiple regression is designed for predicting Y from given levels of X_1, X_2, ..., X_k. It is not, in general, appropriate for the estimation of the values of the constant b_o and the regression coefficients b_1, b_2, ..., b_k as they exist in the hypothetical infinite population corresponding to the sample."

For the pigmeat production models already presented we can justify fairly simply the use of a single equation model, (see section 4.4). However, as already indicated, the detection of, and adjustments for, autocorrelated errors may not be easily

(18) "Statistical Theory in Research", R.L. Anderson and T.A.Bancroft, McGraw Hill, N.Y., 1952,pp.182-86.

(19) Tintner op.cit, p.85.

made; and estimation procedures when independent variables are subject to error require restrictive assumptions.

It does appear however that Least Squares estimates may often be the only estimates of structural coefficients that can be obtained at reasonable cost. They are often, therefore, used as such even though the dangers of so doing may be well known. Though mindful of these dangers, this study has of necessity been no exception to this situation.

CHAPTER IV

A MEAL:WHEY PRODUCTION FUNCTION FOR PIGS - ESTIMATION PROCEDURE

The rationale behind research into production relationships for pigmeat in New Zealand; conceptual production models based on knowledge of the biological process; and statistical estimation problems; have been considered in successive chapters. We now derive production functions for pigs fed rations of barley meal and whey. The present chapter considers general characteristics of the data used for this derivation.

4.1 Source of Data

Data was made available from a meal:whey feeding trial conducted by the Pig Husbandry Department at the Massey College Piggery. A full report of this trial which commenced in September 1960 is given in Appendix I.

Pigs from three litters were fed on nine rations, three pigs per ration, allowing three blocks, each consisting of nine litter mates. The quantity of each ration fed was based on liveweight according to the scales shown in the report. Pigs were placed on trial individually as they reached 48-50 lbs liveweight, and thereafter they were weighed twice weekly. However, the smallest regular interval at which observations on liveweight and feed consumption were available was weekly. No pig had any food for at least 10 hours prior to being weighed on Monday mornings, thus reducing to some extent errors of liveweight measurement, i.e., due to gutfill. Weekly data was based therefore on Monday's observations.

Pigs were fed individually and refusals and the occurrence of scouring on any day was noted. Actual food consumption rather than food offered was recorded.

Pigs remained on trial until at least 180 lbs liveweight. The data recorded for each pig is summarised in Appendix II, and may be opened out to be read with this and following chapters.

4.2 Characteristics of the Trial Data

Some important characteristics of the data generated by this trial are now noted:

(a) Restricted feeding: The rations used have been fed at restricted rates. The portion of the meal:whey plane covered by this trial is shown in fig.1 where each treatment is represented by a ration line. The way in which consumption occurred through time is illustrated in fig.2. The area labelled "9wks." for example, includes meal and whey totals consumed after nine weeks on trial, for each of the ration lines in fig.1. Because restricted feeding rates have been used these areas overlap to some extent.

We have already noted that different liveweight gains may be expected for pigs fed the same total quantity of meal and whey, but in different time periods. Where this characteristic of the data has been important in determining

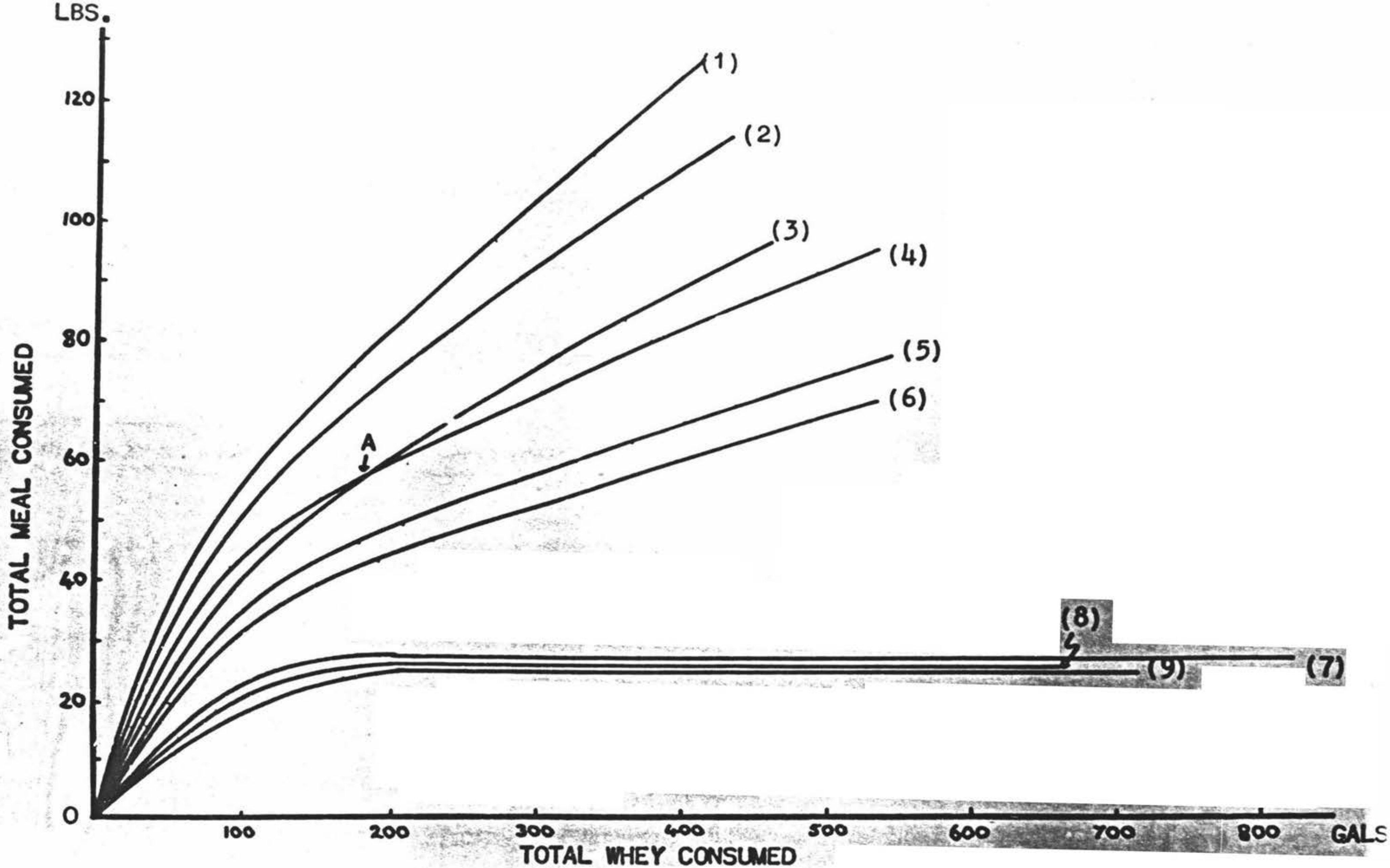


FIG.1 RATION LINES FOR MEAL-WHEY LEVELS TRIAL

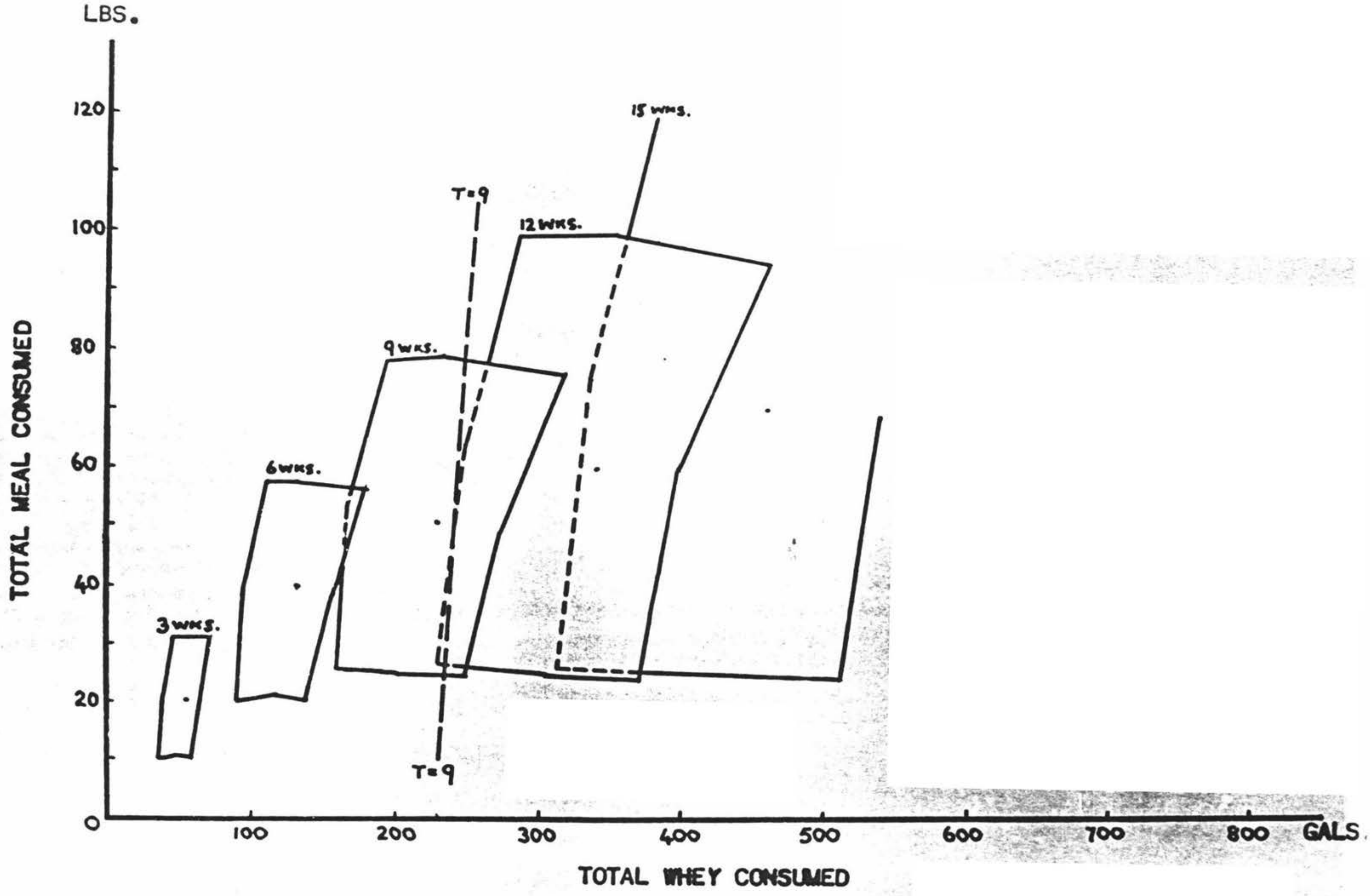


FIG. 2 RATES OF FEEDING FOR MEAL-WHEY LEVELS TRIAL

liveweight gains, the total liveweight gain and feed consumption model (or "cumulative" model), should be estimated in the form

$$Y = f(M, Wh, T).$$

Where this characteristic has not been important the model

$$Y = f(M, Wh)$$

may provide an adequate description of the data. The differences in feeding rates, shown in fig.2, make the function

$$T = f(M, Wh)$$

quite meaningless.

The figure shows, quite clearly, that differences in the <u>rate</u> of feeding, can result in the same total feed quantity being consumed in 9 or 12 weeks. The function $T = f(M, Wh)$, which makes no reference to rate of feeding, is simply inadequate.

Restricted feeding rates do not alter the rationale of the incremental model.

(b) Interdependence of growth rate and quantity of feed fed: This characteristic of the data is important when we consider the replication of treatments. True replication is where individuals have been treated exactly alike. Under conditions of restricted feeding we can consider animals that have moved out along a single ration line together to have been treated exactly alike; i.e., all animals on a given treatment always consume the same quantity of meal and whey in the same time period(1). In this meal:whey levels trial, whey was fed on a scale that changed continuously with the liveweight of each

(1) This is different from <u>ad lib</u>. feeding where individuals treated alike are those feeding <u>ad lib</u>. on a given ration.

animal. Thus animals on the "same treatment" may in fact have been fed at different rates due to the inherent variability in growth rate between pigs, and the feeding system adopted. Differences in rate of feeding may accentuate growth rate differences, so that after a given period of time, animals on the "same treatment" will have consumed different quantities of feed and have made different liveweight gains. Under these circumstances liveweight differences after any given time on trial cannot be attributed to the inherent variability of the experimental material alone, i.e., we have no real estimate of experimental error.

In feeding trials the condition of independence (in the statistical sense) between feeding rate and growth rate, necessary to obtain an estimate of experimental error, may be achieved by predetermining rates of feeding as a function of time.[2] Each ration could also be fed at rates given by more than one function of time (to different pigs, of course), in order to obtain a more efficient estimate of the result of feeding given quantities of meal and whey in different time periods.

(c) Experimental data: This pig production data was generated under experimental conditions. Different results may be expected under group feeding and actual farm conditions. These differences would be important in a critical appraisal of the usefulness of results obtained to management processes associated with pigmeat production. No such assessment has been attempted

(2) In actual fact a true estimate of experimental error could not be obtained where animals remain "on trial" continuously. Although the same quantity of feed can be fed in the same time to pigs under these conditions, it is impossible to guarantee that these pigs will be all the same liveweight at the beginning of each period.

in this thesis.

(d) Time series data: The data consists of successive observations made at weekly intervals, on each of twenty-seven pigs. The information on any one pig therefore constitutes a time series. The length of time series varied from 11 weeks for pig No.8 (on Treatment III) to 33 weeks for pig No.19 (on Treatment VII).

(e) Rejected data: Pig No.23 (on Treatment VIII) did not complete the trial due to paralysis of the hind legs and a spinal abscess. Data from this pig was not therefore used in estimating production functions. Pig No.9 (on Treatment III) and pig No.27 (on Treatment IX) had markedly worse growth rates than other animals on the same treatments. It was not felt however that this data was sufficiently unrepresentational of the variance in growth rates of pigs that could occur in practice to justify rejection. A total of 491 observations made at weekly intervals on 26 pigs was therefore available for estimatin meal:whey production functions. In passing we note that this gives a maximum of 491 and a minimum of 26, degrees of freedom.

4.3 Preliminary Analysis

In regression analysis the total sum of squares of deviations about the mean of the dependent variable may be subdivided into the reduction of this total due to fitting the regression line and the remainder, commonly called "deviations from regression".

Where more than one independent observation of the dependent variable is made at each level of the controlled factor(s), i.e., where treatments are replicated, an estimate of experimental error may be obtained via an analysis of variance. In addition, regression analysis may be applied to the mean of the observations for each treatment (level of independent variables), thus dividing the sum of squares due to treatments into treatment regression and deviations of treatment means from regression (commonly called "lack of fit" sum of squares). We have therefore:

TOTAL SUM OF SQUARES = TREATMENT S.S. + ERROR S.S.

TREATMENT S.S. = REGRESSION S.S. + LACK OF FIT S.S.

Where treatments are truly replicated the regression fitted to all observations yields:

TOTAL S.S. = REGRESSION S.S. + DEVIATIONS S.S.

and

DEVIATIONS S.S. = LACK OF FIT S.S. + ERROR S.S.

By fitting the regression line to the treatment means an estimate of lack of fit sum of squares is obtained, thus allowing us to subdivide the total deviations from regression as shown above.

Where lack of fit sum of squares differs significantly from the estimate of experimental error, we suspect that the fitted regression inadequately expresses the relationship between the variables concerned. This test is therefore useful in deciding questions about the suitability of different forms of regression equations and the inclusion of so-called explanatory variables.

Where observations are not replicated the data does not provide a true estimate of experimental error. Where no a priori estimate of experimental error variance is available deviations from regression cannot be subdivided as shown and use made of the test mentioned.

We suspect already that the data from the meal:whey levels trial was not perfectly replicated. To gain some idea as to whether or not this was so, the total weight gains and meal and whey consumptions after 11 weeks were recorded for each of the 27 pigs involved.(3)

An analysis of variance assuming replication of each treatment was first calculated for the liveweight gain data:

Table I : Analysis of Variance(4) (9 treatments, 3 blocks)

Source	d.f.	S.S.	M.S.	F
Blocks	2	183	91.50	2.04 NS
Treatments	8	6659	832.38	18.60 ***
Error	16	716	44.75	
Total	26	7558		

A simple linear regression model:

$$\hat{Y} = a + bM + cWh_{D.M.}$$

where M = lbs. meal, and Wh = lbs. D.M. whey, consumed in the 11 week period, was then fitted to the 27 observations. The following analysis resulted:

(3) Data for pig No.23 was used in this analysis. See Appendix IIIA for the figures used.

(4) The following convention will be used to denote levels of significance: Significance at 0.1% level = ***; at 1% level = **; Significance at 5% level = *; Significance at 10% level = 10%; Not Significant at 10% level = NS.

Table II : Analysis of Variance, Regression Model

Source	d.f.	S.S.	M.S.	F
Regression	2	7029	3514.50	159.45 ***
Deviations	24	529	22.04	
Total	26	7558		

The important feature of these two analyses of the same data is that regression sum of squares (7029) exceeds treatment sum of squares (6659). This situation is only possible where pigs within treatments have consumed differing quantities of meal and whey that have affected liveweight, i.e., we do not have true replication. We cannot therefore derive separate estimates of "lack of fit" and "experimental error" from this data. Our best estimate of experimental error is obtained from total deviations from regression.

In estimating production models from this experimental data regression analysis has therefore been applied to all observations. No test for lack of fit of regression models in the sense indicated above was available.

4.4 Pigmeat Production Models

The general characteristics of these models have already been discussed.

The cumulative models that suggest themselves under restricted feeding are:

$$\hat{Y} = f(M, Wh, T) \qquad (4.1)$$

or

$$\hat{Y} = f(M, Wh) \qquad (4.2)$$

These models describe the results of the production process under a given set of conditions rather than ellucidating the causal relationships involved. This was seen in Chapter 2 where we noted that knowing the total meal and whey to feed and the time in which to feed it, does not determine uniquely the expected liveweight gain.

The incremental model:[5]

$$\left.\begin{aligned} y_i &= f(Wt_{i-1}, m_i, wh_i) \\ \sum_{i=1}^{n} y_i &= Y'_n \end{aligned}\right\} \qquad (4.3)$$

on the other hand has been derived from knowledge of the biological nature of pigmeat production, and is an attempt to specify a structural relationship.

Because models (4.1) and (4.2) provide mathematical descriptions of the result of the production process, there is no problem of indentification. The particular variable chosen as the dependent variable will hinge upon the problem setting of interest. Although these two models may give satisfactory prediction, their ability to provide valid estimates of structural parameters is open to question. It is felt intuitively that so-called "structural relationships" should not be capable of explanation via any other underlying relationship.

The incremental model attempts to specify such an underlying relationship and could therefore provide valid estimates of structural coefficients. In this model, liveweight gain in the ith subperiod, y_i, is the dependent variable, and is affected by the independent variables. The first of these, total liveweight at the beginning of the ith period (end of the

---

(5) See Appendix II for notation.

i-1th period), Wt_{i-1}, is obviously independent of liveweight gain in this period. Feed consumption, m_i, wh_i, must also occur before liveweight gain results. The importance of this lag between consumption and production will determine to some extent the accuracy of this structural model. Given a suitable subperiod we have no problems of identification and the relationship can be estimated as a single equation. Incremental models based on weekly and fortnightly subperiods will be considered.

4.5 Comparison of Production Models

Perhaps the most important part of this study should be the comparison of the production models postulated. This comparison could be made under the headings: "Predictive Power of the Models", and "Validity of Regression Coefficients as Estimates of Structural Parameters".

As we wish to compare incremental and cumulative production models, it is appropriate to consider the relationship between the dependent variables of these two models, y_i and Y_i, respectively.

For a single pig we have:

$$Wt_o = Wt_o^* + \epsilon_o$$

where Wt_o is the observed liveweight at the beginning of the trial period, Wt_o^* is the expected value (including expected level of gutfill) of liveweight at this time, and ϵ_o is an error term associated with Wt_o^*. Similarly, for observed liveweight at the end of the ith subperiod we have:

$$Wt_i = Wt_i^* + \epsilon_i.$$

Now observed liveweight gain from the start of the trial to the end of the ith subperiod, Y_i, is given by:

$$Y_i = Wt_i - Wt_o = (Wt_i^* - Wt_o^*) + (\epsilon_i - \epsilon_o)$$

or

$$Y_i = Y_i^* + (\epsilon_i - \epsilon_o)$$

where Y_i^* is the expected value of liveweight gain. Similarly for liveweight gain from the start of the trial to the end of the i-1th subperiod we may write:

$$Y_{i-1} = Y_{i-1}^* + (\epsilon_{i-1} - \epsilon_o).$$

Observed liveweight gain in the ith subperiod, y_i, is given by:

$$y_i = Y_i - Y_{i-1} = (Y_i^* - Y_{i-1}^*) + (\epsilon_i - \epsilon_{i-1})$$

$$= y_i^* + (\epsilon_i - \epsilon_{i-1})$$

In making a comparison of the cumulative and incremental models we are especially interested in the properties of the errors associated with the respective dependent variables. The importance of assumptions made about the error term have been discussed in sections 3.1 and 3.6.

In pigmeat production, an animal may have a growth rate that is inherently better or worse than average. This animal will reach liveweights that are <u>increasingly</u> better or worse than expected liveweights of pigs fed on the same ration. In this case errors will not be independently distributed, i.e. $E(\epsilon_i \epsilon_j) \neq 0$, $i \neq j$, but will be positively autocorrelated through time.

It is possible that positively autocorrelated errors follow some autoregressive scheme, i.e. ϵ_i may be generated by the Markoff scheme:

$$\epsilon_i = \rho \epsilon_{i-1} + \eta_i \qquad (4.4)$$

where,
$$\left.\begin{aligned} E(\eta_i) &= 0 \\ E(\eta_i^2) &= \sigma^2 \\ E(\eta_i \eta_j) &= 0, \quad i \neq j \end{aligned}\right\} \qquad (4.5)$$

Regression analysis where the error term has autoregressive properties has been discussed by Cochrane and Orcutt.[6] Let us now consider properties of the error term associated with the cumulative and incremental series when the errors, ϵ_i, are generated by (4.4).

(a) **Cumulative series.** Let us consider for example:

$$Y_3 = Y_3^* + (\epsilon_3 - \epsilon_o)$$

Now from (4.4) above, and when $\rho = 1$, we have:

$$\begin{aligned} Y_3 &= Y_3^* + (\epsilon_2 + \eta_3 - \epsilon_o) \\ &= Y_3^* + (\epsilon_1 + \eta_2 + \eta_3 - \epsilon_o) \\ &= Y_3^* + (\epsilon_o + \eta_1 + \eta_2 + \eta_3 - \epsilon_o) \\ &= Y_3^* + \sum_{i=1}^{3} \eta_i \end{aligned}$$

or, in general:

$$Y_n = Y_n^* + \sum_{i=1}^{n} \eta_i$$

or
$$Y_n = Y_n^* + \lambda_n$$

where
$$\lambda_n = \sum_{i=1}^{n} \eta_i$$

Where the properties of η_i are given by (4.5), we have:

$$\begin{aligned} E(\lambda_i) &= 0 \\ E(\lambda_i^2) &= i\sigma^2 \\ E(\lambda_i \lambda_j) &= j\sigma^2 \quad , i \neq j,\ i > j. \end{aligned}$$

(6) Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," D.Cochrane and G.H.Orcutt, Jnl.Amer.Stats.Assn.Vol.44, 1949, p.32.

In this case errors associated with the cumulative series are positively autocorrelated, and are heteroskedastic.

(b) Incremental series. Consider:

$$y_i = y_i^* + (\epsilon_i - \epsilon_{i-1})$$

Substituting for ϵ_i from (4.4) we have:

$$y_i = y_i^* + (\rho-1)\epsilon_{i-1} + \eta_i$$

Thus, where $\rho = 1$, the errors associated with the incremental series are random and independently distributed with equal variance, as given by (4.5).

Where $\rho \neq 1$, and $\rho \neq 0$, properties of the error terms associated with the incremental and cumulative series are less easy to determine.

An iterative method for obtaining autoregressive - least - squares estimates when the value of ρ is not known is given by Fuller and Martin.(7)

Let us assume now that errors associated with liveweight of a single pig are random and independently distributed, and that:

$$E(\epsilon_{i-1}) = E(\epsilon_i) = E(\epsilon_{i+1}), \text{ etc.} = 0$$

Consider the incremental series, the general term of which is written:

$$y_i = y_i^* + (\epsilon_i - \epsilon_{i-1}).$$

Where ϵ_i is higher than average (i.e. could be due to gutfill), we expect the value y_i to be higher than average. Because the errors are assumed to be random, with expectations all zero,

(7) "The Effects of Autocorrelated Errors in the Statistical Estimation of Distributed Lag Models," W.A.Fuller and J.E.Martin, J.Farm Econ.Vol.43,Feb.1961,p.71-82.

we also expect a lower than average value for liveweight gain in the i+1th period, y_{i+1}, where:

$$y_{i+1} = y_{i+1} + (\epsilon_{i+1} - \epsilon_i).$$

Similarly where ϵ_i is lower than expectation we would expect a higher than average value for weight gain in the subsequent period. Under these circumstances we could obtain negative autocorrelation of the errors associated with the incremental series. In this case efforts should be made to minimise liveweight measurement errors.

This theoretical discussion forms a necessary basis to questions about the validity of regression coefficients, from the cumulative and incremental production models, as estimates of structural parameters.

Another factor affecting the question of validity of regression coefficients as estimates of structural parameters is errors in the independent variables. We will assume that measurement errors associated with feed consumption were negligible under these trial conditions. In the incremental model, however, we cannot assume that the independent variable: "total liveweight at the end of the previous period", is observed without error. Regression coefficients estimated by Least Squares and associated with this variable are therefore likely to be biased estimates of structural parameters.

4.6 Scouring

Where scouring affects feed conversion efficiency the relationship between liveweight gain and feed consumption may be strengthened by allowing for this factor. An analysis of variance (Table III) indicates that significant treatment differences occur for the number of days on which scouring was noted over the whole period that pigs were on trial.[8]

Table III : Analysis of Variance. Treatment Differences : Number of Days Scouring Occurred over approx. Liveweight Range: 48-180 lb.

Source	d.f.	S.S.	M.S.	F
Blocks	2	2234	1107	2.23 NS
Treatments	8	23278-153=23125	2890	5.81 **
Error	16-1=15	7458	497	
Total	25	32970		

In this case any adjustment for scouring in the production model is really a means of eliminating from the dependent variable the part of the treatment effect that is attributable to the effect of the treatments on scouring.[9] A significant adjustment for scouring in the production model must therefore be interpreted with care, i.e. for predictive purposes it would be necessary to explore further the relationship between treatments and scouring. The difficulty is that propensity to scour may be dependent on the ration and therefore knowing what the liveweight gains would be if no scouring occurred is

(8) See Appendix IIIB for this data. The technique for dealing with missing data in randomised block designs - see: "Statistical Methods" G.W.Snedecor, Iowa State College Press, Ames,Iowa,1957,p.310; has been usedto derive an estimate for pig No.23. The analysis of variance presented in Table III has been adjusted accordingly.

(9) Williams, E.J. op.cit. Ch.7.

not really relevant. If scouring affected liveweight gain significantly it would be necessary to be able to predict the level of scouring, i.e., as some function of level of feeding, before a production model that included a term for scouring could become operational.

4.7 Degrees of Freedom

This is the problem of the effective number of independent observations we have on any one pig as described in section 3.5. It arises because we have a number of observations on each animal. The minimum number of degrees of freedom is 26 as this is the number of pigs involved in the trial, while the maximum is 491 - the total number of observations. On this basis we define a "strict" and a "lenient" criterion of significance based on 26 and 491 degrees of freedom respectively.

The series: total liveweight gain, total meal consumed, total whey consumed, and time, are highly autocorrelated for each pig. Calculation of the effective degrees of freedom to use for testing the significance of relationships between these series (10) indicates the seriousness of the problem. A typical example may be given. From the 16 weekly observations on total weight gain and total whey consumption for pig No.16 we obtain:

Autocorrelation coefficient of lag 1 for the weight gain series : $r_1 = 0.99903$

Autocorrelation coefficient of lag 1 for the whey consumed series : $r_1' = 0.99966$

---

(10) G. Tintner *op.cit.*, pp.243, 248.

The variance of the empirical correlation coefficient between these two series : V = 0.98881

The effective number of observations is : $n' \doteq 2$ [11]

Similar results are obtained from the series: liveweight gain and total meal consumed, and liveweight gain and time.

For the cumulative model therefore it is suspected that more weight should be placed on the strict test of significance than on the lenient criterion.

The position for the incremental model is illustrated for pig No.16 and the two series : liveweight gain in each week, and total liveweight at the beginning of the week.

Autocorrelation coefficient of lag 1 for the liveweight gain series : $r_1 = -0.22129$

Autocorrelation coefficient of lag 1 for the total liveweight series : $r_1' = 0.98472$

The variance of the empirical correlation coefficient between these two series : V = 0.04209

The effective number of observations is $n' \doteq 25$ (which is greater than the number of observations N = 16).

For the weekly incremental model therefore, it is suspected that it would be satisfactory to base tests of significance on the lenient criterion.

Unfortunately an exact test of the effective degrees of freedom where a number of time series is used in regression analysis is unknown to the author.

---

(11) The correlation between two series of two observations is necessarily ± 1. This supports Yule's finding that the distribution of the correlation between two autocorrelated series tends to be U shaped with a majority of the correlations near ± 1.

4.8 Estimation Procedure

Least Squares regression analysis in this study has been greatly facilitated through co-operation with the Applied Mathematics Laboratory, Department of Scientific and Industrial Research, Wellington, in the use of Treasury's augmented Type IBM 650 computer. A generalised computer programme for the fitting of data to mathematical equations was available and is referred to as TRAP.(12) TRAP output includes the matrix of sums of squares and products for the terms of the regression equation. Although easy to run, TRAP is time consuming, so to minimise the number of runs in this study, the most complicated regressions were analysed first. The resulting sum of squares and products were then available for calculating all the possible simpler relationships among the variables - either on a desk calculator or by using less time consuming matrix operation programmes with the IBM 650.

The regressions calculated may be subdivided into:

(1) Cumulative models

(2) Incremental models

 (a) weekly subperiods

 (b) fortnightly subperiods.

Results and discussion of regression analyses for the derivation of pigmeat production functions from the particular experimental data mentioned are given in the next chapter.(13)

---

(12) See Appendix IV(A) for a description of this programme.

(13) See Appendix IV(B) for particulars about the availability of this experimental data.

C H A P T E R V

RESULTS OF REGRESSION ANALYSES

Results presented in this chapter are based on weekly observations for each of 26 pigs.

5.1 Cumulative Models

The basic relationship studied is given in conceptual form as:

$$Y = f(M, Wh, T, S).$$

A polynomial equation of second order was fitted to the 491 observations using the regression programme TRAP. A second order polynomial with four independent variables uses 15 degrees of freedom, leaving 476 degrees of freedom for lenient criterion and 11 degrees of freedom for strict criterion tests of significance. The intercorrelations for the terms of this regression equation are given in Table I, while the estimated regression coefficients, and their corresponding "t-test" values calculated by the ratio: $\frac{b_i}{s_{bi}}$, (where s_{bi} is the standard error of the regression coefficient b_i), are given in Table II. The "t-test" values are given in preference to the standard errors for ease of interpreting significance levels. In addition, as we have noted in section 3.3, the ratio $\frac{b_i^2}{c_{ii}}$

TABLE I : Intercorrelations for Terms of Second Order Polynomial : Y = f(M,Wh,T,S)

	TERM														
	M	Wh	T	S	M^2	Wh^2	T^2	S^2	MWh	MT	MS	WhT	WhS	TS	Y
M	1	.49	.29	*	.96	*	*	*	.82	.86	.49	*	*	*	.61
Wh		1	.93	.49	.22	.94	.87	.42	.68	.59	.51	.92	.49	.49	.88
T			1	.32	*	.84	.95	*	.63	.65	*	.93	.31	.33	.87
S				1	*	.51	*	.93	*	*	.87	.40	.96	.97	.34
M^2					1	*	*	*	.82	.86	*	*	*	*	.55
Wh^2						1	.86	.51	.55	.44	.48	.94	.56	.56	.71
T^2							1	*	.52	.53	*	.98	*	*	.72
S^2								1	*	*	.75	*	.98	.98	.24
MWh									1	.96	.39	.54	*	*	.84
MT										1	*	.51	*	*	.81
MS											1	*	.81	.84	.46
WhT												1	.44	.44	.75
WhS													1	.99	.32
TS														1	.33

* correlation < .30

gives the sum of squares for b_i independent of all other regression coefficients. This value may be tested against the error mean square in the usual way by the F-ratio. However, as the t-test and F-ratio values are in fact the same test, only the former will be tabulated.[1]

TABLE II : Second Order Polynomial Regression : Y = f(M,Wh,T,S)

	TERM	REGRESSION COEFFICIENT	t-test Value	
			Lenient Criterion 476 d.f.	Strict Criterion 11 d.f.
	CONST.	-1.687		
First Order Terms	M	0.455	17.76 ***	2.70 *
	Wh	0.169	13.52 ***	2.05 10%
	T	2.107	6.10 ***	0.93 NS
	S	-0.201	2.99 **	0.45 NS
Second Order Terms	M^2	0.898×10^{-4}	0.41 NS	0.06 NS
	Wh^2	-0.222×10^{-3}	22.08 ***	3.36 **
	T^2	-0.294	14.32 ***	2.18 10%
	S^2	-0.525×10^{-2}	7.25 ***	1.10 NS
	MWh	-0.155×10^{-3}	1.04 NS	0.16 NS
	MT	-0.329×10^{-2}	0.79 NS	0.12 NS
	MS	-0.252×10^{-2}	2.83 **	0.43 NS
	WhT	0.137×10^{-1}	20.58 ***	3.13 **
	WhS	0.356×10^{-3}	1.22 NS	0.18 NS
	TS	0.191×10^{-1}	1.90 10%	0.29 NS

(1) The t-test values are computed by TRAP for 476 degrees of freedom; a simple transformation is required to derive these values for 11 d.f.

The analysis of variance for this regression is given in Table III.

TABLE III : Analysis of Variance, Standardised Form, Strict Criterion.

Source	d.f.	S.S.	M.S.	F
Total Regression	14	0.99079	0.07077	84.25 ***
Linear terms	4	0.04235	0.23559	280.46 ***
Extra due to Quadratic terms	10	0.04844	0.00484	5.76 **
Deviations	11	0.00921	0.00084	
TOTAL	25	1.00000		

$R^2 = 0.99079$ for Total Regression.

5.2 Discussion

The pertinent features of these results are:

(1) The relative significance of the linear terms. All linear terms are highly significant for the lenient criterion as judged by the t-test. For the strict criterion, however, Time and Scouring terms are not significant. Also for the strict criterion no second order terms involving Scouring are significant. On the other hand, second order terms involving Time are significant using the strict criterion.

(2) The analysis of variance indicates that the "explanation" afforded by the linear terms is significantly added to by inclusion of the quadratic terms.

(3) Looking at the intercorrelation matrix for the terms of this regression equation, we see that Whey and Time are highly

multicollinear. Where terms are functionally related, i.e. Wh and WhS, high correlations are to be expected.

Conclusions that may be drawn from these results are that relatively little precision should be lost by dropping the terms involving Scouring from the regression model.

Where the regression model is to be used for predictive purposes only, Time should probably be included as indicated by the significance of some of the terms involving this variable. Where, however, estimates of structural parameters are required, the level of multicollinearity between Whey and Time leads to some doubts about including the latter variable in the regression model. Some doubt has already been expressed as to the validity of using the cumulative model to obtain estimates of structural coefficients. In order to examine this question more closely it would be desirable to test the residuals from regression for autocorrelation. This aspect will be considered in more detail when comparing the cumulative and incremental models.

No test for heteroskedasticity was possible as the regression had not been subdivided into periods.

5.3 Alternative Algebraic Forms

The opportunity was also taken to estimate the cumulative model in terms of the Log and Square Root functions. These models did not appear to have any advantage over the more general Polnnomial form, and are presented in Appendix V.

5.4 Weekly Incremental Models

Weekly incremental models have been calculated using whey consumption measured in gallons and pounds dry matter. Detailed

TABLE IV : Intercorrelations for Terms of Second Order Polynomial : $y_i = f(Wt_{i-1}, m_i, wh_i,$

	TERM													
	Wt_{i-1}	m_i	wh_i	s_i	Wt_{i-1}^2	m_i^2	wh_i^2	s_i^2	$Wt_{i-1}m_i$	$Wt_{i-1}wh_i$	$Wt_{i-1}s_i$	m_iwh_i	m_is_i	wh_is_i
Wt_{i-1}	1	-.45	.84	*	.99	*	.80	*	*	.94	*	*	*	*
m_i		1	-.42	-.36	-.42	.95	*	*	.84	-.42	*	.83	*	-.57
wh_i			1	.52	.81	-.42	.99	*	*	.95	.57	*	*	.57
s_i				1	*	*	.56	.96	*	.45	.97	*	.46	.97
Wt_{i-1}^2					1	-.41	.78	*	*	.94	*	*	*	*
m_i^2						1	-.38	*	.70	-.41	*	.70	*	*
wh_i^2							1	.53	*	.95	.61	*	*	.61
s_i^2								1	*	.42	.96	*	*	.97
$Wt_{i-1}m_i$									1	*	*	.96	*	*
$Wt_{i-1}wh_i$										1	.52	*	*	*
$Wt_{i-1}s_i$											1	*	.43	.99
m_iwh_i												1	*	*
m_is_i													1	.39
wh_is_i														1

* correlation < 0.3

BLE IV : Intercorrelations for Terms of Second Order Polynomial : $y_i = f(Wt_{i-1}, m_i, wh_i, s_i)$

	TERM														
	Wt_{i-1}	m_i	wh_i	s_i	Wt_{i-1}^2	m_i^2	wh_i^2	s_i^2	$Wt_{i-1}m_i$	$Wt_{i-1}wh_i$	$Wt_{i-1}s_i$	m_iwh_i	m_is_i	wh_is_i	y_i
	1	-.45	.84	*	.99	*	.80	*	*	.94	*	*	*	*	-.12
		1	-.42	-.36	-.42	.95	*	*	.84	-.42	*	.83	*	-.57	.59
			1	.52	.81	-.42	.99	*	*	.95	.57	*	*	.57	.15
				1	*	*	.56	.96	*	.45	.97	*	.46	.97	-.06
2					1	-.41	.78	*	*	.94	*	*	*	*	-.13
						1	-.38	*	.70	-.41	*	.70	*	*	.54
							1	.53	*	.95	.61	*	*	.61	.10
								1	*	.42	.96	*	*	.97	-.10
m_i									1	*	*	.96	*	*	.60
wh_i										1	.52	*	*	*	0
s_i											1	*	.43	.99	-.06
i												1	*	*	.66
													1	.39	.12
i														1	-.06

* correlation < 0.3

results for a second order polynomial regression for the model:

$$y_i = f(Wt_{i-1}, m_i, wh_i, s_i)$$

where whey consumed per week is measured in gallons are first presented.

The matrix of correlations between the terms of this equation are presented in Table IV, the regression equation is given in Table V, together with t-tests, and the usual analysis of variance is presented in Table VI.

TABLE V : Second Order Polynomial Pegression : $y_i = f(Wt_{i-1}, m_i, wh_i, s_i)$

TERM	REGRESSION COEFFICIENT	t-test Value
		Lenient Criterion 476 d.f.
CONST.	-0.400	
Wt_{i-1}	-0.929×10^{-1}	3.50 ***
m_i	0.607	3.40 ***
wh_i	0.611	6.93 ***
s_i	-0.118	0.48 NS
${Wt_{i-1}}^2$	0.223×10^{-3}	1.17 NS
${m_i}^2$	-0.315×10^{-2}	0.34 NS
${wh_i}^2$	-0.412×10^{-2}	1.29 NS
${s_i}^2$	-0.319×10^{-1}	1.29 NS
$Wt_{i-1}m_i$	0.281×10^{-2}	1.90 10%
$Wt_{i-1}wh_i$	-0.508×10^{-3}	0.34 NS
$Wt_{i-1}s_i$	0.813×10^{-2}	3.23 **
m_iwh_i	-0.136×10^{-1}	2.45 *
m_is_i	-0.243×10^{-1}	1.31 NS
wh_is_i	-0.217×10^{-1}	2.11 *

TABLE VI : Analysis of Variance (Lenient Criterion)

Source	d.f.	S.S.	M.S.	F
Total Regression	14	1612.265	115.162	52.95 ***
Linear terms	4	1506.326	376.582	173.14***
Quadratic terms	10	105.939	10.594	4.87***
Deviations	476	1035.521	2.175	
TOTAL	490	2647.786		

$R^2 = 0.60891$

The evidence available suggests that the observations made at weekly intervals on a single pig are to a large degree statistically independent, hence significance tests for incremental models have been based on the lenient criterion.

5.5 Discussion

The tests of significance suggest that liveweight at the end of the previous week, and meal and whey consumed during the week, are the major variables affecting liveweight gain in any week. A significantly better fit of the regression model to the data is obtained by including second order terms. The matrix of intercorrelations indicates some danger of multicollinearity between liveweight at the end of the previous week and gallons of whey fed during the week. We cannot assume that liveweight is measured without error and this fact alone may result in Least Squares estimates of regression coefficients being biased. This danger is increased when explanatory variables subject to measurement error are highly correlated.[2]

(2) See E.O.Heady and J.Dillon, op. cit., Ch.4, p.135.

Apart from minimising errors of measurement, already a feature of the experiment which we are considering, the author knows of no reasonable way of allowing for this danger.

Apart from these factors, the model is seen to provide a reasonable, if not highly satisfactory, explanation of variation in liveweight gain per week.

5.6 Whey: Dry Matter and Gallons

Simple linear regressions with whey measured as pounds dry matter and as gallons are given:

$$\hat{y}_i = b_o \underset{(\pm .0036)}{-0.0286Wt_{i-1}} + \underset{(\pm .0238)}{0.5062m_i} + \underset{(\pm .0127)}{0.1777wh_i}; \quad R^2 = 0.5569$$

.... (gallons)

$$\hat{y}_i = b_o' \underset{(\pm .0033)}{-0.0254Wt_{i-1}} + \underset{(\pm .0235)}{0.4966m_i} + \underset{(\pm .0193)}{0.2793wh_{D.M.i}}; \quad R^2 = 0.5665$$

.... (dry matter)

The correlation between Wt_{i-1} and wh_i is 0.84, while that between Wt_{i-1} and $wh_{D.M.i}$ is 0.81.

It can be seen that the regression equation with whey consumed per week measured on a dry matter basis is slightly more accurate in terms of the proportion of total variance explained: R^2. In addition, the correlation with liveweight is slightly lower for the dry matter model. However, no great significance is attached to these small differences. For practical purposes the two models will be considered as equivalent.

The respective regression coefficients for whey allow us to calculate the dry matter content of this foodstuff as approx. 6.4%. This agrees closely with the recorded average dry matter content of 6.2% for whey used in this trial.

5.7 Comparison of Cumulative and Incremental Models Calculated from Weekly Observations

(a) Predictive Power: The proportion of variance explained by regression, R^2, is often taken as a guide to the predictive worth of a regression equation. The respective R^2 values for regressions presented have been given in the text. The R^2 for the cumulative and incremental models are repeated in the following table:

TABLE VII : R^2 Values for Second Order Polynomial Models

$Y = f(M, Wh, T, S)$	$R^2 = 0.991$
$y_i = f(Wt_{i-1}, m_i, wh_i, s_i)$	$R^2 = 0.609$

Thus 99.1% of the variance in the dependent variable is explained by the second order polynomial cumulative model, but only 60.9% of the variance in the dependent variable is explained by the corresponding incremental model.

This apparent difference in predictive power alone might lead us to place more reliance on the cumulative model.

It should be noted, however, that the comparison is only on the basis of proportion of variance explained. In actual fact, the dependent variables: "Total Weight Gain after i Weeks", Y_i, and "Weight Gain in the ith Week", y_i, are not directly comparable. The relative variance of these two variables differs widely as can be seen from their respective total sum of squares about the mean value.

TABLE VIII : Relative Variance of Y_i and y_i

Variable	No. of Observations	Corrected Sum of Squares
Y_i	491	660829.30
y_i	491	2647.79

On these grounds it is not reasonable to compare the predictive power of the cumulative and incremental models by their respective R^2 values. One method of carrying out such a comparison would be to predict total liveweight gains, $\hat{Y}_i'$, using the incremental model, i.e.:

$$\sum_{i=1}^{n} \hat{y}_i = \hat{Y}'_n ,$$

where $\hat{Y}'_n$ is the predicted liveweight gain at the end of n weeks. These predicted liveweight gains could then be compared with the observed liveweight gain series. In the usual way, the difference between observed and predicted values could be obtained, and the sum of squares of these deviations corresponds to those obtained through fitting the cumulative model. The proportion of total variation in the liveweight gain series explained by the incremental model may then be obtained to give the "effective R^2" of this model.

Where the series of differences between observed weekly gain (y_i) and predicted weekly gain ($\hat{y}_i$) is available, the required series of differences between total liveweight gain (Y_i) and predicted gain ($\hat{Y}_i'$), necessary to calculate the "effective R^2" of the incremental model, may be relatively easily obtained.

The deviations $(y_{ij} - \hat{y}_{ij})$ are cumulated separately for each pig, (where the j subscript has been introduced to distinguish between pigs), taking account of sign, and in the same order through time as the observations were made. The total after each deviation has been accounted for is noted. When this has been completed for each pig, the total sum of squares calculated from this entire series corresponds to the deviations sum of squares:-

$$\sum_{j=1}^{k} \sum_{i=1}^{n} (Y_{ij} - \hat{Y}'_{ij})^2$$

for n observations on each of k pigs.

The "effective R^2" may then be calculated from the relationship:

$$R'^2 = \frac{\text{TOTAL S.S. } (Y_i) - \text{DEVIATIONS S.S.}}{\text{TOTAL S.S. } (Y_i)}$$

The relationship between the two series may be shown more clearly:

$$Wt_1 - Wt_o = Y_1 = y_1 \quad \text{where } Wt_o = \text{initial liveweight.}$$

$$Wt_2 - Wt_o = Y_2 \quad \text{and} \quad Y_2 - Y_1 = y_2$$

$$Wt_3 - Wt_o = Y_3 \quad \text{and} \quad Y_3 - Y_2 = y_3. \quad \text{etc.}$$

Thus we have:

$$Y_1 - \hat{Y}'_1 = Y_1 - \hat{y}_1 = y_1 - \hat{y}_1 = e_1$$

$$Y_2 - \hat{Y}'_2 = (y_1 + y_2) - (\hat{y}_1 + \hat{y}_2) = (y_1 - \hat{y}_1) + (y_2 - \hat{y}_2)$$

$$= e_1 + e_2$$

etc.

The series of observed values, predicted values and differences, for all 491 observations, was calculated by TRAP for the incremental model

$$\hat{y}_i = f(Wt_{i-1}, m_i, wh_{D.M.i})$$

estimated as a second order polynomial. This model is not strictly comparable with the cumulative second order model:

$$\hat{Y} = f(M, Wh, T, S)$$

as whey is measured in gallons in this model and in pounds dry matter in the first. A term for scouring is also included in the cumulative model. However, results already presented would not lead us to expect these differences to be unduly important. Details of this incremental regression are given in Table IX.

TABLE IX : Weekly Incremental Model, Whey in D.M. Second Order Polynomial.

TERM	REGRESSION COEFFICIENT	t-test Lenient Criterion 4?1 d.f.
CONSTANT	-0.608	-
Wt_{i-1}	-0.648×10^{-1}	2.69 **
m_i	0.630	3.48 ***
$wh_{D.M.i}$	0.788	6.49 ***
Wt^2_{i-1}	0.380×10^{-4}	0.28 NS
m_i^2	-0.497×10^{-2}	0.53 NS
$wh_{D.M.i}^2$	-0.153×10^{-1}	2.77 **
$Wt_{i-1}m_i$	-0.270×10^{-4}	0.02 NS
$Wt_{i-1}wh_{D.M.i}$	0.127×10^{-2}	0.86 NS
$m_i wh_{D.M.i}$	-0.439×10^{-2}	0.67 NS

$R^2 = 0.5902$

The series of differences $(y_{ij} - \hat{y}_{ij})$ for this regression equation were manipulated in the way described above, to calculate the sum of squares of deviations from the liveweight gain series. The results obtained were:

Calculated deviations S.S. = 16132.31

$$\therefore \text{"Effective } R^2\text{"} = \frac{660829.30 - 16132.31}{660829.30}$$

i.e. $R'^2 = 0.976$

The actual R^2 for the cumulative model = 0.991

On the basis of this comparison we conclude that the effective predictive power of the incremental model is almost equivalent to that of the cumulative model.

A feature of this comparison requires further investigation:

(1) Prediction of $\hat{y}_i$ from the incremental model depends in some part upon knowing the liveweight of the animal concerned at the beginning of the ith week. This value, Wt_{i-1}, could be replaced for practical purposes by:

$$Wt_o + \sum_{j=1}^{i-1} \hat{y}_j \qquad (5.1)$$

where Wt_o is the initial liveweight at the beginning of the trial period. This value (5.1) however, is subject to predictive error. The importance of this additional error should be investigated.

(b) Estimates of Structural Parameters: Autocorrelation of the error term was discussed in section 3.6. Where errors are autocorrelated the method of Least Squares does not give efficient estimates of the population regression coefficients. The problem of carrying out the Durbin and Watson test(3)

---

(3) J.Durbin and G.S.Watson, Biometrika, op.cit., p.159-78.

for autocorrelation of the residuals from regression where the analysis has been based on more than one time series has also been mentioned. An attempt, however, has been made to gain some idea of the seriousness of autocorrelation in this data by considering deviations from regression for individual pigs. When calculating the Durbin and Watson "d-statistic" for individual pigs two problems arise:

(1) When the regression equation has been estimated from observations on a number of animals, it is not necessary that the sum of the deviations for any one pig will equal zero. In this case the sum of squared residuals from regression will be greater than if their sum about this same regression line was zero. The "d-statistic" has been derived by Durban and Watson on this latter assumption. Using the "inflated" sum of squared residuals for any individual pig could well result in a downward bias to the "d-statistic" so calculated.(4) An effect equivalent to adjusting the regression parallel to its original position, so that the sum of the residuals for a single pig is equal to zero, may be obtained by correcting the sum of squared residuals for the mean in the usual way. A proof has not been derived to show that this adjustment does in fact remove possible bias from the "d-statistic" calculations. Nevertheless, calculations have been made using this 'intuitive' adjustment.

(2) The "d-statistic" can be conveniently calculated when only first order terms have been included in the regression equation. In this case errors may well be autocorrelated because

---

(4) We have $d = \frac{\Sigma(\Delta z)^2}{\Sigma \Delta z^2}$; where $\Sigma \Delta z^2$ is the residual sum of square

the wrong form of the equation has been used. The linear regression model for the cumulative data used in these calculations is:

$$\hat{Y} = 8.587 + \underset{(\pm .070)}{0.506M} + \underset{(\pm .010)}{0.136Wh}; \quad R^2 = 0.93$$

The variables Time and Scouring have been omitted; the latter because of its relative inadequacy as an explanatory variable.

Because multicollinearity exists between the variables Whey and Time, the latter has been dropped in order to obtain a statistically more reliable estimate of their joint effects.

The weekly incremental regression used is

$$\hat{y}_i = 2.345 - \underset{(\pm .0036)}{0.0286Wt_{i-1}} + \underset{(\pm .0238)}{0.5062m_i} + \underset{(\pm .0127)}{0.1777wh_i}; \quad R^2 = 0.56$$

Calculated values of the "d-statistic" for three pigs are presented in Table X.

TABLE X : Durbin and Watson "d-statistic" Values

Pig No.	Treatment	No. Observations	d-statistic(5)	
			Cumulative	Weekly
No. 3	I	16	0.5318	2.0395 (1.9605)
No.11	IV	21	0.2147	2.4814 (1.5186)
No.22	VIII	21	0.0429	1.4365 (2.5635)

These results indicate that residuals from the linear cumulative regression equation are highly positively auto-correlated. The weekly incremental model, however, is seen

(5) It should be noted that significant adjustments for the mean value of the deviations about regression were only made for the cumulative model. Had these adjustments not been made the d-statistics would have been even lower than those shown.

to have residuals that are relatively independently distributed. Where it is desired to test for negative autocorrelation of the residuals, the statistic (4-d) is used. These values are presented in brackets for the weekly incremental model in Table X; no serious negative autocorrelation is indicated.

These tests suggest that autocorrelation of the residuals is much less serious in the incremental model than in the cumulative one.

5.8 Fortnightly Incremental Model

Regression equations based on fortnightly weight gains and feed consumption were calculated from the experimental data. The main feature of this analysis was that the total production period was divided into three subperiods on the basis of liveweight, thus:

PERIOD I : Start of trial (approx. 50 lb) $\longrightarrow$ 93 lb.
PERIOD II : 94 lb $\longrightarrow$ 136 lb.
PERIOD III : 137 lb $\longrightarrow$ Finish of trial (approx. 180 lb)

In addition, the regression for the total production period was calculated. In each case the second order polynomial was fitted to the variables.

$$y_i = f(Wt_{i-1}, m_i, wh_i, s_i).$$

The regression coefficients for each of the subperiod equations is listed in Appendix VI; standardised analyses of variance for each period are given in Table XI. The regression coefficients for the total production period equation are given in Table XVI.

TABLE XI : Analyses of Variance; Fortnightly Regression Equations.

Period	Source	d.f.	S.S.	M.S.	F
I $R^2 = 0.83$	Linear terms	4	0.7650	0.1912	68.30***
	Quadratic	10	0.0644	0.0064	2.30*
	Deviations	61	0.1706	0.0028	
	TOTAL	75	1.0000		
II $R^2 = 0.85$	Linear terms	4	0.8140	0.2035	88.90***
	Quadratic	10	0.0327	0.0033	1.43NS
	Deviations	67	0.1533	0.0023	
	TOTAL	81	1.0000		
III $R^2 = 0.74$	Linear terms	4	0.6798	0.1700	44.03***
	Quadratic	10	0.0652	0.0065	1.69 10%
	Deviations	66	0.2550	0.0039	
	TOTAL	80	1.0000		
All Periods $R^2 = 0.78$	Linear terms	4	0.7252	0.1813	181.30***
	Quadratic	10	0.0552	0.0055	5.50**
	Deviations	224	0.2196	0.0010	
	TOTAL	238	1.0000		

These analyses of variance indicate that relatively little extra precision is gained through addition of the quadratic terms in each of the subperiods, but the quadratic terms are highly significant when we are estimating over the whole range.

The average substitution rates of meal for whey have been calculated from the linear regression fitted to each period, and are given in the following table:

TABLE XII : Substitution Rates: Meal for Whey ($\frac{d.wh}{d.m}$)

Period	Reg.Coef.Meal (lbs)	Reg.Coef.Whey (gals)	$\frac{d.wh}{d.m}$*
I	0.4703	0.2742	1.72
II	0.4418	0.2480	1.78
III	0.4800	0.1506	3.19

(*Absolute values for $\frac{d.wh}{d.m}$ have been tabulated.)

The substitution rate of meal for whey is seen to increase (i.e., more whey is required to substitute for 1 lb. of meal) from Period I to III. Thus in times of whey shortage heavier pigs should be fed more meal and less whey. One pound of meal fed to heavy pigs (say above 140 lbs.) will release approx. 3 gallons of whey which can be fed to other pigs. Each additional pound of meal fed to light pigs (say 50-100 lbs), however, releases only approx. 1.75 gallons of whey. These results apply to the rations used in the particular feeding trial we are concerned with.

Subdivision into three production periods allows an interesting analysis that confirms an earlier hypothesis: that growth rate is affected by feeding rate and liveweight together. The following analysis of variance indicates that there has been no significant difference in mean fortnightly growth rate between periods.

TABLE XIII : Analysis of Variance : Mean Growth Rates

Source	d.f.	S.S.	M.S.	F
Between Periods	2	61.069	30.534	1.90 NS
Within Periods	236	3792.169	16.068	
TOTAL	238	3853.238		

Mean fortnightly liveweight gains are now adjusted by covariance analysis for quantities of meal and whey consumed per fortnight in each period. This adjustment is given in Table XIV.

TABLE XIV : Analysis of Covariance I - Adjustment for Meal and Whey Consumed per Fortnight

Source	d.f.	d.f.'	Deviations S.S.	M.S.	F
PERIOD I	75	73	300.605		
PERIOD II	81	79	343.113		
PERIOD III	80	78	427.050		
Within Periods Reg.		230	1070.768	4.6555	
Reg. Coefs.		4	25.825	6.4562	1.39NS
Common Reg.	236	234	1096.593	4.6863	
Adjusted Means		2	271.752	135.8760	28.99***
Total Regression	238	236	1368.345		

The difference between meal and whey regression coefficients for the three periods is tested by $F = \frac{6.4562}{4.6555} = 1.39$ which is not significant at the 10% level with 4 and 230 degrees of freedom. It is therefore valid to make an adjustment to the mean growth rate in each period, by taking account of the common regression of liveweight gain per fortnight on meal and whey consumed. The significance of this adjustment is tested by $F = \frac{135.876}{4.6863} = 28.99$ with 2 and 234 degrees of freedom, and is found to be significant at the 0.1% level of probability. Thus, after adjusting for meal and whey consumption in each period we have significant between period differences in mean fortnightly growth rate. The analysis of covariance is now supplemented to account for differences in liveweight between periods. The results are presented in Table XV.

TABLE XV : Analysis of Covariance II - Adjustment for Meal and Whey Consumed and Liveweight Between Periods

Source	d.f.	d.f.'	Deviations S.S.	M.S.	F
PERIOD I	75	72	282.630		
PERIOD II	81	78	310.783		
PERIOD III	80	77	419.351		
Within Periods Reg.		227	1012.764	4.4615	
Reg.Coefs.		6	50.989	8.4982	1.90 10%
Common Reg.	236	233	1063.753	4.5655	
Adjusted Means		2	19.358	9.6790	2.12 NS
Total Regression	238	235	1083.111		

Thus, after allowing for the effect of liveweight on mean fortnightly growth rate there are no significant differences between periods. This is in agreement with the original analysis of variance (Table XIII). Thus the same fortnightly growth rate can be maintained at higher liveweights only by increased feeding rates. This conclusion is in accordance with the reasoning given in derivation of the incremental production model.

5.9 Weekly and Fortnightly Incremental Models

Consideration of the weekly and fortnightly incremental models may allow us to draw some conclusions about the underlying population incremental model. In particular, we should be interested in the length of the incremental period. The likely importance of length of incremental period has already been discussed in section 2.3. Before comparing which of the two incremental models best conforms to the underlying population relationship, a definite hypotheses as to the characteristics of this relationship must be made.

In order to decide which of the two incremental models estimated best fits the production data, they could be compared on a predictive basis as was done for the weekly incremental and cumulative models. Time, however, has not permitted a detailed comparison of the weekly and fortnightly models on this basis.

A visual comparison of the two models may be made with the aid of Table XVI.

TABLE XVI : Regression Coefficients for the Weekly and Fortnightly Models

Term	Weekly Model	t-statistic (476 d.f.)	Fortnightly model (All periods)	t-statistic (224 d.f.)
CONSTANT	-0.400		-0.256	
Wt_{i-1}	-0.929×10^{-1}	3.50 ***	-0.200	3.83 ***
m_i	0.607	3.40 ***	0.597	3.45 ***
wh_i	0.611	6.93 ***	0.604	6.85 ***
s_i	-0.118	0.48 NS	-0.176	0.68 NS
Wt_{i-1}^2	0.223×10^{-3}	1.17 NS	0.477×10^{-3}	1.30 NS
m_i^2	-0.315×10^{-2}	0.34 NS	-0.166×10^{-2}	0.37 NS
wh_i^2	-0.412×10^{-2}	1.29 NS	-0.202×10^{-2}	1.34 NS
s_i^2	-0.319×10^{-1}	1.29 NS	-0.706×10^{-2}	0.53 NS
$Wt_{i-1}m_i$	0.281×10^{-2}	1.90 10%	0.291×10^{-2}	1.99 *
$Wt_{i-1}wh_i$	-0.508×10^{-3}	0.34 NS	-0.382×10^{-3}	0.28 NS
$Wt_{i-1}s_i$	0.813×10^{-2}	3.23 **	0.875×10^{-2}	3.49 ***
m_iwh_i	-0.136×10^{-1}	2.45 *	-0.724×10^{-2}	2.67 **
m_is_i	-0.243×10^{-1}	1.31 NS	-0.411×10^{-2}	0.39 NS
wh_is_i	-0.217×10^{-1}	2.11 *	-0.129×10^{-1}	2.46 *
	R^2 = 0.6089 Total Variation = 2647.786 Variation Explained by Regression = 1612.266		R^2 = 0.7804 Total Variation = 3853.238 Variation Explained by Regression = 3006.985	

The regression coefficients for the fortnightly model are slightly more accurately determined in general and with less than half the degrees of freedom of the weekly model.

5.10 Discussion of Results

A comparison of the predictive power of the weekly incremental and cumulative regression equations indicates that there is little to choose between these production models.

Tests for autocorrelation of the residuals on individual pigs using linear weekly incremental and cumulative models were made. The results of these tests are subject to some reservation, but they indicate that residuals from the cumulative model, for individual pigs, are likely to be positively autocorrelated. The residuals from incremental regression, however, appeared to be independently distributed. The weekly incremental model is therefore more satisfactory from a statistical point of view, than the cumulative model. This property is likely to carry over to incremental models based on different time periods. Whether incremental models should be based on weekly, fortnightly, three-weekly, or some other period, is not clear. There is little obvious difference, however, between the weekly and fortnightly regression equations estimated. There is also the possibility of sacrificing information when less observations are used, as is the case in estimating say a three-week incremental model versus a weekly model.

Incremental models have the added advantage over cumulative models in that more meaningful information is provided on the production process. Thus, given a total quantity of meal and whey and a total time in which to feed these quantities, the cumulative model allows us to predict a value for expected liveweight gain. It does not, however, tell us what ration line to use in achieving this result, in this sense the expected liveweight gain is not uniquely determined. In addition, the production functions estimated in cumulative form in this study only refer to liveweight gain from 48-50 lb liveweight. For practical purposes it is therefore necessary to know the total quantities of meal and whey that have already been fed at any time subsequent to this if we are to predict future liveweight. The incremental models, however, allow us to specify ration lines and to predict liveweight gains according to the rate of feeding. In order to predict liveweight gain in any future period a knowledge of present liveweight and future rates of feeding is required. It is reasonable to assume that present liveweight is likely to be easier to obtain under practical conditions than quantities of feed consumed in the past.

Additional evaluation of a production model should involve tests in actual practice. The production functions presented in this thesis could be tested using data from the second half of the meal:whey levels trial, conducted at the Massey College Piggery in 1961. Time has not been available to carry out these tests in this study.

Smith[2] has given the relationship between frozen carcass weight, Y, and liveweight, X, for pigs as:

$$Y = 0.79X - 4.4(\pm 3.1)$$

This relationship allows us to calculate price of pigmeat on a liveweight basis:

pigmeat = 15.8d per lb. liveweight.

6.1 Cumulative Model

We will consider first the cumulative production function given in Table II, Ch.5; derived from the meal:whey levels trial already described. This model includes the explanatory variables Time and Scouring.

The occurrence of scouring has been shown to vary significantly with treatments used in the experiment, but adjustments to liveweight gains due to this factor have been relatively insignificant. For this reason scouring should be dropped from a relationship to be used in economic analysis. Thus we see that from an operational point of view scouring should have been omitted from the production function. However, in the computing-time available it was not possible to estimate the cumulative production model:

$$Y = f(M, Wh, T).$$

In the absence of this model, the level of scouring has been set to the mean level ($\bar{S}$ = 8.8 days) and the original production

(2) "Relationship Between Liveweight and Carcass-Weight Increments of Pigs", D.M. Smith, N.Z.J.Sc. and Tech.; sec.A, Vol.38, No.8, Aug. 1957, p.803.

C H A P T E R V I

PRODUCTION ECONOMICS ANALYSIS

Given a production function and the prices of the variables oncerned, the principles of production economics can be used to aximise net returns.

Cumulative and incremental production models pertaining o liveweight gain from 'weaning',(1) have been derived for meal nd whey feeding of pigs. We ignore the cost of rearing pigs o 'weaning' weight (though this could be the subject of a eparate analysis). Given prices of meal, whey and pigmeat, we ay maximise net returns in some sense for fattening pigs. In his chapter, production economics principles are applied to umulative and incremental production models in turn, and the sefulness of the results is discussed.

For illustrative purposes, the following prices re assumed:

meal	=	3.4d per lb.
whey	=	1.5d per gal.
pigmeat	=	20d per lb. carcass weight.

(1) For ease of exposition, we refer to 'weaning' in this chapter as the initial liveweight (actually 48-50 lbs.) of pigs used in the meal:whey levels trial.

function (Table II in Ch.5) adjusted to give:

$$\hat{Y} = -3.866 + 0.433M + 0.173Wh + 2.275T + 0.0001M^2 - 0.0002Wh^2 - 0.2940T^2 - 0.0002MWh - 0.0033MT + 0.0137WhT \quad (6.1)^{(3)}$$

It should be recognised that (6.1) is not the Least Squares estimate of the relationship; but it is only inaccurate in so far as the mean level of scouring does not occur and will serve for illustrative purposes.

6.1.1 Unrestricted Maximisation of Net Revenue Per Pig

Production economics principles applied to the cumulative production function, (6.1), and using the prices given, allow us to predict the optimum feed and time combination and liveweight gain, for a single pig.[4]

Predicted net revenue per pig, $\hat{R}$, for liveweight gain from 'weaning' may be written:

$$\hat{R} = 15.8\hat{Y} - 3.4M - 1.5Wh \quad (6.2)$$

where $\hat{R}$ is measured in pence.

Substituting in (6.2) for $\hat{Y}$ (given by (6.1)) we have:

$$\hat{R} = -61.083 + 3.441M + 1.233Wh + 35.945T + 0.0016M^2 - 0.0032Wh^2 - 4.6452T^2 - 0.0032MWh - 0.0521MT + 0.2165WhT \quad (6.3)$$

(3) In this and subsequent equations in this chapter subscripts from variables have been dropped. Thus, $\hat{Y}$ is predicted liveweight gain from weaning when M pounds of meal and Wh gallons of whey have been consumed at the end of T weeks.

(4) This information on a single pig can easily be extended to a litter basis.

To maximise predicted net revenue per pig, R, we want:

$$d\hat{R} = \frac{\delta\hat{R}}{\delta M}.dM + \frac{\delta\hat{R}}{\delta Wh}.dWh + \frac{\delta\hat{R}}{\delta T}.dT = 0 \qquad (6.4)$$

and,

$$dR^2 < 0 \qquad (6.5)$$

Equation (6.4) defines the stationary point of the net revenue function (6.3). If (6.5) holds, the stationary point is a maximum. To calculate the co-ordinates of the stationary point for (6.3) we solve the following set of simultaneous equations:-

$$\frac{\delta\hat{R}}{\delta M} = 3.441 + 0.0032M - 0.0032Wh - 0.0521T = 0$$

$$\frac{\delta\hat{R}}{\delta Wh} = 1.233 - 0.0032M - 0.0063Wh + 0.2165T = 0$$

$$\frac{\delta\hat{R}}{\delta T} = 35.945 - 0.0521M + 0.2165Wh - 9.2900T = 0 \qquad (6.6)$$

The co-ordinates of the stationary point found by the solution of the set of simultaneous equations, (6.6), are:

M = 208.4 lbs

Wh = 906.5 gals

T = 23.8 weeks.

Predicted liveweight gain from 'weaning', obtained by substituting these values for M, Wh and T in (6.1) is 212.2 lbs. Predicted net revenue, calculated from (6.3), is £5.35 per pig, or £0.225 per week.

The most important feature of this stationary point is that it is well outside the levels of meal and whey consumption explored in the particular feeding trial from which the production relationships have been derived.[5]

---

(5) In addition it could be shown that this stationary point is a saddle point exhibiting increasing returns in the meal direction.

This may be seen from fig.1 which illustrates the limits of meal and whey consumption in this trial. To be realistic then it is necessary to maximise (6.3) subject to certain side restrictions that will ensure we remain within the limits of meal and whey consumption explored.

6.1.2 Restricted Maximisation of Net Revenue Per Pig

Considerable difficulty might be experienced in attempting to decide upon restrictions that would confine attention to that portion of the meal, whey, time space explored in the trial. This is because each meal:whey combination may be fed in a number of different time periods, as shown by fig.2, Ch.5. However, in maximising net revenue from a single pig, any given quantity of meal and whey should be fed in the time period that maximises liveweight gain from this quantity of feed. In this case, no account is taken of competition for fattening accommodation, i.e. extra fattening time is considered to have no opportunity cost. Liveweight gain from 'weaning' as a function of meal and whey consumption and time, is given by (6.1). This production function allows us to calculate the total time in which a given quantity of feed should be consumed to maximise liveweight gain (and hence net profit per pig) from this quantity of feed. This time is given by solving:

$$\frac{d\hat{Y}}{dT} = 0,$$

where

$$\frac{d^2\hat{Y}}{dT^2} < 0$$

Now, from (6.1) we have:

$$\frac{d\hat{Y}}{dT} = 2.275 - 0.588T - 0.0033M + 0.0137Wh = 0$$

and $\frac{d^2\hat{Y}}{dT^2} = -0.588.$

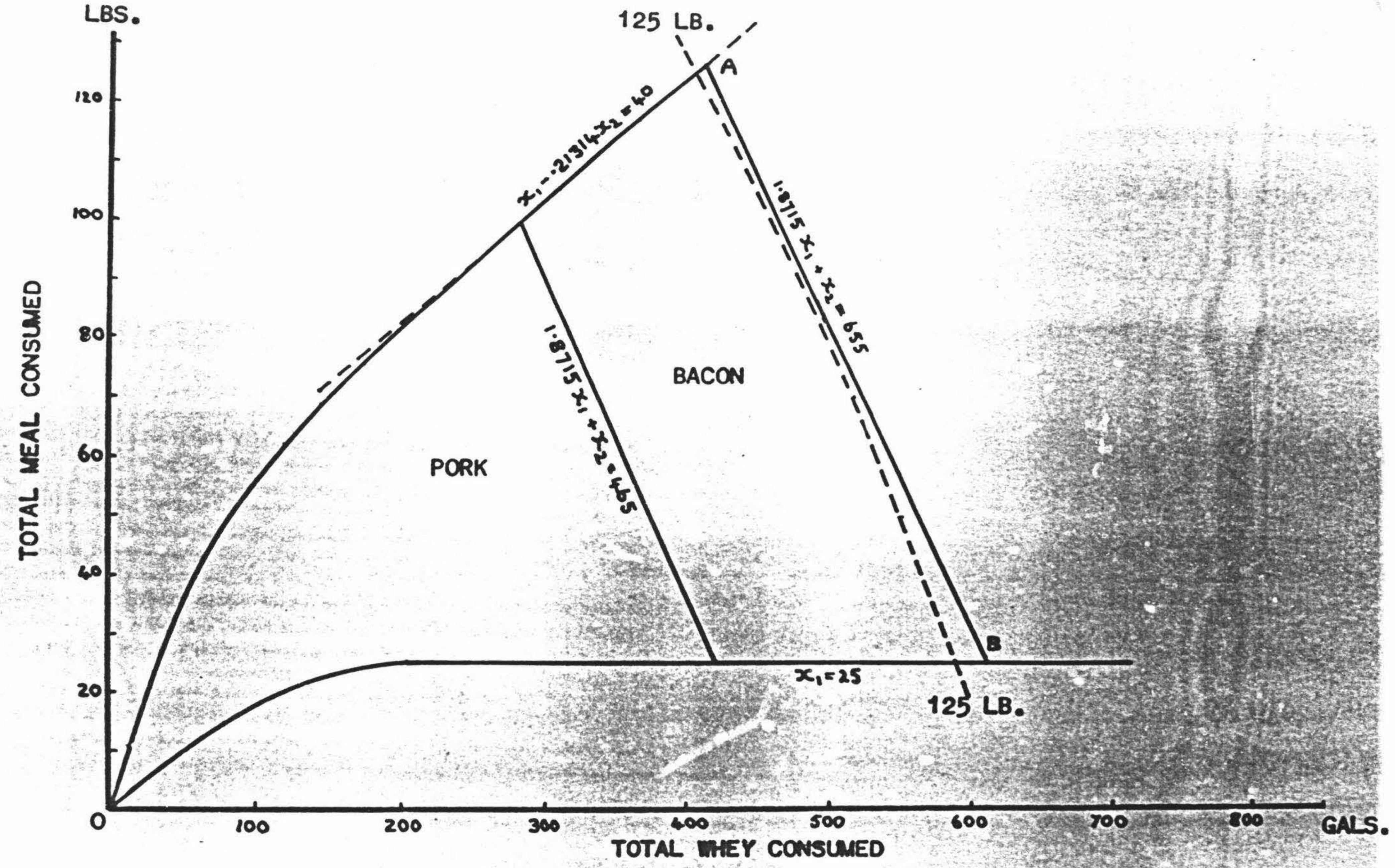


FIG.1 LINEAR RESTRICTIONS ON BACON PRODUCTION

Thus, $\frac{d\hat{Y}}{dT} = 0$ gives the optimum time in which to feed any given quantity of meal and whey.[6]

This relationship may be rewritten:

$$T = 3.8690 - 0.0056M + 0.0233Wh. \qquad (6.7)$$

Where we feed so as to maximise liveweight gain from any M:Wh combination we may substitute for T, given by (6.7), in the predicted net revenue function per pig (6.3). This substitution results in (6.8) and gives the predicted net revenue per pig, $\hat{R}'$, when meal and whey is fed according to (6.7).

$$\hat{R}' = 8.4530 + 3.2360M + 2.0708Wh + 0.0016M^2 - 0.0006Wh^2 - 0.0044MWh \qquad (6.8)$$ [7]

The problem has now been reduced to maximising (6.8) subject to suitable side restrictions on total quantities of meal and whey that can be consumed.

The price of 20d/lb carcass weight, used in this analysis, applies at present times to pigs of bacon rather than pork weight. Pigs in the carcass weight range : 110-140 lbs, are classed as baconers; this corresponds to a liveweight range of approx. 145-180 lbs. As the production models derived in this study refer to liveweight gain from 'weaning' (48-50 lb liveweight), a weight gain, Y, of 95-130 lb corresponds to pigs of bacon weight.

The feasible area for bacon production from meal and whey rations used in the particular feeding trial we are concerned with,

---

(6) The quantities of meal and whey that should be fed in T = 9 weeks for maximum weight gains from these combinations, are shown in fig.2,Ch.5.

(7) We note in passing that (6.8) could equally as well have been obtained by taking the condition $\frac{\delta \hat{R}'}{\delta T} = 0$ in (6.6), solving for T, and substituting for T in (6.3). Equations (6.3) and (6.8) therefore have the same stationary point, (apart from rounding errors).

is approximated by the convex set of four linear inequalities:

$$x_1 \geqslant 25 \quad (6.9)$$

$$1.8715x_1 + x_2 \geqslant 465 \quad (6.10)$$

$$1.8715x_1 + x_2 \leqslant 655 \quad (6.11)$$

$$x_1 - .21314x_2 \leqslant 40 \quad (6.12)$$

as illustrated in fig.1, and where x_1 = total meal consumed in lbs., and x_2 = total whey consumed in gals. Equations (6.10) and (6.11) approximate the predicted 95 lb and 130 lb liveweight gain contours when meal and whey is fed according to (6.7). Equations (6.9) and (6.12) approximate the feeding rations used that delimit the portion of the meal:whey plane explored. It is only necessary that these four approximations should be accurate for the area of interest. From fig.1 it would appear that this condition is reasonably well fulfilled.

The recently developed general solution to the problem of maximising a quadratic function subject to linear inequalities may be used to solve the problem of maximising (6.8) subject to the inequalities (6.9) to (6.12).[(8)] (In this simple two dimensional case, cut and try methods could be used to explore values of the net revenue function in the area of interest). If the maximum is found to lie on the restriction (6.10), which divides bacon from pork production, we should need to explore the area for pork production after adjusting the net revenue function for pork price (approx. 17d per lb carcass weight).

The computational procedure for maximising (6.8) subject to inequalities (6.9) to (6.12) is given in Appendix VII. The

---

(8) "The Maximisation of a Quadratic Function of Variables Subject to Linear Inequalities", W. Candler and R. Townsley. In Press.

maximum maximorum co-ordinates are:

$$M = 25.0 \text{ lbs}$$

$$Wh = 608.2 \text{ gals},$$

corresponding to point B in fig.1. Predicted net revenue at this point is £4.375 per pig when this quantity of meal and whey is fed in 17.9 weeks as given by (6.7). This gives a predicted liveweight gain of 128.4lbs, (from (6.1)), and a predicted net revenue of £0.244 per week.

6.1.3 Maximum Net Revenue Per Unit Time

Equation (6.3) predicts total net revenue per pig, $\hat{R}$, from given quantities of meal and whey fed in a given time. Predicted average net revenue per week, $\hat{R}*$, is given by:

$$\hat{R}* = \frac{\hat{R}}{T}$$

Thus $\hat{R}* = T^{-1}\hat{R}$

$$= T^{-1}(-61.083 + 3.441M + 1.233Wh + 35.945T + 0.0016M^2 - 0.0032Wh^2 - 4.6452T^2 - 0.0032MWh - 0.0521MT + 0.2165WhT) \quad (6.13)$$

To maximise predicted average net revenue per week, $\hat{R}*$, we want:

$$d\hat{R}* = \frac{\delta\hat{R}*}{\delta M}.dM + \frac{\delta\hat{R}*}{\delta Wh}.dWh + \frac{\delta\hat{R}*}{\delta T}.dT = 0 \quad (6.14)$$

$$\text{and} \quad d\hat{R}*^2 < 0 \quad (6.15)$$

Equation (6.14) defines the stationary point of the revenue function (6.13). If (6.15) holds, the stationary point is a maximum. To calculate the co-ordinates of the stationary point we could solve the set of simultaneous equations for M, Wh, and T, given by:

$$\frac{\delta\hat{R}^*}{\delta M} = 0$$

$$\frac{\delta\hat{R}^*}{\delta Wh} = 0$$

$$\frac{\delta\hat{R}^*}{\delta T} = 0$$

If the stationary point so found was feasible we could check for a maximum. If the stationary point was not feasible, or was not a maximum (i.e. a saddle point), steps would have to be taken to maximise (6.13) subject to suitable side restrictions, as was illustrated in 6.1.2 and Appendix VII. In either event, a single meal, whey time (M^*, Wh^*, T^*) combination, that maximises predicted average net revenue per week, $\hat{R}^*$, is obtained.

6.2 Weekly Incremental Model

The way in which the weekly incremental model allows us to predict feeding rations that maximise net revenue over a period of time, and per pig, will now be illustrated. The model which will be used for this purpose is one neglecting scouring and where whey consumption has been measured in terms of pounds dry matter (see Table IX, Ch.5). Thus we have:

$$\begin{aligned}\hat{y} = & - 0.60813 - 0.06485Wt + 0.63005m + 0.78828wh \\ & + 0.000038Wt^2 - 0.004973m^2 - 0.015279wh^2 \\ & - 0.000027Wtm + 0.001267Wtwh - 0.004395whm \qquad (6.16)\end{aligned}$$

where the i subscripts have been suppressed.

The mean drymatter content of whey used in the meal:whey levels trial, from which (6.16) was estimated, was 6.2%. Thus, 1 gallon of whey (approx. 10 lbs) contains approx. 0.62 lbs D.M., and whey at 1.5d/gallon gives a price of 2.4d/lb D.M.

Predicted net revenue per week, $\hat{r}$, for a single pig is therefore given by:

$$\hat{r} = 15.8\hat{y} - 3.4m - 2.42wh \qquad (6.17)$$

Substituting in (6.17) for $\hat{y}$ we have:

$$\hat{r} = -9.60845 - 1.02463Wt + .00600Wt^2 + m(6.55479 - .00043Wt) + wh(10.03482 + .02002Wt) - .07857m^2 - .24141wh^2 - .06944mwh \qquad (6.18)$$

The quantity of meal and whey to feed each week that maximises predicted net returns per week from a single pig is given by solving simultaneously:-

$$\frac{\delta\hat{r}}{\delta m} = 0$$

$$\frac{\delta\hat{r}}{\delta wh} = 0$$

where $d\hat{r}^2 < 0.$

From (6.18) we have:

$$\frac{\delta\hat{r}}{\delta m} = 6.55479 - .00043Wt - .15715m - .06944wh = 0$$

$$\text{i.e.} \quad m = 44.71146 - .00271Wt - .44189wh \qquad (6.19)$$

and,

$$\frac{\delta\hat{r}}{\delta wh} = 10.03482 + .02002Wt - .48282wh - .06944m = 0$$

$$\text{i.e.} \quad m = 144.50850 + .28829Wt - 6.95290wh \qquad (6.20)$$

Solving (6.19) and (6.20) simultaneously we obtain:

$$\left.\begin{aligned} wh &= 15.78818 + .04469Wt \\ m &= 34.73482 - .02246Wt \end{aligned}\right\} \qquad (6.21)$$

Now, $$d\hat{r}^2 = \begin{bmatrix} dm, dwh \end{bmatrix} \begin{bmatrix} -.15715 & -.06944 \\ -.06944 & -.48282 \end{bmatrix} \begin{bmatrix} dm \\ dwh \end{bmatrix}$$

i.e. $d\hat{r}^2 < 0.$

Thus at any given liveweight, equations (6.21) allow us to predict the quantities of meal and whey that should be fed in the following week in order to maximise net returns in that week. In passing we note that these relationships specify a decreasing proportion of meal:whey in the optimum ration as liveweight increases.

Where pigs are fed according to (6.21), the incremental model may be used to maximise net revenue per pig. Substituting for m and wh from (6.21) in (6.18) we obtain:

$$\hat{r} = .00645Wt^2 - .72350Wt + 202.66805 \qquad (6.22)$$

From inspection of (6.22) we see that $\hat{r} = \infty$ when $Wt = \infty$. In addition, (6.22) equated to zero has no real roots; thus there is no real liveweight where net revenue per week equals zero.

These results indicate that to maximise net revenue per pig, animals should be fed according to equations (6.21) until they reach infinite liveweight! As pigs were only taken to 180 lb liveweight in the meal:whey levels trial, we may assume that they should be fed until they are at least this liveweight in order to maximise net revenue per pig.

Consideration of the optimum rates of feeding (6.21) for any liveweight in the range 50-180 lbs, as studied in the meal:whey levels trial, indicates that the meal feeding rate so obtained is far in excess of any used in this particular trial. For pigs of 150 lbs liveweight, the rates of feeding given by (6.21) are:

m = 31.4 lbs per week

wh = 22.5 lbs.D.M. = 36.3 gals. per week.

This rate of whey feeding has been achieved in the trial for 150 lb pigs, but the maximum rate of meal feeding at this liveweight was 7 lbs per week. In general, specified rates of whey feeding at low weights (50-60 lbs) are in excess of those used in the trial, the reverse being true at higher liveweights (175 lbs). The level of meal feeding specified is far in excess of any used in the trial. Even when the price of whey is reduced to zero, the level of meal feeding employed for maximum net returns per week is in excess of any level actually used. The danger (and, indeed, probability) is that these levels of feeding cannot be achieved because of stomach restrictions. If stomach capacity could be expressed as some function of liveweight and quantities of meal and whey that could be fed per week, it might be possible to maximise (6.18) subject to this restriction in order to obtain optimum feasible feeding rates.

In the absence of this information relating to stomach capacity, the best we can do is to compare rations actually used in this meal:whey levels trial. This comparison indicates that average net returns per week were a maximum for pigs fed on Treatment No.3, i.e. along ration line (3), fig.1, Ch.5; with an average figure of £0.267 per pig, per week. However, this analysis has necessarily been limited to the one particular rate at which each ration was fed in this trial. Thus, if the ration corresponding to Treatment No.1 could be fed in an average of 13 weeks rather than 16 weeks, without affecting adversely total liveweight gain from the same quantity of meal and whey, we could expect an average net revenue per week of approx. £0.296 rather than £0.240.

Production economics analysis based on the weekly incremental model cannot therefore be made entirely meaningful until stomach capacity relationships in the form suggested are available. There should be no real difficulties in obtaining this relationship from suitable experimental data.

6.3 Discussion of Production Economics Analyses

Subject to suitable restrictions on quantities of meal and whey that may be consumed we may make the following observations:

(a) Maximum net revenue per pig. Both the cumulative and weekly incremental models allow us to maximise net revenue per pig. The cumulative model gives the total meal and whey combination, and the total time in which this quantity of feed should be consumed in order to maximise net revenue per pig. The weekly incremental model allows us to calculate optimum weekly meal and whey feeding rates as a function of total liveweight at the beginning of the week, and the total liveweight at which net revenue per pig is a maximum. A direct comparison between the two approaches was not possible because of the unrealistic feeding rates specified by the incremental model.

Although both models allow us to maximise net revenue per pig, the operational difference between them is the amount of information obtained in each case. The cumulative analysis results in a single optimum meal, whey, time combination; the problem of what feeding schedule to use in order to attain this factor combination and predicted liveweight gain, still remains. The incremental model however defines a specific optimum feeding plan in terms of quantity of meal and whey to be fed each week. Operationally therefore, the incremental model would be favoured, subject to the specification of feasible feeding rates.

(b) Maximum net revenue over time. Where pig production is carried out on a continuous basis, the objective will be maximum net revenue over time rather than per litter of pigs.

The cumulative model may be modified in order to calculate the factor combination (M*,Wh*,T*) that maximises average net revenue per week, as discussed in 6.1.3. Where factor and product prices are stable, net revenue over time will be maximised by feeding successive litters so as to maximise average net revenue per week, as given by the above (M*,Wh*,T*) factor combination.

The weekly incremental model allows us to calculate the weekly meal and whey feeding rates that maximise net revenue per week. There is no simple way of calculating the period of time over which average net returns per week are maximised, from the incremental model. (This is because the model is discrete rather than continuous in time.) However, to maximise net revenue over a given number of weeks from a single pig, we should feed so as to maximise net revenue in each week.

The advantage the cumulative analysis has over the weekly incremental model in predicting the fattening period, T*, to maximise average net revenue per week, is largely nullified where factor and product prices vary. Consider the case where we wish to maximise net revenue from a pig over a four week period. We assume that the pig will be sold at the end of the 4th week at a price of P*d/lb. The prices for meal and whey during each week are:
P_{m1}, P_{m2}, P_{m3}, P_{m4}, and P_{w1}, P_{w2}, P_{w3}, P_{w4}, respectively.
Using P* as the price of pigmeat and P_{m1} and P_{w1} as prices of

meal and whey in the first week, we maximise predicted net revenue in this week. The optimum quantities of meal and whey to feed in the first week, and total liveweight at the beginning of this week allow us to predict weight gain for this week. This process is repeated for the three remaining weeks to obtain a feeding schedule for each week that maximises net returns for the four week period. This flexibility in the selection of optimum rations is not available from the cumulative type analysis.

It would appear likely that the cumulative and incremental analyses have some complementary features. Under conditions of stable prices the cumulative model could be used to specify the M*, Wh*, T*, combination that maximises net revenue over time, while the incremental model could be used to specify optimum quantities of meal and whey to be fed each week, subject to the restriction that total quantities M* and Wh* are consumed at the end of time T*. The existence of a complementary relationship in this respect has not been investigated in this study.

6.4 Linear Programming

Production economics analyses of the type illustrated in this chapter may be useful in making the sorts of short term decisions discussed in section 1.6.

In theory, linear programming should provide an ideal procedure for optimising intermediate term planning of piggery operations. However, where the main supply of feedstuff is subject to seasonal fluctuations (i.e. whey), and where a large number of alternative feeding plans could be specified even for one type of pigmeat production (i.e. bacon), the resulting computing burden can be substantial. Little effort is required to specify a linear programme for pigmeat production which exceeds the capacity of computing equipment currently available in New Zealand, as was illustrated in section 1.7.

If we know the value of meal and whey each week, and the price of pigmeat at the end of some specified production period, we may use the weekly incremental model to predict an optimum feeding schedule over this time period. A virtue of linear programming is that it imputes values to scarce resources, hence we have the possibility of an iterative approach to the problem. Optimum rations, predicted from the incremental model and based on zero price for whey and the market price for meal (we will assume no restrictions on the purchase of meal), could be incorporated in a linear programme that was within the capacity of available computing equipment. No specific allowance for 'buying meal' in the linear programme need be made as this would already be taken into account in the optimum rations, which would be specified in the linear programme in terms of whey requirements only. Net revenue for each activity would be adjusted for the quantity of meal used. The maximisation of net revenue for this problem will result in values being imputed to scarce resources. In particular values will be imputed to whey supply in each week. We may now use the weekly incremental model to predict optimum rations based on these values for whey in each subperiod. These rations should be optimum in the sense that they will maximise income for the given production period and under existing valuations of feed supplies.

Rations specified by this procedure would be substituted for the original set in the linear programme. This procedure could be repeated until a stable optimum was reached.

6.5 Conclusions

Production economics analysis of cumulative and incremental production models results in basically similar information in each case. However, information obtained from the weekly incremental model is more operational in that optimum feeding schedules are specified. In addition, the feasibility of using linear programming for intermediate term planning of piggery operations would seem to be considerably enhanced by an iterative approach using optimum feeding schedules specified by an appropriate incremental model.

APPENDIX I

REPORT ON MEAL AND WHEY LEVELS TRIAL 1960/61

Supplied by Mr. A.C. Dunkin, Senior Lecturer
in Pig Husbandry, Massey College.

Object: To study the effect of different scales of whey feeding each fed in conjunction with three different levels of meal supplementation on growth rate, efficiency of food conversion and carcass quality of pigs killed at bacon weight.

Trial Design: 3 x 3 factorial, with 3 pigs on each treatment in each of two successive seasons. In each season, 3 blocks, each consisting of nine littermates, were selected.

Animals Used: Weanling pigs out of 3 L.W. sows and sired by Berkshire boars were used:-

vis. ex Natalie, 0258 by Cicero, litter weaned at 6 weeks (pens 20-28)
ex Lois, 41/7 by Brk. boar, litter weaned at 6 weeks (pens 1-9)
ex Leila, 0458 by Cicero, litter weaned at 3 weeks (pens 10-19)

At 56 days, the weight of the pigs in these 3 litters were:-

ex Natalie	32-42 lb	(average 39.3 lb)
ex Lois	32.5-39 lb	(average 34.3 lb)
ex Leila	39.5-50 lb	(average 42.9 lb)

Pre-Experimental Treatment

1. The pigs from Natalie's and Lois' litters had been on "weaner" mixture fed at the rate of 1½ lb per pig daily plus whey from the 53rd day of age. At the 59th day, the daily meal allowance was reduced to 1¼ lb and the whey increased to 1 gal. per pig.

The selected pigs were moved to the Test House on the

62nd day. At this time the daily meal ration was changed to 1 lb barley meal plus ½ oz boneflour, while the whey was rationed according to scale B.

2. Because of their heavier weight the pigs from Leila's litter were moved to the Test House on the 55th day. Until then, they had been receiving 1½ lbs "weaner" meal mixture per pig daily plus whey. On installation in the Test House, for those pigs over 46 lb, the meal was changed immediately to 1 lb barley meal and whey according to scale B. The remainder continued on 1½ lb "weaner" meal for 3 days, and were then changed on 1 lb barley meal plus ½ oz boneflour.

<u>Experimental Treatments</u>

Treatment	lb barley meal daily	Whey according to Scale
I.	1½ ⟶ 1	A
II.	1½ ⟶ 1	B
III.	1½ ⟶ 1	C
IV.	1 ⟶ ½	A
V.	1 ⟶ ½	B
VI.	1 ⟶ ½	C
VII.	½ ⟶ nil	A
VIII.	½ ⟶ nil	B
IX.	½ ⟶ nil	C

Management: The whey feeding scales used are set out below.

WHEY SCALES (lbs. per day)

Liveweight	A	B	C
40	12	15	18
60	20	25	30
80	28	35	42
100	34	42	51
120	40	50	60
140	44	55	66
160	48	60	72
180	52	65	78

Pigs were placed on trial individually as they reached 48-50 lbs liveweight. Thereafter they were weighed twice weekly and their allowances of meal and whey were adjusted.

From the beginning of the third week after the first pig came on trial, any whey remaining in the trough at 8.00 p.m. on Sundays was removed. By this means it was ensured that no pigs had had any food for at least 10 hours prior to being weighed on Monday mornings.

Following the weighing at which pigs reached 84-85 lbs, the daily meal allowance was reduced by 2 oz. This was followed by further reductions of 2 oz every 5 days until the daily meal allowance had been reduced by a total of 8 oz, that is, the meal reduction was completed in 15 days.

Initially, whey was fed twice daily but as the capacity of the buckets containing the individual whey rations was exceeded this was increased to three feeds and eventually to four feeds daily.

From the start of the trial until 85 lb. liveweight, the full ration of whey was not always given in cases where pigs were scouring badly. Subsequent to this weight, however, the full ration was offered, regardless of the consistency of the dung.

Pigs remained on trial until at least 180 lb liveweight. On the last day before the final liveweighing and prior to dispatch to the bacon factory, the fourth feed of whey for the day was omitted and any whey remaining in the trough at 5.00 p.m. was removed. This ensured that the pigs were fasted for at least 13-14 hours before their final liveweighing.

APPENDIX II

VARIABLES USED AND THEIR NOTATION.
OBSERVATIONS MADE ON EACH OF 26 PIGS.

VARIABLE	NOTATION
Total liveweight gain from the start of the trial to the end of the ith week (lbs). Thus, $Y_i = Wt_i - Wt_o$; where Wt_i = total liveweight at the end of the ith week, and Wt_o = liveweight at the beginning of the trial.	Y_i
Total meal consumed from start of trial to end of ith week (lbs).	M_i
Total whey consumed from start of trial to end of ith week (gals).	Wh_i
Total whey consumed from start of trial to end of ith week (lbs Dry Matter).	$Wh_{D.M.i}$
Number of weeks from the start of the trial.	T
Total number of days on which scouring was observed to the end of the ith week.	S_i
Liveweight gain in the ith week: $y_i = Y_i - Y_{i-1}$	y_i
Total liveweight at the beginning of the ith week (end of the i-1th week).	Wt_{i-1}
Meal consumed in the ith week (lbs).	m_i
Whey consumed in the ith week (gals).	wh_i
Whey consumed in the ith week (lbs Dry Matter).	$wh_{D.M.i}$
Number of days on which scouring was observed in the ith week.	s_i

Note: Y_i is liveweight gain from the start of the trial to the end of the ith week, and y_i is liveweight gain from the end of the i-1th week to the end of the ith week. Thus when we talk about meal or whey fed in the ith week or weight gain in the ith week, the period referred to is from the end of the i-1th week (beginning of the ith week) to the end of the ith week.

The cumulative model:

$$Y = f(M, Wh, T, S)$$

may be interpreted as follows: Expected liveweight gain, Y, in any period, T, is a function of total meal, M, and whey, Wh, consumption, the time taken to consume this feed, T, and the total number of days on which scouring was observed in this period.

APPENDIX III

(A) Total liveweight gain (Y); Total meal (M) and whey ($Wh_{D.M.}$) consumption data after 11 weeks for 27 pigs.[1]

LITTER NUMBER	VARIABLES			TREATMENT
	Y(lbs)	M(lbs)	$Wh_{D.M.}$(lbs)	
1	81	96	154	I
2	93	95	160	
3	86	93	161	
1	96	93	197	II
2	98	95	196	
3	103	94	199	
1	102	92	252	III
2	114	91	265	
3	82	99	137	
1	70	60	141	IV
2	68	61	137	
3	66	62	136	
1	77	63	176	V
2	83	58	190	
3	84	58	192	
1	94	58	229	VI
2	98	57	227	
3	87	58	219	
1	50	43	126	VII
2	52	26	129	
3	51	28	127	
1	69	26	167	VIII
2	64	25	170	
3	72	23	179	
1	68	23	208	IX
2	76	25	210	
3	59	26	188	

(1) Data from Pig No.23 was available at 11 weeks although this pig did not complete the trial.

(B) Number of days on which scouring was observed for each pig over the entire trial period (approx. 48-180 lb liveweight range).

TREATMENT

BLOCK	I	II	III	IV	V	VI	VII	VIII	IX
1	3	0	7	3	35	4	2	72	111
2	0	25	0	0	5	14	1	27*	34
3	0	16	20	5	33	41	1	39	148

*Pig No.23: Estimated number of days scouring given by technique for handling missing data in randomised block designs: G. W. Snedecor "Statistical Methods" 1957, p.310.

APPENDIX IV

(A) Summary of TRAP - Multiple Regression Analysis Programme for IBM 650 Calculator with Magnetic Tape Programming(1)

Author of Programme: J.E. Nichols(2)
Additional Participants: R.Y. Seaber and R.A. Stewart.

TRAP is a generalised, flexible, regression analysis programme for the augmented Type IBM 650 computer with magnetic tape. The mathematical procedure is standard Least Squares method. The equations finally solved by the computer are linear although many non-linear and transcendental functions may be included. This is accomplished by transforming the data as it is originally fed to the computer. Twenty-six transformations are provided for by TRAP. Examples of these for the variable x are:

$$x + a,\ x - a,\ ax,\ \frac{a}{x},\ e^x,\ x^a,\ \mathrm{Log}\ x.$$

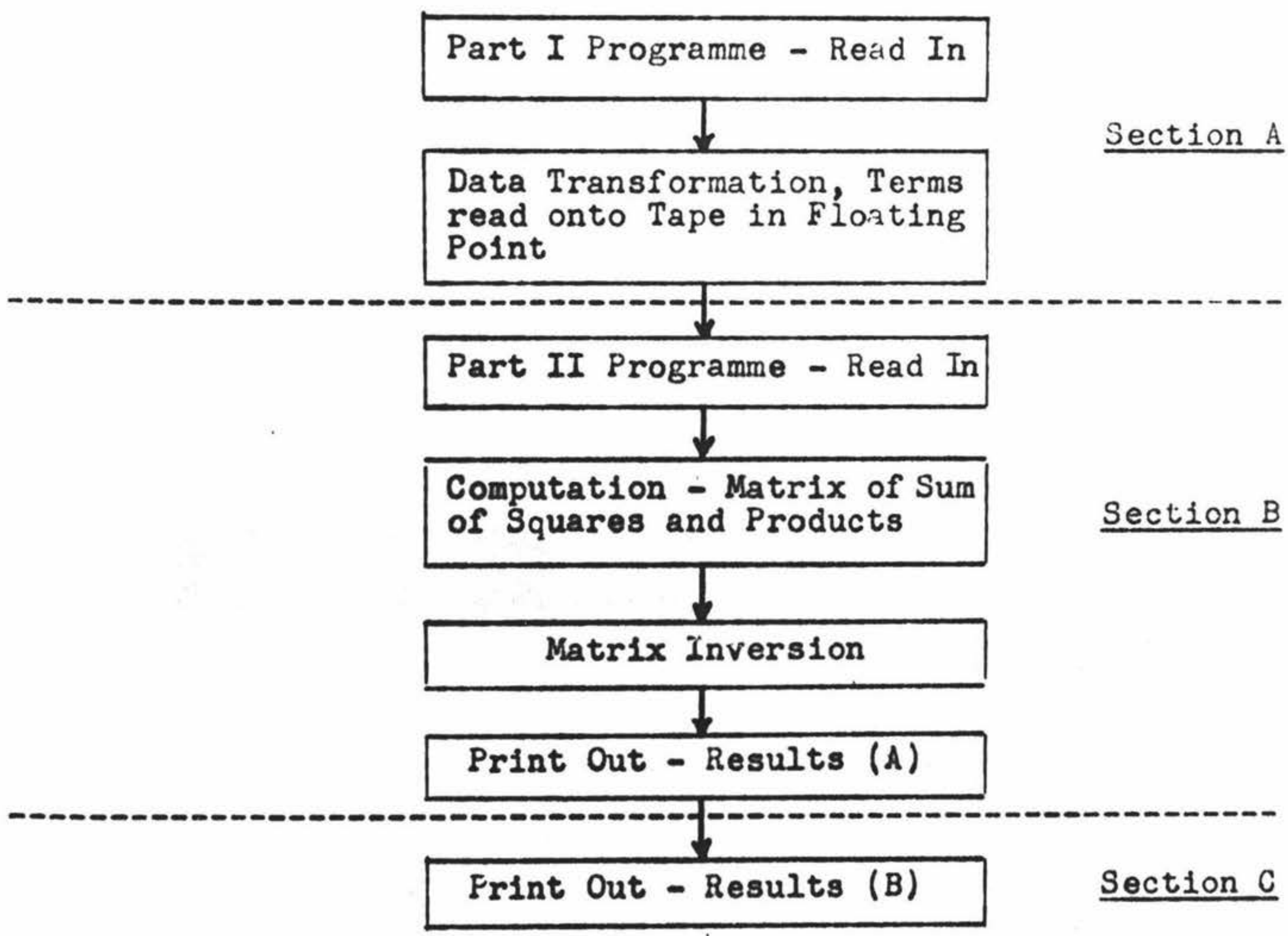
The programme is divided into two parts. The input for Part I is the original data which is transformed as required and expressed in floating decimal point form on magnetic tape. This output is then the input for Part II. The Part II output is the desired result. The original data can consist of up

---

(1) TRAP was made available from the Applied Mathematics Laboratory's library of programmes.

(2) Of the Shell Oil Company, Houston Research Laboratory.

to 32 variables for each observation, any of which may be positive or negative. Up to nine dependent variables may be correlated using the same function in one machine operation, provided there are no more than twenty-six terms, including dependent variables, in the equation to be fitted. Alternative relationships between the original variables may be obtained by separate machine runs. The dependent and independent variables, the transformations to be used, and the terms of the fitted equation for a particular machine run, are given in the first fourteen input cards. A maximum of 999 observations is provided for by TRAP. A modified flow chart for TRAP is now given:



Results (A) consist of:

(a) Problem identification.

(b) The original matrix of sum of squares and products.

(c) The inverse matrix.

(d) Values of the regression coefficients.

(e) The total variation for each dependent variable.

(f) Variation by regression for each dependent variable.

(g) Value of R^2 for each dependent variable.

(h) Degrees of freedom.

(i) F-test for each term.

(j) t-test for each term.

Results (B) are optional and consist of:

(a) A table of the observed and calculated values for each dependent variable and their differences.

Sections A, B and C require approximately equal machine time for completion.

The weekly incremental model (491 observations):

$$y_i = f(Wt_{i-1}, m_i, wh_i, s_i)$$

estimated as a second order polynomial, required approx. 2 hours to complete all three sections of TRAP. Considerable machine time was therefore saved by foregoing section C (or Results (B)) of TRAP, as was done for the majority of functions estimated in this study.

(B) Data Availability

The data cards No.15 onwards, contain the observations of the variables used in the study. Each card may contain information on up to 16 variables, at the rate of 5 digits per variable. The first 4 digits give significant figures of the variable concerned while the last digit indicates the position of the decimal point. Thus 3.0 may be represented as -

30003, 03002, 00301, or 00030.

Variables are numbered 1 to 16 consecutively according to their position on the card.

Two sets of data cards, one each for the weekly and fortnightly models are available from the Department of Agricultural Economics and Farm Management, Massey University College, on request. The layout of these cards is given in the following two tables:

TABLE I : Weekly Information (491 cards)

Variable No.	Variable Identification
1	Identification Code
2	Y_i
3	y_i
4	m_i
5	M_i
6	$wh_{D.M.i}$
7	$Wh_{D.M.i}$
8	wh_i
9	Wh_i
10	s_i
11	S_i
12	Wt_{i-1}
13	T
14	-
15	-
16	-

TABLE II : Fortnightly Information (239 cards)

Variable No.	Variable Identification
1	Identification Code
2	y_i
3	Wt_{i-1}
4	m_i
5	wh_i
6	s_i
7	Y_i
8	M_i
9	Wh_i
10	S_i
11	T
12	-
13	-
14	-
15	-
16	-

The identification code 04121 is interpreted as follows:

04 - fourth pig;

12 - the twelfth observation.

The last digit is ignored.

APPENDIX V

(A) Details of Cumulative model estimated in the form $Log\hat{Y} = a + b_1 LogM + b_2 LogWh + b_3 LogT + b_4 LogS$.

TERM	Regression Coefficient	t-test (486 d.f.)
Constant	-0.216	
LogM	0.323	35.39***
LogWh	0.668	22.28***
LogT	-0.747×10^{-1}	2.09*
LogS	0.958×10^{-3}	0.19 NS

$R^2 = 0.9813$

(B) Details of Cumulative Model estimated in the form:

$$\hat{Y} = a + b_1\sqrt{M} + b_2\sqrt{Wh} + b_3\sqrt{T} + b_4\sqrt{S} + b_{11}M + b_{22}Wh$$
$$+ b_{33}T + b_{44}S + b_{12}\sqrt{MWh} + b_{13}\sqrt{MT} + b_{14}\sqrt{MS}$$
$$+ b_{23}\sqrt{WhT} + b_{24}\sqrt{WhS} + b_{34}\sqrt{TS}$$

TERM	Regression Coefficient	t-test (476 d.f.)
Constant	-20.998	
$\sqrt{M}$	-2.556	3.88***
$\sqrt{Wh}$	-1.271	1.22 NS
$\sqrt{T}$	37.230	6.03***
$\sqrt{S}$	0.124	0.14 NS
M	0.557	9.08***
Wh	-0.637	7.05***
T	-30.553	8.83***
S	-0.506	7.68***
$\sqrt{MWh}$	-0.183	1.79 10%
$\sqrt{MT}$	0.814	1.31 NS
$\sqrt{MS}$	-0.124	1.48 NS
$\sqrt{WhT}$	9.088	8.39***
$\sqrt{WhS}$	0.409	3.21**
$\sqrt{TS}$	-1.477	2.09*

$R^2 = 0.9891$

A P P E N D I X   V I

TABLE I : Fortnightly Incremental Model - Period I.

TERM	Regression Coefficient	t-Test (61 d.f.)
Constant	-19.470	
Wt_{i-1}	0.576	1.84 10%
m_i	0.558	1.44 NS
wh_i	0.256	1.06 NS
s_i	0.769	0.51 NS
Wt_{i-1}^2	-0.108×10^{-1}	3.15**
m_i^2	0.558×10^{-2}	0.67 NS
wh_i^2	-0.155×10^{-1}	3.57***
s_i^2	-0.944×10^{-3}	0.01 NS
$Wt_{i-1}m_i$	0.560×10^{-4}	0.01 NS
$Wt_{i-1}wh_i$	0.203×10^{-1}	2.84**
$Wt_{i-1}s_i$	-0.680×10^{-1}	1.05 NS
$m_i wh_i$	-0.586×10^{-2}	0.95 NS
$m_i s_i$	0.298×10^{-1}	0.37 NS
$wh_i s_i$	0.719×10^{-1}	1.09 NS

$R^2 = 0.829$

TABLE II : Fortnightly Incremental Model - Period II.

TERM	Regression Coefficient	t-Test (67 d.f.)
Constant	14.710	
Wt_{i-1}	-0.505	1.46 NS
m_i	0.226	0.46 NS
wh_i	0.616	2.21*
s_i	0.123	0.16 NS
Wt_{i-1}^2	0.281×10^{-2}	1.67 10%
m_i^2	-0.194×10^{-4}	0 NS
wh_i^2	0.435×10^{-4}	0.01 NS
s_i^2	-0.285×10^{-1}	0.97 NS
$Wt_{i-1}m_i$	0.467×10^{-2}	1.10 NS
$Wt_{i-1}wh_i$	-0.315×10^{-2}	0.97 NS
$Wt_{i-1}s_i$	0.446×10^{-2}	0.67 NS
$m_i wh_i$	-0.442×10^{-2}	0.75 NS
$m_i s_i$	-0.514×10^{-2}	0.23 NS
$wh_i s_i$	-0.724×10^{-2}	0.53 NS

$R^2 = 0.847$

TABLE III : Fortnightly Incremental Model - Period III

TERM	Regression Coefficient	t-Test (66 d.f.)
Constant	-44.112	
Wt_{i-1}	0.392	0.58 NS
m_i	0.136	0.20 NS
wh_i	0.582	1.34 NS
s_i	0.168	0.16 NS
${Wt_{i-1}}^2$	-0.133×10^{-2}	0.58 NS
${m_i}^2$	-0.501×10^{-2}	0.43 NS
${wh_i}^2$	-0.120×10^{-2}	0.35 NS
${s_i}^2$	0.252×10^{-3}	0.01 NS
$Wt_{i-1}m_i$	0.751×10^{-2}	1.68 10%
$Wt_{i-1}wh_i$	-0.102×10^{-2}	0.34 NS
$Wt_{i-1}s_i$	0.839×10^{-2}	1.35 NS
$m_i wh_i$	-0.962×10^{-2}	1.96 10%
$m_i s_i$	-0.315×10^{-2}	0.19 NS
$wh_i s_i$	-0.170×10^{-1}	1.80 10%

$R^2 = 0.745$

APPENDIX VII

MAXIMISATION OF A QUADRATIC FUNCTION SUBJECT TO LINEAR INEQUALITIES

From chapter 6 we may rewrite the linear inequalities (6.9) to (6.12) as equalities thus:

$$-25 = x_3 \qquad\qquad - x_1 \tag{7'.1}$$

$$-465 = x_4 \qquad - 1.8715x_1 - x_2 \tag{7'.2}$$

$$655 = x_5 + 1.8715x_1 + x_2 \tag{7'.3}$$

$$40 = x_6 + x_1 - .21314x_2 \tag{7'.4}$$

We wish to maximise

$$\hat{R} = 8.4530 + 3.2360x_1 + 2.0708x_2 + .00158x_1^2 - .00063x_2^2 - .00442x_1x_2$$

subject to restraints (7'.1) to (7'.4), and

$x_j \geqslant 0 \quad (j = 1,2, \ldots, 6).$

In solving this problem we may conveniently utilize the well known simplex layout for solving linear programmes.[1] Table I gives the first tableau.

TABLE I : Statement of Linear Restraints

	B	x_3	x_4	x_5	x_6	x_1	x_2
x_3	-25	1				-1	0
x_4	-465		1			-1.8715	-1
x_5	655			1		1.8715	1
x_6	40				1	1	-.21314

(1) "Linear Programming Methods", E.O. Heady and W.V. Candler, Iowa State College Press, Ames, Iowa, 1958, Ch.3, p.61.

Both x_3 and x_4 are at negative levels violating the condition $x_j \geqslant 0$. The first step therefore is to obtain any feasible solution. This may be achieved by bringing x_1 and x_2 into the plan to give Table II.

TABLE II : First Feasible Plan

	B	x_3	x_4	x_5	x_6	x_1	x_2
x_3	103.39229	1		.15236	.71485		
x_4	760.83250		1	1	0		
x_2	414.71383			.71486	-1.33784		1
x_1	128.39229			.15236	.71485	1	

The value of $\hat{R}$ at this point (A in fig.1,ch.6) is 964.85163. We now express income as a function of each of the non-basic variables in turn. Each unit of x_6 supplies 1.33784 units of x_2 and requires .71485 units of x_1. With x_1 and x_2 at the levels shown we may write therefore:

$$\begin{aligned}\hat{R} = 8.45300 &+ 3.23600(128.39229 - .71485x_6) \\ &+ 2.07080(414.71383 + 1.33784x_6) \\ &+ .00158(128.39229 - .71485x_6)^2 \\ &- .00630(414.71383 + 1.33784x_6)^2 \\ &- .00442(128.39229 - .71485x_6)(414.71383 + 1.33784x_6)\end{aligned}$$

which simplifies to

$$\hat{R} = 964.85163 + .01967x_6 + .00229x_6^2.$$

Now $\frac{d\hat{R}}{dx_6} = .01967 + .00458x_6 = 0$

gives $x_6 = -4.29480$; while $\frac{d^2\hat{R}}{dx_6^2} = .00458$.

Thus for x_6, predicted net revenue, $\hat{R}$, is minimised at $x_6 = -4.29480$. The present level of x_6 is zero; we can increase income therefore by making x_6 as large as possible. This is done by replacing x_3 with x_6 in Table II to give Table III.

TABLE III : Reintroduction of x_6 into the Basis

	B	x_3	x_4	x_5	x_6	x_1	x_2
x_6	144.63490	1.39889		.21314	1		
x_4	760.83250	0	1	1			
x_2	608.21218	1.37149		1			1
x_1	25.00000	-1		0		1	

The value of predicted net revenue, $\hat{R}$, at this point (B in fig.1,ch.6), is 1049.972. Net revenue cannot be increased by reintroducing x_3 as we would only replace x_6. Revenue is now expressed in terms of the remaining non basic variable x_5, as was done previously for x_6, to give

$$\hat{R} = 1049.972 - 1.19385x_5 - .00063x_5^2$$

now $$\frac{d\hat{R}}{dx_5} = -1.19385 - .00126x_5 = 0$$

gives $$x_5 = -947.5; \quad \text{while } \frac{d^2\hat{R}}{dx_5^2} = -.00126.$$

Thus revenue is maximised in the x_5 direction by making $x_5 = -947.5$. However, $x_5 \geqslant 0$, so to maximise revenue we make x_5 as small as possible, i.e., $x_5 = 0$. As x_5 is at zero level in Table III, we are at a local optimum.

The next step is to express predicted net revenue as a function of all non basic variables, i.e.,

$$\hat{R} = 1049.972 - 1.61006x_3 - 1.19385x_5 + .00765x_3^2 - .00063x_5^2 + .00206x_3x_5.$$

Now

$$\frac{d\hat{R}}{dx_3} = -1.61006 + .01530x_3 + .00206x_5$$

$$\frac{d\hat{R}}{dx_5} = -1.19385 - .00126x_5 + .00206x_3.$$

It can easily be shown that the local optimum already found is the maximum maximorum in the region given by the original constraints plus the additional restraints:

$$.01530x_3 + .00206x_5 \leqslant 2(1.61006) \qquad (7'.5)$$

$$.00206x_3 - .00126x_5 \leqslant 2(1.19385) \qquad (7'.6)$$

We wish then to explore the remainder of the original region of interest. In order to restrict ourselves to this region we cannot simply reverse the sense of the inequalities (7'.5) and (7'.6) and add them to the original tableau. Candler and Townsley have shown that a reasonable restriction which ensures neither of the restraints (7'.5) and (7'.6) are violated can easily be obtained. In our example we first set $x_5 = 0$, and solve (7'.5) and (7'.6) for x_3, i.e.,

$$x_3 \leqslant 210.465$$

$$x_3 \leqslant 1156.830.$$

The maximum feasible value of x_3 is therefore 210.465. Now setting $x_3 = 0$ and solving for x_5 we have:

$$x_5 \leqslant 1560.136$$

$$x_5 \leqslant +\infty$$

Thus, the maximum feasible value of x_5 is 1560.136. The desired restriction is then:

$$\frac{x_3}{210.465} + \frac{x_5}{1560.136} \leqslant 1$$

or $$7.4128x_3 + x_5 \leqslant 1560.1360 \qquad (7'.7)$$

This restraint may easily be rewritten in terms of x_1 and x_2 to give:

$$5.54x_1 - x_2 \leqslant 2029.81 \qquad (7'.8)$$

Addition of (7'.8), with the inequality reversed, to the original restraints gives us a new problem with a smaller set of feasible values. In fact, when this is done, no feasible solution can be obtained, hence the solution given by Table III was the maximum maximorum for the initial area of interest, i.e.,

$\hat{R}$ = 1049.972 pence (£4.375) per pig.

M = 25.0 lbs.

Wh = 608.212 gals.

APPENDIX VIII

A REVIEW OF LITERATURE ON THE SELECTION OF EXPERIMENTAL DESIGNS FOR RESPONSE SURFACE ESTIMATION

Because regression analysis can be applied to almost any scatter of observations, and some results obtained, little interest has been shown until relatively recently in experimental designs especially suited for regression studies. In 1951 a fundamental article by Box and Wilson[1] considered the selection of experimental designs so that regression estimates from the resulting observations had certain desirable properties. The derivation of experimental designs according to desired properties of the regression equation fitted to these experimental points has been developed by Box et.al. A review of these papers is now given. Liberal use has been made of the terminology and notation from these papers; the numbers indicating references in the text refer to the publications listed at the end of this appendix.

8'.1 Introduction

Suppose we have k variables whose levels are denoted by $X_1, X_2, \ldots, X_k$, on which depends the expected level of some response, η, in accordance with an unknown relationship:

$$\eta = f(X_1, X_2, \ldots, X_k). \qquad (8'.1)$$

Suppose that in order to explore this relationship, N experiments are performed. The uth of these experiments consists in adjusting the factor levels to a certain set of k predecided values, X_{1u}, X_{2u}, ..., X_{ku}, and of observing a response Y_u.

8'.1.1 The Problem

The problem of experimental design discussed is that of choosing the N sets of levels at which the observations are to be made.

8'.1.2 Notation

A set of standardised levels for the variables is defined:

$$x_{iu} = \frac{(X_{iu} - \bar{X}_i)}{S_i}, \text{ where } S_i = \left\{ \sum_{u=1}^{N} \frac{(X_{iu} - \bar{X}_i)^2}{N} \right\}^{\frac{1}{2}} \qquad (8'.2)$$

For these standardised levels therefore:

$$\sum_{u=1}^{N} x_{iu} = 0 \quad \text{and} \quad \sum_{u=1}^{N} x_{iu}^2 = N \qquad (8'.3)$$

It is often convenient to view the problem geometrically and to regard (8'.1) as defining a surface, referred to as the response surface. We assume that in the limited region of immediate interest, (8'.1) can be represented by a polynomial of degree d, so that the expected response at the uth point, η_u, in terms of the standardised variables is assumed to be

$$\eta_u = \beta_0 x_{ou} + \beta_1 x_{1u} + \dots + \beta_k x_{ku} + \beta_{11} x_{1u}^2 + \dots + \beta_{kk} x_{ku}^2 + \dots + \beta_{12} x_{1u} x_{2u} + \dots + \beta_{k-1,k} x_{k-1,u} x_{ku} + \beta_{111} x_{1u}^3 + \dots \text{ etc.} \qquad (8'.4)$$

We can obtain Least Squares estimates b_o, b_1, etc. of the coefficients β_0, β_1, etc. by fitting (8'.4) by regression to the N observed values: $Y_1, Y_2, \ldots, Y_u, \ldots, Y_N$, (provided of course, that N is sufficiently large).

A design which includes k variables and allows us to determine all constants up to order d, will be called a k-dimensional design of order d. In a polynomial equation of degree d there are $\binom{k+d}{d}$ terms, so that for a k-dimensional design of order d, the number of experimental points, N, must be at least $\binom{k+d}{d}$. If only $\binom{k+d}{d}$ points are observed, then (8'.4) will fit exactly. To obtain an estimate of error we need at least $\binom{k+d}{d} + 1$ observations.

8'.1.3 Requirements

Which properties of a design are considered to be most important, or desirable, is bound to be to some extent subjective. Box and Hunter[4] list the following 'desirable' properties of an experimental design of order d:-

(a) The design should allow the approximating polynomial of degree d (tentatively assumed to be representationally adequate) to be estimated with satisfactory accuracy within the region of interest.

(b) The design should allow a check to be made on the representational accuracy of the assumed polynomial.

(c) The design should not contain an excessively large number of experimental points.

(d) The design should lend itself to 'blocking'.

(e) The design should form a nucleus from which a satisfactory design of order d + 1 can be built in case the assumed degree of polynomial proves inadequate.

For any <u>linear model</u>, such as (8'.4) in which there are L unknown coefficients, the N equations of the N experimental points may be written in matrix notation as:

$$\eta' = X_i \beta'_i \qquad (8'.5)$$

where i denotes the order of the polynomial, and where:

X_i is an NxL matrix, called the matrix of independent variables, giving the levels of the independent variables at successive observations, and

β'_i is a Lx1 vector of population parameters, $\beta_o, \beta_1, \ldots, \beta_L$, and η' is a Nx1 vector of expected values of the dependent variable.

The NxL matrix X_i provides a programme of the N experiments to be performed and is called the <u>design matrix</u>, and is denoted by D.

If the observed values found at the N experimental points of a particular design D, are represented by a vector $\underline{y}$, and $E(\underline{y}) = \eta$, Least Squares estimates $\underline{b}_i$ of β_i are given by:

$$\underline{b}'_i = (X'_i X_i)^{-1} X'_i \underline{y}' \qquad (8'.6)$$

The vector of predicted values of the dependent variable, $\hat{\underline{y}}'$, is given by:

$$\hat{\underline{y}}' = X_i \underline{b}'_i \qquad (8'.7)$$

The variances and co-variances of the Least Squares estimates are the elements of the matrix:

$$E(\underline{b}_i - \beta_i)(\underline{b}_i - \beta_i)' = (X'_i X_i)^{-1} \sigma^2 \qquad (8'.8)$$

(where σ^2 is the population variance parameter).

... tions, including the

supposition that the mathematical model (8'.5) exactly represents the true situation, the method of Least Squares gives unbiased estimates, $\underline{b}_i$, of β_i, (i.e. $E(\underline{b}\ ) = \beta_i$), and an unbiased estimate of $(N-L)\sigma^2$, provided by:

$$(\hat{\underline{y}} - \underline{y})'(\hat{\underline{y}} - \underline{y}) = \underline{y}'\underline{y} - \underline{b}_i'X_i'X_i\underline{b}_i \qquad (8'.9)$$

The method of Least Squares reduces to a minimum the sums of squares of deviations, $(\hat{\underline{y}} - \underline{y})'(\hat{\underline{y}} - \underline{y})$, between the observed values, $\underline{y}$ and the values, $\hat{\underline{y}}$, given by the fitted function (8'.7).

It is also convenient to consider here the case where, contrary to supposition, the mathematical model (8'.5) is inadequate and in fact L_j further terms, $X_j\beta_j$, are needed to ensure an adequate representation of response so that:

$$\eta = X_i\beta_i + X_j\beta_j \qquad (8'.10)$$

then the estimates given by (8'.6) are biased, for:

$$E(\underline{b}_i) = \beta_i + A\beta_j \qquad (8'.11)$$

where $A = (X_i'X_i)^{-1}X_i'X_j$ is an LxL_j matrix of bias coefficients which has been called the alias matrix.(1) In this situation the residual sum of squares is also biased.

The problem of choosing a "best" design for the fitting of a model given by (8'.5) has usually been interpreted as that of satisfying the requirement that D should be so chosen that the coefficients β_i are separately estimated with minimum variance.

For a particular design D, the method of Least Squares gives estimates having smallest variances. Different designs can have, of course, different variance estimates that are minimum for each design. Amongst all possible designs then, we wish to select the one(s) that give estimitates that are minimum minimorum variance estimates. In reference (2) a theorem is proved (for the case where the variables in the matrix X_1 are functionally independent and the diagonal elements of $X_1'X_1$ are fixed by the definition of the problem) that the requirement of minimum variance is satisfied by so choosing D that the matrix $X_1'X_1$ is diagonal. Such an arrangement may be called an <u>orthogonal design</u>.

In the present context it is only in the case of designs of first order that the variables are functionally independent and that all the diagonal elements of $X_1'X_1$ are fixed by the definition of the problem. For this reason the obove theorem is directly helpful only in the derivation of first order designs.

8'.2 Multifactor Designs of First Order

Suppose the true regression plane in the region considered is:

$$\eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \tag{8'.12}$$

As before the N equations (8'.12) may be written in matrix notation:

$$\eta' = X_1 \beta_1'$$

where $X_1 = [\underline{u}' \vdots D]$ and each element of the column vector $\underline{u}'$ is unity. Now, if $\underline{y}'$ is a Nx1 column vector of observations made at the N experimental points, and providing X_1 is of rank k+1 (which implies D of rank K), then separate linear estimates b_i of each of the β_i's may be calculated.

The variances and covariances for the estimates, b_i, are given by the matrix:

$$(X_1'X_1)^{-1}\sigma^2. \qquad (8'.13)$$

In order to minimise the variances of the estimates, b_i, we need to choose the design matrix D, so that the diagonal elements of $(X_1'X_1)^{-1}$ are minimised (subject of course to the standardisation restrictions (8'.3)).

Where the b_i are all mutually uncorrelated we have:

$$(X_1'X_1)^{-1} = \begin{bmatrix} \sum_{u=1}^{N} x_{1u}^2 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \sum_{u=1}^{N} x_{2u}^2 & & & & \\ \cdot & & \cdot & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \cdot & \\ 0 & & & & & \sum_{u=1}^{N} x_{ku}^2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} N & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & N & & & & \\ \cdot & & \cdot & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \cdot & \\ 0 & & & & & N \end{bmatrix}^{-1}$$

$$= N^{-1}I$$

Hence, the variance-covariance matrix of the b_i, (8'.13) is given by:

$$(X_1'X_1)\sigma^2 = N^{-1}I\sigma^2$$

and, $\quad Var(b_i) = N^{-1}\sigma^2$ for all i

and, $\quad Cov(b_i, b_j) = 0$ for $i \neq j$.

Since the variables $x_1, x_2, \ldots, x_k$ are functionally independent, and since $\sum_{u=1}^{N} x_{iu}^2 = N$ $(i = o, 1, 2, \ldots, k)$, the diagonal elements of $X_1'X_1$ are fixed by definition of the problem.

The smallest variance theorem(2) referred to in 8'.1.3, therefore leads at once, to the conclusion that a best design matrix, D, is one given by $N^{\frac{1}{2}}O$, where O is an orthogonal matrix with the elements of its first column all equal. Thus we have $X = N^{\frac{1}{2}}O$, and $X_1'X_1 = NI$.

Therefore first order designs of optimum precision for up to $k = N-1$ factors in N experiments may be obtained from <u>any</u> orthogonal matrix O with elements in the first column all equal, and $X_1 = N^{\frac{1}{2}}O$. These designs fall within the class described in 8'.1.3 as orthogonal designs.

Geometrically these designs consist of N points, at the vertices of an N-1 dimensional regular simplex if $k = N-1$, or the projections onto a space of k dimensions of the vertices of the N-1 dimensional regular simplex if $k < (N-1)$. The arbitrariness in the choice of D corresponds to the fact that the simplex may be taken in any orientation, i.e. it is a feature of orthogonal first order designs that under rotation the variances and covariances remain constant. This class of designs includes the factorials and fractional factorials. These latter designs are of special value because they are easy to carry out, they allow the adequacy of the first degree representation to be checked and the nature of departures from it to be readily identified, they form natural nuclei which can be augmented to form designs of higher order, and they are readily arranged

The construction and properties of such designs are discussed in reference (2).

8'.2.1 Bias From Higher Order Terms

Although the variances and covariances under orthogonal rotation of the design remain constant, the magnitude and arrangement of the possible biases which might occur due to the inadequacy of the planar approximation are affected by the orientation of the design. If, however, we have no prior knowledge concerning the relative importance of particular second-order terms, no arrangements which are dramatically worse or better than others can be expected to arise as a result of rotation of the designs.

When, on the other hand, something is known of the type of approximating second-degree equation to be expected, it may be possible to reduce bias by suitable rotation of the design. The particular case of designs which are such that only b_o is biased by quadratic terms ('Type B' designs) are discussed by Box, in reference (2), as examples of designs where suitable rotation gives unbiased estimates of effects: $b_1, b_2, \ldots, b_k$.

8'.3 Second Order Designs

For designs of order higher than the first, the quantities x_o; $x_1, \ldots, x_k$; $x_1^2, \ldots, x_k^2$; $x_1x_2, \ldots, x_{k-1}x_k$; x_1^3, etc. are not all functionally independent and hence a diagonal matrix of sums of squares and products for the 'independent' variables is impossible since, unless the x_{iu} are all zero, certain sums of products such as those between x_i^2 and x_o, and between

x_i^2 and x_j^2 are necessarily positive.

Box and Hunter [4] show that orthogonal second order designs of a sort can be obtained. Examples of such designs are the factorials with more than two levels and the orthogonal composite designs given by Box and Wilson[1].

Box and Hunter also note that $N^{-1}X'X$ may be viewed as a matrix of moments of the design. The condition of orthogonality (i.e., that the moment matrix is diagonal) may result in the choice of some moments that are not necessarily good. In the sense that we have chosen an orthogonal design we are associating the term 'good' with the fact that all the effects are uncorrelated. It is far from clear in this case that this criterion is necessarily 'good'. This is because the choice of an orthogonal second order design involves the magnitude of the diagonal elements, and this choice decides the relative precision with which linear, quadratic and interaction coefficients are estimated.

It may be noted, for example, that, for the 3^k factorial design in conventional scaling, the variances of the estimates for the quadratic coefficients are twice as large as those for the interaction coefficients. This was pointed out by Box and Wilson (1951),[1] and an intuitive attempt was made to reduce this apparent unbalance, by the introduction of composite designs. The design matrix for a three factor composite design can be written:-

$$
D = \begin{array}{r} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \end{array}
\begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\
\begin{bmatrix}
1 & 1 & 1 \\
1 & 1 & -1 \\
1 & -1 & 1 \\
1 & -1 & -1 \\
-1 & 1 & 1 \\
-1 & 1 & -1 \\
-1 & -1 & 1 \\
-1 & -1 & -1 \\
\alpha & 0 & 0 \\
-\alpha & 0 & 0 \\
0 & \alpha & 0 \\
0 & -\alpha & 0 \\
0 & 0 & \alpha \\
0 & 0 & -\alpha \\
0 & 0 & 0
\end{bmatrix}
\end{array}
$$

Where the magnitude of $|\alpha|$ may be so chosen that the design is orthogonal, <u>or</u> that the variances for second order effects are made equal.

Where the surface can be represented by an equation of 2nd degree, Box and Hunter [4,5] show how the variances of, and correlations between, second order coefficients estimated from a 3^2 factorial design change as the design is rotated. The condition of orthogonality refers to orthogonality in a particular orientation, and this property is in general lost on rotation of the design.

8'.3.1 Summary

...h we seen so far that in attempting to develop a class of optimum experimental designs for estimating the coefficients in a general k-variable second order model the terms "independent variable", "orthogonality", and "minimum variance" take on new aspects.

For first order designs the criteria of minimum variance and orthogonal estimates are simultaneously obtained by requiring that

$$\sum_{u=1}^{N} x_{iu} = 0, \quad \sum_{u=1}^{N} x_{iu}^2 = N \quad \text{and} \quad \sum_{u=1}^{N} x_{iu}x_{ju} = 0$$

For second order models orthogonal estimates may require the choice of unrealistic design moments and give estimates of unequal variances.

Alternatively, we may select designs that estimate all quadratic effects with equal variance, but not necessarily minimum variance; minimum variance estimates on the other hand, will be correlated as in the case of composite designs where we increase the magnitude of $|\alpha|$.

We are also aware of the important fact that the variance - covariance structure of an experimental design is not necessarily independent of the orientation of the design with respect to any fixed set of co-ordinate axes. (This contrast with first order orthogonal designs, where the variance - covariance structure of an experimental design is independent of the orientation of the design with respect to any fixed set of co-ordinate axes.)

The approach described so far has been concerned only with the accuracy of estimation for <u>individual</u> coefficients, and apart from designs of first order, this does not lead to any unique class of solutions.

8.4 Rotatable Designs

We now consider the estimation of complete equations, and not the estimation of individual coefficients and their variances.

The variance of the estimated response $\hat{Y}$ at any point $(x_1, x_2, \ldots, x_k)$ can be readily obtained. For example, a second order model fitted to the nine observations provided by the 3^2 design would give the following fitted equation

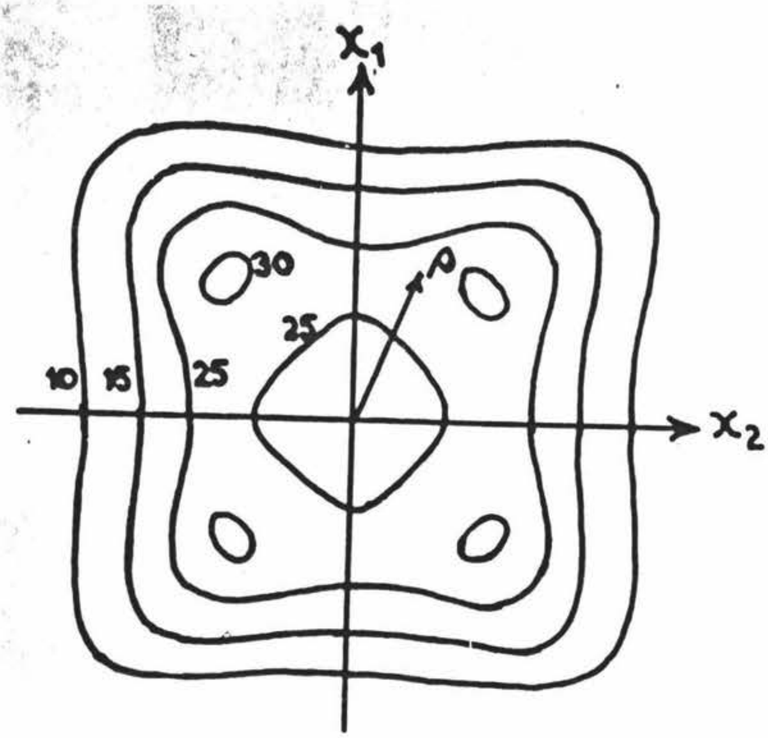
$$\hat{Y} = a + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2$$

and at any point (x_1, x_2) the variance of the estimate $\hat{Y}$ is given by

$$V(\hat{Y}) = V(a) + x_1^2V(b_1) + x_2^2V(b_2) + x_1^4V(b_{11})$$
$$+ x_2^4V(b_{22}) + x_1^2x_2^2V(b_{12}) + 2x_1^2Cov(a,b_{11}) + 2x_2^2Cov(a,b_{22})$$
$$= \sigma^2(5 - 3x_1^2 - 3x_2^2 + 2x_1^4 + 2x_2^4 + x_1^2x_2^2)$$

where in conventional scaling and orientation the 3^2 factorial design gives "inherent" variances of the coefficients a, b_i, b_{ii}, and b_{ij}, placed on a "per observation" basis equal to $5\sigma^2$, σ^2, $2\sigma^2$, and σ^2 respectively. All the covariances are zero except between a and b_{ii} which has covariance equal to $-2\sigma^2$ on a "per observation" basis. One way of decreasing the variance of the estimates is obviously to replicate the experimental points.

The <u>information</u> at any point (x_1,x_2) is defined as the reciprocal of the variance of that point. It is now possible to completely determine an information pattern for any experimental design. The contours of equal information provided by the 3^2 factorial are illustrated:



This contour diagram indicates that the 3^2 factorial design provides a greater concentration of information at a given distance ρ from the centre of the design in some orientations than in others.

Since an experimenter will in general have no initial idea as to where the most interesting portion of the response will ultimately lie, it would seem advisable to use an experimental design that provided the same information on all points equidistant from the centre of the design.

Ideally then, the experimental design in any orientation would provide information contours that were circles centred at the origin, and for designs with three or more independent variables the contours of constant information should be spheres or hyperspheres. These designs are called "<u>Rotatable Designs</u>".

A basic discussion on the derivation of rotatable designs, the design, is given by Box and Hunter.[4] Moment conditions, which a design of any given order must satisfy to obtain constant precision on spheres centred at the origin of the design, are derived; and the problem of finding arrangements of points which satisfy these conditions is considered.

8'.4.1 Rotatable Designs of Order One

Rotatability is a property of first order orthogonal designs. If it happens, contrary to assumption, that terms of second order are not negligible, bias is introduced. However, by selecting a design for which all the third order moments are zero, bias is eliminated in the estimates of the linear coefficients in every orientation of the design.

Specific designs of this sort are discussed by Box and Wilson[1] under the name "first order designs of type B". They can be obtained by duplicating with reversed signs any orthogonal first order design. The two-level factorial designs and many of the fractional factorials are also examples of particular orientations of designs of this sort.

8'.4.2 Rotatable Designs of Order Two

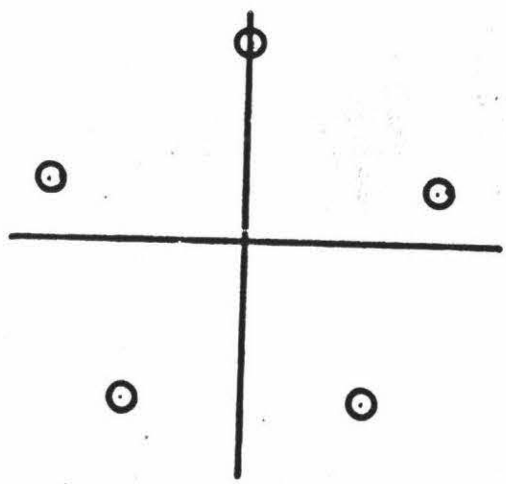
Rotatable second order designs, for any number of independent variables, can be obtained by using the vertices of regular figures, or combinations of regular figures, with one or more points at the centre of the design array. The addition of more than one point at the centre of any rotatable design provides (on the usual assumption that the variances of all determinations are equal) a valid estimate of experimental error.

For $k = 3$ independent variables one rotatable second order design is provided by combining a cube (the 2^3 factorial) with an octahedron (a "star design") to provide a design with 14 peripheral points. This design must of course have at least one centre point. This design is the analogue of the central composite design. For rotatability, the axis arms of the star design should be $\alpha = 2^{k/4}$. (An example of determining the coefficients in a second order model for $k = 2$ independent variable from a hexagon design with four centre points, for a total of $N = 10$ experiments, is given by Box and Hunter.[5])

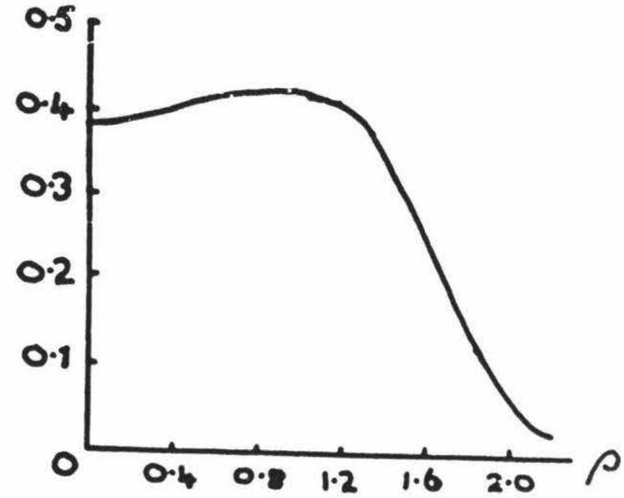
Where, contrary to assumption, third order constants exist, the expectations $E(b_o)$ and $E(b_i)$ will be biased. However, the expectations $E(b_{ii})$ and $E(b_{ij})$ will be unbiased in every orientation by selection of a design for which fifth order moments are zero.

8'.4.3 The Information Profile

One other important aspect of a rotatable design is the information profile. This profile indicates how the information changes as we increase the distance ρ from the centre of the design. In a rotatable design the information at a given distance ρ from the centre is the same no matter what the orientation.



THE PENTAGON DESIGN



INFORMATION PROFILE (3 CENTER P

Additional points at the centre of a second order rotatable design can be used to change the information profile of the design. For example, by placing a total of three experimental points at the centre of the pentagon design the information profile remains nearly constant over the interval $0 \leqslant \rho \leqslant 1$. Rotatable designs for which the information at $\rho = 0$ equals that at $\rho = 1$ are called <u>Uniform Information</u> designs, and can be constructed for any number of variables.

8'.4.4 Summary

The discussion so far has been concerned with two major approaches to the problem of selecting experimental designs that will fulfil the requirements given in section 8'.1.3.

The first approach was concerned only with the accuracy of estimation for individual coefficients of the estimating equation. Apart from designs of first order, this did not lead to any unique class of solutions.

The second approach was where the interest was directed at the complete estimation equation. As we have seen, Box and Hunter have derived conditions such that experimental designs in any orientation and of any given order, provide information contours that are concentric about the origin of the design.

A third approach to the problem of selecting experimental designs within the requirement context, section 8'.1.3, and where interest has been directed at the complete estimation equation, is described by Box and Draper.(6)

8'.5 Box and Draper Approach to Selection of a Response Surface Design

8'.5.1 Introduction

Most approaches to the theory of experimental design have been concerned only with the errors arising from "sampling" variation. The fitted equation is assumed to be capable of providing a perfect representation and the expectation of a fitted value $\hat{y}(\underline{x})$, is supposed to equal the expectation of the true value, $\eta(\underline{x})$. In practice there are two possible sources of discrepancy between the true function and the fitted equation. The first occurs because of sampling error ("variance error"), and the second because of the inadequacy of the graduating function ("bias error").

A "region of immediate interest" R, is assumed within a larger "operability region" O - within which it is possible to carry out experiments. If a particular model is assumed, and variance error is the only kind of discrepancy considered, then to obtain good representation over R, we ought to take as large a design as possible, covering the whole operability region O. However, this result is only reached because the decreased ability of the simple graduating function to represent the real relationship as wider and wider regions of the space of the variables are considered, is ignored. What is required then is some way in which the apparent added precision obtainable by making the design larger may be balanced against the loss of representational accuracy, to give optimum size properties of the design. The distribution of design points in the space of the variables can then be decided later.

In order that a polynomial of specific degree d_1 should represent the true function of degree d_2 ($d_2 > d_1$) within the region R as well as possible (Requirement (a) section 8'.1.3), it would be desirable to choose the design so as to minimise

$$J = \frac{N}{\sigma^2}\int_R E\left[y(\underline{x}) - \eta(\underline{x})\right]^2 d\underline{x} \Big/ \int_R d\underline{x} \qquad (8'.14)$$

where $d\underline{x} = dx_1, dx_2, \ldots, dx_k$. J is the mean squared deviation from the true response, averaged over the region R and normalized with respect to the number of observations and the variance. N is the total number of observations and σ^2 is the experimental error variance. The division into "variance error" and "bias error" can be illustrated by rewriting:

$$J = V + B$$
$$= \frac{N\Omega^{-1}}{\sigma^2}\int_R V\left[y(\underline{x})\right] d\underline{x} + \frac{N\Omega^{-1}}{\sigma^2}\int_R \left\{E\left[y(\underline{x})\right] - \eta(\underline{x})\right\}^2 d\underline{x} \qquad (8'.15)$$

where

$$\Omega^{-1} = \int_R d\underline{x}$$

Box and Draper interpret requirement (b), section 8'.1.3, that the design should allow a check to be made on the representational accuracy of the assumed polynomial in the following way:

It is supposed that a test for lack of fit is to be made by the use of analysis of variance in which the residual sum of squares

$$S_R = \sum_{u=1}^{N} (\hat{y}_u - y_u)^2$$

is compared with the experimental error variance, which is either assumed to be known exactly σ^2, or obtained from some independent estimate S^2; and where v is the number of degrees of freedom on which the residual sum of squares is based.

The null hypothesis tested in the analysis of variance is that $E(S_R)/v = \sigma^2$. This hypothesis is most commonly tested using the F-ratio. If this ratio reaches some a priori level of significance (usually 5%), the hypothesis is rejected. At this level of significance, we know that the probability of rejecting a true hypothesis is 5%. The probability that the hypothesis is rejected, when the hypothesis is in fact not true is called the "power of the test".[7] Thus, where $F_{.05}$ is the value of the ratio necessary for significance at the 5% level, we have,

$$\text{Power of the Test} = Pr\left\{\text{hypothesis is rejected} \mid \text{hypothesis is not true}\right\}$$

$$= Pr\left\{S_R/v\sigma^2 > F_{.05}\right\}$$

$$= Pr\left\{S_R > v.\sigma^2 F_{.05}\right\}$$

In general our aim should be to make the power of the test as large as possible.

---

(7) "Statistics - an Introduction", Donald A.S.Fraser, Ch.10, p.247-49. John Wiley & Sons Inc., 1960.

Box and Draper argue that where it is <u>known</u> that the hypothesis is not true, (i.e., that the fitted polynomial of degree d, is not the true polynomial) and in the particular instance where v is assumed fixed, an objective that increases the "power of the test" will be to make the expectation of S_R large.

However, where the order of the true polynomial is unknown (and it could be of the order selected (d_1) to approximate the surface) we should maximise the precision of the test. Minimising the variance of S_R,

$$Var(S_R) = E\left\{\left[S_R - E(S_R)\right]^2\right\}$$

should allow the test ratio

$$S_R / v.\sigma^2$$

to be determined with minimum variance. This should result in the "best" estimate of the test ratio, which is intuitively at least, the desired property of this ratio when the order of the true function is unknown.

8'.5.2 Application to First Order Designs

For first order designs, Box and Draper show that setting third order moments equal to zero assists in minimising J. The optimal choice of the spread of the design depends on the ratio of variance contribution to bias contribution. Where no bias exists, i.e., we believe implicitly in the adequacy of the first order polynomial fitted, the bounds of the design should be allowed to extend as far as possible. At the other extreme, B alone is minimised, and conditions that limit the bound of the optimum design within the region R, are obtained. A consideration of intermediate cases leads to optimum designs that are closely similar to those obtained when bias alone is minimised, except for the case where variance contribution is largely or completely dominant. Minimisation of the mean squared deviation from the true response (J) thus leads to the selection of designs that have certain moment and size properties

These properties are discussed by Box and Draper, for the cases where the true polynomial is a quadratic and where the true polynomial is of any order d_2, ($d_2 > d_1$, where d_1 is the order of the approximating polynomial). In this latter case, and where Averaged Squared Bias only is minimised, the conclusion is reached that the design points should probably be spread evenly over the region R.

Box and Draper give one particular way of generating first order designs, that satisfy the design requirements, and where concern is shown only for bias from second order terms. Designs satisfying requirement (a) section 8'.1.3 must be first order orthogonal with third order moments zero; these are the designs of type B, discussed by Box and Wilson (1951)[1]. A particular class of such designs are the two level fractional factorials, where no two factor interaction is confounded with a main effect, and where the complete design matrix of 2k points (where k factors are tested) is obtained, by replicating the first k points with reversed signs.

Requirement (b), section 8'.3.1, is that departures from the assumed model, which occur because the (assumed) true function is quadratic rather than linear, should be readily detectable. Two level factorials in which k factors are tested in 2k trials by replicating a square kxk orthogonal matrix with reversed signs, necessarily have points all equispaced from the origin, and provide quite sensitive tests for departure from linearity. If the basic design is to be modified by extra points, these should all be added at the origin, to give the greatest increase in the sensitiveness of the test for departure from linearity. Additional points at the centre also give an estimate of the error variance, thus making possible tests of departure from the linear model based solely on the internal evidence supplied by the design. Box and Draper limit their discussion to the selection of optimal first order designs.

8'.6 Discussion

The stimulus for this work on experimental designs for the estimation of response surfaces was provided in the original paper by Box and Wilson,[1] where a method of sequential experimentation was outlined for the attainment of optimum conditions. The method involved a series of small trials, the location of each one depending on the direction of "steepest ascent" calculated from the previous trial. The efficiency of this sequential approach to finding optimum factor combinations depends to a large extent on the accuracy of the response surface estimates gained from each trial carried out.

This approach is not suited in general to the exploration of agricultural relationships where information on a large portion of the production surface of interest is required. This is the case in pigmeat production where we wish to be able to predict liveweight gains from a number of alternative rations.

Pigmeat production however is a sequential process. Thus we do have the possibility of deriving optimum designs at a number of points in the production process. In the case of the incremental model, experimental treatments would be the particular weekly rates of meal and whey feeding employed at any given liveweight. The optimum location of these treatment could be derived for a given experimental area of interest and order of the polynomial thought sufficient to represent the response in this area, according to the methods derived by Box et.al. This procedure could be repeated at suitable liveweight intervals to obtain a design that had optimum properties, in the sense that the regression fitted to these points, i.e. the incremental model, had desired properties.

Fig.2 Ch.5, indicates that the experimental design used in the meal:whey levels trial corresponded roughly to a 3x3 factorial in each time period. A likely disadvantage of this trial was that each ration was fed at only one rate. It may be possible to use the design principals reviewed in this chapter to select arrangements of points for determining the effect of feeding selected rations at different rates.

Time and the scope of this present study have not allowe detailed exploration of this interesting and potentially useful approach to the selection of experimental designs for response surface estimation of the type encountered in agriculture.

References cited:

(1) "On the Experimental Attainment of Optimum Conditions", G.E.P. Box and K.B. Wilson. Journal of the Royal Statistical Society, Vol.13, No.1, 1951, p.1.

(2) "Multifactor Designs of First Order", G.E.P. Box.
[illegible] 49-57.

(3) "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples", G.E.P. Box. Biometrics, Vol.10, 1954, pp.16-60.

(4) "Multifactor Experimental Designs for Exploring Response Surfaces", G.E.P. Box and J.S. Hunter. The Annals of Mathematical Statistics, Vol.28, No.1, 1957, pp.195-241.

(5) "Experimental Designs for the Exploration and Exploitation of Response Surfaces", G.E.P. Box and J.S. Hunter. p.138 in Experimental Designs in Industry, V. Chew, Ed. John Wiley and Sons, Inc., 1958.

(6) "A Basis for the Selection of a Response Surface Design", G.E.P. Box and N.R. Draper. Journal of the American Statistical Association, Vol.54, pp.622-654, 1959.