MASSEY UNIVERSITY

THE APPLICATION OF MULTIPLE REGRESSION ANALYSIS TO THE VALUATION OF RURAL LAND

by

R.V. HARGREAVES

Occasional Paper No. 8
DEPARTMENT OF AGRICULTURAL ECONOMICS
AND FARM MANAGEMENT, MASSEY UNIVERSITY,
PALMERSTON NORTH, NEW ZEALAND.

MAY 1974

THE APPLICATION OF MULTIPLE REGRESSION ANALYSIS TO THE VALUATION OF RURAL LAND

by

R.V. Hargreaves

May 1974

Occasional Paper No. 8.
Department of Agricultural Economics
and Farm Management, Massey University,
Palmerston North, New Zealand.

Preface

Interest in valuation research at Massey University has increased following the appointment of a Lecturer in Rural Valuation in 1972. In this preliminary study Mr. Hargreaves examines the application of multiple regression analysis to the valuation of dairy farms within four counties in the Manawatu region.

This study should be of interest to all valuers, but particularly those involved with the valuation of farm land. In addition to discussing the most important factors (variables) which determined or influenced the selling prices of the farms an attempt has been made to predict the selling prices by the use of the regression equations. The regression equations have generally "explained" between eighty and ninety per cent of the variation in selling prices between farms.

It is hoped that the results of this study will encourage further work in this field since multiple regression analysis appears to be potentially a very useful tool to assist the rural valuer in his work.

A.R. Frampton
Professor of Agricultural Economics
and Farm Management

Acknowledgements

I am indebted to Mr H.A. Witty, former Deputy Valuer General, and to Mr W. Bartosh, District Rural Valuer Palmerston North for their co-operation and assistance, enabling me to obtain from Valuation Department records much of the farm sales data that was required.

I would also like to thank Professor Townsley for his assistance in the analysis of the results, and his helpful comments during the preparation of this report.

R.V. Hargreaves

Contents

1. INTRODUCTION

Valuers are frequently criticised for not being objective in their approach. Standard text books on valuation concede that valuation is not an exact science and that subjective elements in the form of considered opinions play an important part in the valuation process. The situation is further confused by the lack of general agreement amongst valuers on a workable definition of value. Until valuers can agree on this crucial definition there are likely to be as many valuations for a property as there are definitions of value.

A definition of value that is economically sound and applicable in practice is given by Ratcliffe[1] who argues that the valuer must attempt to predict the most probable selling price of the property. The introduction of probability to the concept of value is an important step forward in valuation theory and recognises the conditions of uncertainty that the valuer faces.

Traditionally students of valuation have been exposed to three approaches to valuation, namely the sales approach, the cost approach, and the productive approach. Greatest reliance is usually placed on the sales approach because it simulates most closely what is actually occurring in the market place. Both the productive approach and the cost approach embody market elements.

As the sales approach basically involves comparing like with like the greatest practical problem in using this approach has always been the lack of homogeneous sales information. Each parcel of real estate is said to be unique and the valuer is immediately confronted with the problem of adjusting heterogeneous sales information in such a way that it can be compared with the property being valued. In order to minimize the variation between sales valuers have typically used only a small number of the total

---

1 Richard U. Ratcliffe, Valuation for Real Estate Decisions, 1972, published by Democratic Press, see Chapter 3.

sales available as comparables. Ring[2] provides a detailed method of analysing a small number of sales that uses weighting factors to take account of variation between sales. It is unlikely however that many valuers go into this detail in a sales analysis.

The problem that the valuer faces if he tries to use a large number of heterogeneous sales is that it becomes very difficult to know how to handle the variations between properties. Using traditional methods the valuer is not capable of doing the calculations involved in comparing say 100 properties similar in all respects except for three variables. It is at this stage that, rather than ignoring sales information, the valuer can turn to the use of statistical techniques. The use of statistical methods to assist in the valuation process is not new; researchers were looking at the possibilities as early as the 1920's.[3] The technique used is multiple regression analysis. The early workers were faced with a large amount of difficult calculation, and the method was abandoned because it was too time consuming. It was not until the advent of the modern computer that valuers began to look again at regression analysis as a valuation tool. Regression analysis is not confined to valuation and is a widely used method of estimating the quantitative nature of relationships among variables. The technique is commonly used in scientific research by workers in fields such as medicine, economics, and biology.

Although valuation researchers in New Zealand have been somewhat behind the United States in the use of this technique, an article appeared in the Valuer as early as 1968.[4] The Valuation Department have published a comprehensive report on their work in this field up to 1972.[5] More recently the computer has been used to assist the valuers doing the five yearly revision of Porirua city.

---

2 Alfred A. Ring, The Valuation of Real Estate, 1970, published by Prentice Hall, see Appendix I.

3 G.C. Haas, Sale Prices as a Basis for Farm Land Appraisal, Minn. Agr. Exp. Sta., Tech. Bull. 9, 1922, p. 3.

4 M. Hildebrant, "The Value of Computers in Valuing", N.Z. Valuer 1 Dec. 1968, Vol. 20, No. 8.

5 Valuation Department, Research Paper 72-3, Valuing By Computer.

2. APPLICATIONS OF REGRESSION ANALYSIS TO RURAL VALUATION

In 1965 a major study was carried out by Davis in California. Davis[6] showed that regression analysis could be of considerable assistance to the rural valuer. Papers have continued to appear periodically in various journals pointing out the usefulness of regression analysis in rural valuation work. Progress from the theory to the practical application of this technique has been more rapid in urban valuation work than in rural valuation work.

The main reason why the urban valuer, particularly the valuer of residential housing, has been able to make more use of regression analysis than his rural counterpart is the large volume of sales information available in the towns and cities. For maximum statistical significance, regression analysis requires a sufficiently large number of observations or sales. Unfortunately sales information in the rural field is usually relatively scarce, and thus it is more difficult to develop useful predicting equations.

Sales data can be expanded by either taking a wider geographical area or a wider time series of sales information. Both of these methods have problems as in the first instance expanding the area may result in so much variation between sales that the predictive ability of the regression equation is reduced and in the second instance the market forces may radically change over a longer period of time.

---

6 Irving F. Davis Jnr., A Statistical Approach to Real Estate Value with Applications to Farm Appraisal, 1965, published by State of California Division of Real Estate.

2.1 Objectives of the Study

The following objectives were defined at the outset of the study.

(1) The analysis of selected variables contributing towards the selling price and the ranking of these in order of importance.

(2) The formulation of a regression equation for predicting farm sale prices and the evaluation of the equation.

(3) The investigation of the application of this technique to the practical valuation problem.

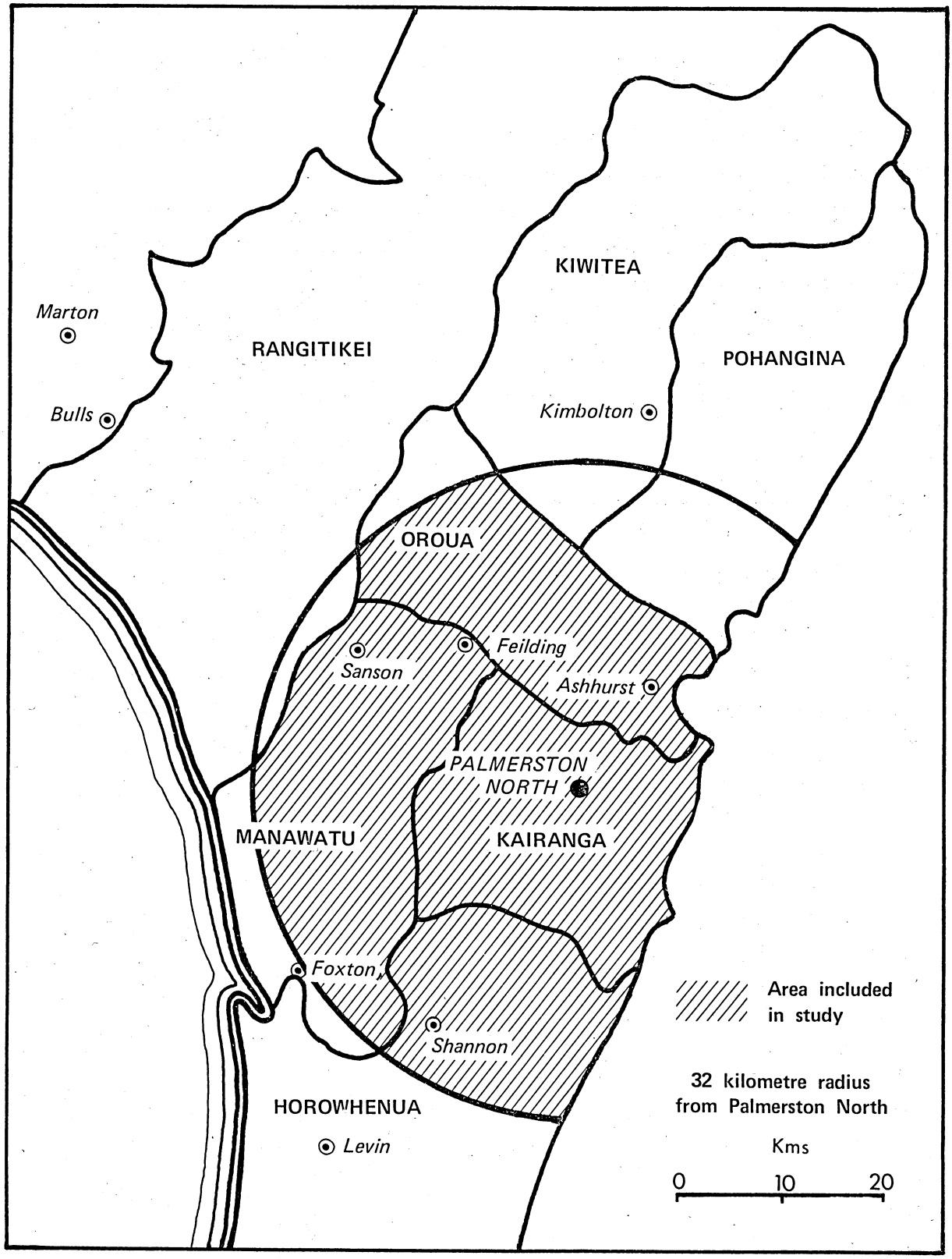
2.2 The Area Studied

This study was confined to looking at dairy farm sales data in the Manawatu area. For the purposes of this study dairy farm land included land being dairyed on at the time of sale and land where the highest and best use was considered by the Valuation Department classification to be dairying. A number of partial farms resulting from subdivision and amalgamation were included in the sample. Dairy farms were selected because of the volume of sales information, and the relative similarity between dairy farms in comparison to other types of farming systems such as fattening farms or hill country store sheep farms.

The farm sales data was drawn from within a 32 km radius of Palmerston North city including all of Kairanga and Manawatu Counties, and parts of Oroua and Horowhenua Counties. (See plan on page 6).

The physical details on the farms that had been sold were collected from the Valuation Department field slips and included sales over a three year period (December 1969 - December 1972).

PLAN SHOWING COUNTIES OF THE MANAWATU INCLUDED IN THE STUDY



2.3 Data Collection

Given the objectives of the study and having defined the geographical area to be included the next step was to collect information on farm sales. A property data collection sheet was designed so that as much information as possible could be recorded.* The data collection sheet went beyond the typical data that is shown in a standard valuation to include over 100 separate items of information about each farm sale. It was realised at the outset that probably less than ten variables would show the required level of significance when the regression was run, but to omit variables at the outset on a purely subjective basis could have prejudiced the results of the study.

As the data collection proceeded, it became obvious that insufficient information was available to adequately quantify some of the variables. For example some of the information required on page 2 of the data collection sheet could not have been obtained without interviewing the vendors. In this instance it was decided to simplify page two and when the data was finally tabulated only one variable relating to page two was used.

Although every effort has been made to include the variables that the author felt were important it is possible that some important variables have been overlooked because of the lack of readily available information about them.

Another major decision that had to be made when the data was tabulated for punching onto computer cards was how to handle the contour, soil type, and drainage factors. It was decided to use number scales based around the Storie Index[7] to take care of the soil factors. An alternative system that has not been explored in this study would have been to use a system of dummy variables to eliminate the subjective nature of the number scales. The practical problem of using dummy variables is that there is a real danger of losing statistical significance, due to the large number of variables necessary to quantify the soil factors using up too many degrees of freedom in the equation.

---

* See Appendix I.

7 R. Earl Storie, "Revision of the Soil-Rating Chart", Calif. Agr. Exp. Sta., Berkeley, 1959.

2.4 The Stepwise Regression Procedure

Stepwise regression is a technique that enables the user to determine the variables which are the most statistically significant. The computer is instructed to search out the independent variables that come within a pre-determined tolerance level and to fit these variables to the regression equation. Variables that do not conform with tolerance levels are discarded and colinearity is avoided since the computer avoids selecting independent variables that are significantly related.

In this study all twenty-three independent variables were included for each stepwise regression that was run. The final equations showed that from two to six independent variables were statistically significant.

Confidence Levels

The confidence level that was used for all equations was $F = 4$. This means that variables had to be significant at the five per cent level in order to be entered into and to remain in the equation during the stepwise regression procedure. There is only a five per cent chance that a variable that has no relation to price is included in the equations.

2.5 Variables Used in the Study

Dependent Variable.

Sale Price per Hectare: The actual sale price was deflated to a common base using the consumer price index. This is thought to remove some of the variation in rural land prices due to depreciation in the purchasing power of the dollar. Work done by Halstead[8] indicates that the consumer price index is unlikely to explain all the variation over time in rural land price. Specific independant variables namely factory milk fat payout and deflated building values/hectare, were included in the equation in an attempt to further explain price variations over time.

The Independent Variables.

(1) Area: Farm size was expressed in hectares. This variable proved to be important in several of the estimating equations that were developed. Area seemed to have a negative contribution in the sense that the larger the farm size the lower the price per hectare. Reasons for this could include the fact that the per hectare value of the buildings is likely to decrease with farm size, the capital constraints imposed on most buyers of larger farms, and the less intensive production from larger farms.

(2) Vendor Finance: Expressed as the percentage of the total purchase price that the vendor left in the property on mortgage. Accurate information on this variable was difficult to gather, and it was considered that searching the titles in the Lands and Deeds Office to confirm information would not have been warranted. Vendor finance was found to be important in the equation that was developed for the Kairanga County. In this particular instance every

---

8 Valuation Department Research Papers 72-2, Rural Real Estate Market in New Zealand 1950-1969.

percentage increase in vendor finance has a positive contribution to the price per hectare.

(3) Farm Amalgamations: This variable included properties that did not directly adjoin the purchasers property. The criterion used was to ask the question whether the purchaser owned other farm land in the locality. If yes = 1; If no = 0. Surprisingly perhaps, the market did not appear to pay a premium for amalgamated properties. A possible explanation is the large number of amalgamations in the sample. At this time State Advances Corporation were encouraging amalgamations and the dairy industry was just coming out of a period of depressed product prices.

(4) Distance to Palmerston North: Expressed as kilometres from the centre of the city. Location proved to be a very significant variable. A typical figure for location would be that price per hectare would fall $30 on average for every kilometer away from the city. This result is in line with economic theory and the variable could have possibly been further tested by considering non-linear functions such as the driving time to the centre of the city.

(5) Road Frontage: Expressed in metres. This variable was included as an attempt to consider the potential of the property for subdivision into ten acre blocks. The yield of ten acre blocks will be maximized on farms with a large amount of road frontage and subdivision costs would be kept to a minimum. Road frontage was significant in one of the predicting equations that was developed, possibly substituting for area in this particular equation.

(6) Subdivision: Expressed in whether ten acre subdivisions were permitted in the County at the date of sale. If yes = 1; if no = 0. This variable was not significant in any of the equations that were developed. Subdivision was still permitted in some of the Counties during the period in question.

(7) Modified Storie Index: An index of soil quality based on factors A and B under the Storie index. It is acknowledged that the Storie index may have limited application in New Zealand, and the numbers assigned to soil classes are subjective. Not withstanding this, the variable that was developed has been shown to be significant, and further work on soil productivity indexes for New Zealand would be helpful for the valuation profession.

(8) Number of Troughs and Dams: Expressed in the total number of troughs and dams on the property. This variable was introduced in an attempt to evaluate the added value of the water supply. The variable did not appear to be significant in the equations.

(9) Adjusted Milk Fat per Hectare: Expressed as the production of milk fat per hectare that the farm would produce on its own. This eliminated the effect of runoffs on production. This variable was shown to be highly significant, reinforcing what valuers have known for a long time, the price paid for land must relate to its productivity. On a dairy farm milk fat production is the most readily available index of production.

(10) Town Milk: Did the farm have a town milk quota? (Yes = 1; no = 0). This variable did not show up as being significant. Only two farms in the sample had town milk quotas and in the Manawatu a quota is usually not fully transferable.

(11) Assessed Value of Buildings per Hectare: Expressed either in the value of the buildings per hectare on the latest government valuation or the value on the sales analysis slip done by the Valuation Department. This figure was deflated to a common base using New Zealand Institute of Valuers model figures. There are subjective elements involved with this variable, but it is unfortunately not possible to eliminate these as the sample size did not allow enough independent variables to quantify all the variables associated with the buildings. This variable appeared

to be highly significant in the majority of the equations that were developed.

(12) Cowshed: Expressed as whether the farm studied had a herringbone cowshed. A dummy variable was used to ask the question, "does the farm have a herringbone cowshed?" (Yes = 1; no = 0). In one equation this variable was shown to be significant.

(13) Number of Paddocks: Expressed as the total number of paddocks on the property. This variable did not appear to be significant, probably because most dairy farms have a similar number of paddocks in order to maintain a typical three week grazing rotation.

(14) Central Race: Expressed as whether the farm had a central race? (Yes = 1; no = 0). It was felt that this variable may help to measure the state of development on the farm but it did not show up as being significant. Most dairy farms in the sample did have a central race.

(15) Distance to the Nearest Town: Expressed in kilometres. This variable was significant in several of the equations that were developed. Although distance to Palmerston North appeared to be the overriding locality factor, given that two farms were the same distance from the city then distance to the nearest town is important.

(16) Contour Index: Expressed in terms of factor C in the Storie classification. The contour classifications were the same as shown on the Valuation Department field slips. The scoring system used was as follows:

Flat	=	1.0
Undulating (ploughable)	=	.8 - .9
Undulating (non ploughable)	=	.6 - .7
Hills	=	.4 - .5

This variable did not appear to be significant, possibly because the variation in soil types tends to also explain the variation in contour.

(17) House: Expressed in terms of whether the property had at least one house? (Yes, no). This variable did not appear to be significant.

(18) Area Developed: Expressed as the percentage of the farm in permanent pasture or crop. Possibly because most of the farms in the sample were highly developed, this variable did not appear to be significant. In addition, there will be a correlation between production and the area developed.

(19) Excess Drainage Factor: The Storie index did not appear to discriminate enough against sand dune country and this additional factor was introduced. This factor was scored as follows:

Sand ridges and sand hills	= .3
Dry sand flats	= .6
Silt loam, clay loam	= 1.0

(20) Drainage Index: Expressed in terms of factor X in the Storie Index. This variable attempted to take into account both the natural and induced drainage characteristics of the soil. An example of the scoring system is as follows:

Dry sand ridges	= 1.0
Well drained silt loam	= .9

Neither of the above drainage variables appeared to be significant.

(21) Distance to the Beach: Expressed in kilometres as the direct distance to the west coast.

(22) Distance to the Hills: Expressed in kilometres as the direct distance to the Tararua or Ruahine Ranges. These two variables were an attempt to highlight climatic differences since rainfall increases nearer the hills. Neither variable appeared to be significant.

(23) Payout: Expressed as the factory milk fat payout in the season that the farm was sold. The work done by Johnson showed that buyers are predominantly influenced by the prices that prevailed in the year immediately preceding

purchase when formulating their expectations.[9] Payout did not appear as a significant variable in this study possibly because the relatively short time period studied did not allow significant trends to emerge.

3. RESULTS

The sales information that had been assembled from the four counties offered a number of alternatives for developing predicting equations. The alternatives that have been explored in this study, include a regression on all the data, and separate regressions based on the data from different counties, different farm sizes, and different time periods. It was hoped that a pattern would emerge showing certain variables as consistently being important in the majority of the equations developed. This would result in greater confidence in the use of these variables.

At first a regression was run on all the data. Due to the large amount of variation in the sample the predictive ability of this equation turned out to be quite low. More importantly, however, this equation did produce a list of variables that could be used as a 'bench mark' to evaluate subsequent equations.

---

9 R.W.M. Johnson, Trends in Rural Land Prices in New Zealand 1954-1969, Lincoln College Agricultural Research Unit, Technical Paper No. 4.

3.1 Table I

Regression Equation for Predicting Dairy Farm Sale Prices From 85 Farm Sales in Four Counties of the Manawatu Region (using all data)

Independent variable	Units of measurement	Regression coefficient
1. Distance to Palmerston North	kilometres	-45.516
2. Road frontage	metres	11.453
3. Adjusted milk fat	kilos/hectare	4.485
4. Distance to nearest town	kilometres	-32.905
5. Deflated building assessment	dollars/hectare	5.60

Constant = 654.918 Degrees of Freedom = 80

$R^2 = .7208$

$$\frac{\text{Standard Error of Estimate}}{\text{Average Selling Price/Hectare}} = 20.3\%$$

This means that the equation explained 72 per cent of the variation in the selling prices in the sample and that 72 per cent of the predictions were within 20 per cent of the actual selling price.

In order of importance the variables ranked as follows:

		Beta Coefficient
1.	Distance to Palmerston North	-.496
2.	Deflated buildings assessment	.3234
3.	Adjusted milk fat production	.3129
4.	Road frontage	.2202
5.	Distance to nearest town	-.1381

A further indication of the predictive ability of the equation can be seen from Table 2 and Figure 1.

Table 2

Accuracy of Predictions for Five Variable Regression Equation for all Manawatu Dairy Farm Sales Data

Residuals as a Percentage of price %	Cumulative Percentage of 80 Estimates %
5	15
10	40
15	54
20	72
25	82
30	85
35	92
40	96
45	97
50	97
55	100

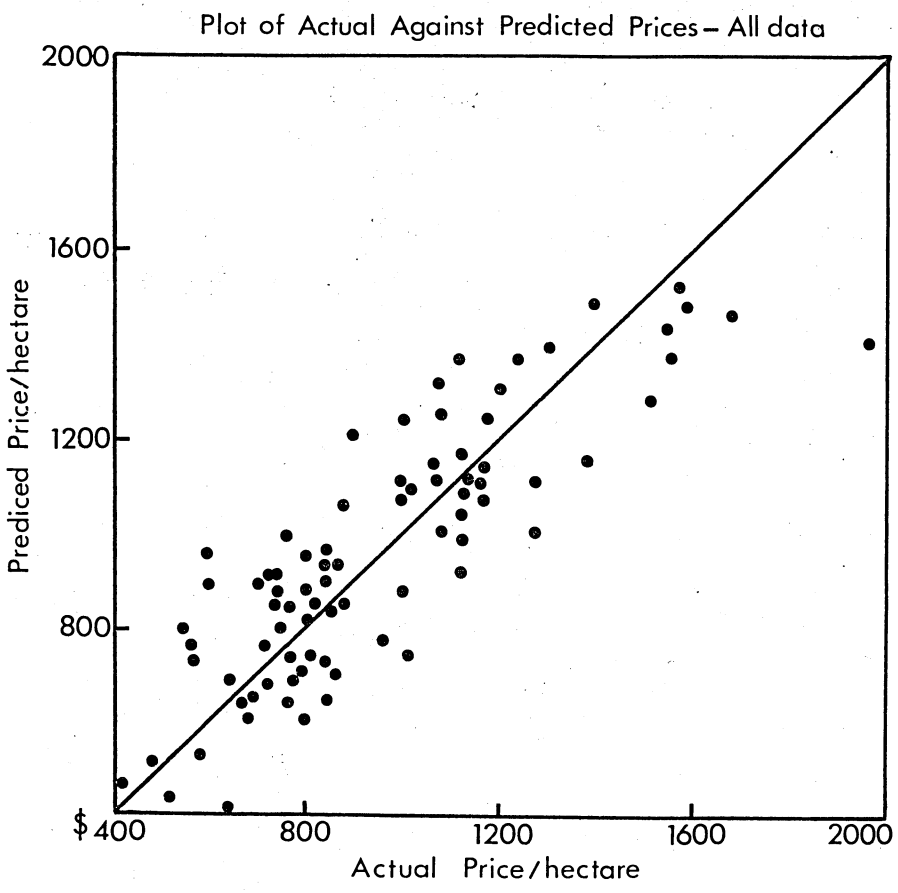
The second regression to be run included all the data from the Manawatu County over the three year time span that the data included. The Manawatu County has quite a large variation in the dairy farms within it. On the eastern boundary are the fertile river flats adjacent to the Oroua river, the central and northern part of the county is typically clay country, and the southern and western part of the county is mainly sand country of varying productivity.

As there were only twenty three observations in this sample, care must be taken when interpreting the results of the regression since the degrees of freedom in the equation are close to the minimum. Three of the variables that had shown up in the equation developed for all the data reappeared in this equation. These were the buildings variable, the production variable, and road frontage. The locality variables, namely distance to Palmerston North, and distance to the nearest town did not appear significant in this regression. Possible reasons for the absence

of locality in the equation would be that location becomes less important the further away a farm is from the city and that location may not be a linear function. For example driving time to Palmerston North may be just as important to farm buyers in Manawatu County as is the actual distance to the city. The relatively static nature of the populations of the small towns in Manawatu County make it unlikely that urban land pressures will have a significant effect on farm land prices, and this may account for the fact that distance to the nearest town is not shown to be significant in the equation.

Although one would not expect that road frontage would be a significant variable, the reason for the appearance of this variable appears to be the correlation between road frontage and farm size.

Figure 1:



3.2 Table 3

Regression Equation for Predicting Dairy Farm Sale Prices in the Manawatu County (23 farm sales)

Independent variable	Units of measurement	Regression coefficient
1. Road frontage	metres	9.236
2. Adjust milk fat	kilos/hectare	3.782
3. Deflated buildings assessment	dollars/hectare	3.045

Constant = 83.216 Degrees of freedom = 20

$R^2 = .8812$

$$\frac{\text{Standard Error of Estimate}}{\text{Average Selling Price/Hectare}} = 12.9\%$$

It can be seen that this equation represents an improvement over the initial equation development in Table 1. The three variables explain 88 per cent of the selling price, and 78 per cent of the predictions are within 15 per cent of the actual selling prices. All the predictions were within 35 per cent of the selling price.*

In order of importance the variables ranked as follows:

		Beta Coefficient
1.	Deflated building assessment	.5472
2.	Adjusted milk fat,	.4417
3.	Road frontage	.2242

* See Appendix II.

Table 4

Accuracy of Predictions for Three Variable Regression Equation for Manawatu County Sales Data

Residual as a percentage of price	Cumulative percentage of 23 estimates
%	%
5	21
10	56
15	78
20	87
25	91
30	95
35	100

Table 4 and Figure 2 give a further indication of the predictive ability of the equation.

One of the difficulties that arises when data over a three year time period is used is that the real estate market forces are likely to alter over time. Unless there is a variable or variables in the equation that take account of the market forces over time then the usefulness of the equation will be reduced. In order to test the variables relative to time it was decided to run a separate regression on thirty six farm sales in the Kairanga and Horowhenua Counties that had occurred during 1972. By confining the time period of the data to one year it was hoped that variation in the real estate market forces over time could be minimized. Farms were drawn from Kairanga and Horowhenua Counties because there was insufficient data to run a separate regression on Horowhenua County and Kairanga appeared to be the most comparable area with the northern part of the Horowhenua County.

The six variables that were fitted in this equation included three of the variables that had been shown to be important in the initial equation that was developed. The two locality variables and the buildings variable were the same as the initial

equation and in addition area, the type of cowshed, and the soil type appeared to be significant. It is interesting to note that soil type was apparently more significant than milk fat production per hectare in this regression. There is obviously a relationship between these two variables, and possibly the demand for land suitable for horticultural purposes has emphasised the importance of soil type. While there is often not a great deal of difference between milk fat production per hectare on the clay and silt soils, there is a definite preference by horticulturists for the silt and to a lesser extent the silty peat soils. Given that horticulture is a more intensive use of the land than dairying so the price per hectare is likely to be higher.

Area is thought to appear as an important variable in this equation due to the demand for small holdings close to the city as well as the reasons outlined in the discussion on the Manawatu County equation.

It is not known why the type of cowshed should appear to be important in this equation as this variable does not appear to be significant in any of the other equations.

Confining the data to a period of one year has produced a regression equation that is more accurate at predicting the sale prices than the equation that has been developed for Kairanga County using all the data over three years. This can be seen by comparing the results in Table 5 with Table 7.

Figure 2:

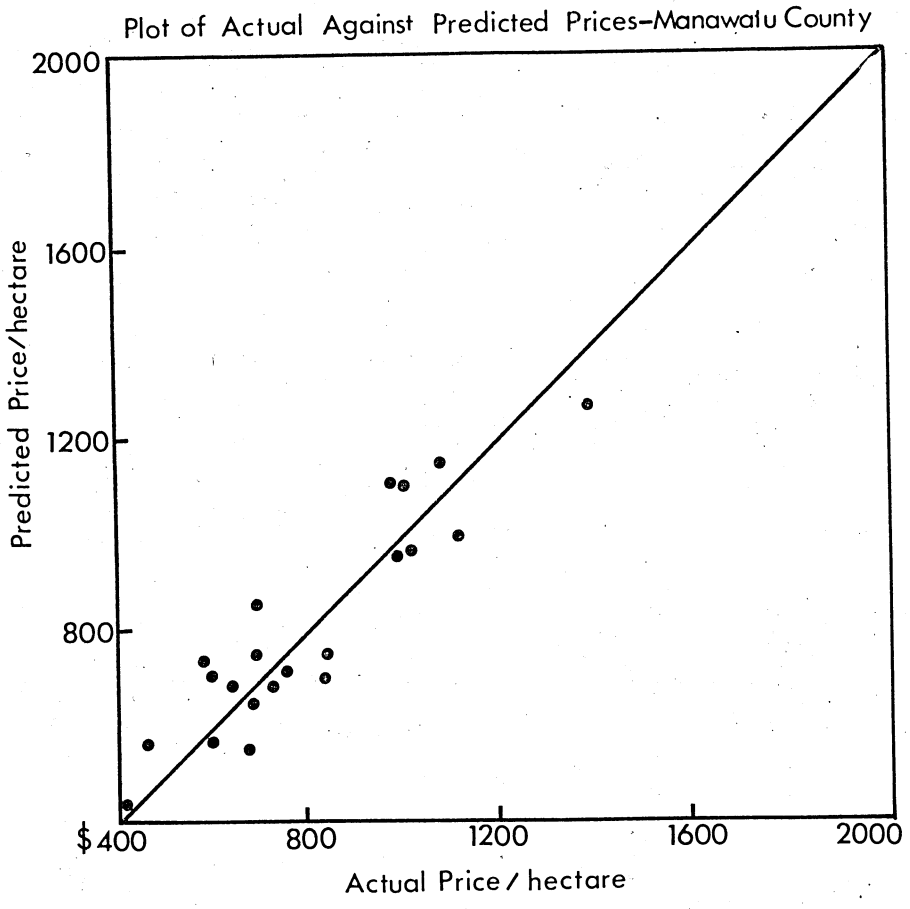
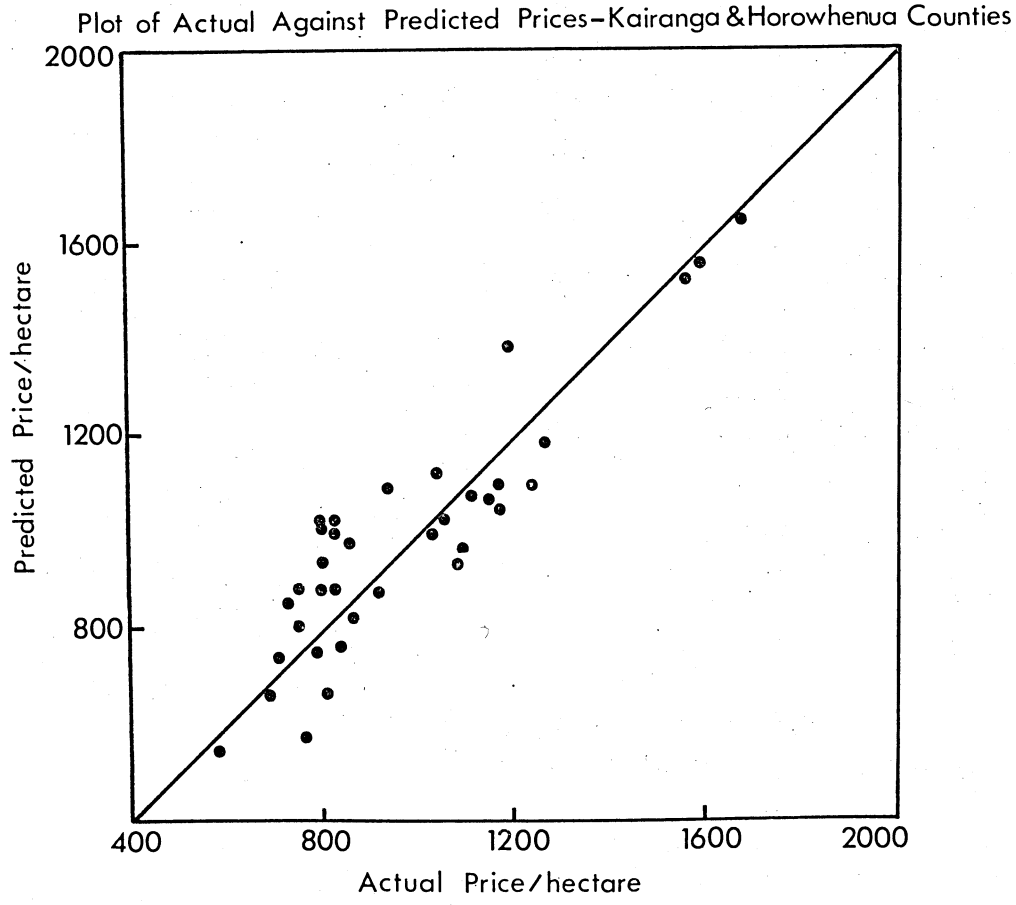


Figure 3:



3.3 Table 5

Regression Equation for Predicting Dairy Farm Sale Prices in the Kairanga and Horowhenua (northern) Counties. (36 farm sales 1972)

Independent variable	Units of measurement	Regression coefficient
1. Area	hectares	-5.745
2. Distance to Palmerston North	kilometres	-32.347
3. Storie Index	0 - 1.0	1623.7
4. Cowshed	yes = 1, no = 0	319.648
5. Distance to nearest town	kilometres	-27.752
6. Deflated buildings assessment	dollars/hectare	6.820

Constant = 378.593 Degrees of Freedom = 30

$$R^2 = .8664$$

$$\frac{\text{Standard Error of Estimate}}{\text{Average Selling Price/Hectare}} = 12.36\%$$

The above equation explained 86 per cent of the selling price; all the predictions were within 30 per cent of the actual selling price and 70 per cent within 15 per cent of the actual selling prices.* See Table 6.

In order of importance the variables ranked as follows:

		Beta Coefficient
1.	Distance to Palmerston North	-.8076
2.	Deflated Buildings Assessment	.4693
3.	Area	.3625
4.	Storie Index	.3152
5.	Cowshed	.2859
6.	Distance to nearest town	-.2811

Table 6 and Figure 3 give a further indication of the predictive ability of this equation.

* See Appendix III.

Table 6

Accuracy of Predictions for Six Variable Regression Equation for Kairanga and Horowhenua (northern) Counties

Residual as a percentage of price	Cumulative percentage of 36 estimates
5	39
10	61
15	70
20	92
25	94
30	100

The Kairanga County had the largest volume of sales of the counties included in the study, and all the data for this county was used as the basis for a regression equation.

Kairanga County surrounds the city of Palmerston North and is the Manawatu county most subject to the various forms of land pressure caused by urban population growth, rural subdivisions, and horticultural uses. For these reasons it was anticipated that developing a useful equation for the county would be difficult. No attempt was made to exclude observations from the data that were close to the city boundary and that had ultimate urban subdivision potential as it was hoped that the variables in the equation would be able to explain these influences. Four of the five variables that were entered in the equation had already appeared in the other equations that were developed. As expected from the results that appeared in the regression on the combined 1972 Horowhenua and Kairanga data, buildings, locality, and soils were all shown to be significant. Road frontage appeared again and is thought to be correlated with the variable for area.

The other variable that appeared was vendor finance, the proportion of the purchase price that the vendor left in the

property on mortgage. The data shows that vendor finance seemed to be more prevalent in Kairanga County both in terms of the number of cases involving vendor finance and the proportion of vendor finance. This may be the reason for the significance of the vendor finance variable in the equation.

The results are shown in Table 7.

3.4

Table 7

Regression Equation for Predicting Dairy Farm Prices Within the Kairanga County From 35 Farm Sales

Independent Variable	Units of measurement	Regression coefficient
1. Vendor finance	Percentage of price left of mortgage	2.973
2. Road frontage	metres	16.555
3. Storie Index	Score 0 - 1.0	2064.26
4. Distance to nearest town	kilometres	-129.82
5. Deflated building index	dollars/hectare	7.12

Constant = -576.64 Degrees of Freedom = 30

$$R^2 = .7861$$

$$\frac{\text{Standard Error of Estimate}}{\text{Average Selling Price/Hectare}} = 14.9\%$$

This equation explained 78 per cent of the selling price with 88 per cent of the predictions within 20 per cent of the actual selling prices.*

---

* See Appendix IV.

In order of importance the variables ranked as follows:

		Beta Coefficient
1.	Deflated building assessment	.4685
2.	Distance to nearest town	-.4399
3.	Storie Index	.3853
4.	Road frontage	.2845
5.	Vendor finance	.1926

Table 8 and Figure 4 give a further indication of the predictive ability of this equation.

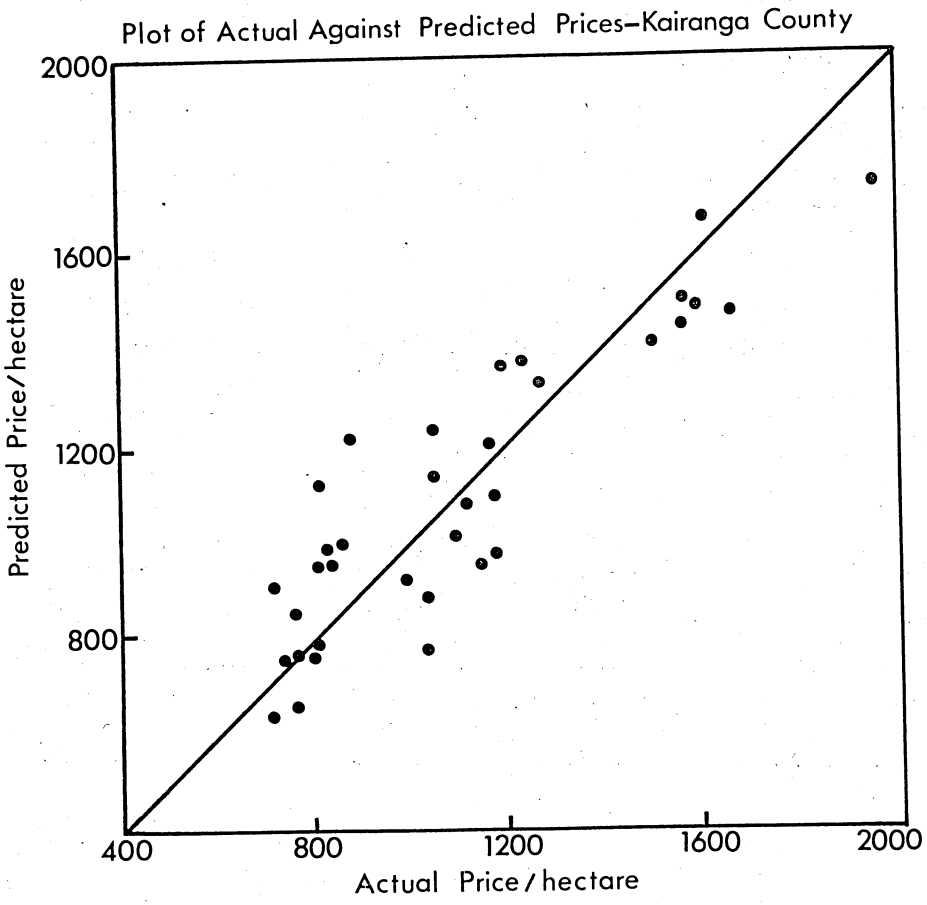
Table 8

Accuracy of Prediction for Five Variable Regression Equation for Kairanga County

Residuals as a percentage of price	Cumulative percentage of 35 estimates
%	%
5	22
10	48
15	62
20	88
25	91
30	93
35	93
40	100

In the four equations that had been developed no attempt had been made to stratify the data into groupings relating to farm size. It was felt that farms of similar sizes should have many common features and that this would reduce the number of variables needed to explain the selling price as well as improving the predictive ability of the equation developed.

Figure 4:



The data were divided into four groups:

(1) Farms 0 - 20.25 ha
(2) Farms 20.25 ha - 40.5 ha
(3) Farms 40.5 ha - 60.75 ha
(4) Farms 60.75 ha.

In the case of the second group, which had 37 observations, about 70 per cent of the selling price was explained by a four variable equation. In the third group, which had 25 observations, 74 per cent of the selling price was explained by a two variable equation. There were insufficient observations in groups 1 and 4 for a meaningful result.

Ranking of Variables Farms 20.25 - 40.5 ha

(1) Distance to Palmerston North
(2) Area
(3) Distance to nearest town
(4) Contour index.

Ranking of Variables Farms 40.5 - 60.75 ha

(1) Storie Index
(2) Number of troughs.

The results of these two regressions do not appear to be as useful as the previous equations that have been developed. It is probable that grouping by size alone results in the problem of having farms with very different soil types in the same general group. For example a 100 acre farm on silt loam may be more comparable in terms of the resource mix with a 200 acre sand country farm than a 100 acre sand country farm.

Possible solutions to this grouping problem would be to group farms on the basis of total production, or group them on a basis of similar soil types. The second alternative has been taken care of to some extent by arranging the data by counties as in the case of the Kairanga County. The first alternative has not been explored in this study.

The four variables that were shown to be significant in the equation for farms from 20.25 - 40.5 ha were consistent with the significant variables in the first four equations. The two locality variables were shown to be the most significant influence on the dependant variable. In addition area was still important as was the contour index. The contour index is partially correlated with the Storie Index since the farms that are not flat had either clay, or sand soils. Thus, in addition to explaining the variation in the contour between farms in the sample, contour index also explains some of the variation in soil types.

Only two variables appeared to be significant in the regression equation developed for farms from 40.5 - 60.75 ha. The Storie Index is a number scale ranking the soil types and because there were wide variations in the productivity of the soil

types within the sample it is not surprising that this variable appeared to be significant. It is more difficult to explain why the number of troughs was the other variable included in this equation. A possible explanation for this is that the number of troughs will generally be correlated with the state of development on a farm and the number of paddocks the farm is subdivided into. A number of the larger farms included in this sample were not as intensively subdivided or developed as typical smaller farms.

Several of the equations that have been developed come close to predicting ninety per cent of the variation in selling prices between properties in the sample. The standard error of estimate is not less than $50 per acre in any of the equations that have been developed. This means that 68 per cent of the time the equation is likely to be within $50 per acre of the actual selling price. While this result is encouraging it is not thought to be as accurate as the results that a skilled valuer would achieve using conventional methods.

3.5 Use of the Equations

The method of using the equation for an actual computer prediction is shown in Table 9.

Table 9

Estimated Price per Hectare of a 28.512 ha Opiki Dairy Farm Using a Six-Variable Multiple Linear Regression Equation Developed From Thirty Six Farm Sales in the Kairanga and Horowhenua Counties

Variable used	Coefficient	Data	Estimate
1. Area	-0.9418	70.425 acres	-66.349
2. Distance to Palmerston North	-21.0900	17.00 miles	-358.53
3. Storie Index	657.3848	1.0	657.384
4. Cow shed	129.4124	0	
5. Distance to nearest town	-18.0933	8 miles	-144.746
6. Deflated value buildings	1.118	64.01 $1 per acre	71.564
Constant Term			153.2766
Estimated value per acre			$312.6
per hectare		$772.12	
Actual sale price per hectare		$794.35	
Standard error of estimate		± 12.36 per cent.	

The above prices are deflated by the consumer price index and will need to be multiplied by a factor (1.2433) in order to equate actual sale price.

Once a regression equation is developed for an area and providing that the valuer has confidence in the equation then it is a relatively simple matter to use the equation in the field. Of the six variables included in the equation in Table 9, four involve straight forward measurement. Valuers are already very familiar with the techniques for valuing buildings, and provided that uniform costings and depreciation rates are used then there should be little problem with measuring this variable. To achieve

uniformity with the soils variable would necessitate valuers becoming familiar with a soils rating chart.

One of the greatest benefits that is likely to accrue to the rural valuer as a result of regression analysis is that an objective measurement is possible for the locality and productive influences. For example in Table 9 the equation tells us that for every mile we move away from Palmerston North the average value per acre falls by $21. Similarly the average price per acre falls by $18 for every mile we move away from the nearest town. Although rural valuers presently place considerable reliance on the price paid per kilo of milk fat for dairy farms it is difficult for them to isolate the locality influence from the production without the use of techniques such as regression analysis.

The equations were also used during 1973 to check conventional valuations carried out as part of the Rural Valuation course work at Massey University. It was found that although the equations were useful in early 1973 the rapid increases in rural land values during 1973 meant that the equations were generally underpredicting prices later in the year. This result clearly showed that the 1972 equations were not capable of handling the time element in times of rapidly rising land values.

One approach to this problem would be to constantly upgrade the equations by using the most recent sales information. This would need to be coupled with an independent variable that explained the change in buyers' expectations over time. The consumer price index and the butterfat payout price do not adequately explain changes in buyers' expectations, and further research work will need to be done on this aspect in the future.

4. CONCLUSIONS

Regression analysis is a useful tool for the valuer to use to assist in the appraisal process. The results of this study help to point out the important variables that contribute to Dairy Farm prices in the Manawatu. In the equations that were developed the variables relating to location, productivity, and the value of the buildings were consistently shown to be important. The actual contribution of each variable to the selling price is estimated by the regression coefficients.

It is significant that between eighty and ninety per cent of the variations in selling price is explained by five or six variables. Once the important variables have been identified by the valuer it may well be that he should spend a proportionately greater amount of time considering these variables. For example if location accounts for fifty per cent of the variation in selling price between the properties being considered then time spent evaluating the location factor may be more worthwhile than time spent evaluating a factor which is common to all farms in the sample.

Having to put the farm sales data into a form acceptable to the computer helped to identify areas where the data was subjective and areas where further research work is required. The variation in selling price that has not been explained by the equations could be reduced by using additional statistical techniques such as factor analysis to produce more homogeneous groupings within the original sample.

The effect of time on the sales data has not been satisfactorily explained by the equations that have been developed since they have not been able to cope with the recent phenomenal increases in rural land prices. One of the aspects that will need to be looked at in the future is a variable relating to the amount of surplus liquidity in the economy.

In addition a variable relating actual production with potential production may warrant inclusion in any future study.

Although such a variable could not be entirely objective it would help to overcome the problems associated with differing standards of management between farms.

Care must be taken when interpreting the results of regression analysis. Unless the valuer is trained in the relatively sophisticated statistical techniques that are used he should work with a statistician when interpreting the results. The valuer is an integral part of the process because his knowledge of the practical situation is essential to help select the variables that are considered.

It is unlikely that regression analysis will have a place in the smaller private valuation practice until such time as valuers have access to a central data bank containing recent sales information. The information contained in the data bank would not need to be any more confidential than that which can be presently obtained from the New Zealand Institute of Valuers sales sheets and searching titles in the Lands and Deeds Office. Ideally both government and private valuers would have access to the same data bank.

Regression analysis may well have a place in the mass appraisal of farm land for rating purposes, particularly for the prediction of the land value.

References cited

(1) Davis, Irving F., Jr., A Statistical Approach to Real Estate Value with Applications to Farm Appraisal, published by the State of California Division of Real Estate, 1965.

(2) Haas, G.C., "Sale Prices as a Basis for Farm Land Appraisal", Minn. Agr. Exp. Sta., Tech. Bull. 9, 1922, p. 3.

(3) Hildebrant, M., "The Value of Computers in Valuing", N.Z. Valuer, Dec. 1968, Vol. 20, No. 8.

(4) Johnson, R.W.M., "Trends in Rural Land Prices in New Zealand 1954-1969", Lincoln College Agricultural Research Unit, Technical Paper No. 4.

(5) Ratcliffe, R.U., Valuation for Real Estate Decisions, published by Democratic Press, 1972.

(6) Ring, A.A., The Valuation of Real Estate, published by Prentice Hall, 1970.

(7) Storie, R.E., "Revision of the Soil-Rating Chart", Calif. Agr. Exp. Sta., Berkeley, 1959.

(8) Valuation Department,"Rural Real Estate Market in New Zealand 1950-1969", Research Paper 72-2, 1972.

(9) Valuation Department, "Valuing By Computer", Research Paper 72-3, 1973.

APPENDIX I

P R O P E R T Y D A T A - Sheet 1

Code		
		IDENTIFICATION
1.		File No. Vendor _____
2.		Assessment No. Purchaser _____
3.		Date of Sale County _____
4.		Consumer Price Index at Time of Sale Govt. Val. (as at _/_/_) V.I. _|_|_|_|_|_|_
5.		Sale Price per acre U.V. _|_|_|_|_|_|_
6.		Total Sale Price L.V. _|_|_|_|_|_|_
7.		Butterfat Payout in season of sale C.V. _|_|_|_|_|_|_
		Legal Description _____
		_____
		_____ Aerial Photo Ref. _____
		Name of Road _____
		FINANCE
8.		First mortagee (% of Total Sale Price) Amount: _____
9.		First mortgage Interest Rate: _____
10.		First mortgage Term: _____ Mortgagee: _____
11.		Second mortgage (% of Total Sale Price) Amount: _____
12.		Second mortgage Interest Rate: _____
13.		Second mortgage Term: _____ Mortgagee: _____
14.		Third & Subsequent mortgages (% of total Sale Price) Amount: _____
15.		Third & Subsequent mortgages Interest Rates (Av.) _____
16.		Third & Subsequent mortgages, Term (Av.) ____ Mortgagee: _____
17.		Cash Transaction (Yes, No) _____ _____
18.		Vendor Finance (%)
19.		Element of collateral security (specify)

PROPERTY DATA - Sheet 2.

Code		
		VENDOR (Apparent motivation for Sale)
20.		Ill Health (Yes, No)
21.		Urban Encroachment (Yes, No)
22.		Forced Sale (Yes, No)
23.		Amount Sold (i.e. % of total holding)
24.		Vendor buying another Farm or Additional Land (Yes, No)
25.		Family Transaction
26.		Going Concern Sale (State Amount) $ _____
27.		Length of Time property on the Market (months)
		PURCHASER
28.		Age
29.		Previously Farming in District (Yes, No)
30.		Will personally supervise OR farm property (Yes, No)
31.		Businessman (Yes, No)
32.		Syndicate (Yes, No)
33.		Intends to lease or use sharemilker (Yes, No)

P R O P E R T Y D A T A - Sheet 3.

Code		
		LOCATION
34.		Sealed road (miles)
35.		Intersection (miles)
36.		Dairy Co. (miles)
37.		School (miles)
38.		Bus (Yes, No)
39.		Nearest town
40.		Population of town
41.		Nearest city
42.		Population of city
43.		Nearest grocery shop
44.		Nearest saleyards
		SUBDIVISION
45.		Road frontage (parts mile)
46.		Number of separate titles
47.		Proximity to nearest 10 acre block subdivision (miles)
48.		Distance to beach (miles)
49.		Price per acre for 10 acre blocks in area (Ave. comparable)
50.		Ten acre subdivision permitted (Yes, No)

PROPERTY DATA - Sheet 4.

Code		LAND AND CLIMATE						
		acres						
		Soil Type	Swamp	Flats	Undulating (P)	Undulating	Hills	Other

51. _____ % Developed

52. _____ Modified Storie Index Score:

53. _____ Total % of farm in permanent grass or crop

54. _____ % in Plantation (millable) Area:

55. _____ Age of Plantation

56. _____ Rainfall (inches or m.m.)

57. _____ % of property requiring further drainage Acres:

58. _____ Irrigation system available (Yes, No)

59. _____ Potential for irrigation high (Yes, No) (Proximity to river etc.)

60. _____ Altitude

61. _____ Typical effective drainage costs/acre

P R O P E R T Y D A T A — Sheet 5.

Code		
		WATER SUPPLY
62.		Natural water available in sufficient quantities on property
63.		Relies on outside water supply
64.		Number of troughs
65.		Number of paddocks without water
66.		Effective age of water supply system
67.		High pressure system
68.		Gravity fed low pressure system
		PRODUCTIVITY
69.		Milking cow stock units/acre Total: _____ (3 year ave.)
70.		Other cattle stock units/acre Total: _____ (3 year ave.)
71.		Milkfat produced/acre Total: _____ (3 year ave.)
72.		Land devoted to cash crop (acres _____ Value: _____
73.		Feed normally purchased Value: _____ Quantity: _____
74.		Adjusted milkfat/acre (Runoffs)
75.		Estimated optimum milkfat prod./acre
76.		Income per acre from stock sales Total: _____

PROPERTY DATA - Sheet 6.

Code		
		MANURE
77.		Cwt. applied per acre (P, K, N) (Ave.)
78.		Cwt. applied per acre (Lime)
		TYPE OF SUPPLY
79.		Town milk (Yes, No)
80.		Quota (gallons)
81.		Glaxo supply (Yes, No)

P R O P E R T Y D A T A - Sheet 7.

Code		BUILDINGS						
		Type	Sq.Foot	Cost New	Age	Depreciation and Obsolescence Factor	Assessed Value	Other Factors

82. _____ Assessed Value at time of Sale/Acre: Total _____

PROPERTY DATA - Sheet 8.

Code

OTHER STRUCTURAL IMPROVEMENTS

Fences	Length	Type	Condition	Value/chain	Total
Internal	.....	.....	.....	.....	.....
	.....	.....	.....	.....	.....
Road	.....	.....	.....	.....	.....
Boundary	.....	.....	.....	.....	.....

83. Total Assessed Value of fencing/acre

84. Ave. sized paddock (acres)

85. No. paddocks (total)

86. Central Race (Yes, No)

87. Superior Layour (Yes, No)

88. Average Layout (Yes, No)

89. Below Average Layout (Yes, No)

90. Other Structural Inputs. Specify: e.g. (wintering pad) .....

91. .....

92. .....

DISTRICT

93. Market Activity at Date of Sale (active, mod., slow)

94. Average production per acre for district

P R O P E R T Y D A T A - Sheet 9.

Code		TENURE		
95.		Freehold (Yes, No)		
96.		Leasehold (Yes, No)	Lessor	
97.		Lessors Interest ($)(assessed)	Term	
98.		Right of Renewal (Yes, No)	From	
			Expiry	
			Annual Rent	
			Rental Value	

APPENDIX II

Comparison of the Predicted Values per Hectare with the Actual Sale Prices From Twenty Three Farm Sales in the Manawatu County

	Actual sale price	Predicted price	% Difference
1.	1029.4	954.16	-7.3
2.	590.0	747.17	+26.0
3.	618.4	578.2	-6.4
4.	692.09	848.9	+22.5
5.	402.6	417.4	+3.6
6.	478.6	546.6	+14.0
7.	835.3	697.2	-16.2
8.	1012.2	989.2	-2.2
9.	1002.8	1111.7	+10.8
10.	327.02	298.6	-8.3
11.	1393.8	1259.9	-9.5
12.	771.8	736.8	-4.5
13.	741.5	721.9	-2.3
14.	1124.09	998.1	-11.2
15.	986.51	1092.2	+10.5
16.	671.34	726.4	+8.1
17.	640.7	701.7	+9.2
18.	1100.8	1161.64	+5.3
19.	829.9	722.72	-12.8
20.	703.2	720.49	+2.4
21.	692.8	645.9	-6.4
22.	671.6	543.8	-18.8
23.	301.5	397.9	+31.9

APPENDIX III

Comparison of the Predicted Values per Hectare With the Actual Sale Prices from Thirty Six Farm Sales in the Kairanga and Horowhenua Counties

	Actual sale price	Predicted price	% Difference
1.	833.8	764.7	-8.0
2.	804.2	656.7	-18.1
3.	803.4	879.5	+2.8
4.	1117.1	962.3	-14.3
5.	794.3	772.1	-2.4
6.	710.1	701.2	-1.2
7.	778.05	878.3	+12.6
8.	749.1	791.6	+5.6
9.	949.2	1080.3	+13.8
10.	927.9	902.04	-2.6
11.	584.6	560.44	-3.8
12.	1268.3	1095.4	-13.6
13.	1095.9	960.08	-12.1
14.	1198.4	1385.4	+15.4
15.	1175.9	1073.2	-8.6
16.	802.5	933.9	+15.4
17.	731.6	864	+17.9
18.	1066.7	1072.7	+.4
19.	1159.4	1075.4	-7.2
20.	1283.1	1193.9	-6.9
21.	810.9	972.4	+19.8
22.	861.7	824.4	-4.3
23.	829.4	1023.8	+23.2
24.	774.5	564.3	-27.1
25.	801.02	881.5	+9.8
26.	1124.8	1100.8	-2.0
27.	731.3	746.3	+2.0
28.	1598.8	1568.6	-1.8
29.	1177.2	1077.6	-8.4
30.	2016.2	1872.01	-7.1
31.	1687.01	1653.6	-1.9
32.	1577.09	1556.3	-1.25
33.	1046.5	1027.7	-1.7
34.	1044.3	1119.6	+7.1
35.	844.4	991.4	+17.3
36.	809.9	1041.1	+28.4

APPENDIX IV

Comparison of the Predicted Values per Hectare with the Actual Sale Prices from Thirty Five Farm Sales in the Kairanga County

	Actual sale price	Predicted price	% Difference
1.	1518.5	1417.5	-6.65
2.	1059.6	1152.5	+8.7
3.	1589.1	1500.03	-5.59
4.	1003.06	927.2	-7.38
5.	1615.8	1675.4	+3.6
6.	1198.4	1387.6	+15.79
7.	1175.9	965.02	-17.9
8.	802.5	786.9	-1.9
9.	731.6	735.5	+.5
10.	1066.7	1245.6	+16.7
11.	1159.4	1188.07	+2.4
12.	1283.1	1338.49	+4.3
13.	810.9	961.07	+18.5
14.	861.7	1006.2	+16.7
15.	778.05	775.8	-.3
16.	829.4	968.9	+16.8
17.	774.5	672.3	-13.2
18.	801.02	782.2	-2.3
19.	1124.8	1085.3	-3.5
20.	731.36	635.2	-13.1
21.	1598.8	1487.1	-6.9
22.	1177.2	1110.7	-5.6
23.	2016.2	1764.5	-12.4
24.	1687.01	1491.1	-11.6
25.	1577.09	1469.4	-6.8
26.	1046.4	782.0	-25.2
27.	1044.3	878.5	-15.8
28.	844.4	990.2	+17.2
29.	809.9	1115.4	+37.7
30.	744.2	907.4	+21.9
31.	1177.6	961.07	-18.3
32.	1249.0	1371.8	+9.8
33.	1105.07	1018.6	-7.8
34.	762.4	857.8	+12.5
35.	891.4	1235.2	+38.6