Forecasting Farm Performance: Simulating Non-Normal Distributions.

Peter R. Tozer and Geoff Kaine*

Abstract

Accurate forecasting of farm performance is essential to efficient policy development and allocation of resources to farmers most in need of assistance. To accurately predict farm performance researchers need to consider the distribution(s) of critical variables beyond the mean and variance of these variables. Given that the distributions of many variables are not normal, and are often correlated, a technique to simulate farm performance must incorporate these two features. The method proposed in this study can be used to reproduce non-normally distributed variables that are correlated, these variables are then employed to predict farm performance. The simulation results show that the distributions of critical variables do impact on the measures of farm performance, and that policy advisors need to consider the skewness and kurtosis, along with the mean and variance, of distributions when analysing farm performance data and making policy advice.

Contributed paper presented to the Australian Agricultural and Resource Economics Society Conference, University of Melbourne, February 13-15, 1996.

---

* Project Officer and Senior Project Director, The Rural Development Centre, University of New England. This project was supported by the Commonwealth Government's core grant to The Rural Development Centre, administered by the Department of Primary Industry and Energy. The authors wish to acknowledge the assistance of Garry Griffith, NSW Agriculture, Armidale.

Background

Australian rural industries are constantly adjusting to changes in the farming environment, both economic and physical. As changes occur in this environment, and farm businesses adapt to those changes, the financial performance of farms within industries varies. Since variations in the financial performance of farms can have important implications for government policy, the ability to reliably forecast the financial performance of farms and rural industries is a critical factor in the development, formulation and resourcing of economic and social policies.

In many instances forecasts of the distribution of performance across farms in industries are needed because the performance of farms meeting certain criteria is of particular interest. For example, farms that are experiencing severe financial pressure are the focus for assistance provided through the Rural Adjustment Scheme. Hence, forecasts of the proportion of farms that are likely to be experiencing financial stress, and the severity of that stress, would be valuable. Unfortunately, the quantitative techniques that are most commonly used for forecasting are not well suited to constructing forecasts about the range or distribution of farm performance.

Broadly speaking, quantitative approaches to forecasting can be classified into two categories (see Allen 1994, for a comprehensive review of quantitative forecasting methods). One approach is to construct econometric or time series models using time series data and to use these models for predictive purposes. Examples of this approach are Just *et al.*, (1991) and Johnson (1992). A major limitation of this approach is that the distribution of outcomes across farms in an industry usually cannot be predicted. This is because time series data is usually only available for industry aggregates rather than for individual farms. Hence, forecasts can only be constructed for the industry as a whole. The second approach is to develop forecasts from mathematical programming models of representative farms. Examples of this approach are Parton (1979), Francisco (1980) and McClintock *et al.*, (1991). The applicability of this approach is limited by the difficulty and expense of constructing models representing different types of farms in different circumstances so that the entire range of farms in an industry are adequately represented.

Distributions, Correlations and Simulation

Most of the statistical procedures that have been developed for constructing quantitative models have been developed to 'explain' or 'predict' variables that are normally distributed. Normally distributed variables are symmetric about their means and can be completely summarised by their mean and variance. Algebraic equations describing such variables are relatively easy to manipulate which means that procedures for estimating the relationships between such variables are relatively easy to formulate and implement. Furthermore, the simplicity of these variables means that the derivation of statistical tests which can be used to draw inferences about the significance of relationships between such variables is relatively straight-forward.

The popularity of quantitative models such as regression analysis is due, apart from their simplicity, to the fact that the assumption of normality can be justified by appeal to Central Limit Theorems (Mittelhammer, forthcoming). Basically, these theorems demonstrate that as the number of observations on an independent random variable approaches infinity, the probability distribution of that variable approaches a normal distribution function (Larson 1982). Therefore, the assumption of normality can be justified on the grounds that the actual observations available for a particular variable are, in principle, drawn from an infinite process.

In many cases, however, the assumption that the sample of values available for a variable are drawn from a population of values that are independently and normally distributed cannot be justified. In particular, the assumption is invalid when the population of values for a variable is known not to be normally distributed, as is often the case for cross-sectional data. For example, the population of farms in a particular industry is known and is a number somewhat less than infinity. Consequently, given a sufficiently large sample of these farms, the nature of the distribution of variables across the population can be reliably inferred. Often, the result is not consistent with normality. If this is the case, the procedures that have been developed for use with normally distributed data cannot be legitimately employed, unless the data can be appropriately transformed.

Distributions can depart from normality in a number of ways. For example, data may be uni-modal but exhibit varying degrees of skewness and kurtosis. Alternatively, data may have more than one mode. In these circumstances the distribution is no longer normal and cannot be summarised by mean and variance alone. The distribution may be skewed and will be kurtotic. Skewness and kurtosis both affect the reliability of some of the tests for assessing the significance of variables or regressions (Pearson and Please 1975). Skewness is the third moment of a probability distribution function, following the mean and variance. Skewness measures the degree of asymmetry of a distribution. Kurtosis is the fourth moment of a distribution and is a measure of the thickness of the tails of a distribution relative to the thickness of a normal distribution.[1] The degree of skewness and kurtosis can affect the validity of tests for the significance of relationships between variables leading to incorrect conclusions (Pearson and Please 1975).

When variables are not normally distributed the specification of relationships between them becomes difficult. While such variables can be easily simulated independently of each other, techniques for quantifying the relationships between such variables are available for only a few specialised types of distributions (see Judge *et al.* 1980). This is due to the mathematical complexity of the distributions. Consequently, multivariate models for predicting the behaviour of related, non normal variables are relatively uncommon and tests for the statistical significance of relationships in such models are not well developed.

In most cases of statistical analysis more than one variable is considered, which leads to multivariate analysis, however in these cases it is also assumed that the variables are uncorrelated. But in reality it could be reasonably argued that most variables within economic analyses are correlated to some degree, some highly correlated, other correlations approaching zero. For example, a farmer's income is highly correlated to the prices paid and received for inputs and outputs from their farming system, but it could also be correlated to the size of the individual enterprises or the equity of the farmer in her/his business. Therefore, the assumption of zero correlation is not realistic, but this assumption is included with others in many analyses, which could lead to incorrect conclusions and inferences of some models.

The use of the simulation technique to predict future performance of any economic entity, whether it is the whole economy, a sector of the economy or a single unit within the economy, is based on several assumptions concerning the distribution of the density function from which random numbers are drawn from to generate the simulated output, and also the correlation of variables within a multivariate simulation model. Given that most simulations are based on the assumption of the distribution of the population, from which the random numbers are drawn, following a normal distribution pattern with mean zero and variance σ^2. Which is a reasonable enough assumption given that the third and fourth moments of most distributions are not supplied, or calculated, in data collection. However, if these moments are not estimated, and random numbers are drawn from a population not following a similar distribution to that of the sample population, then the results of such simulations may be misleading, or incorrect.

[1]For a normal distribution kurtosis always equals 3.0 (Anderson et al 1977). Higher or lower values can be obtained for non-normal distributions (Greene 1993).

Another problem as mentioned earlier is the assumption of zero correlation between variables in a multivariate analysis. However, the generation of multivariate non-normal random numbers requires more information concerning the correlations between variables, which may not be available from data sources or initial data collection, hence it could be assumed that there is zero correlations between variables, and the use of such uncorrelated variables could once again lead to incorrect model specification.

Since the distributions of the variables that we are interested in predicting are non-normal, and are correlated, the commonly used forecasting techniques such as regression analysis and time series analysis cannot be employed unless these distributions can be transformed into approximately normal distributions. This would be difficult given the nature of these distributions.

We chose to explore a method for simulating non normal, correlated distributions developed by Vale and Maurelli (1984). Their simulation method involves two stages. In the first stage a procedure proposed by Fleishman (1978) is employed to define the non-normally distributed variables as a polynomial function of a standard normal distribution. In the second stage, a procedure described by Kaiser and Dickman (1962) is employed to derive a weighting scheme which will allow the correlations between the original distributions to be reproduced. These stages are described in detail in the following section.

Approximating non-normal distributions

Fleishman (1978) proposed a technique that could be used to generate univariate non-normal random numbers from a polynomial power function of normally distributed random numbers. He proposed that a standardised, non-normally distributed random variable, Y, could be approximated by the following function where X denotes a variable that follows a standard normal distribution. The constants in the function, denoted by the letters a, b, c, and d, and which we will term 'power function coefficients', are calculated from the mean, variance, skewness and kurtosis of the non-normal distribution:

(1) $Y = a + bX + cX^2 + dX^3$ where $X \sim N(0,1)$

The constants in the equation above are derived as follows. Consider the following moment generating function (from Kendall, Stuart and Ord 1987, 78):

(2) $$\mu_{2r} = \frac{\sigma^{2r}}{2^r}\frac{(2r)!}{r!}, \qquad r \geq 0$$

Where μ_{2r} is the even moment of order r about the mean, and σ^{2r} is the variance of a normal distribution raised to the rth power. A solution for this function exists only for even moments around the origin. This function is used to derive expressions for the constants in equation (1) by substituting the values of the moments of the normally distributed variable, X, to obtain expressions for the moments of the non-normal variable, Y. Hence, for the first moment:

(3) $$\mathcal{E}(Y) = a + b\mathcal{E}(X) + c\mathcal{E}(X^2) + d\mathcal{E}(X^3)$$

Where $\mathcal{E}(.)$ denotes the 'expected value' or mean Since X is a standard normal variable on substituting (3) into (1) we find that:[2]

(4) $$\mathcal{E}(Y) = a + c.$$

Since Y has been standardised it follows that :

(5) $$a = -c$$

A similar procedure of substitution is used to solve for the second, third, and fourth moments of Y to yield the following equations as measures of these moments (Fleishman 1978):

(6) $$b^2 + 6bd + 2c^2 + 15d^2 - \sigma^2 = 0$$
(7) $$2c(b^2 + 24bd + 105d^2 + 2) - \gamma_1 = 0$$
(8) $$24[bd + c^2(1 + b^2 + 28bd) + d^2(12 + 48bd + 141c^2 + 225d^2)] - \gamma_2 = 0$$

where σ^2 is the variance of Y, and γ_1 and γ_2 are the skewness and kurtosis of the variable Y.

The solutions to this system of equations yield values for the constants in equation (1). In principle, the non-normal distribution Y can be reproduced by substituting random values drawn from a standard normal into equation (1), given the solutions to equations (6), (7) and (8). In essence, Fleishman's (1978) procedure allows non-normal

[2]For a standard normal distribution $\mathcal{E}(X) = \mathcal{E}(X^3) = 0$, and $\mathcal{E}(X^2) = 1$.

distributions to be simulated using an appropriate transformation of a normal distribution. This transformation into univariate non-normal random numbers is the limit of the usefulness of this technique, it cannot be used to generate multivariate non-normal random numbers, where the multivariate numbers have specified correlations.

To preserve the correlations between the non-normal distributions a matrix decomposition process is employed. Kaiser and Dickman (1962) demonstrated that principal component analysis can be used to reproduce a correlation matrix. In essence, principal components is a regression technique for decomposing a group of variables into 'components' which are weighted linear combinations of the original variables (Tabachnick and Fidell 1989). Usually, the technique is employed to collapse a large number of variables into a smaller, more manageable number of composite variables. Kaiser and Dickman (1962) demonstrated that, given a matrix of correlations between standard normal variables, the application of principal components to such a matrix yielded 'components' that could be employed to generate standard normal distributions with correlations identical to those in the original matrix. Of course, being a regression procedure this technique could only be applied when variables were normally distributed.

Vale and Maurelli (1983) combined the techniques of Kaiser and Dickman (1962) and Fleishman (1978) to provide a method of generating multivariate non-normal distributions with given correlations. To obtain the desired correlations between the non-normal distributions an 'intermediate' correlation matrix must be specified. These 'intermediate' correlations are correlations that have been adjusted to compensate for the transformation from standard normal to non-normal distributions.

The calculation of the 'intermediate' correlations is as follows. Let x be a vector of standard normal variables as below:

(9) $x' = [1, X, X^2, X^3]$

These variables are transformed into non-normal distributions using the power function weights in equation (2.1) above. In matrix notation let:

(10) $w' = [a,b,c,d]$.

The non-normal variables are the product of these two vectors:

(11) $Y = w'x$

Let $r_{Y_iY_j}$ $(i \neq j)$ be the correlation between two standardised non-normal variables Y_i and Y_j with corresponding normally distributed variables X_i and X_j. Since the variables are standardized the correlation between Y_i and Y_j is equal to their expected cross product, as the variances are equal to one (Vale and Maurelli 1983):

$$(12) \quad \begin{aligned} r_{Y_iY_j} &= E(Y_i'Y_j') \\ &= E(w_j' x_i x_j' w_j) \\ &= w_j' R w_j \end{aligned}$$

Where R is the expected cross product matrix of x_i and x_j'. The expected product matrix, R, is determined by using moment generating functions as described in Johnson and Kotz (1972) and Larson (1982). The product of equation (12) yields the following equation, which when solved for $\rho_{X_1X_2}$, the 'intermediate' correlation between the standard normal variables X_i and X_j:

$$(13) \quad r_{Y_1Y_2} = \rho_{X_1X_2}(b_1b_2 + 3b_1d_2 + 3d_1b_2 + 9d_1d_2) + \rho^2_{X_1X_2}(2c_1c_2) + \rho^3_{X_1X_2}(6d_1d_2)$$

Principal components analysis is applied to the resulting 'intermediate' correlations to derive component weights as demonstrated by Kaiser and Dickman (1962). These weights, in conjunction with a sample of values drawn from standard normal distributions, yield a set of normally distributed variables with correlations equal to those of the intermediate correlations table. These normally distributed variables are then transformed into the appropriate non-normal distributions with the desired correlations by applying the power function coefficients estimated from equation (1).

Simulation Model of Farm Performance

The model and data used in this study are derived from the 1992-93 farm surveys conducted by the Australian Bureau of Agriculture and Resource Economics. The Bureau reports a range of measures of financial performance, as well as the variables from which these measures are derived, for beef cattle enterprises in the Northern Territory in Tables F11 and F12 of the Farm Survey Report (ABARE 1994, 152-153). These measures, and the variables used to calculate them, (A summary of these variables is presented in Table 1), constitute the model we wish to construct. The model itself is summarised in Table 2. We use Vale and Maurelli's method to reproduce the distributions of the variables that are used to calculate measures of financial performance. That is, variables such as cash receipts, costs, capital appreciation, interest, rent and so on . The distributions of the financial performance measures themselves, measures such as profit, equity, and rate of return, are derived from the variables by means of identities as shown in Table 2. We will use the term *variables* to refer to those distributions we reproduce using Vale and Maurelli's method and the term *measures* to refer to those distributions we derive as identities.

Table 1: Distributions of selected financial variables for beef enterprises in the Northern Territory (1992-93).

Variable	Mean	Variance	Skewness	Kurtosis
Trading stocks	85362.170	454450.065	1.423	6.282
Capital appreciation	-224702.255	422018.013	-2.456	6.207
Change in debt	8093.081	120158.432	1.432	3.242
Cash costs	572560.251	635205.739	2.031	3.861
Depreciation	85900.322	126763.380	3.588	13.636
Farm debt	573690.301	1030207.220	2.371	4.970
Farm capital	3057622.016	4197574.725	2.917	8.624
Cash receipts	599614.636	790898.441	2.423	5.825
Interest payments	74249.897	124816.645	2.225	4.187
Off-farm income	12495.962	23584.949	2.163	3.457
Operator labour	35154.525	17900.320	0.055	-0.546
Rent	4527.479	5649.253	1.887	2.913

ABARE provided data on the distributions of the variables and financial performance measures that they report. As confidentiality provisions prevent the release of the actual data they collect, they supplied five percentile values and maximum and minimum values for each variable.[3] They also supplied estimates of the correlations between the measures. 5000 random data observations were generated for each variable using an algorithm for sampling from a segmented cumulative density function (Anderson 1983).[4] This artificial sample was employed to obtain estimates of the mean, variance, skewness and kurtosis of each measure.

[3]We later discovered that there were anomalies in the data provided with respect to the financial performance measures which we were unable to rectify. See the discussion later in this section.
[4]This program was tested by comparing histograms of the simulated data with the percentile values provide by ABARE. The simulated data produced by the program accurately reproduced the original distributions.

Table 2: Summary of Simulation Model

		Total Cash Receipts[5]
minus		Total Cash Costs
	=	**Farm Cash Income**
plus		Build-up in Trading Stocks
minus		Depreciation
minus		Operator and Family Labour
	=	**Farm Business Profit**
plus		Rent
plus		Interest
	=	**Profit at Full Equity**
divided by		Farm Capital July 1
	=	**Rate of Return (excluding Capital Appreciation)**
		Profit at Full Equity
plus		Capital Appreciation
	=	**Profit at Full Equity (including Capital Appreciation)**
divided by		Farm Capital July 1
	=	**Rate of Return (including Capital Appreciation)**
		Farm Capital July 1
plus		Capital Appreciation
	=	**Farm Capital June 30**
minus		Farm Business Debt June 30
	=	**Farm Business Equity June 30**
divided by		Farm Capital June 30
	=	**Farm Equity Ratio**

The calculation of the power function weights requires the simultaneous solution of equations (6), (7) and (8). Brown's Newton-based algorithm was implemented to derive the solutions to a set of non-linear equations (Todd and Roe 1978). In most cases we could apply Fleishman's technique directly. In some instances, however, the measures exhibited combinations of skewness and kurtosis that were too extreme to be reproduced by this technique (see Fleishman 1978, p526). We experimented with various non-linear transformations of the data to obtain less extreme combinations of

[5] Measures in bold type are 'endogenous', variables not in bold type are 'exogenous'.

skewness and kurtosis which could be reproduced. In general a logarithmic transformation yielded satisfactory results. Two distributions, interest payments and farm debt, were found to be tri-modal. Since the procedure we are using can only be applied to uni-modal data we were forced to 'split' the data for these two distributions into three separate parts and to calculate separate power function weights for each part.

The accuracy of the estimates was tested by substituting them into equation (1) and comparing the moments of the resulting distributions with those of the original distributions. In nearly all cases this test showed that the estimated distributions closely resembled the original distributions. In only two cases, farm business debt and interest paid, was the result unsatisfactory and we believe that this reflects the trimodality of these two distributions.

The next step in the process was to calculate the intermediate correlations between all variables. As discussed previously, these correlations are necessary to compensate for the transformation of data from a normal distribution to a non-normal distribution. Brown's algorithm was employed to solve Equation (13).[6] The resulting intermediate correlations were analysed using principal components decomposition to obtain factor weights which would enable the multivariate correlations in the data to be reproduced.

Having obtained estimates of the power function weights, the intermediate correlations and the factor weights from the principal components analysis we could now construct the simulation model. Fifteen samples of 5000 observations drawn from standard normal distributions were generated. These samples were multiplied by the factor weights from the principal component analysis to produce a series of normally distributed variables with correlations equal to that of the intermediate correlations. The application of the estimated power function coefficients to these variables provided the estimated set of non-normally distributed random variables with correlations and moments approximating those required.

The procedure required some modification to incorporate the separate segments of the interest paid and farm debt distributions. We took the approach of randomly sampling from one of the three segments in accord with the probability of a value falling within each segment. Taking farm debt for example, 5 per cent of observations in the original sample were zero, 20 per cent had values corresponding to the range of the second segment and the remaining 75 per cent had values falling within the range of the third segment. Consequently, to obtain an aggregate distribution for farm debt we sampled 5

[6]This program was tested using the example provided by Vale and Maurelli (1984). The estimates provided by the program were identical to those reported by these authors.

per cent, 20 per cent and 75 per cent of values from the first, second and third segment respectively.

In Appendix 1 the original and estimated distributions of the financial variables are graphed. An inspection of these figures reveals that the simulated distributions closely resemble the original distributions. The only obvious difference between the distributions is that, in some cases, the simulated values are substantially higher than the original values at the extreme upper end of the distributions. This reflects the sampling of extreme values from the tails of the normal distribution. Recall that the original distributions supplied by the Bureau were specified as percentiles with minimum and maximum values which were derived from survey estimates. Consequently, the original distributions are truncated approximations to the population distributions. Since the simulated distributions are derived from sampling non-truncated normal distributions extreme values will be simulated that exceed the minimum or maximum values of the original distributions. In other words, the tails of the simulated distributions would be expected to be 'longer' than those of the original distributions. This means that the skewness and kurtosis of the simulated distributions will tend to be marginally higher than is the case for the original distributions. In all cases less than two per cent of the simulated values exceeded the maximum (or minimum) values specified in the original data.

In Table 3 the moments of the original and estimated distributions are compared. In most cases the results are satisfactory with the estimated mean and variance for most distributions being within ten per cent of the original values. More substantial errors were obtained for the mean and variance of farm debt and interest payments. We believe these errors reflect errors in sampling and aggregating from three different components for these two distributions (see discussion above).

The results for skewness and kurtosis are difficult to assess. On the one hand, although the absolute errors are small in most instances they are reasonably large in relative terms. On the other hand, reasonably large changes to the third and fourth moments of distributions have only marginal impacts on the shape of distributions. Note that the largest absolute errors are generally associated with those variables where values were simulated which were substantially different from the original values at the extreme ends of the distributions (appreciation, farm debt and capital, depreciation, interest, rent and off-farm income for instance). If the simulation model were constrained such that the maximum values produced by the model are consistent with the maximum values provided in the original data then these errors could be easily and substantially reduced. We chose not to construct such a model as we believed that

Table 3: Comparison of the moments of the original and estimated distributions for the financial variables.

Variable	Estimated Mean	Estimated Variance	Estimated Skewness	Estimated Kurtosis
Trading stocks	85495.944	454479.731	1.168	5.010
Capital appreciation	-224779.36	422027.359	-3.200	14.672
Cash costs	572713.212	635123.61	1.661	2.706
Change in debt	8045.255	120168.974	1.318	2.322
Depreciation	85893.022	126791.555	1.995	3.922
Farm debt	669962.454	956780.432	1.213	0.061
Opening farm capital	3958202.71	4197854.55	4.534	29.127
Cash receipts	599441.574	790971.035	1.989	4.942
Interest	88715.161	111342.934	0.822	-0.941
Off-farm income	12405.81	23578.891	3.379	20.555
Operator labour	35148.524	17894.021	0.030	-0.535
Rent	4526.192	5648.513	1.892	5.298

Variable	Mean (per cent error)	Variance (per cent error)	Skewness (absolute error, per cent error)		Kurtosis (absolute error, per cent error)	
Trading stocks	0.16	0.01	-0.26	-17.92	-1.27	-20.25
Capital appreciation	0.03	0.00	-0.74	30.29	8.47	136.38
Change in debt	-0.59	0.01	-0.11	-7.96	-0.92	-28.38
Cash costs	0.03	0.01	-0.37	-18.22	-1.16	-29.91
Depreciation	-0.01	0.02	-1.59	-44.40	-9.71	-71.24
Farm debt	16.78	-7.13	-1.16	-48.84	-4.91	-98.77
Opening farm capital	0.02	0.01	1.62	55.43	20.50	237.74
Cash receipts	0.03	0.01	-0.43	-17.91	-0.88	-15.16
Interest	19.48	-10.79	-1.40	-63.06	-5.13	-122.47
Off-farm income	-0.72	-0.03	1.22	56.22	17.10	494.59
Operator labour	-0.02	-0.04	-0.03	-45.45	0.01	-2.01
Rent	-0.03	-0.01	0.00	0.26	2.39	81.87

errors in the original data that were supplied, which are discussed below, rendered developing such a model problematic.

In Table 4 the predicted values of the financial performance measures are presented. We have not presented the original distributions for the financial performance measures in the figures or reported errors in predicting the moments of the distributions for these measures because we discovered errors in the original data that was provided to us. For example, the mean of farm cash income can be shown algebraically to be equal to the difference in the means of farm cash receipts and farm cash costs. A quick calculation will show that this is the case for the simulated data (that is \$26 728). The expected mean for farm cash income using the original data for cash receipts and cash costs is \$27 054 yet the mean of the distribution for farm cash income that was supplied is -\$16 336. Similar errors are apparent in the original data supplied for measures such as farm profit and profit at full equity. Clearly, these errors in the original data mean that the data supplied on other measures such as rates of return and equity ratios are suspect. Consequently, the accuracy of the simulation model cannot be validly assessed by comparison with the original data. Hopefully the data that was supplied for the financial variables does not contain similar errors.

In Table 5 the original and simulated correlations are compared for the financial variables. The simulated correlations are reported in the upper diagonal of the table. The values in the lower diagonal are the residual errors, that is, the differences between the simulated and the original correlations. On the whole the results are quite reasonable, especially considering that logarithmic transformations of some of the variables would have attenuated many of the correlations. Approximately 59 per cent of the residual errors between the original and the simulated correlations are less than 0.10. Only 14 per cent of the residuals are greater than 0.20 and these are associated with farm debt and rent payments. Given the errors in the original data we did not believe that it was worthwhile comparing the original and simulated correlations for the financial measures. However, as the correlations do provide some information on the linkages and dependencies between the various financial variables and measures we have reported the simulated correlations between all the variables and measures contained in the model in Appendix 2.

Table 4: Estimated moments of the financial performance measures.

	Estimated Mean	Estimated Variance	Estimated Skewness	Estimated Kurtosis
Farm income	26728.362	277029.709	2.277	9.788
Farm profit	-8817.240	359587.011	1.320	8.513
Profit at full equity	84424.113	356527.941	1.263	8.302
Profit at full equity (including capital appn)	-140355.695	501435.433	-1.140	7.490
Closing farm capital	2833422.37	3988612.960	4.671	30.812
Net rate of return	-0.777	25.850	-0.524	11.305
Net rate of return (including capital appn)	-7.793	28.721	-0.488	8.548
Farm equity	2163460.451	4094776.454	4.268	27.660
Equity ratio	47.954	221.609	-2.115	697.985
Debt servicing ratio	0.334	23.990	-47.770	3185.935

Simulation Experiments

To analyse the effects of changes in the moments of the distributions on farm performance we conducted a number of experimental simulations. First, we simulated the effect of a ten per cent increase and a ten percent decrease in mean total cash costs. For comparison we also simulated the effect of a ten per cent increase in farm cash receipts. The results of these experiments are summarised in Table 6. A ten per cent increase in cash costs is forecast to result in approximately a 12 per cent increase in the number of enterprises with negative farm incomes and negative profits. The proportion of farms with a debt servicing ratio between zero and one is forecast to fall by approximately 16 per cent. Note that the impacts of a ten per cent reduction in mean farm cash costs are not the exact reverse of the impacts of an increase in costs. This asymmetry is a reflection of the skewness of the distributions. A ten per cent increase

Table 5: Comparison of the original and estimated correlations for the financial variables.

	Trading stocks	Capital appreciation	Change in debt	Cash costs	Depreciation	Farm debt	Opening farm capital	Cash receipts	Interest	Off-farm income	Operator labour	Rent
Trading stocks	1.0000	-0.3052	-0.0624	0.1349	0.4733	-0.1578	0.3278	-0.1742	-0.1143	0.0179	0.0469	-0.0905
Capital appreciation	0.1216	1.0000	-0.2678	-0.6597	-0.6780	0.4937	-0.7524	-0.5433	-0.3086	0.1768	0.0757	-0.4429
Change in debt	0.1942	-0.0043	1.0000	0.1942	0.1356	0.4809	0.15[illegible]	0.1013	0.3053	0.0726	-0.2839	0.1566
Cash costs	0.0604	0.0702	0.0730	1.0000	0.7249	0.3676	0.7340	0.8732	0.[illegible]682	-0.2228	-0.2002	0.6193
Depreciation	0.0399	0.12[illegible]	-0.1083	0.1164	1.0000	0.0515	0.7910	0.5580	0.0345	-0.1517	-0.0798	0.4074
Farm debt	-0.0834	0.3125	-0.0870	-0.2269	-0.0469	1.0000	0.2491	0.3058	0.9577	0.0030	0.1273	0.1155
Opening farm capital	0.0065	0.2187	-0.047[illegible]	0.0408	0.0142	-0.2238	1.0000	0.719[illegible]	0.0463	-0.2141	-0.1699	0.5456
Cash receipts	0.0009	0.0289	0.11[illegible]4	0.0746	0.1654	0.1886	0.0167	1.0000	0.0755	-0.2439	-0.2141	0.7394
Interest	0.0228	0.1914	-0.0[illegible]	-0.1217	0.0579	0.6195	0.0581	0.0513	[illegible]00	0.0155	0.2013	-0.0712
Off-farm income	-0.1095	0.0257	0.0[illegible]	0.0009	0.1594	-0.0830	-0.0278	-0.1080	-0.0416	1.0000	-0.2641	-0.2411
Operator labour	0.0278	0.0626	0.[illegible]9	[illegible]567	0.0009	0.0539	0.0886	0.1690	-0.0651	0.0598	1.0000	-0.1859
Rent	0.1585	0.[illegible]381	0.0284	[illegible]270[illegible]	0.3282	0.0743	0.1698	0.2095	0.0510	-0.0630	-0.2108	1.0000

Note: Values below diagonal are estimated correlations; values above diagonal are residual errors.

in farm cash receipts is predicted to have largely the same impact as a ten per cent decrease in cash costs. These results suggest that changes in mean revenues and costs tend to have a more than proportionate impact on critical financial measures.

Table 6: Predicted effects of changes in means of receipts and cash costs.

	Base simulation (%)	10% increase in mean receipts (% change)	10% decrease in mean cash costs (% change)	10% increase in mean cash costs (% change)
Farms with negative farm income	60.4	-11.8	-12.9	12.8
Farms with negative farm profit	55.4	-12.3	-12.7	14.5
Debt servicing ratio equal or greater than zero and less than one	42.6	16.4	17.9	-17.4
Farms 'at risk'	21.3	-11.5	-12.1	10.3

Note: 'At risk' farms are defined as those with negative farm cash income and equity less than 70 per cent.

We also simulated the effects of a ten per cent increase and decrease in the variance of farm cash receipts which could be interpreted as the result of different seasonal conditions. The results of this simulation are reported in Table 7. Again, the predicted impacts on measures of farm performance are asymmetrical. Since the distribution of farm cash receipts has a positive mean and skewness an increase in variance can be expected to have a negative impact on profit. A ten per cent increase in the variance of farm cash receipts is predicted to raise the number of farms with negative income by four per cent. However, the number of farms earning negative profits only changes marginally. This is due to the fact that the simulated correlation between cash receipts and farm business profit (0.27) is substantially weaker than the simulated correlation between cash receipts and farm cash income (0.66).[7] Basically, the lower correlation with business profit attenuates the impact of the change in the variance of cash receipts. The proportion of farms with debt servicing ratios between zero and one is forecast to fall by five per cent. In contrast, a ten per cent decrease in variance is predicted to reduce the number of farms with negative farm cash income by 12 per cent and to reduce the numbers of farms 'at risk' by almost 13 per cent. These results indicate that the financial performance of beef cattle operations in the Northern Territory are just as

[7] See Appendix A for a tabulation of the simulated correlations between all the variables and measures in the model.

sensitive to changes in the variance of farm revenues as they are to changes in mean farm revenue.

Table 7: Predicted effects of changes in distribution of receipts.

	Base simulation (%)	10% increase in variance of receipts (% change)	10% decrease in variance of receipts (% change)	Normally distributed receipts (% change)
Farms with negative farm income	60.4	3.9	-12.2	-33.6
Farms with negative farm profit	55.4	1.0	0.9	-21.2
Debt servicing ratio equal or greater than zero and less than one	42.6	-5.0	14.1	45.6
Farms 'at risk'	21.3	4.3	-12.7	-40.2

Note 'At risk' farms are defined as those with negative farm cash income and equity less than 70 per cent

We also evaluated the effect of assuming that farm cash receipts were normally distributed. For this simulation the mean and variance of cash receipts, and the correlation between receipts and the other variables in the model were unchanged. The results of this experiment are also summarised in Table 7. The number of farms with negative incomes and profits falls by 34 per cent and 21 per cent respectively. The number of farms with debt servicing ratios between zero and unity increases by 46 per cent and the number of farms 'at risk' is reduced by 40 per cent. These results suggest that invoking the assumption of normality in order to simplify analysis can have a major impact on analytical results and any policy implications drawn from such results. In this particular instance, invoking this assumption leads to a substantial under-estimation of the numbers of farms that may be experiencing financial difficulty.

Finally, we simulated the effect of a ten per cent change in the skewness and kurtosis of cash receipts. As mentioned earlier, reasonably large changes in the values of these moments have only marginal impacts on the shape of distributions. We found that a ten per cent change in the these moments produced results which were almost identical to those obtained in the base simulation. In other words, relatively small changes in the skewness and kurtosis of cash receipts are predicted to have little, if any, impact on farm cash income, profit and so on. The differences between this result and the results obtained when cash receipts were treated as being normally distributed suggests that the effects of changing skewness and kurtosis are not linearly related to those changes.

Small changes have negligible effects while larger changes have, increasingly, proportionately greater impacts.

In our opinion these results indicate that the simulated model is capable of reproducing the original data with a high degree of accuracy. Unfortunately, the errors in the original data that were supplied prevents a proper evaluation of the true accuracy of the techniques that we have used. Despite the inadequacies of the data, in our view, the results suggest that this approach to simulating distributions is powerful and well worth exploring

Conclusion

At present there is little in the way of formal, systematic procedures for generating forecasts of the distribution of variables from cross-sectional data. Consequently, quantitative predictions about how changes in the farming environment might affect the numbers of farms in an industry that are experiencing severe financial stress are difficult to make. Yet such predictions would be valuable for policy makers and others who need information about the way in which changes in commodity prices, interest rates r seasonal conditions might affect specific segments of the farm population such as those on low incomes or with high debts.

In this study a new approach to forecasting which can be used to make short term predictions about the distribution of the financial performance of farms was explored. The technique developed by Vale and Maurelli (1983) was adapted to provide short-term predictions of the distribution of financial performance across farms in the beef industry in the Northern Territory. In principle, the method can be employed to create forecasts of financial performance of agricultural enterprises in other industries, either on a state by state or national basis.

The approach we have explored in this study has a number of appealing features. First, the approach generates or replicates data distributions. In other words, the method is designed to generate predictions of the distribution of farm income for example rather than point forecasts of 'average' farm income. Second, the approach is specifically designed for use with data which are not normally distributed. The distributions of many physical and financial variables are likely to be asymmetrical around the mean, that is, non-normal. Third, the approach can be applied when distributions are correlated. The method provides a means of generating formal quantitative forecasts of financial performance which preserves the relationships (correlations) between the variables that determine financial performance. Fourth, the approach could, in principle, be integrated with econometric, programming or informal approaches.

This method has the potential to allow more accurate and detailed assessments to be made about the way in which changes in commodity prices, interest rates or seasonal conditions might affect specific segments of the farm population. The approach could, for example, be employed to construct a more detailed model in which the distributions of indicators of physical performance such as livestock sales and the various components of cash costs were reproduced. Total farm cash receipts and cash costs would be derived as identities in such a model. This would allow the impacts of quite specific changes in revenues and costs to be analysed.

An interesting and valuable avenue for further research is to validate the accuracy of this approach to making forecasts. The analysis we have presented here could be repeated using data from another year with distributions exhibiting different characteristics from those used in this work (for example, a year in which prices were higher or farm revenues and costs were more variable). The results of such an analysis could be used to assess the accuracy of the predictions obtained in the simulation experiments we have conducted.

In conclusion, in our view, the results we have obtained indicate that the approach developed by Vale and Maurelli (1983) can be employed to obtain useful insights into the effects that changes in the business environment can have on the distribution of financial performance across farms in an industry. Consequently, we believe that the method warrants further exploration and refinement.

References

A.B.A.R.E. (1994). *Farm Surveys Report 1994*. A.B.A.R.E. Canberra.

Allen, P.G. (1994). "Economic forecasting in agriculture", *International Journal of Forecasting* 10(1994), 81-135.

Anderson, J.R. (1983). *Methods and Programs for Analysis of Risky Gross Margins.* Miscellaneous Publications No. 5. University of New England, Armidale.

Anderson, J.R., Dillon, J.L. and Hardaker, B. (1977). *Agricultural Decision Analysis.* Iowa State University Press. Ames. Iowa.

Fleishman, A.L. (1978). "A method for simulating non-normal distributions", *Psychometrika* 43(4), 521-31

Francisco, E.M (1980). *Regional Rainfalls, Crop Yields and Farm Incomes in New South Wales - A Simulation Approach.* APMAA Report No. 9. Department of Agricultural Economics and Business Ma. .gement. U.N.E. Armidale.

Greene, W.H. (1993). *Econometric Analysis*. (2nd ed.). MacMillan. New York.

Johnson, D. (1992). "An investigation of changes to pricing arrangements in the dairy industry using a general equilibrium model - simulations of Orani-f-milk." Paper presented to the 36th Annual meeting of the Australian Agricultural Economics Society. Canberra.

Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley and Sons. New York.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H. and Lee, T-C. (1985). *The Theory and Practice of Econometrics* 2nd ed. John Wiley and Sons. New York.

Just, R.E., Zilberman, D. and Rausser, G.C. (1991). "Modelling policy reform in the U.S. wheat and feed grain sectors." in G.C. Rausser (ed.). *GATT Negotiations and the Political Economy of Policy Reform* Springer-Verlag, Heidelberg.

Kaiser, H.F. and Dickman, K.(1962). "Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix", *Psychometrika* 27(2), 179-82.

Kendall, M., Stuart, A. and Ord, J.K. (1987). *Kendall's Advanced Theory of Statistics*. (5th ed.) Charles Griffin and Co. London.

Larson, H.J. (1982). *Introduction to Probability Theory and Statistical Inference*. (3rd ed.).John Wiley and Sons. New York.

McClintock, D.N., Ahmadi-Esfahani, F.Z., and Johnston, J.H. (1991). "A Simulation of the grain producers' decision problem at harvest", *Review of Marketing and Agricultural Economics*. 59(3), 271-282.

Mittelhammer, R.C. (forthcoming) *A Graduate Primer in Mathematical Statistics for Business and Economics Students*. Springer-Verlag, New York.

Parton, K.A. (1979). *An Appraisal of Policy Evaluation Methodologies in an Agricultural Sector Linear Programming Model: Australian Wool Corporation Reserve Price Policies and the APMAA Model.* APMAA Report No. 14. Department of Agricultural Economics and Business Management. U.N.E. Armidale.

Pearson, E.S. and Please, N.W. (1975). "Relation between the shape of population distribution and the robustness of four simple test statistics", *Biometrika* 62(2), 223-41.

Tabachnick, B.G. and Fiddell, L.S. (1989). *Using Multivariate Statistics*. 2nd ed. Harper and Row. New York.

Todd, R. and Roe, T. (1978). *An Introduction to Iterative Methods for Solving Simultaneous Equation Models: Illustrations from Economics*. Staff paper P78-11, Department of Agricultural and Applied Economics. University of Minnesota. St. Paul.

Vale, C.D. and Maurelli, V.A. (1983). Simulating multivariate non-normal distributions", *Psychometrika* 48(3), 465-71

Appendix 1

Figure A1.1: Original and estimated distributions for build-up in trading stocks.

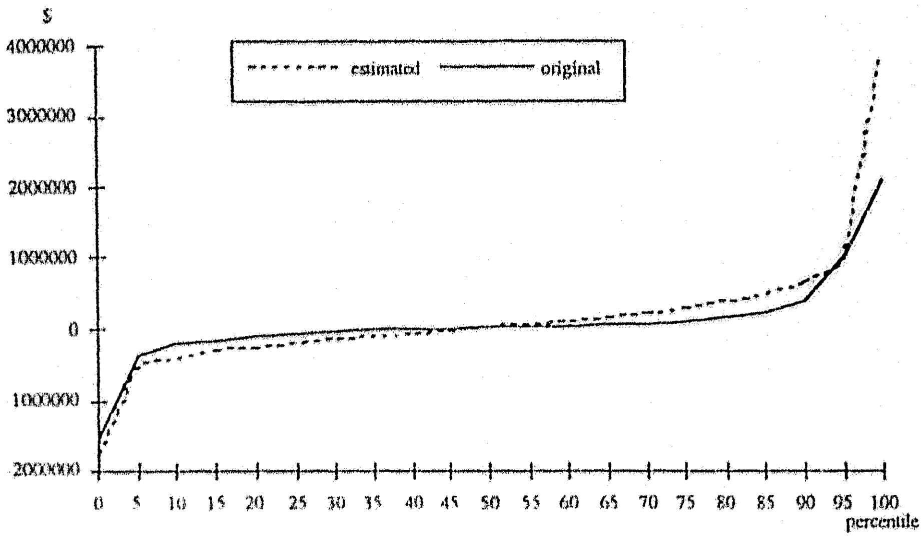


Figure A1.2: Original and estimated distributions for capital appreciation.

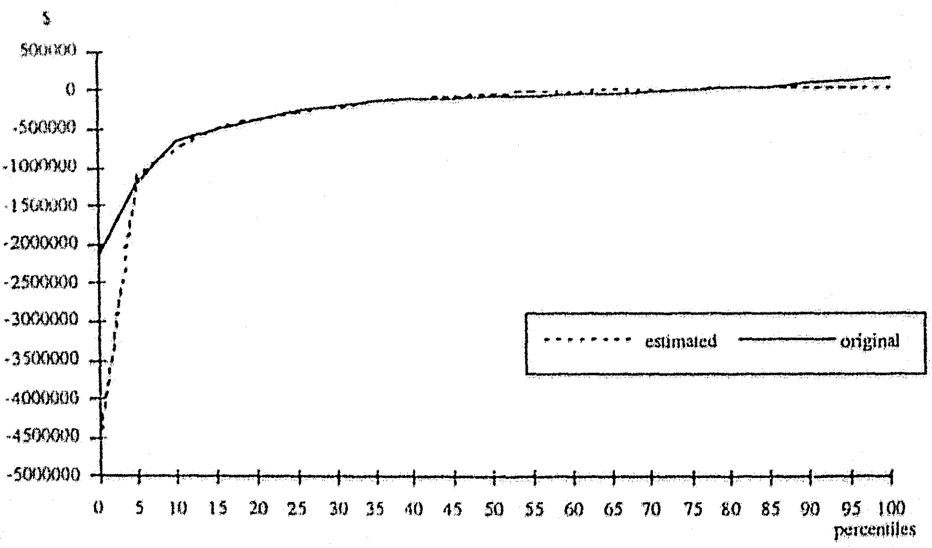


Figure A1.3: Original and estimated distributions for change in debt.

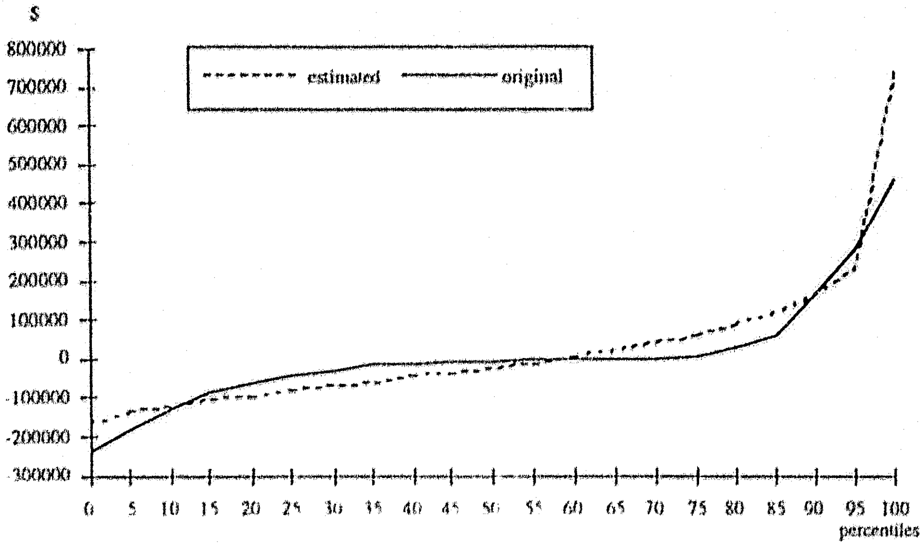


Figure A1.4: Original and estimated distributions for cash costs.

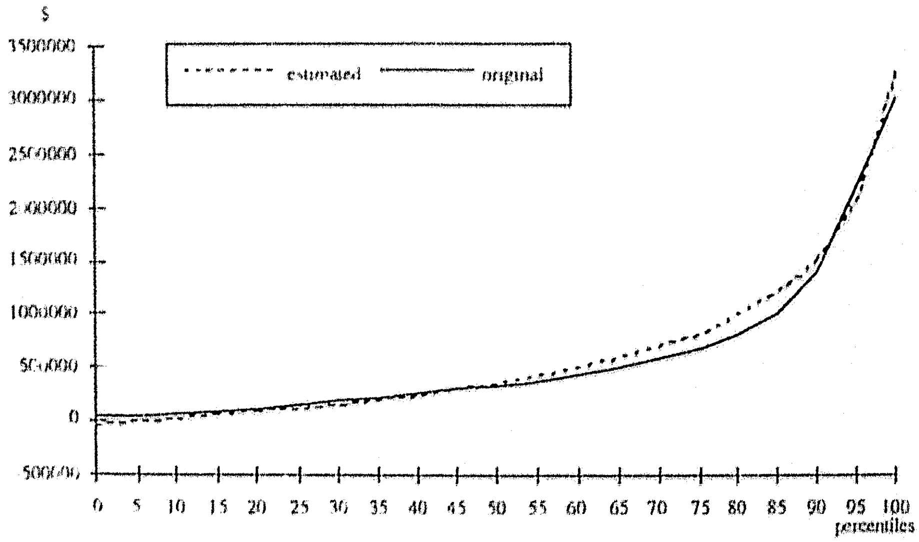


Figure A1.5: Original and estimated distributions for depreciation.

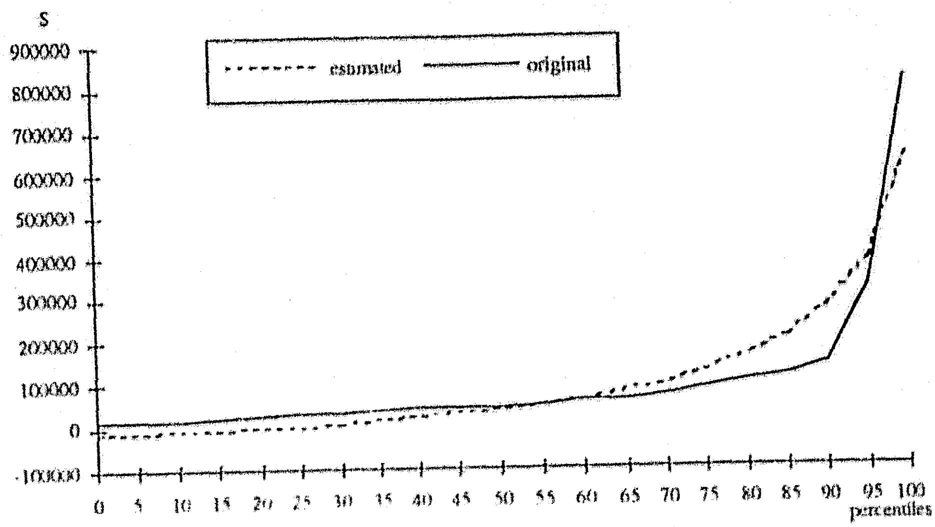


Figure A1.6: Original and estimated distributions for farm debt.

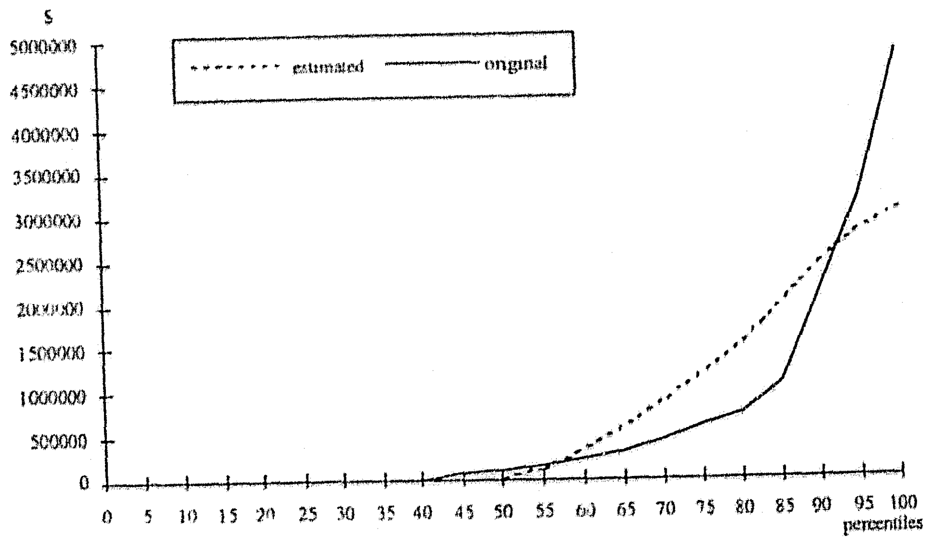


Figure A1.7: Original and estimated distributions for farm capital (opening).

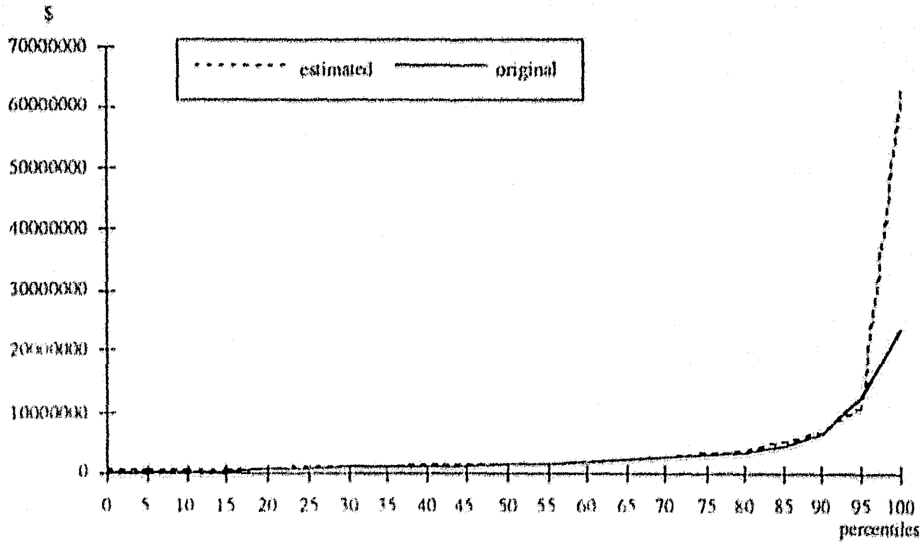


Figure A1.8: Original and estimated distributions for cash receipts.

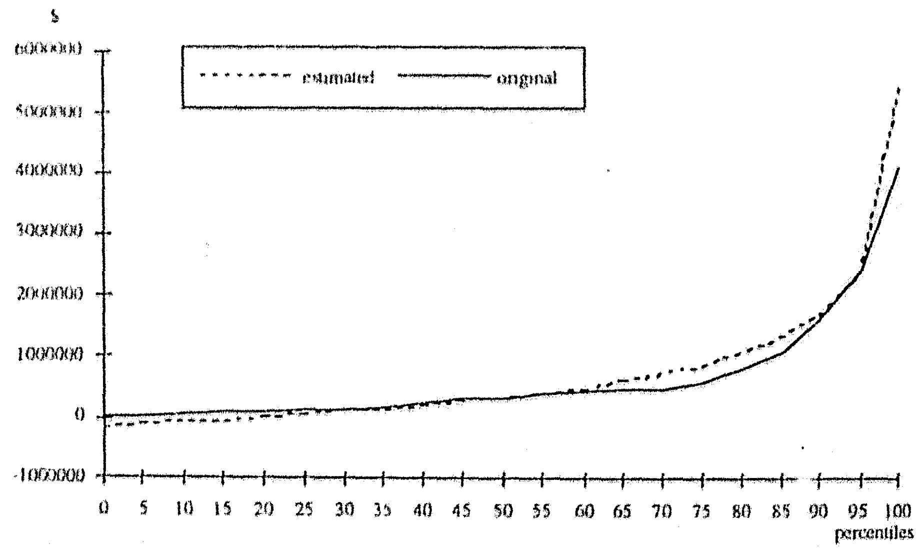


Figure A1.9: Original and estimated distributions fc - interest payments.

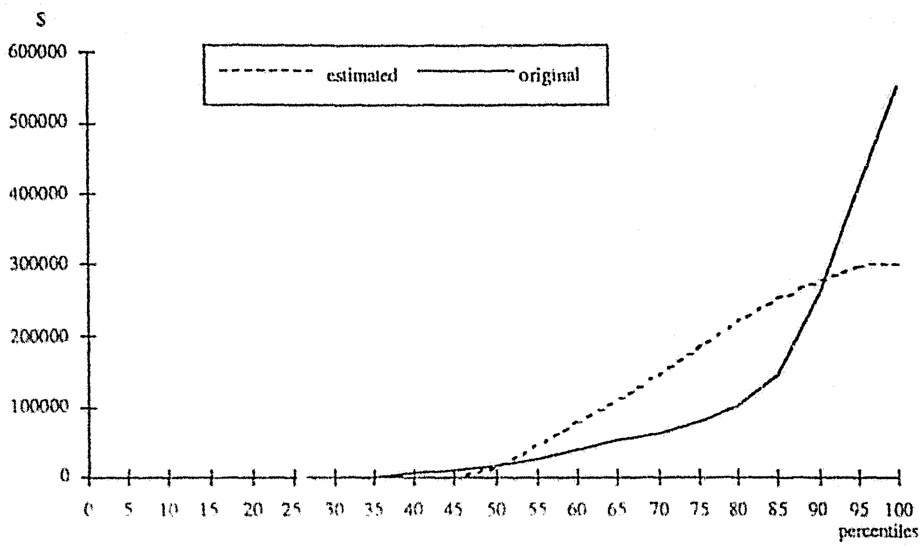


Figure A1.10: Original and estimated distributions for off-farm income.

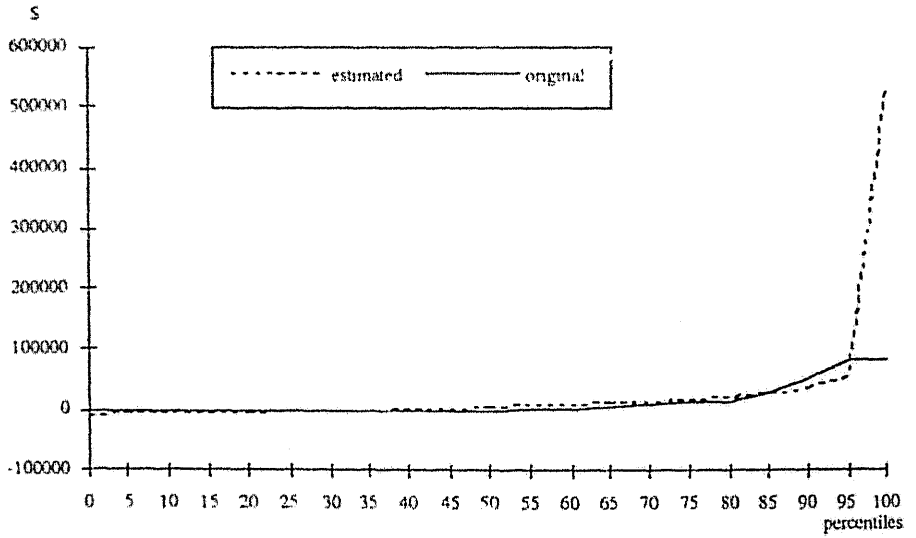


Figure A1.11: Original and estimated distributions for operator labour.

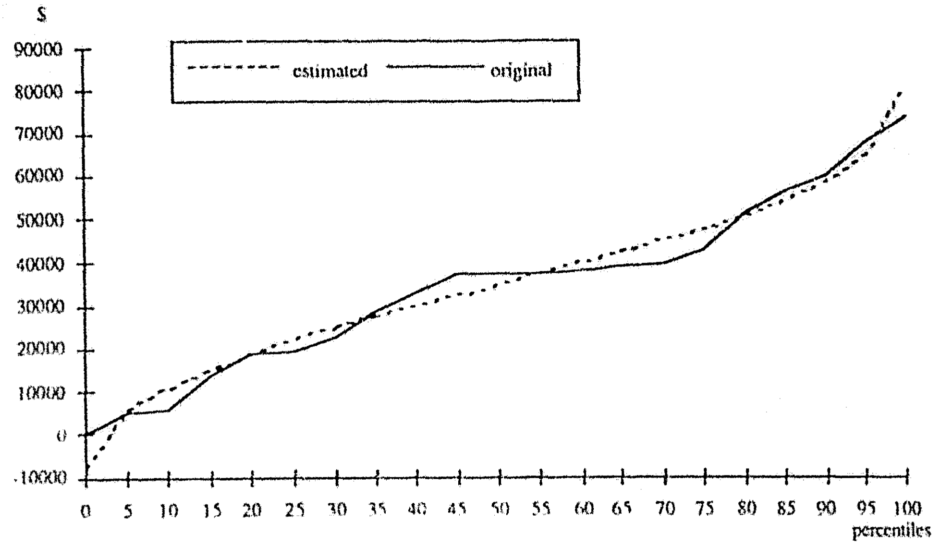
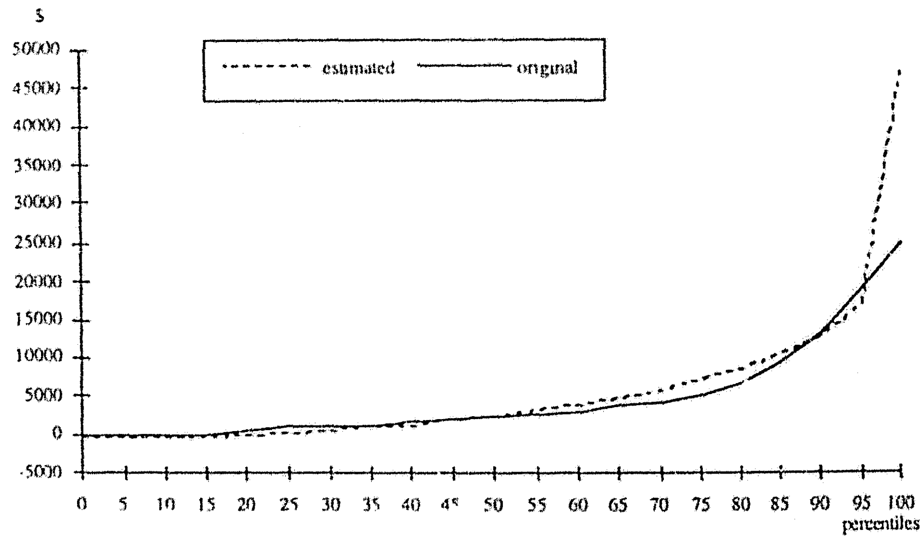


Figure A1.12: Original and estimated distributions for rent paid.



APPENDIX 2: Correlations of variables and measures

	Trading stocks	Capital appreciation	Change in debt	Cash costs	Depreciation	Farm debt	Opening farm capital	Cash receipts	Interest	Off-farm income	Operator labour	Rent	Income	Debt serv. ratio	Profit	Profit at full equity	Profit at full equity inc. cap.app.	Net rate of return	Net rate of return inc. cap.app.	Closing farm capital	Equity	Equity ratio
Trading stocks	1.0000																					
Capital appreciation	-0.1836	1.0000																				
Change in debt	-0.2566	-0.2721	1.0000																			
Cash costs	0.1953	-0.5395	0.2672	1.0000																		
Depreciation	0.5132	-0.5551	0.0273	0.8413	1.0000																	
Farm debt	-0.2412	-0.1812	0.3939	0.1407	0.0046	1.0000																
Opening farm capital	0.3213	-0.5337	0.1116	0.7743	0.8052	0.0253	1.0000															
Cash receipts	0.0134	0.5144	0.2167	0.9478	0.7234	0.1172	0.7363	1.0000														
Interest	-0.1371	-0.1172	0.2405	0.0465	-0.0234	0.3382	[illegible]	0.0242	1.0000	–												
Off-farm income	-0.0916	0.2025	0.0276	-0.3512	-0.3111	-0.0800	-0.2419	-0.3519	-0.0261	1.0000												
Operator labour	0.0191	0.1383	0.3831	-0.3569	-0.2064	0.0734	-0.2585	-0.3831	0.1362	-0.2043	1.0000											
Rent	0.0610	-0.4810	0.1850	0.8893	0.7356	0.0412	0.7134	0.9459	-0.0202	-0.3041	-0.3967	1.0000										
Income	-0.4095	-0.1171	0.0063	0.4135	0.1368	0.0123	0.3261	0.6823	-0.0373	-0.1836	0.2734	0.6706	1.0000									
Debt servicing ratio	0.0148	-0.8042	0.0600	0.0049	0.0119	0.0353	0.0044	0.0016	0.0352	0.0053	0.0113	0.0155	-0.0066	1.0000								
Profit	0.7665	-0.1734	[illegible]	0.2565	0.4117	0.3607	0.3863	0.3065	0.2007	-0.1179	-0.1630	0.3630	0.2183	0.0039	1.0000							
Profit at full equity	0.7313	-0.1797	-0.2345	0.3176	0.4196	-0.1979	0.3772	0.3318	0.1096	-0.1515	-0.1302	0.3757	0.2151	0.0199	0.9517	1.0000						
Profit at full equity inc cap app	0.3653	0.7146	-0.3957	0.2704	[illegible]	0.2926	[illegible]	0.1971	0.0207	0.0627	0.0238	0.1378	0.0572	0.0107	0.5644	0.5605	1.0000					
Net rate of return	0.6582	0.0731	-0.3115	[illegible]	[illegible]	[illegible]	0.0597	0.6926	[illegible]	0.1270	0.0305	[illegible]	0.0160	0.0103	0.7584	0.8036	0.5098	1.0000				
Net rate of return inc cap app	0.5860	0.2375	0.3053	0.0354	[illegible]	[illegible]	[illegible]	[illegible]	0.0448	0.0672	0.0315	0.0626	-0.0978	0.0059	0.6093	0.7203	0.7120	0.8816	1.0000			
Closing farm capital	[illegible]	0.4571	0.0887	0.7531	0.1857	[illegible]	[illegible]	0.7205	[illegible]	0.2132	0.2575	[illegible]	0.3306	0.0042	0.3924	0.3992	-0.0959	0.0972	0.1147	1.0000		
Equity	0.3669	[illegible]	-0.0036	0.7007	0.7672	[illegible]	0.9642	0.6745	0.1032	[illegible]	0.2679	0.6742	0.3194	0.0005	0.4535	0.4349	-0.0239	0.1345	0.1665	0.9723	1.0000	
Equity ratio	0.1229	[illegible]	[illegible]	0.0871	0.1065	[illegible]	[illegible]	[illegible]	0.1079	0.0525	0.0771	0.0965	0.0418	-0.0046	0.1535	0.1230	0.0775	0.1185	0.1111	0.0911	0.1666	1.0000