ABARE CONFERENCE PAPER 94.5

Risk and returns in Australian broadacre agriculture

Jane Wright, Stephen Beare, Vivek Tulpulé and Philip Kokic
Australian Bureau of Agricultural and Resource Economics
GPO Box 1563, Canberra 2601

38th Annual Conference of the
Australian Agricultural Economics Society
Wellington, New Zealand, 8–10 February 1994

The objective in this paper is to present a method of estimating farm level probability distributions for income and other variables that uses spatial smoothing of cross-sectional survey data. The method is applied to ABARE's Australian agricultural and grazing industries survey (AAGIS) data for 1978-79 to 1991-92 to obtain farm level measures of mean income and the standard deviation and skewness of income. These quantities provide measures of the overall risk faced by broadacre farm enterprises throughout Australia. Estimated probability distributions for each farm can be expected to vary given the physical and structural diversity of Australian broadacre agriculture.

ABARE project 1227.103

1. Introduction

In this paper a method is developed for assessing farm level risk using ABARE's Australian agricultural and grazing industries survey (AAGIS) data. The technique uses spatial smoothing of cross-sectional farm level data to produce temporal farm level distributions. From these, it is possible to estimate the mean, standard deviation and skewness of financial and production measures such as farm cash income or cropping yields. The results presented here are obtained using the modelling techniques developed in recent ABARE research.

2. Background

Policy changes and risk management

The role of risk management has taken on increased importance over the past decade as a result of changes in Australian economic policies at the national, sector and industry levels. Broad macroeconomic changes likely to have affected the risks facing farmers include the float of the Australian dollar on world currency markets in 1983 and the removal of interest rate ceilings in April 1985. At the same time, a greater variety of financial packages have become available for farmers, enhancing their ability to adjust debt levels to achieve some desired mix of debt and equity financing (Peterson, Dunne, Morris and Knopke 1991, p. 353).

The orientation of government policy in agriculture has continued to move away from compensation and assistance toward policies which provide incentives for effective farm management and adjustment. For example, revisions to the income equalisation deposit scheme were introduced to allow primary producers to better manage income variability. Problems associated with climatic variability led to the formal announcement of a national drought policy (NDP) in August 1992. The policy puts forward the view that drought is part of the uncertainty of Australian agriculture, and encourages a self-reliant approach to managing the risks associated with climatic variability. Similarly, the most recent changes to the Rural Adjustment Scheme remove its emphasis on assistance and debt reconstruction to focus on improving farm productivity, profitability and sustainability (Crean 1992). The new RAS measures give consideration to a farmer's technical, financial and business competence, including risk management.

At the industry specific level, important changes have been introduced in the Commonwealth marketing policies which affect risk. Deregulation of domestic grain marketing arrangements has given farms greater scope to adopt marketing strategies involving varying degrees of risk. The abolition of the Reserve Price Scheme for wool has effectively transferred the onus for managing price risk to individual growers and processors.

The importance of information

Information about the probability distributions associated with financial and production variables is critical to the development of any market for facilitating the management and transfer of farm business risk.

At each stage of the production process farmers, marketing agents and processors are exposed to both price and production risks. Each of these groups will demand risk management services according to their understanding of the extent and nature of the risks they face.

A lack of information concerning risks faced or of the risk management services available may lead to a less than optimal supply of risk management services. Also, financiers must be able to assess the risks associated with providing financial services to farm enterprises. Information about the riskiness of income streams from farming will also help in the assessment of equity financing arrangements for farm enterprises.

In accepting that risk management must be integrated into the physical and financial operation of the farm enterprise, it is clear that information requirements of providers and receivers of risk management services can be highly specific. The type of information required could include: the financial status of the enterprise (for example, the level of farm equity); the managerial capacity of the operator (for example, understanding of the liabilities associated with hedging); and the geographic and climatic conditions affecting the farm operations.

While the ultimate effect of price and production risks on grower returns may be the same, the tools available for managing such risks are quite different. At the same time, the use of any individual risk management instrument is tied to the overall operation of the farm enterprise.

3. Risk

A common definition of risk is elusive, although Barry, Hopkins and Baker (1988) provide the following description of financial and business risk in agriculture:

> Business risks for farmers include (1) production and yield risks; (2) market and price risks; (3) losses from severe casualties and disasters; (4) social and legal risks from changes in tax laws, government programs, trade agreements, and so on; (5) human risks on performance of labour and management; and (6) risks of technological change and obsolescence. Financial risks arise from the financial claims on the firm . . . [and] are further increased by unanticipated variations in interest rates, credit availability, and other changes in loan terms, as well as in leasing terms.

One of the most widespread measures of the 'riskiness' associated with a random variable is its variance. Particularly in the context of mean–variance analysis to analyse behaviour in risky situations. In this type of analysis risk averse agents trade off reductions in the variance of income or wealth with reductions in mean income; that is, a tradeoff would exist between risk and return, with individuals willing to accept a greater level of variability in returns only if the mean rate of return is greater.

Such tradeoffs may be established within the financial and enterprise structure of farms (for example, through the level of debt financing or enterprise diversification), although the tradeoffs may be limited by geographic and climatic conditions. In some areas physical constraints may limit the ability of farms to tradeoff mean rates of return with variability in rates of return. This may partly explain the quite distinct patterns in regional land values seen in Australia (Topp and Beare 1993).

In recent years understanding of the importance of skewness of returns has grown. In particular, Machina (1982), in his development of generalised expected utility theory, has shown that agents might be willing to trade off a mean-preserving spread for an increase in right skewness.

The mean–variance approach has been criticised on a number of grounds. For example, quadratic utility functions (on which mean–variance analysis can be based) displays increasing risk aversion in wealth. Other criticisms have been outlined by Borch (1969). The more general approach of analysis of behaviour under risk using stochastic dominance procedures (see Meyer and Ormiston 1983) can depend on knowledge of the entire probability distribution function. For example, to judge whether one distribution can be obtained from another by a series of mean-preserving spreads.

In this paper, a technique for measuring farm level probability distributions is developed which focuses on the geographic and climatic factors which affect farm performance. The technique makes extensive use of temporal and spatial averaging to develop farm specific distributions. The spatial smoothing aspect of the technique utilises farm locations to estimate local averages. As enterprise opportunities are regionally correlated, enterprise composition is explicitly included in the estimation process. However, the technique does not incorporate managerial and financial considerations (such as operator skill, equity levels, non-farm investments etc.), as these are unlikely to be regionally correlated.

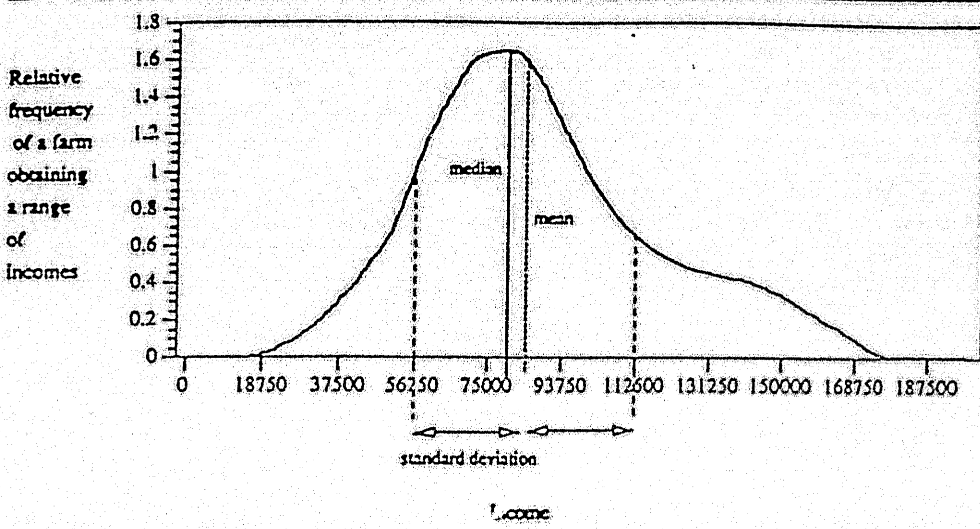
Spatial smoothing algorithms are employed to impute a panel structure for ABARE's farm survey data, and this is subsequently used to generate farm level distribution functions. The functions quantify temporal variation in physical and financial variables, such as receipts, income, rates of return, and crop yields.

From the individual farm distribution functions, locationally specific estimates of mean and variance and skewness of the distributions can be calculated and compared.

An example of a farm's income distribution is shown in figure 1. It shows the relative frequency of a farm obtaining a range of incomes. Also marked on the graph is the mean, standard deviation and the median.

The *median* of the distribution is the measure that has 50 per cent of the distribution below it and 50 per cent above. If the mean is equal to the median then the distribution is said to be symmetric. However, in the diagram the mean lies to the right of the median (or is greater than the median). This suggests that there are a few very large values in the distribution and this is indicated by the long right hand tail. Such a distribution is right skewed. A left skewed distribution has a mean less than the median inducing a long left hand tail. The *skewness coefficient* is a measure of the degree of this skewness. A skewness coefficient of zero indicates a symmetric distribution, a negative skewness coefficient measures a left skewed distribution and a positive skewness coefficient measures a right skewed distribution.

Figure 1: Characteristics of an income distribution



4. Constructing temporal farm level distributions from survey data

The analytical technique used to estimate these income distributions is summarised in this section and a more technical explanation is provided in the appendix. A description of the results obtained by applying these techniques to physical and financial data collected in AAGIS is given in section 5.

Data collected in the Australian agricultural and grazing industries survey (AAGIS) between 1978-79 to 1991-92 were used to estimate the probability distributions obtained here. The aim is to estimate the temporal income distribution of farms using repeated survey data. There is an efficiency/bias tradeoff problem due to the incomplete structure of the data set. Information for a sampled farm is collected for only a few years at most and in any given year data for only a subset of all farms are observed. The proposed imputation process offers efficiency gains of the cost of a potential bias.

Australian agricultural and grazing industries survey data

The AAGIS target population is the broadacre farming sector, which comprises farms engaged mainly in the production of cereals, oilseeds, grain legumes, sheep, wool and beef cattle (excluding small and hobby farms). Farms which are primarily involved in the dairy, sugar, horticultural, cotton, vegetable and other agricultural industries are not covered by this survey.

ABARE's large scale annual survey of broadacre farm financial performance provides consistent data from the late 1970s until the present. In total there were approximately 80 000 farms in 1991-92 AAGIS target population. The target population varies in size over time and the size of the various industries also changes.

The data used in this paper cover the fourteen financial years from 1978-79 to 1991-92. In most years, between 800 and 1000 farms were surveyed. Individual farms usually remain in the survey for more than one year. In total, 4213 farms were surveyed at some time during this period.

In any particular year the population is stratified and the sample is selected in varying proportions from the strata so as to optimise the accuracy of survey estimates. To compensate for the non-uniform structure of the sample a survey weight is calculated and used when forming estimates. This weight indicates roughly how many farms in the population the sample farm represents.

The industry classification of farms may change from one year to the next for various reasons including changes in enterprise composition and receipts. In order to apply the techniques developed in this paper at industry level it is necessary to classify farms to their 'most frequent industry' — that is, to the broadacre industry the farm is most often classified as being a part of during the time the farm was in the sample.

Smoothing

The procedure for estimating farm level probability distribution functions relies on combining time series data of the variable of interest for each farm. However, because farms are rotated out of the sample, there are data limitations for determining the time series data for any particular farm. Spatial smoothing allows the estimation of a time series for the entire period 1978-79 to 1991-92, from the time series of neighbouring farms in the sample. A heuristic explanation of the procedure used to estimate the probability distribution functions is provided here and a more complete technical discussion is offered in the appendix.

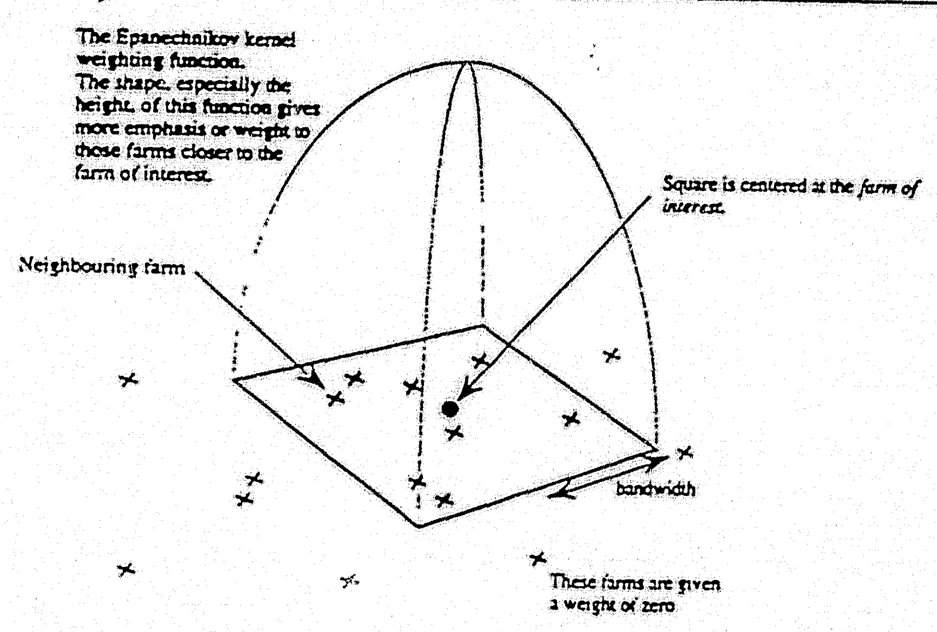
Spatial smoothing is appropriate when sample farms that are 'close together' are more alike in terms of their data values than those that are 'far apart' — that is, the values of 'close' farms are positively correlated. Cross-sectional data obtained from farms whose locations are known are ideally suited to spatial smoothing.

Regional trends are estimated using data mainly from farms that are geographically close to that farm, thus reflecting similarities in the data related to location. Farms that are close to a particular farm have more influence on the local average for that farm since they are given more weight in the calculation of this average.

Figure 2 shows a small 'neighbourhood' of sample farms. The objective is to estimate the income distribution of the *farm of interest* using a weighted combination of the income of its nearest neighbours.

The smoothing method assumes there is a smooth nonlinear relationship between farm incomes and the geographic location of farms that is characterised by the following equation:

Figure 2: Diagram of spatial smoothing procedure



(1) $$Y = f(x) + \varepsilon$$

where Y is the variable of interest, x is the two-dimensional vector specifying the location of the farm, f is a smooth, but nonlinear, function of x with an error term ε.

The general form of the kernel estimator $\hat{f}(x)$ of f is:

$$\hat{f}(x) = \sum_{j=1}^{n} \omega_j(x) Y_j$$

where

$$\omega_j(x) = \frac{\Omega_j K\left(\frac{\|x - X_j\|}{h}\right)}{\sum_{k=1}^{n} \Omega_j K\left(\frac{\|x - X_j\|}{h}\right)}$$

$$K(t) = \begin{cases} c\left(1 - t^2\right) & \text{if } |t| \leq 1 \\ 0 & \text{otherwise for some constant } c. \end{cases}$$

where $\omega_j(x)$ is the smoothing weight of sample observation j at location x. K is the kernel or weighting function and h is the bandwidth of the smoother and Ωj is the sample weight of the surveyed farm.

The kernel function $K(t)$ used here is the Epanechnikov kernel (see Härdle 1990; Nelson and Pope 1991.) As illustrated in the figure, its parabolic shape gives greater emphasis to farms closer to the farm of interest and less to those further away. This function is only applied to farms in the square. Any farms outside this box receive a weight of zero. The box is moved from one farm to the next with the procedure repeated at each location.

The square centred at the farm of interest has a sidelength of twice the bandwidth. The bandwidth is a distance and is subjectively selected according to the desired amount of spatial smoothing. A large bandwidth will tend to smooth out any local variation in farm income while a bandwidth that is too small will only allow a few farms to contribute to the averaging and potentially result in an estimate with too much 'white noise'. The optimal bandwidth lies somewhere between these two extremes.

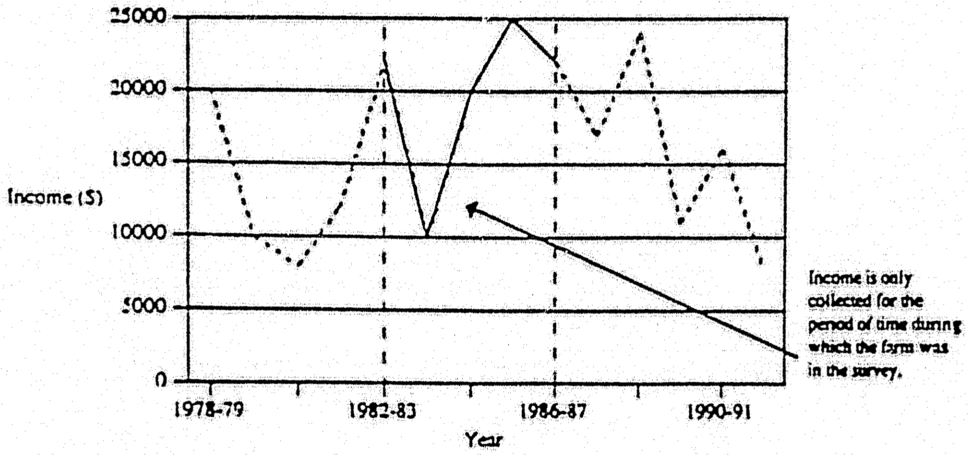
It was necessary to include the sample weight in order to adjust for the unbalanced nature of the sample against known population characteristics.

In years during which a farm is out of sample, the smoothing technique allows the variables of interest for that farm to be imputed. In figure 3, for example, the farm was in the survey between 1982-83 and 1986-87. Smoothing allows mean income, or indeed any other continuous variable, to be estimated for the years during which data were not collected for the farm in question.

The technique used in this paper involves smoothing not just to obtain an estimate of the mean value of the variable Y in any given year but to also give a non-parametric estimate of the probability distribution function from which Y is obtained. Such distribution functions are obtained for each of the 4213 farms for each year 1978-79 to 1991-92. In short, a time series of probability distribution functions for each farm is obtained.

Each farm's fourteen probability distributions are averaged to obtain an estimate of the temporal probability distribution function. Before shifting this distribution function so that it is centred at the mean, its variance and skewness must be adjusted for bias. A significant bias may result in the estimates of these quantities because of the short time that some farms remain in sample. Again refer to the appendix for a more complete technical explanation.

Figure 3: Example of a time series of a farm's income



5. Results

The focus in this section is on the business risks facing farms in the wheat and other crops industry. These farms are mainly involved in the production of cereal grains, coarse grains, oilseeds or grain legumes (ABARE 1993).

To illustrate the extent and nature of business risk facing cropping specialist farms, estimates for a range of financial and physical variables are reported. The financial variables are based on gross cash income; that is, the income available to meet living costs and debt repayment needs after meeting all other farm costs (farm cash income, defined in ABARE 1993 page 71, plus interest paid). All dollar amounts have been converted to 1991-92 dollars based on changes in the consumer price index. It should be noted that the technique outlined above is not specific to income and may be applied to any other continuous variable collected in the survey. Wheat yield, for example, is an AAGIS variable that is well suited to this technique and is also presented in the results.

Only general conclusions are drawn about some of the possible causes of the trends observed.

The results presented below include: the interpretation of risk distributions; maps of regional variation; exploratory data analysis; and preliminary cross-validation estimates. All results refer to producers in the wheat and other crops industry, except the analysis of wheat yields, which is based on the subset of producers in this industry with more than 100 hectares of wheat sown. The variables which are analysed are as follows:

- rate of return adjusted for depreciation for producers in the wheat and other crops industry; that is, the ratio of gross cash income less depreciation of farm improvements, plant and equipment, to the total gross value of capital (ABARE 1993, p. 70);

- the ratio of gross cash income to closing land area for producers in the wheat and other crops industry; and

- wheat yields for producers in the wheat and other crops industry with greater than 100 hectares of wheat sown.

Exploratory data analysis is performed for gross cash income in the wheat and other crops industry while cross-validation estimates are provided for the mapped variables as well as gross cash income.

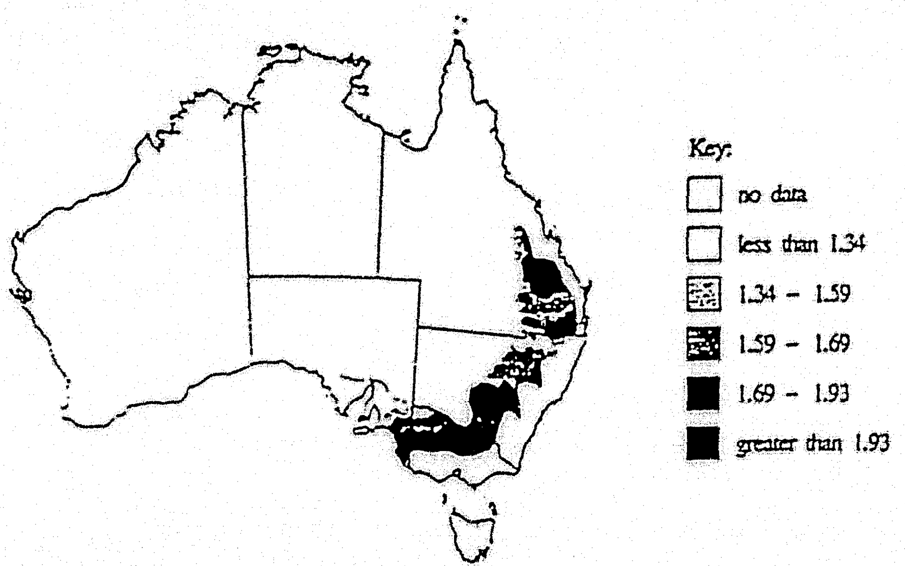
Mapping regional variation in financial and physical variables

The regional variation in income is often obscured in tabulated data. As a result, ABARE developed a method of mapping these regional averages which relies on the farm locations to identify regional changes. Areas around each farm are defined using Thiessen polygons (Cowling and Shafron 1992). The polygons are shaded according to the interval in the key in which they fall into, while each interval contains approximately 20 per cent of farms.

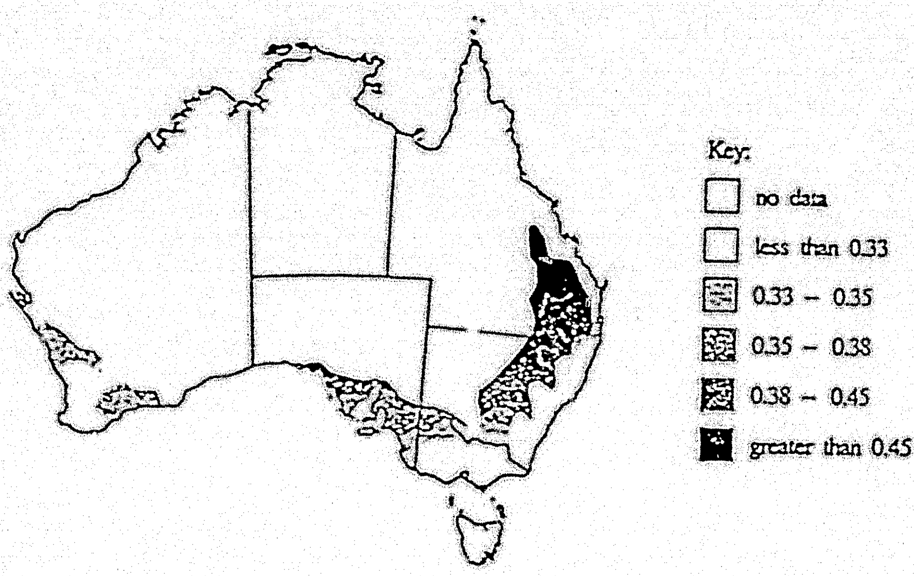
The fourteen-year national average of individual farm mean wheat yields for producers with greater than 100 hectares of wheat sown was estimated to be 1.61 tonnes per hectare. The average variance about the mean was estimated to be 0.62, indicating that there is a 95 per cent probability that the true mean lies between 0.37 and 2.85 tonnes per hectare.

Generally, variations in wheat yield may be due to the variability of rainfall across the continent and warrants further exploration.

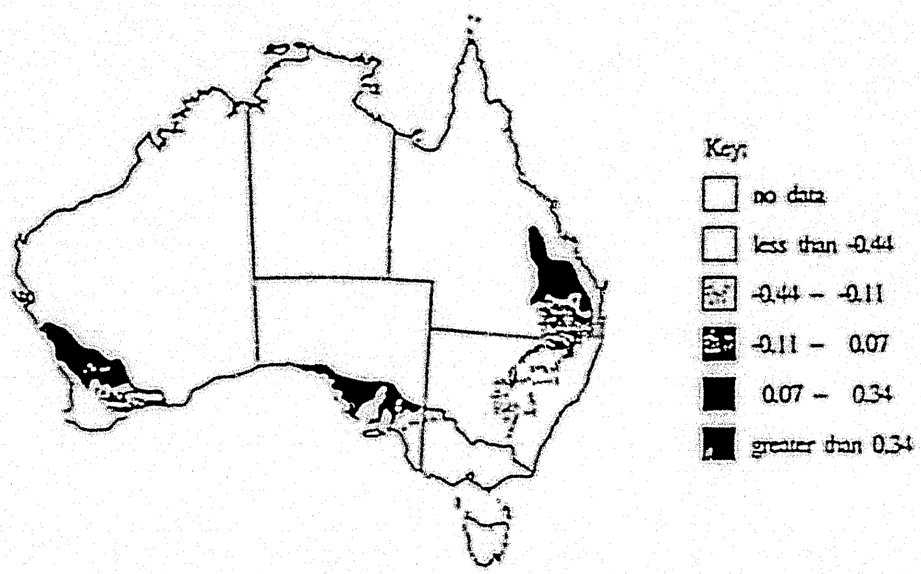
Map 1: Mean wheat yield for producers in the wheat and other crops industry with greater than 100 hectares of wheat sown Local averages for 1978-79 to 1991-92



Map 2: Coefficient of variation of wheat yield for producers in the wheat and other crops industry with greater than 100 hectares of wheat sown Local averages for 1978-79 to 1991-92



Map 3: Skewness coefficient of wheat yield for producers in the wheat and other crops industry with greater than 100 hectares of wheat sown Local averages for 1978-79 to 1991-92

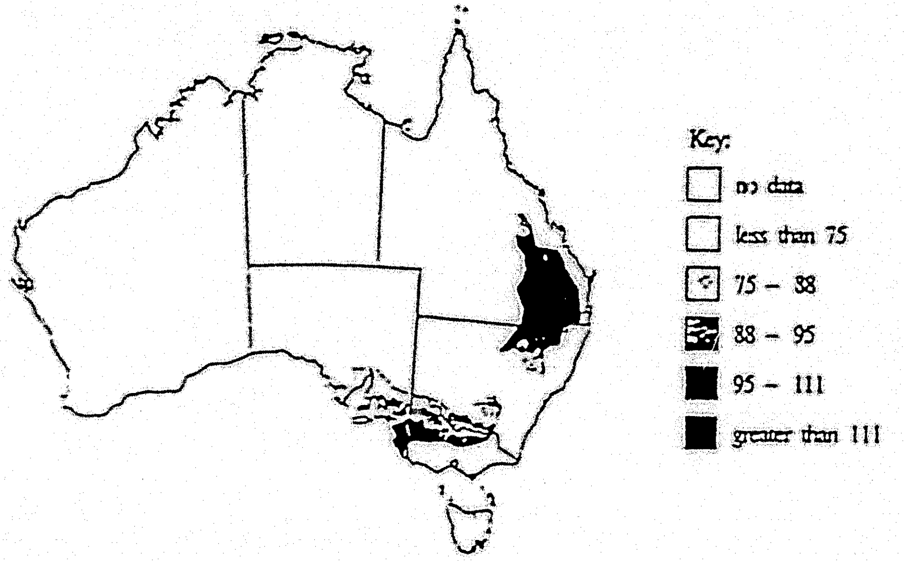


Low mean wheat yield estimates for Western Australia, of around 1.50 tonnes per hectare and below, are perhaps associated with the cropping rotation system and the clearing of new cropping land which is generally less fertile. Higher cropping percentages of around two to four pasture years in every ten years has been attributed to low yield trends within this state. Other factors which affect yields include '. . . constraints from waterlogging . . ., low nitrogen-use efficiency . . ., high levels of root disease . . ., and restricted subsoil rooting . . . (which) have all been attributed to the low performance of cereal crops in Western Australia' (Bureau of Resource Sciences, 1993, p. 71).

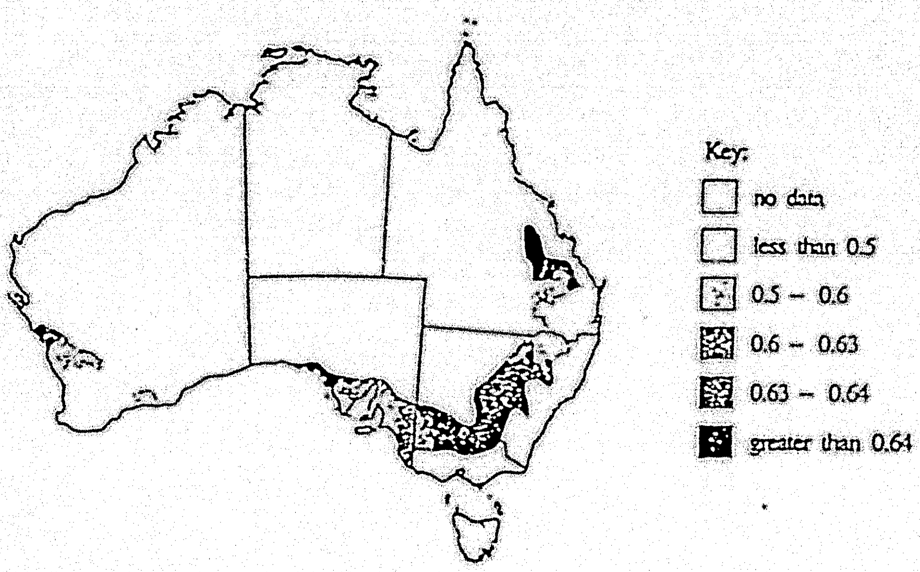
Sandy soils, high cropping percentages (or proportion of cropping rotation plan devoted to crops as opposed to pasture) and consistently low rainfall of 350 mm and below may influence the low mean wheat yield and associated low variation in yield indicated in maps 1 and 2 in the Eyre Peninsula.

Soil conditions of the Wimmera region of Victoria region are conducive to reasonable crop yields. Maps 4 and 5 reflect the high mean wheat yield and corresponding moderate variability for producers in this region.

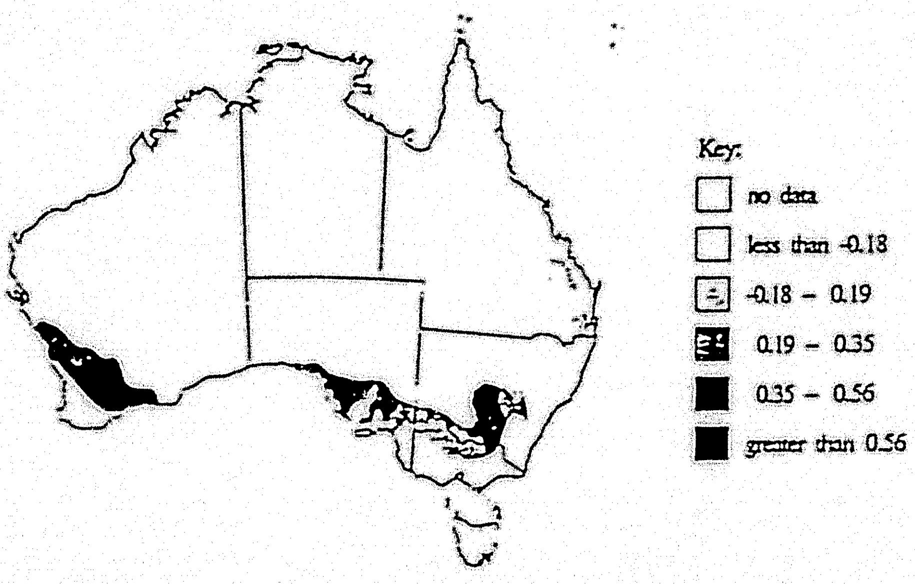
Map 4: Mean of gross cash income per hectare for producers in the wheat and other crops industry Local averages for 1978-79 to 1991-92



Map 5: Coefficient of variation of gross cash income per hectare for producers in the wheat and other crops industry Local averages for 1978-79 to 1991-92



Map 6: Skewness coefficient of gross cash income per hectare for producers in the wheat and other crops industry Local averages for 1978-79 to 1991-92



In general, the drier regions and higher cropping ratios of the northern regions of New South Wales may explain the moderate to low mean and high variability in wheat yield observed in that area. Other soil structure problems, for example reduced stability, biological activity and low fertility, have combined to produce nitrogen deficiencies and low wheat yields (Bureau of Resource Sciences 1993).

High mean wheat yields in the Central Highland of Queensland may '...be attributed to the new crop lands which have been developed from the leguminous brigalow forest' (Bureau of Resource Sciences 1993).

Mean gross cash income per hectare over the fourteen-year period was estimated to be \$95, with a variance about the mean of \$55. This indicates that there is a 95 per cent probability that the true mean lies between –\$15 and \$205 per hectare.

The regional variability in income per hectare illustrated in maps 4 to 6 can be partly explained with reference to agronomic and climatic factors, as well as commodity prices and farm size. For example, in Western Australia, mean income per hectare is low, while the variability is also relatively low. As farms in this state are very large, low returns per hectare are offset by much higher areas sown to crops. The distributions in this part of the state are more right skewed, however, perhaps reflecting the occasional well above average seasons.

In South Australia's Eyre Peninsula, mean cash income per hectare is relatively low, although it increases slightly from north to south. Variability is higher in the northern part of the peninsula, however, as rainfall here is less reliable.

In Victoria, mean cash income per hectare is above average in most parts of the state, although farms in the Wimmera region record slightly higher returns. The variability in per hectare incomes is around average levels, except in the Mallee district where variability is well above average.

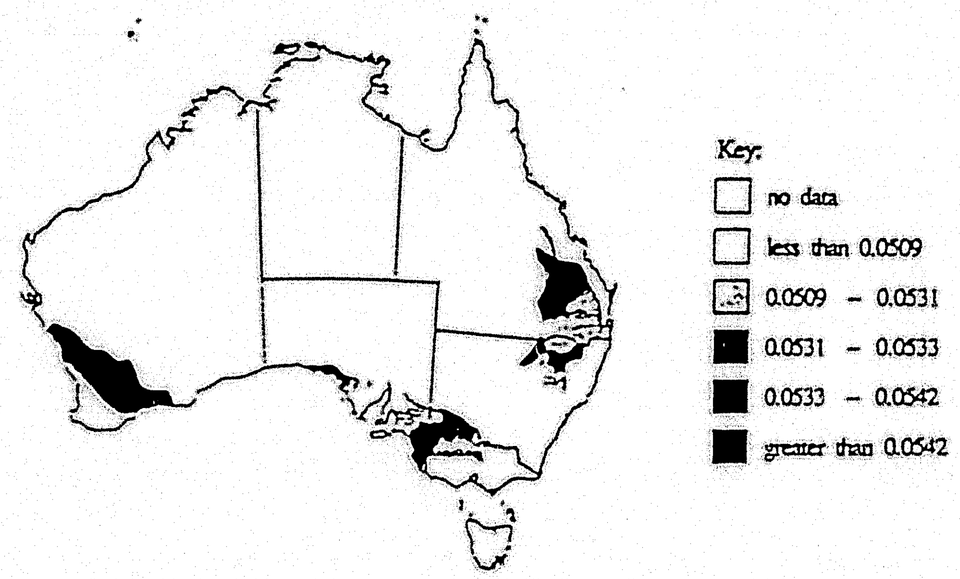
Mean income per hectare in southern New South Wales is average, while variability is very high. Income per hectare is below average in central New South Wales, and more variable. In the northern region incomes per hectare begin to increase and become less variable as the country here is generally more favourable to crop production, and farmers also have the option of growing summer as well as winter crops.

In Queensland, mean income per hectare is well above average in most parts of the state, and displays a relatively low level of variability.

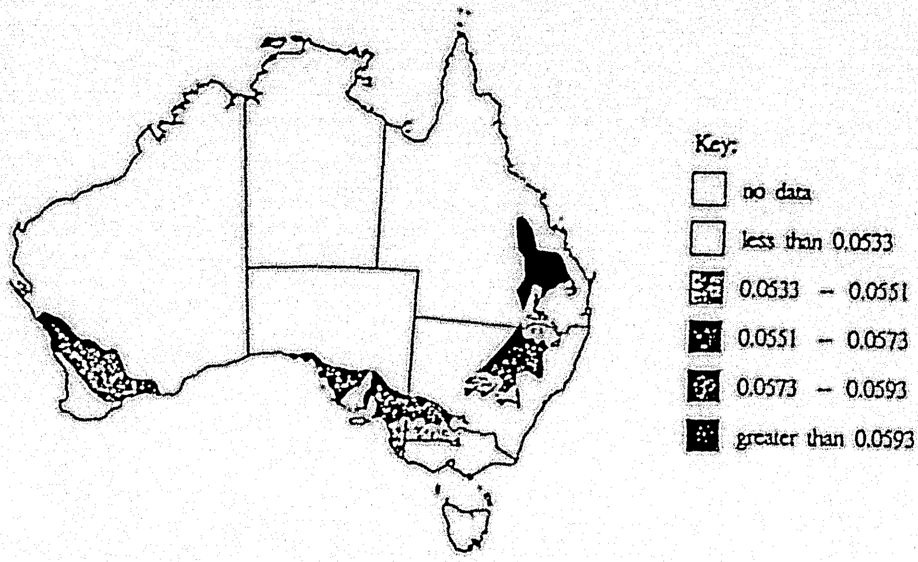
The fourteen-year national average of the individual farm mean rates of return for producers in the wheat and other crops industry was estimated to be 5.4 per cent. The average variance about the mean was estimated to be 5.7 per cent, indicating that there is a 95 per cent probability that the true mean lies between –6 and 16 per cent.

Map 7 shows mean rate of return, adjusted for depreciation, ranging from 5.1 per cent to 5.4 per cent across the country. As a result, it is difficult to explain the small changes in this measure; however, it may be noted that Western Australia has a mean rate of return adjusted for depreciation which is comparatively higher compared to the remainder of Australia, with a relatively high variation and skewness.

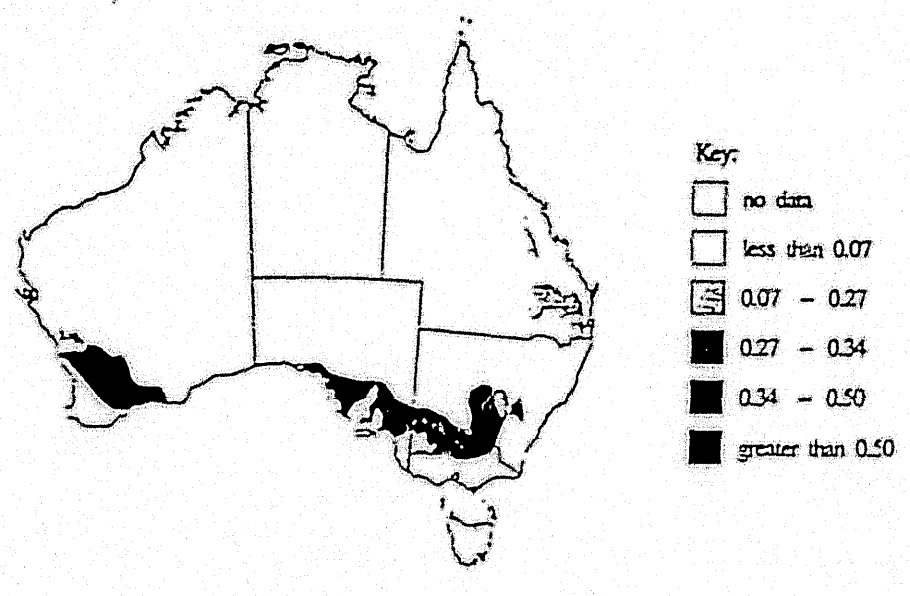
Map 7: Mean rate of return adjusted for depreciation for producers in the wheat and other crops industry Local averages for 1978-79 to 1991-92



Map 8: Standard deviation of rate of return adjusted for depreciation for producers in the wheat and other crops industry Local averages for 1978-79 to 1991-92



Map 9: Skewness coefficient of rate of return adjusted for depreciation for producers in the wheat and other crops industry Local averages for 1978-79 to 1991-92



Exploratory data analysis

Exploratory data analysis using scatter plots is a simple approach to examining the characteristics of the individual farm level distributions. It provides an insight into data relationships and any further analysis that may be performed.

Figure 4 may be used as a reference map. Each region, or group, is designated a marker symbol. These groups are subjectively selected for easy regional identification in the following figures, and are not necessarily statistically significant.

The relationships between mean gross cash income and its corresponding standard deviation and skewness in 1978-79 to 1991-92 are demonstrated in figures 5 and 6. With the exception of Western Australia, the mean versus standard deviation of gross cash income for wheat and other crop producers shows that as the mean increases, standard deviation increases. The pattern occurs at the Australia and regional level and supports the mean variance tradeoff.

Figure 4: Selected groups of producers in the wheat and other crops industry

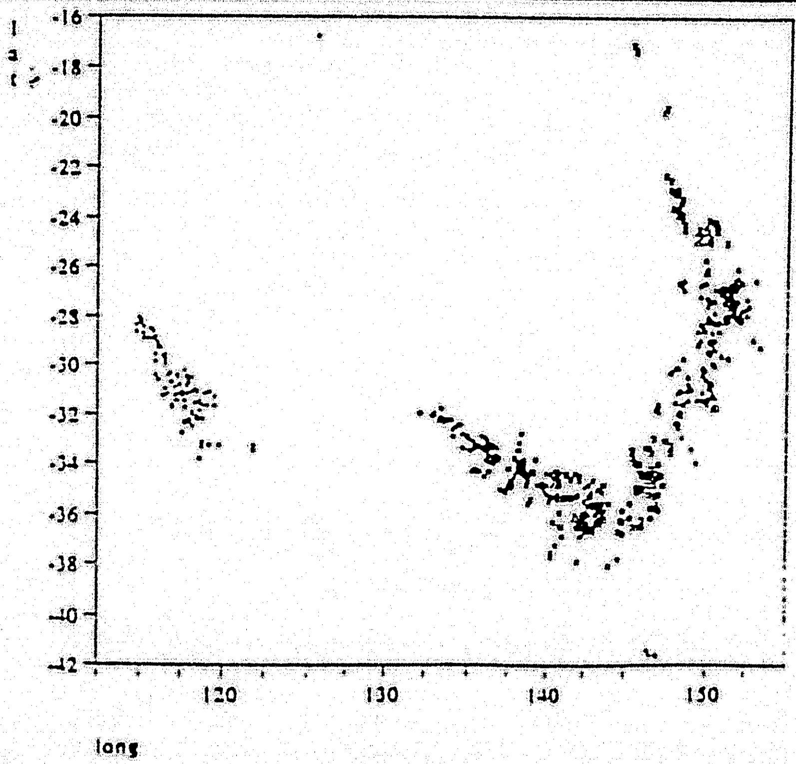
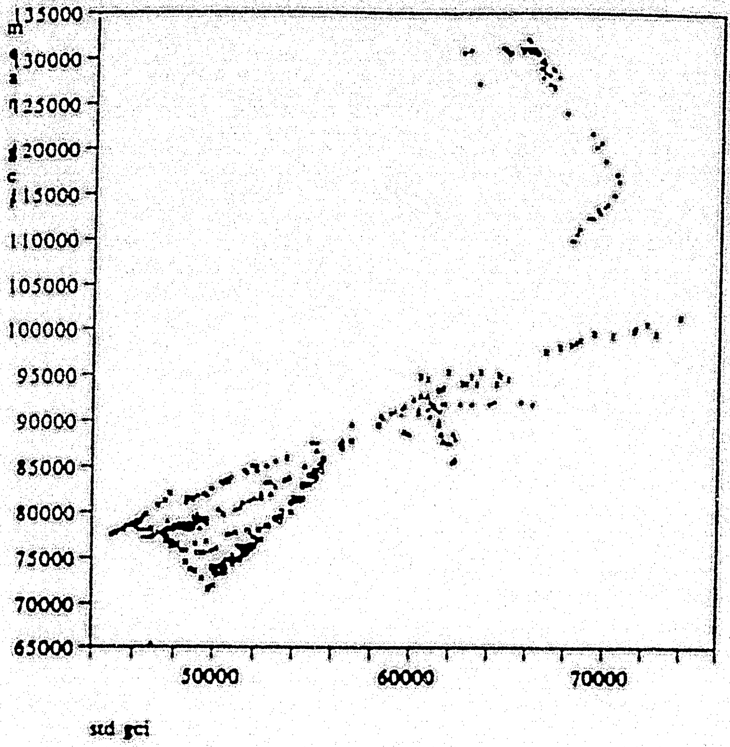


Figure 5: Means versus standard deviation of gross cash income during 1978-79 to 1991-92 for producers in the wheat and other crops industry



The mean versus skewness coefficient of gross cash income for these producers is shown in figure 6. Most regions demonstrate a decreasing mean and corresponding increasing skewness.

Validating the smoothing technique

To examine the accuracy of the technique, mean values estimated from the smoothed distributions were compared with 'raw' farm level estimates. For example, the mean gross cash income from the smoothed farm level distributions was compared with the mean of the 'raw' values for each sample farm, where the mean of the 'raw' values may be calculated from as few as one or two observations.

Smooth means and 'raw' means for gross cash income, rate of return, and wheat yield are displayed in tables 1, 2 and 3 respectively.

Figure 6: Mean versus skewness coefficient of gross cash income during 1978-79 to 1991-92 for producers in the wheat and other crops industry

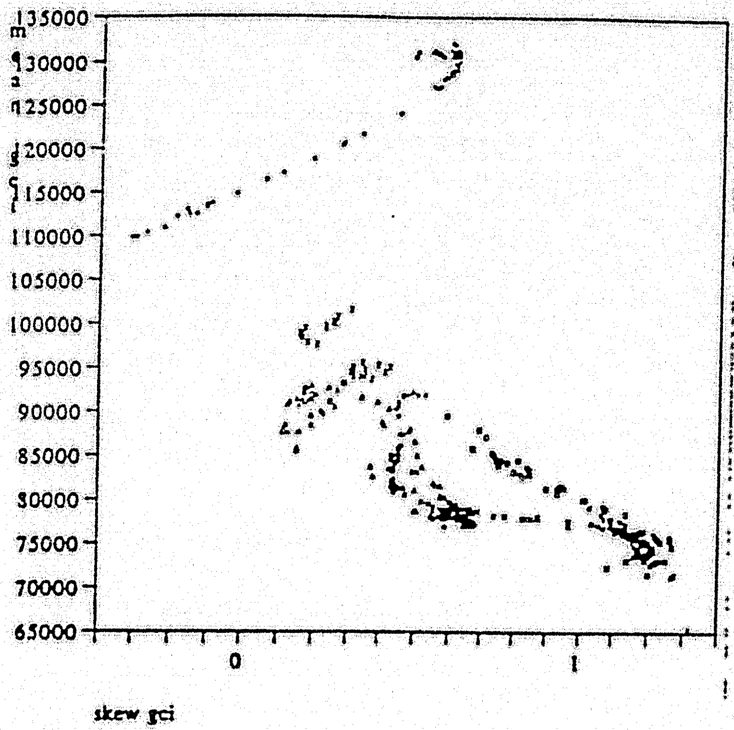


Table 1: Cross-validation of mean gross cash income for producers in the wheat and other crops industry Average per farm

	'Raw' mean gross cash income		'Smoothed' mean gross cash income	
	$		$	
Australia	85 000	(4.0)	86 660	(0.3)
New South Wales	85 170	(11.1)	86 210	(0.8)
Victoria	72 420	(8.4)	74 840	(0.2)
Queensland	79 710	(7.9)	82 510	(0.8)
South Australia	80 360	(8.6)	79 920	(0.5)
Western Australia	124 570	(8.0)	126 420	(0.7)
Tasmania	18 860	(83.0)	65 550	(0.0)

Note: Figures in parentheses are relative standard errors, expressed as percentages of the estimates. See ABARE (1993 p. 72) for a guide to interpreting the relative standard errors.

Table 2: Cross-validation of mean rate of return adjusted for depreciation for producers in the wheat and other crops industry
Average per farm

	'Raw' mean rate of return		'Smoothed' mean rate of return	
	$		$	
Australia	0.058	(6.0)	0.054	(0.2)
New South Wales	0.047	(21.8)	0.051	(0.7)
Victoria	0.059	(11.6)	0.053	(0.2)
Queensland	0.061	(10.7)	0.054	(0.2)
South Australia	0.062	(10.8)	0.052	(0.3)
Western Australia	0.064	(9.0)	0.065	(0.7)
Tasmania	0.032	(137.2)	0.049	(0.0)

Note: Figures in parentheses are relative standard errors, expressed as percentages of the estimates. See ABARE (1993 p. 72) for a guide to interpreting the relative standard errors.

Generally, the smoothed averages are comparable to 'raw' averages values, and show no systematic bias. At the national level, average for the raw mean gross cash income is $85 000 while its corresponding smoothed average is $86 660. The relative standard errors are 4.0 per cent and 0.3 per cent respectively, which indicates the impact the smoothing process in reducing cross-sectional variation.

From table 3, the raw mean wheat yield averages 1.59 tonnes per hectare at the national level while the smoothed mean wheat yield average is also 1.61 tonnes per hectare. Corresponding standard errors are 2.2 and 0.4 tonnes per hectare respectively.

Table 3: Cross-validation of mean wheat yield for wheat and other crop producers with 100 hectares of wheat sown Average per farm

	'Raw' mean wheat yield		'Smoothed' mean wheat yield	
	$		$	
Australia	1.59	(2.2)	1.61	(0.4)
New South Wales	1.75	(5.1)	1.73	(1.2)
Victoria	1.97	(4.4)	1.99	(0.9)
Queensland	1.75	(5.3)	1.70	(0.2)
South Australia	1.42	(4.8)	1.50	(0.7)
Western Australia	1.04	(3.7)	1.09	(0.3)

Note: Figures in parentheses are relative standard errors, expressed as percentages of the estimates. See ABARE (1993 p. 72) for a guide to interpreting the relative standard errors.

6. Conclusion

The technique developed in this paper is used to estimate temporal farm level distributions of physical and financial variables from repeated survey data. Estimates of the mean, variance and skewness of these distributions were calculated and presented in tabular and map form.

The method discussed in this paper appears to provide sensible results with the preliminary cross-validation technique when applied to AAGIS data. However, further work is needed to provide an objective method of selecting an optimum bandwidth in this case. Some of the methods developed for one dimensional smoothing, such as cross validation, may be appropriate in this situation.

In further analysis, it will be important to link variation in farm prices to the estimates of farm income distributions. This will enable the analysis of changes to exchange rates and international market conditions on farm income.

The estimates of mean, standard deviation and income variability will also be useful in obtaining measures of land and farm business values and the risks associated with farm debt.

Appendix: Model development

This section describes how the kernel smoothing techniques are applied to an indicator function of the variable Y in order to estimate a local value of the distribution function. In particular, if in the formulation given in section 4, expression 1 is replaced by:

$$I(Y \leq y) = F(y|x) + \varepsilon,$$

where $F(.|x)$ is the distribution function of Y at the location x and I is the indicator function, then the kernel estimate of the distribution function is:

$$\hat{F}(y|x) = \sum_{j=1}^{n} \omega_j(x) I(Y_j \leq y).$$

Theoretical properties of the Kernel estimator of the unconditional distribution function was examined by Wang (1991). The authors are unaware of any research on non-parametric estimation of the conditional distribution function. However, Jones and Hall (1990) have investigated the related problem of kernel estimation of regression quantiles.

The end result of the smoothing process is a smoothed value of $F(y|x)$ for each farm. In addition, the function $F(.|x)$ can be interpolated to provide smoothed values for locations x where raw data are unavailable. This feature is important in the subsequent analysis, because it allows interpolation of distributions for all farms for every year of the analysis, not just the years in which the farm was in sample.

Estimation of the risk profiles

The major steps involved in estimating farm level income distributions are outlined below. It should be noted that the process is not specific to income and may be applied to any other continuous variable collected in the survey.

Step 1

Let G_{kt} denote the income of sample farm k in year t, $t \in \{1,2,...,T\}$. Calculate an adjusted income value as: $Y_{kt} = G_{kt} - \overline{G}_k$, where $\overline{G}_k$ is the average of the income values G_{kt} calculated over the time the farm was in the survey. Therefore Y_{kt} denotes the mean adjusted income of sample farm k in year t. Temporal distributions are estimated for Y_{kt} and then the means are added back on.

Step 2

The aim is to estimate the values of the income distribution function for each farm at a given set of points, $y_1, y_2, ..., y_r$. When estimating the distribution function of a set $\{X_k\}$ of independent random variables, it is usual to set $y_1 = X_1$, $y_2 = X_2$, and so on. However, this approach will be impractical in this situation because of the large sample sizes involved.

To reduce the amount of computation required to a feasible level, y_i was set equal to the $i/(r+1)$ quantile of the set $\{Y_{kt}\}$, where r is significantly less than the total number of observations in $\{Y_{kt}\}$. The distribution function of each farm was estimated at the same points, $y_1, y_2, ..., y_r$. The larger the r that is chosen, the more accurate the final results will be. However, there is a limit depending on the amount of computing resources available. The reason that this limit exists is that smoothing needs to be performed on indicator functions of these y-values for each of the fourteen years to the locations of the 4213 sampled farms, which is a massive computing task even for small values of r. The value $r = 99$ seemed to provide reasonably accurate estimates of the income distribution functions.

Let y denote any one of the y-values, $y_1, y_2, ..., y_r$, and let $I_{kt}(y)$ be the indicator function:

$$I_{kt}(y) = \begin{cases} 1 & \text{if } Y_{kt} < y, \\ 0 & \text{otherwise.} \end{cases}$$

Step 3

For each year t, spatially smooth the $I_{kt}(y)$ values. If $\bar{I}_{kt}(y)$ denotes the smoothed value then:

$$\bar{I}_{kt}(y) = \sum_{j \in s_t} w_{jkt} I_{jt}(y), \tag{2}$$

where s_t denotes the sample farms in year t and w_{jkt} is the weight of sample farm j in the year t, spatial smooth to the location of farm k. $\bar{I}_{kt}(y)$ is an estimate of the probability that the mean adjusted income is less than or equal to y in year t. A smoothed value is obtained for all farms surveyed at any time in $t \in \{1,2,...,T\}$.

The bandwidths for smoothing were chosen on the following basis. For all values of y the same bandwidth was used, otherwise the weights in (2) would depend on y, which can potentially lead to the inconsistency that for some $y_a < y_b$, $\bar{I}_{kt}(y_a) > \bar{I}_{kt}(y_b)$. Since the sample size varied from one year to the next, different bandwidths, h_t, were used for each year. Although no theory exists for the choice of optimal bandwidth in this case, Silverman

(1986, pp. 84–6) showed that for d-dimensional kernel density estimation, the optimal bandwidth is proportional to the sample size to the power $-1/(d+4)$. Thus the bandwidths were set to:

$$h_t = h / n_t^{1/6},$$

where n_t is the sample size in year t. The value of h was chosen in a subjective manner. A range of different values were tried and the smallest value of h, which gave reasonably smooth appearing maps of the temporal coefficient of variation and skewness coefficient were chosen. A description of how these coefficients were calculated is given in the following steps.

Step 4

The temporal distribution function $\bar{F}_k(y)$ of the mean adjusted income of sample farm k evaluated at y is then obtained by averaging the $\bar{I}_{kt}(y)$ over time. That is,

$$\bar{F}_k(y) = T^{-1} \sum_{t=1}^{T} \bar{I}_{kt}(y).$$

In terms of the original indicator variables:

(3)
$$\bar{F}_k(y) = T^{-1} \sum_{t} \sum_{i \in s_t} w_{ijt} I_{it}(y).$$

Step 5

Before shifting this distribution function so that it is centred at the farm's mean, its variance and skewness must be adjusted. A significant bias may result in the estimates of these quantities because of the short time that some farms remain in sample. The estimate of the p^{th} moment based on the distribution function at (3) is:

(4)
$$\hat{\mu}_{kp} = \sum_{l=1}^{r+1} y_l^{*p} \left\{ \bar{F}_k(y_l) - \bar{F}_k(y_{l-1}) \right\},$$

where $\bar{F}_k(y_0) = 0$; $\bar{F}_k(y_{r+1}) = 1$; y_l^* is the midpoint of the interval (y_{l-1}, y_l); $y_1^* = y_1$ and $y_{r+1}^* = y_r$. In current ABARE research it is shown that for some constants c_{k2} and c_{k3}, and $E(\hat{\mu}_{k2}) \equiv \mu_{k2} c_{k2}$ and $E(\hat{\mu}_{k3}) \equiv \mu_{k3} c_{k3}$. These constants are easily calculated in the process of smoothing described above. Finally, using the technique also describe in Kokic et al.

(1993), the variance and skewness of the distribution function at expression 3 are adjusted. This corresponds to forming new values $y_{k1}, y_{k2}, \ldots, y_{kr}$ from $y_1, y_2, \ldots, y_r$ for farm k so that the second and third moments of $\overline{F}_k^*(y_{k1}), \overline{F}_k^*(y_{k2}), \ldots, \overline{F}_k^*(y_{kr})$ are approximately unbiased, where

$$(5) \qquad \overline{F}_k^*(y_{kl}), \ \overline{F}_k(y_l), \ l = 1, 2, \ldots, r.$$

Step 6

Finally, this distribution function was shifted so that it would be centred at the mean of farm k. Note that from (3) and (4), the estimate of the mean adjusted distribution function at (5) is spatially smooth, whereas the farm's actual mean, $\overline{G}_k$, aren't spatially smooth. Therefore, it was decided to smooth these values first. To be consistent with the approach described above, the same smoother used in expression 1 was used. That is,

$$\bar{\mu}_k = T^{-1} \sum_t \sum_{i \in s_t} w_{ijt} \overline{G}_i.$$

Step 7

The resulting distribution function for farm k was estimated as:

$$F_k(x) = \begin{cases} \overline{F}_k^*(y_{kl}) \text{ if } y_{kl} \le x - \bar{\mu}_k < y_{k,l+1}, \\ 0 \text{ if } x - \bar{\mu}_k < y_{k1}, \\ 1 \text{ otherwise.} \end{cases}$$

Note that although the main application in this paper is to model farm specific variability of income, the combined spatial and temporal modelling technique described above can be applied quite generally to any variable for which time series/cross-sectional data are available. For example, it could be used to model farm specific variability in crop yields or it can easily be generalised for estimating joint temporal distribution functions.

A key feature of the modelling process is that it results in locally smoothed income distribution for any sample farm. These income distributions can be combined, using the average weights described earlier, over any particular group of farms. This results in an estimate of the aggregate risk profile, and associated risk characteristics, of the group of farms.

References

ABARE 1993, *Farm Surveys Report 1993*, ABARE, Canberra.

Barry, J.B., Hopkin, J.A. and Baker, C.B. 1998, *Financial Management in Agriculture*, 4th edn, Interstate Printers & Publishers., Danville, Illinois.

Borch, K. 1969, 'A note on uncertainty and indifference curves', *Review of Economic Studies* vol. 36, no. 1, pp. 1–4.

Bureau of Resource Science 1993, *Trends in Wheat Yields and Soil Fertility in Australia*, AGPS, Canberra.

Cowling, A. and Shafron, W. 1992, Mapping ABARE's survey estimates, ABARE paper presented at the 1992 Sixth Australian ESRI/ERDAS Users' Conference, Adelaide, 9–11 September.

Crean, S. 1992, Federal Government announces new farm adjustment and assistance measures, Media release, Department of Primary Industries and Energy, Canberra, 21 September.

Jones, M.C. and Hall, P. 1990, 'Mean squared error properties of kernel estimates of regression quantiles', *Statistics and Probability Letters*, vol. 10, pp. 283–9.

Härdle, W. 1990, *Applied Nonparametric Regression*, Cambridge University Press, England.

Machina, M. 1982, 'Expected utility analysis without the independence axiom', *Econometrica*, vol. 50, no. 2, pp. 277–323.

Meyer, J. and Ormiston, M. 1983, 'Comparative statics of cumulative distribution changes for the class of risk-averse agents' *Journal of Economic Theory*, vol. 36, no. 1, pp. 153–69.

Nelson, R.D. and Pope, R.D. 1991, 'Density estimation and empirical applications of stochastic dominance', in 1991 *Proceedings of the Business and Economic Statistics*

Section, Proceedings of the Annual Meeting of the American Statistical Association, Alexandria, Vancouver, pp. 271–82.

Peterson, D.C., Dunne, S.H., Morris, P.C. and Knopke, P. 1991, 'Developments in debt for broadacre agriculture', *Agriculture and Resources Quarterly*, vol. 3, no. 3, pp. 349–60.

Silverman, B.W. 1986, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

Topp, V. and Beare, S. 1993, Agricultural land values in Australia: An economic perspective, Paper presented at the Municipal Valuers Conference, Lakes Entrance, Victoria, 26 November.

Wang, S. 1991, 'Nonparametric estimation of distribution functions', *Metrika*, vol 38, pp. 259–67.