The Impact of Irrigation and Soil Type on Sugar Cane Production: A Whole Production Set Approach

Rodney Beard
Dpt of Agriculture
University of Queensland, 4072

Anthony Bloesch
S.V R.C.
University of Queensland, 4072

Trevor Willcox
Bureau of Sugar Experiment Stations
Bundaberg

Abstract

We use a new econometric model of whole production sets to analyse the impact of irrigation and soil type on sugar cane production. The method used models a production function as a conditional probability density function. Considering a production function from this perspective allows us to derive expected profit and yield functions without knowledge of the frontier production function. Further, we are able to estimate the risk associated with various input levels. We review kernel based techniques for estimating conditional probability density functions.

1 Introduction

Production functions are traditionally defined as the maximal output attainable with a given set of inputs. This definition precludes the use of traditional mean based regression methods since we actually seek to estimate limiting values rather than central values. Instead, a number of *ad hoc* methods have been used to estimate frontier production functions [1, 5, 12, 13]. For a representative review see [6].

The frontier production function neglects, for practical purposes, technically inefficient data points. These data points may have an important economic interpretation. For example, expected prouctivity, production risk, minimal productivity and the probability of attaining a particular output for a particular input.

In practice, we are interested in the probability of achieving a particular level of output independent of whether or not it is derived from technically efficient or inefficient production. Interestingly, we can derive a qualitative measure of the distribution of efficiency by looking at the underlying probability density function. For example, a bimodal distribution would indicate a mixture of efficient and inefficient producers but a unimodal distribution would indicate that producers were homogeneously efficient. For decision making purposes, this qualitative measure will usually suffice.

In [3], two of the authors argue that, the use of, the conditional probability density function (cpdf) of the production data is more useful for decision making than traditional approaches.

They discuss both, how the cpdf can be used to support decision making and model the data, and how such a cpdf can be estimated from a data set.

In order to illustrate the advantages of using density methods for modelling production sets, we analyse some sugar cane production data [17]. The production data consists of 212 data points taken from farms in the Fairymead region (Bundaberg) in 1993. Each data point consists of yield per hectare, irrigation per hectare and soil type (since the farms are close together, we assume constant rainfall). We are interested in determining the optimal irrigation level.

2 Production Functions and Technical Efficiency: A Density Approach

Consider the likely components of production data. Some producers will be technically efficient and others not. Further, the data will have a stochastic element due to variation in the economic and physical environment. In fact, there will be a distribution of producers, with varying degrees of efficiency, which will be differentially affected by exogenous environmental factors. For a given data set the relative contribution of these factors will be difficult if not impossible to determine. Traditional approaches derive much of their awkwardness from trying to distinguish between efficient and inefficient producers, in the presence of exogenously derived variation, in an a priori manner. They force the modeller to make unwarranted assumptions such as all producers are efficient or all variability is one sided.

A better way of modeling the data is to view it as being drawn from a conditional probability density function (cpdf). The cpdf is that function which, given an input set, tells us the probability that a given production level will be achieved. Using standard statistical techniques it is possible to estimate this cpdf from a given data set (see section 3).

In practice, modeling the data with a cpdf will be better than traditional approaches since we will not only be able to determine the maximum possible production level but also expected production levels and production risk. To this extent our approach can be seen as a generalization of existing approaches, albeit not a generalization of traditional methods (e.g. in our approach, frontier production functions can be seen as extreme percentile values of the corresponding cumulative probability density functions (cdf) or more usefully the upper mode of the cpdf). Traditionally, we could only predict the maximum attainable production level but using our approach we are able to predict the probability of achieving a given production level and a number of other parameters.

Note, summary information, such as that derived from regression based methods, provides only a partial model of the production set. For decision making purposes, we are not only interested in the hull of the production set but its total structure. This loss of information is illustrated by the fact that while it is possible to derive this summary information from the cpdf the reverse is, not in general, possible. As an example of the benefits of estimating cpdf's over hulls, recent developments in risk analysis, such as stochastic efficiency theory [2, 16, 9], require knowledge of cdf's.

3 Estimating Density Functions

Since there is an extensive statistical literature on nonparametric density estimation we can use efficient and statistically sound methods to model the data with a cpdf. For a representative survey of these techniques see [14, 11, 4, 10, 15]. Below, we confine ourselves to the fundamentals of nonparametric density estimation.

Perhaps, the most ancient and well known method of estimating pdf's is the histogram of [7]. Each bar in the histogram is an estimator for the probability density at its centre. In a histogram all observations in an interval are given equal weight. Often, a better estimate of the probability density at a point x can be gained by weighting points with some other *kernel function*. Whereas the histogram's kernel function is:

$$K(u) = \begin{cases} 1 & \text{if } x - h \leq u \leq x + h \\ 0 & \text{otherwise} \end{cases}$$

other kernel functions (e.g. the Gaussian) typically weight observations inversely to their distance from x. Further, if these kernel functions overlap then we can make use of more of the data set and thus gain a better estimate of the pdf. Rather than estimate the pdf at all points we need only estimate it at a series of *knot points* and interpolate between them with splines, discrete Fourier transforms, etc. In essence, kernel based methods differ primarily in the choices made for such things as knot points, kernel functions and interpolation methods.

Other methods exist for estimating pdf's that are not based on kernel functions. We do not review them here since they are typically more difficult to apply and understand (although often more efficient). The interested reader is referred to the above references for pointers to the literature.

4 Deriving Traditional Production Functions

Given a cpdf $f(y|\vec{x})$ we can derive a cdf $F(y|\vec{x}) = \int_0^\infty f(y|\vec{x})\,dy$. The frontier production function is given by $g(\vec{x}) = \lim_{F(y|\vec{x})\to 1} y$ or more readably, $g(\vec{x}) = \lim_{p\to 1} F^{-1}(\vec{x},p)$ (note: since $F(y|\vec{x})$ is monotonic increasing $F^{-1}(\vec{x},p)$ always exists). Similarly, for a given probability p and input set $\vec{x}$ the associated production level of $g(\vec{x};p)$ is given by $F^{-1}(\vec{x},p)$. The expected production level $\hat{y}(\vec{x})$ is given by $\hat{y}(\vec{x}) = \int_0^\infty y\, f(y|\vec{x})\,dy$ and is a measure of the average efficiency of production. The second moment of the expected production level $s(\vec{x})$ is given by $\int_0^\infty (y - \hat{y}(\vec{x}))^2 f(y|\vec{x})\,dy$ and is a measure of production risk due to variations in technical efficiency and other exogenous factors.

In order to separate variations in technical efficiency from other exogenous factors one or other of the distributions would need to be known. We could then deconvolve the joint pdf to give the other distribution. However, except in extremely fortuitous circumstances, such a deconvolution will not be possible.

5 A Case Study: Sugar Cane Production

In order to illustrate the advantages of using density methods for modelling production sets, we analyse some sugar cane production data [17]. The production data consists of 212 data points taken from farms in the Fairymead region (Bundaberg) in 1993. Each data point consists of yield (CCS and cane) per hectare, irrigation per hectare and soil type (since the farms are close together, we assume constant rainfall). We are interested in determining the optimal irrigation level.

The production data is plotted in figures 1 and 2. Kernel based methods were used to estimate the cpdf, the expected (i.e. conditional mean) yield and the upper hull (i.e. the frontier production function) of the data.

As we can see from figure 1 the frontier production function is a poor summary of the data. If we were to use it for decision making purposes, we could easily conclude that the total cane yield for red volcanic soils is maximal beyond 2.5 megalitres/Ha. However, as the expected yield curve shows, this is misleading – the frontier does not give a good picture of the production set, further a few isolated data points have biased our estimate of the frontier. This bias is to be expected as estimators (kernel based or otherwise) for the hulls of cpdf's are notoriously for varying greatly from sample to sample. Even so, it is worth noting that kernel methods, unlike many traditional methods, can be used to estimate the frontier of nonconvex production sets.

It is worth noting that optimal irrigation levels broadly agree with Bureau of Sugar Experiment Stations (BSES) recommendations based on pan evaporation data.

Interestingly, if price and costs are random variables then it is sufficient to work with the expected price $\bar{p}(y)$ and expected cost $\bar{C}(y, \vec{x})$ [3]. The conditional probability of incurring a loss can then be found, analogously, by determining $\tilde{y}$ such that $\bar{p}(\tilde{y})\tilde{y} - \bar{C}(\tilde{y}, \vec{x}) = 0$ and calculate $F(\tilde{y}|\vec{x}) = \int_{y<\tilde{y}} f(y|\vec{x})$.

In figure 3 we plot gross margin data and expected gross margins. Gross margins (GM) were calculated as:

$$GM = (0.009p_s(CCS - 4) + 0.328)CANE - 75I$$

where p_s is the spot price of sugar, CCS is the sugar content of the cane, $CANE$ is the tonnage of cane, and I is the irrigation level in megalitres.

Note: the optimal (in the sense of maximizing expected profit) irrigation levels conform not only to the BSES recommended levels but also to those suggested by expected yield. The optimal irrigation level suggested by the frontier production functions can be expected to differ considerably from the profit maximizing level. See [3] for a proof that, under uncertainty, to calculate expected profit we need only know the expected production function and not the frontier production function.

Many of the curves plotted in figures 1–3 show asymptotic behaviour. The most likely cause of this is lodging (i.e. large crops of some varieties of cane collapse under their own weight) leading to a diminishing returns effect. Note: the CCS level seems to reach an asymptote much faster than cane tonnage.

None of the curves in figures 1–3 go through the origin. There are two likely causes: firstly, the approximately 500mm of effective rainfall that fell during the 1993 growing season (1 megalitre is equivalent to 100mm of rainfall) and secondly, ground water (this is particularly pronounced in the case of the alluvial and humic gley soil types).

The variability within soil types is most likely to be caused by variations in soil fertility, cane variety, and farm practices (e.g. irrigation method: flood, trickle, spray). Note: due to terrain, certain irrigation methods are more likely to be used on certain soil types (e.g. flood on alluvial). Note: farms in the study containing mixed soil types have been categorized according to the dominant soil type.

The podzolic soils are structurally very similar but vary in fertility (red is the most fertile and grey the least fertile).

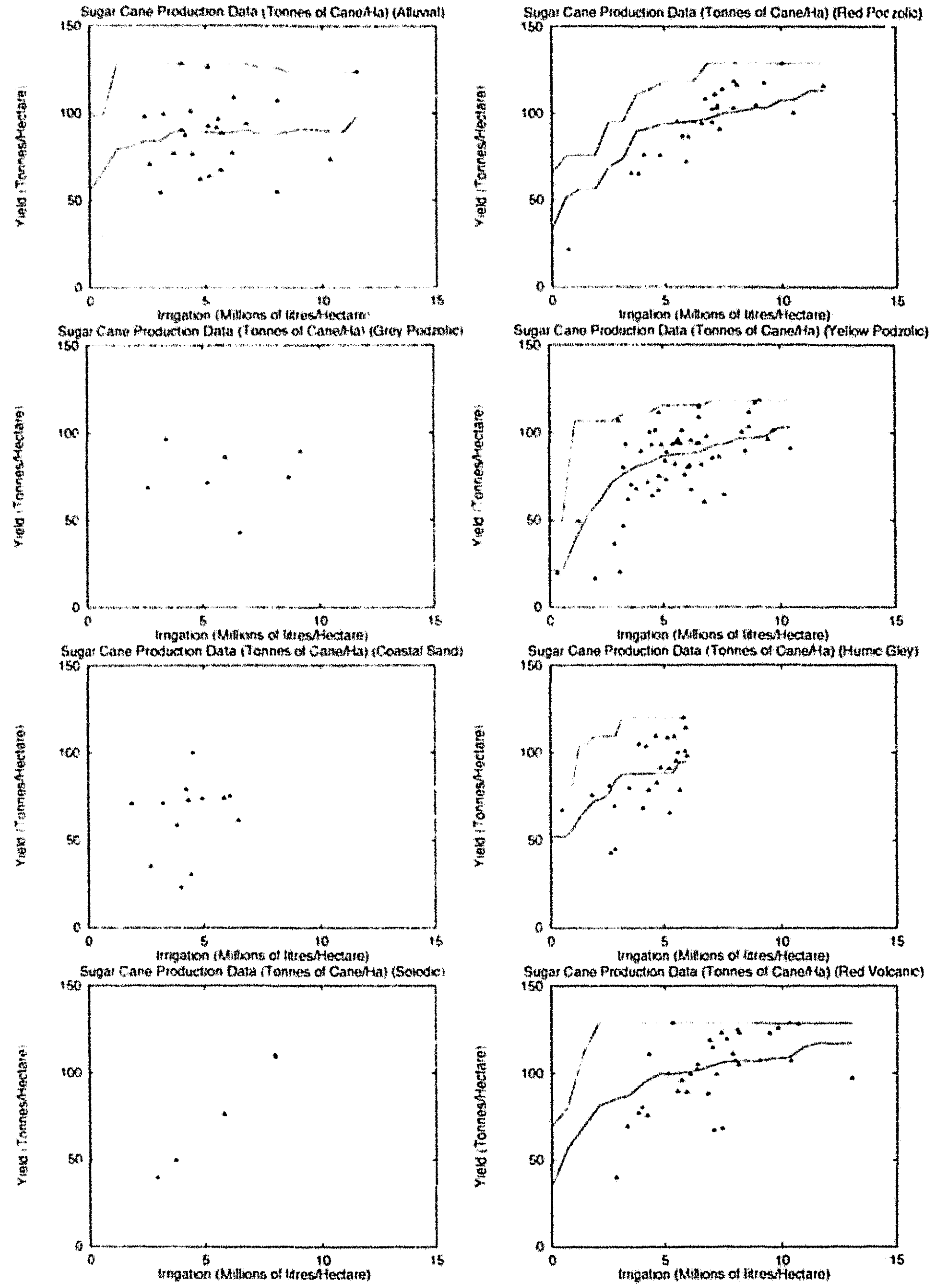

Figure 1: Cane production in tonnes/hectare. The frontier (dashed) and expected (solid) production functions have been plotted, where there was sufficient data.

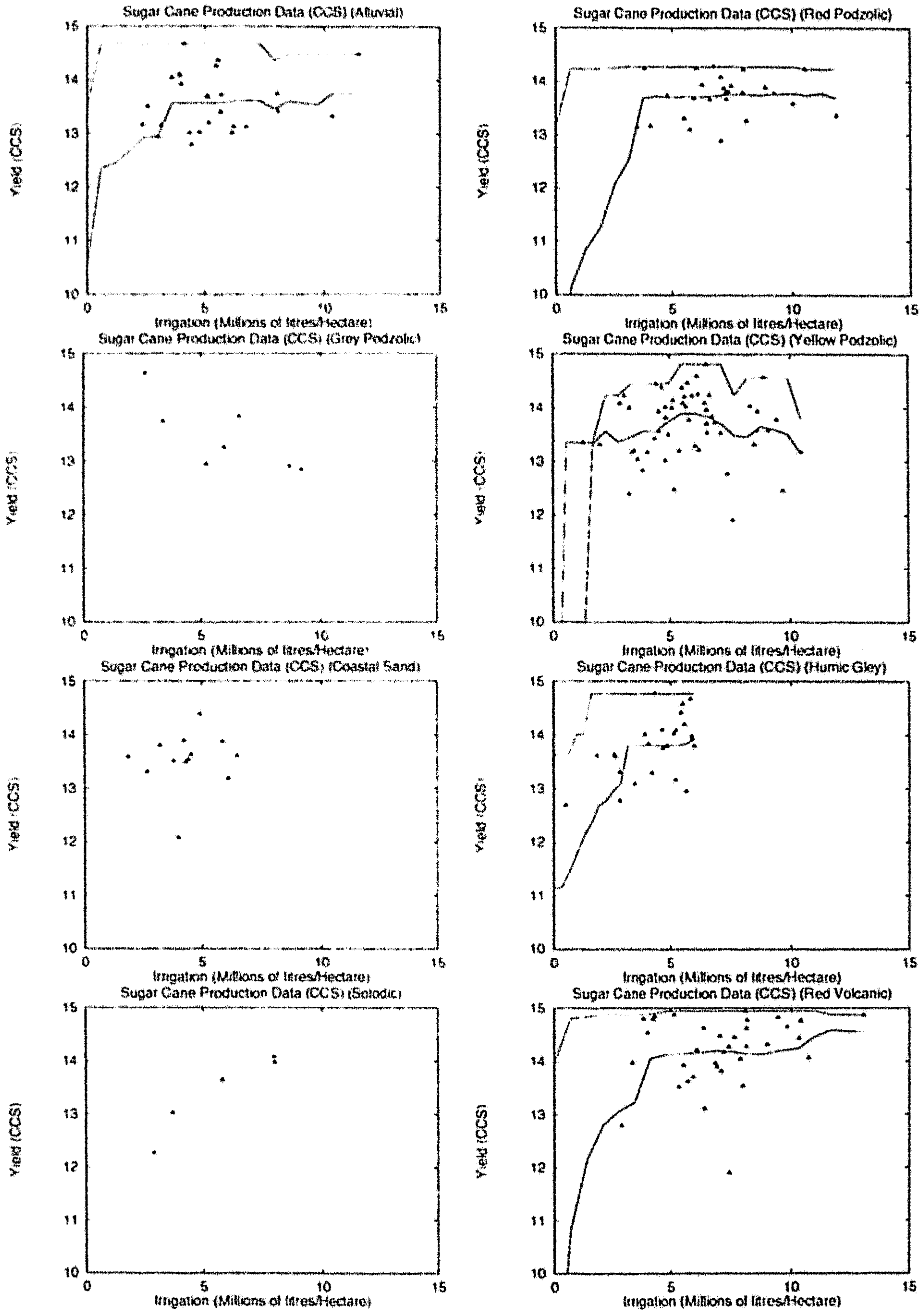

Figure 2: Cane CCS production. The frontier (dashed) and expected (solid) production functions have been plotted, where there was sufficient data.

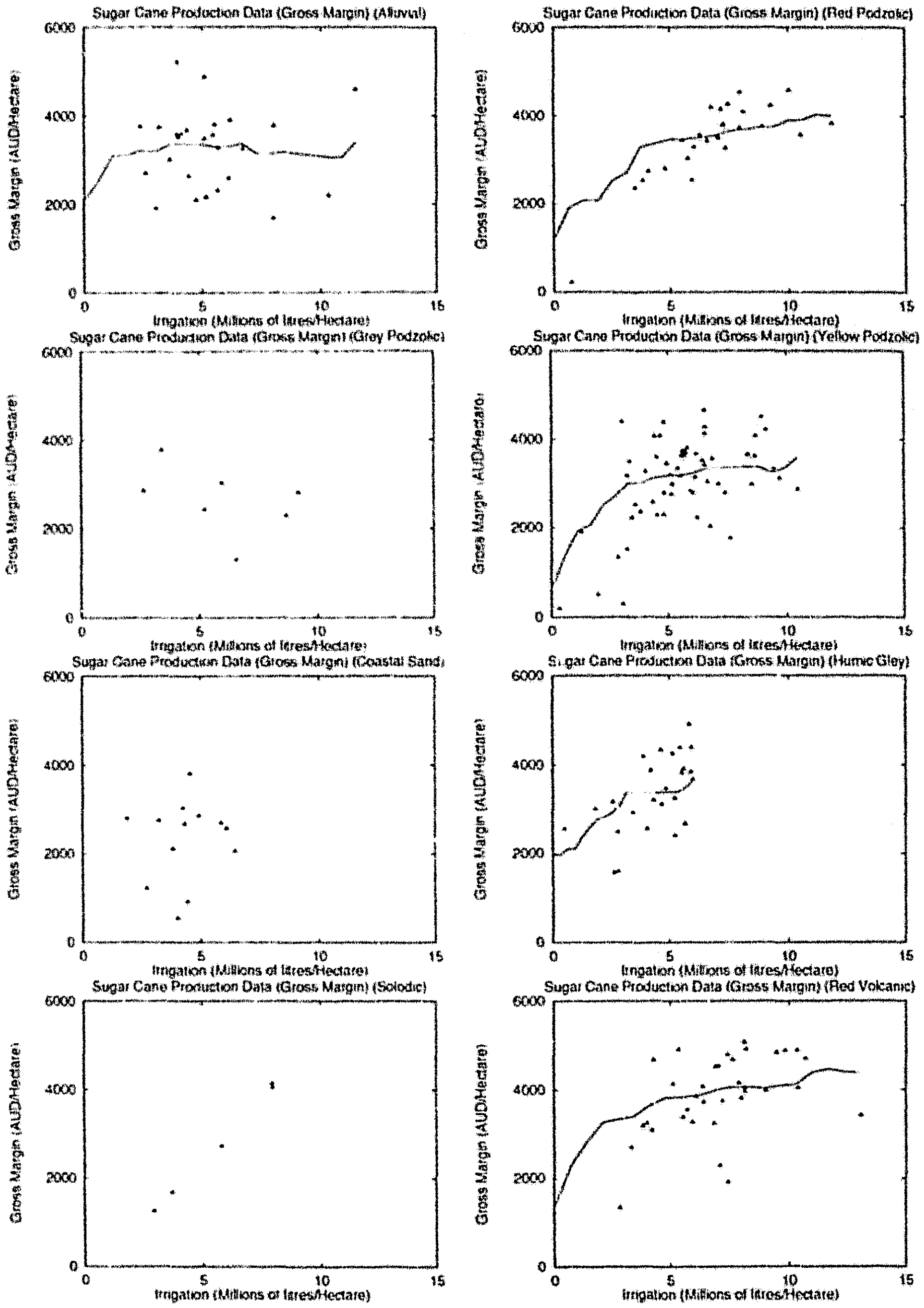

Figure 3: Gross margin of cane irrigation. The expected gross margins have been plotted, where there was sufficient data.

6 Conclusions

We have argued that the cpdf of production data provides more useful information than the frontier production function. From it both the expected yield and frontier production functions can be derived. Also, we have argued (and proved elsewhere [3]) that, in the presence of uncertainty, the expected yield is more useful for decision making than the frontier production function.

We have discussed how kernel methods can be used to estimate the cpdf and thus the expected yield and frontier production functions as well as measures of variability.

We estimate the expected yield and frontier production functions for a variety of soil types as well as gross margins. From our analysis of gross margins, it appears that many sugar cane farmers are underirrigating their crops. This result agrees with recommendations of the BSES, based on pan evaporation data.

It is interesting to note that, as we predict, if farmers were to use frontier production functions rather than expected yield to choose the optimal level of irrigation then they would have a lower expected profit than if they had used the expected yield function.

The solodic soil type shows very interesting behaviour. It would be interesting to obtain more data from solodic soils in order to examine both the relationship between irrigation level and yield, and the amount of variability in yields.

It seems likely that the method of irrigation used will have an impact on yield. Given more data it would be interesting to explore the effects of irrigation methods on yield and profitability.

Increasingly, in risk analysis, complete knowledge of the decision makers utility (profit) function is not assumed. Instead, criteria such as first and second order stochastic dominance, which make use of the cdf of some random variable such as profit, are used. Unfortunately, primitive density estimators, such as the stationary histogram, are used to calculate stochastic efficiency. Since, calculating stochastic efficiency requires integration and reintegration over a pdf, poor initial estimates of the pdf can lead to larger errors than if a better initial estimate had been used (see for example [2]). It would be worthwhile to investigate the use of more powerful estimators, such as kernel estimators, in calculating stochastic efficiency for the univariate categorical case.

In the multivariate mixed variable case the picture becomes more complex. In univariate stochastic dominance theory the simplifying assumption is made that all producers are behaving optimally. Thus inputs, apart from the categories under analysis, need not be considered. It would be interesting to examine the use of multidimensional kernels in stochastic dominance problems. It seems, to the authors, that the use of cdf's in the multidimensional case is not as fruitful an approach as the use of cpdf's and higher order moments.

Acknowledgments

We would like to thank: Bing So, Mal Wegener and Colin Brown for their comments.

References

[1] S. N. Afriat. Efficiency estimation of production functions. *International Economic Review*, 13(3):568–598, October 1972.

[2] Jock R. Anderson, John L. Dillon, and Brian Hardaker. *Agricultural Decision Analysis.* Iowa State University Press, Ames, Iowa, 1977.

[3] Anthony Bloesch and Rodney Beard. Estimation of frontier production and other related functions by density methods. Submitted for publication.

[4] Luc Devroye and Laszlo Györfi. *Nonparametric Density Estimation: The L_1 View.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1985.

[5] M. J. Farrell and M. Fieldhouse. Estimating efficient production under increasing returns to scale. *Journal of the Royal Statistical Society Series A*, 125(2):252–267, 1962.

[6] Finn R. Førsund, C. A. Knox Lovell, and Peter Schmidt. A survey of frontier production functions and of their relationship to efficiency measurement. *Journal of Econometrics*, 13:5–25, 1980.

[7] J. Graunt. *Natural and Political Observations Made Upon the Bills of Mortality.* 1662. Reprinted in [8].

[8] J. Graunt. *The Earliest Classics.* Pioneers of Demography. Gregg, Farnborough, 1973.

[9] Jack Hirshleifer. *The Analytics of Uncertainty and Information.* Cambridge Surveys of Economic Literature. Cambridge University Press, Cambridge, 1992.

[10] Samuel Kotz and Norman L. Johnson, editors. *Encyclopedia of Statistical Sciences.* John Wiley & Sons, 1982.

[11] David W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1992.

[12] W. D. Seitz. The measurement of efficiency relative to a frontier production function. *American Journal of Agricultural Economics*, 52(4):505–511, November 1970.

[13] W. D. Seitz. Productive efficiency in the steam electric industry. *Journal of Political Economy*, 79(4):878–886, July 1971.

[14] B. W. Silverman. *Density Estimation for Statistics and Data Analysis.* Monographs on Statistics and Applied Probability. Chapman and Hall, London, 1986.

[15] James R. Thompson and Richard A. Tapia. *Nonparametric Function Estimation, Modelling, and Simulation.* SIAM, Philadelphia, 1990.

[16] G. A. Whitmore and M. C. Findlay, editors. *Stochastic Dominance: An Approach to Decision-making Under Risk.* Lexington Books, Lexington, Massachusetts, 1978.

[17] Trevor Willcox. *Australian Canegrower*, 21'st March 1994.