



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

**Simulating multivariate distributions with sparse data:
a kernel density smoothing procedure**

James W. Richardson

Department of Agricultural Economics, Texas A&M University

Gudbrand Lien

Norwegian Agricultural Economics Research Institute, Oslo, Norway

J. Brian Hardaker

School of Economics, University of New England, Armidale, Australia

Poster paper prepared for presentation at the International Association of
Agricultural Economists Conference, Gold Coast, Australia,

August 12-18, 2006

*Copyright by James W. Richardson, Gudbrand Lien, and J. Brian Hardaker. All rights reserved.
Readers may take verbatim copies of this document for non-commercial purposes by any means,
provided that this copyright notice appears on all such copies.*

Simulating multivariate distributions with sparse data: a kernel density smoothing procedure

Abstract

Often analysts must conduct risk analysis based on a small number of observations. This paper describes and illustrates the use of a kernel density estimation procedure to smooth out irregularities in such a sparse data set for simulating univariate and multivariate probability distributions.

JEL classification: Q12; C8

Keywords: stochastic simulation; smoothing; multivariate kernel estimator, Parzen

1. Introduction

All businesses face risky decisions, such as: enterprise mix, marketing strategies, and financing options. Agricultural producers face production risks from weather, pests, and uncertain input responses so production risk remains a significant consideration for agricultural business managers. Hence, risk should be explicitly considered in studies of agricultural production choices (e.g., Reutlinger, 1970; Hardaker et al., 2004).

Agricultural businesses are often best studied in a system context, implying a need to cast the risk analysis in a whole-farm context. Several methods exist for whole-farm analysis under uncertainty. One frequently and increasingly used approach is whole-farm stochastic simulation. In stochastic simulation, selected input variables or production relationships incorporate stochastic components (by specifying probability distributions) to reflect

important parts of the uncertainty in the real system (Richardson and Nixon, 1986; Hardaker et al., 2004). If stochastic dependencies between variables are ignored for ease in modelling, the distribution of the output variable(s) may be seriously biased (Richardson et. al., 2000).

Probabilities for whole-farm stochastic simulation should ideally be based on reliable farm-level histories for the stochastic variables. In practice, the required historical data may not be available for the farm to be analyzed. In particular, there will seldom be extensive and wholly relevant historical data for all stochastic variables. The typical situation is that the available, relevant data are too sparse to provide a good basis for probability assessment (e.g., Hardaker et al., 2004, ch. 4.).

In the case of sparse data the underlying probability distribution which generated the observations cannot be easily estimated. For example, three samples of sparse data were generated at random from a normal distribution with a mean of 100 and a standard deviation of 30 (or $N(100,30)$) and depicted in Figure 1. The known distribution is depicted as a probability distribution function (PDF) with the sparse samples shown as points on the X axis. With such few data points one cannot reliably estimate the frequency that a value will be observed. It is evident from the three sparse samples that it is extremely difficult to consistently estimate both the form of the probability distribution and the parameters which generated the sample.

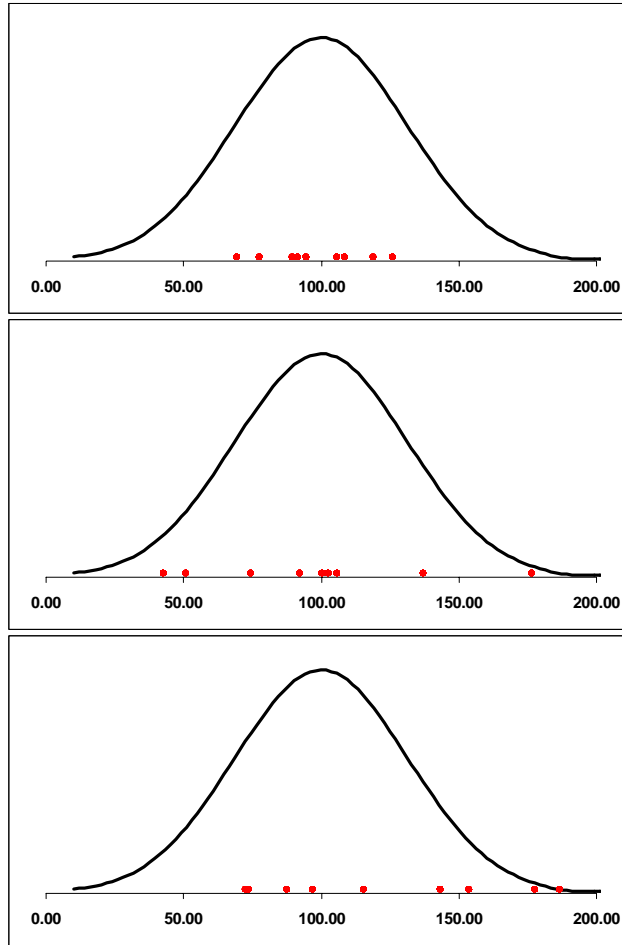


Figure 1. Example of three sparse samples generated at random from a normal distribution with mean 100 and standard deviation 30.

Hardaker and Lien (2005) suggest that in the case of sparse data, one should use the best probability judgments about the uncertain variables so the analysis can proceed. Sparse historical data are obviously useful in forming such judgments but only in conjunction with careful assessment of their relevance and of how to accommodate the limitations of the data in seeking to make the probability estimates more consistent. There are often irregularities in sparse data due to sampling errors, and it is useful to smooth them out by fitting a distribution (Schlaifer, 1959, ch. 31; Anderson et al., 1977, pp. 42-44). Further, it is usually reasonable to assume that the upper and lower bounds of a distribution, if they exist, will be more extreme than those observed from a sparse data set, due to the under sampling of the tails endemic in

most sampling regimes. It may be possible to use expert opinion to assess the likely upper and lower bounds of the distribution, and sometimes such opinions can be used to identify other features of the distribution. Such assessments and augmentations can be included before smoothing is attempted.

Although the kernel density smoothing estimator is extensively used in many fields to estimate probability distributions prior to simulation, it is not widely used in agricultural economics simulation studies. The aim of this paper is to illustrate the use of a kernel density estimation procedure to smooth out irregularities in a sparse data set for simulating univariate and multivariate probability distributions.

2. Alternative smoothing procedures

It is generally safe to assume that the marginal cumulative density function (CDF) of some continuous uncertain quantity is a smooth sigmoidal curve (Anderson et al., 1977; Hardaker et al., 2004). So one smoothing option is to fit a parametric probability distribution (such as the normal) to the sparse data based on statistics computed from the sample. As demonstrated in Figure 2 this procedure may not produce a reasonable approximation of the underlying distribution. In Figure 2 the assumed normal distribution estimated from the sparse samples' moments was plotted along with the known distribution which generated the sparse sample of nine dots. The simulated distribution very closely follows and smoothes out the sparse sample points. However, the smoothed distribution estimated from the sparse sample's moments would significantly under sample the tails and over sample the mid-section of the known distribution.

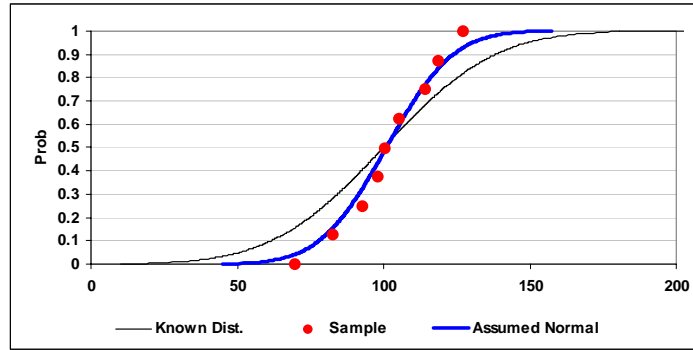


Figure 2. The CDF of a sparse sample of nine points plotted against the known CDF and the assumed normal CDF.

A better option might be to use distribution-fitting software such as BestFit or Simetar to fit a parametric distribution and test its appropriateness in simulation of the random variable (e.g., Richardson, 2004; Hardaker et al., 2004). However, it may be unsafe to assume that a sparse sample conforms adequately to some unimodal parametric distributional form. Moreover, with sparse data it is often impossible to statistically invalidate many reasonable parametric distribution forms (Ker and Coble, 2003), making choice among them difficult. Yet different functional forms may lead to quite different results in simulation, especially in the tails of the distributions.

Another option is to plot the fractile values, and to interpolate linearly between the empirical point estimates to get a continuous empirical probability distribution (Richardson et al., 2000), as illustrated in Figure 3. Obviously, this method does not fully smooth out the CDF, especially with small samples and additionally it underestimates the frequency with which the end points are simulated (Figure 3). Therefore a better, although more subjective, choice may be to draw a smooth curve approximating these points by hand (Schlaifer, 1959, ch. 31; Hardaker et al., 2004, pp. 69-71). Hand smoothing a CDF curve through or close to the plotted fractile estimates gives the opportunity to incorporate additional information about the shape and location of the distribution, which is not always possible with computer-based approaches. Nonparametric estimation methods can be used as a formalization of hand

smoothing. Several spline-fitting methods exist, such as penalized least-square and piecewise polynomial (e.g., cubic) least-square procedures (see, for example, Silverman, 1985; Yatchew, 1998).

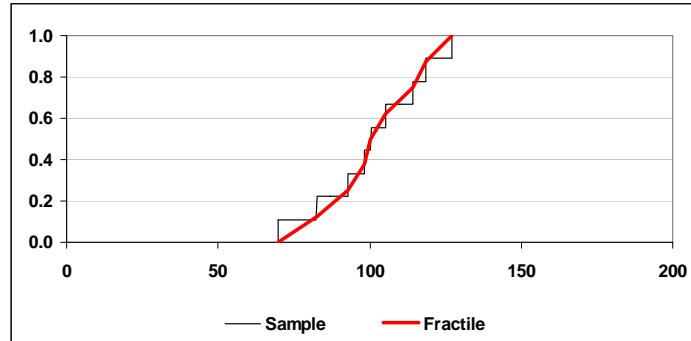


Figure 3. CDF of a sparse sample of nine points plotted as a discrete empirical vs. the fractile values for the sample.

Instead of minimizing the sum of squared residuals, the classical nonparametric kernel density estimation method weights observations based on relative proximity to estimate a probability. The basic idea is that one slides a weighting window along the number scale, and then estimate of the probability at a given point depends on a pre-selected probability density. The smoothed estimate is a result of the individual observations that are weighted based on their relative positions in the window. In that way the kernel estimator is analogous to the principle of local averaging, by smoothing using evaluations of the function at neighbouring observations (Yatchew, 1998). Several kernels are available that control the shape of the local distribution, e.g., uniform, triangular, Parzen, Epanechnikov, and Gaussian. A smoothing parameter, also called the window or bandwidth, dictates the weighting frequency of smoothing, i.e. how much influence each observation point has. A large bandwidth will put more significant weight on many observations while a small bandwidth will only put weight on neighbouring observations in calculating the probability estimate at a given point. The bandwidth can also be varied from one point observation to another. The smoothed curve in Figure 4 shows the CDF for a Parzen kernel smoothed distribution of the same sparse sample

used in Figures 2 and 3. The Parzen kernel fit a smooth curve through the sample points rather than linearly interpolating between the sample points and does not in this case under sample the frequency observed for the end points.

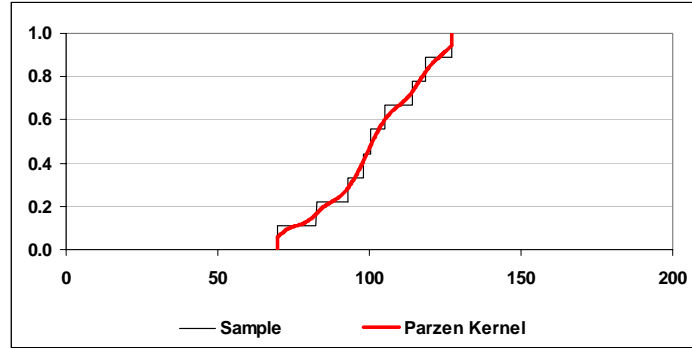


Figure 4. Smoothed Parzen kernel distribution CDF and a linear interpolated CDF for a sparse sample drawn from a known normal distribution.

3. Application of kernel distribution to sparse samples

Three sparse samples of crop yields are summarized in Table 1. The data were analyzed with Simetar to estimate the parameters for alternative kernel distributions and simulated using an empirical distribution estimate of the 500 points generated by the Parzen kernel smoothing procedure (Parzen, 1962).¹ The Parzen kernel was selected after determining that it produced the smallest root mean square error (RMSE) of residuals ($\hat{\epsilon}_i = F(x_H) - F(x_j)$) between the histogram kernel cumulative probabilities, $F(x_H)$, and those for the j th kernel, $F(x_j)$, in the following list of kernels: (1) cosines, (2) Cauchy, (3) double exponential, (4)

¹ In practice, in cases with such small samples as nine observations it is very important to evaluate the usefulness of the data before using them in a stochastic simulation model. The first step should be to consider how representative and relevant these data are likely to be to describe the uncertain quantity to be modelled. Is there a need to correct for any trend, due, for example, to technological change or to price responses? One should check whether the seasonal conditions in these nine years were reasonably representative of the general climatic conditions in the area. If not, it would be appropriate to assign differential probabilities to the observations. And the possible bias in the data needs to be addressed. If the simulation model represents commercial farming conditions and the data are from an experiment, it might be necessary to account for the fact that experimental yields are often higher than farm yields. In this illustration with synthetic simulated data we assume that any such adjustments are unnecessary.

Epanechnikov, (5) Gaussian, (6) Parzen, (7) quartic, (8) triangle, (9) triweight, and (10) uniform.

Table 1. Sparse samples of yields for three crops with suggested subjective minimums and maximums

	Yield 1	Yield 2	Yield 3
1991	6,870.6	6,897.1	31,430.0
1992	6,070.6	5,467.6	40,800.0
1993	6,323.5	6,111.8	42,650.0
1994	4,688.2	5,226.5	30,740.0
1995	2,717.6	4,905.9	31,520.0
1996	3,658.8	5,502.9	27,290.0
1997	5,358.8	3,811.8	26,230.0
1998	5,670.0	4,710.0	27,390.0
1999	3,800.0	5,910.0	19,500.0
Subj. Max	8,700.0	8,600.0	49,000.0
Subj. Min	1,600.0	1,800.0	15,000.0

Three sparse samples were used to demonstrate how a Parzen kernel density can be used to smooth actual samples. The samples represent nine years of actual crop yields (Table 1). Expert opinion can be used to augment a sparse sample by adding minimum and maximum values that extend the observed distribution. Once the augmented minimum and maximum are specified, the Parzen kernel can be used to estimate the parameters to simulate the augmented distribution. The three sparse samples of yield data were augmented using expert opinion (Table 1) and simulated using the smoothing effect of a Parzen kernel (Figure 5). The resulting CDFs from simulating the random variable with 500 iterations suggest that the augmented distributions were indeed smoothed in the process. For all three distributions the Parzen kernel sampled the extreme points with about the same frequency as they were observed in the original sample which results in vertical end points in the CDFs. The remaining frequency is spread over the remaining points in the sample using the bandwidth chosen for the Parzen kernel. For all three distributions the Parzen kernel smoothed over the effect of outliers and produced a smoothed approximation of the CDF for the distributions.

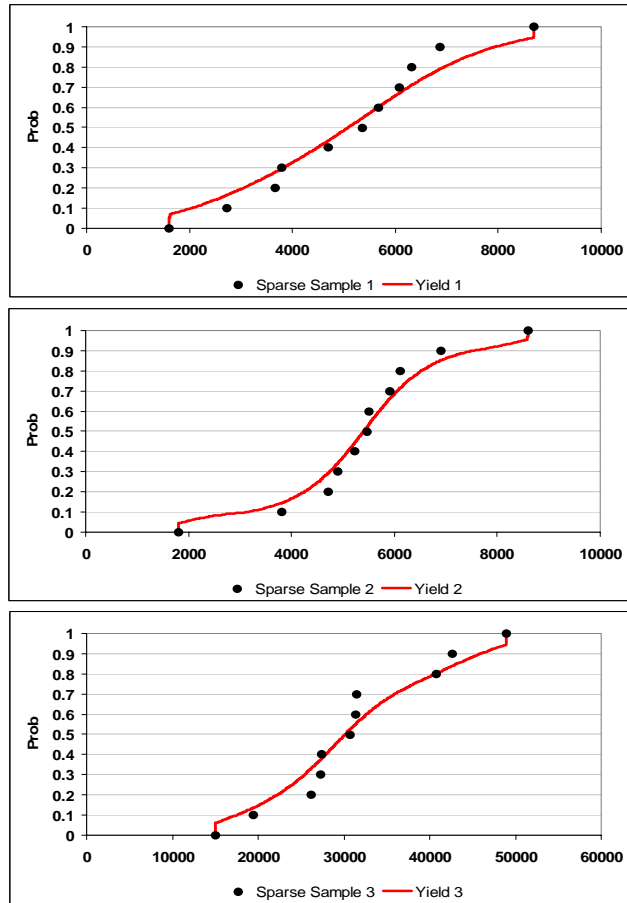


Figure 5. Examples of the Parzen kernel to simulate three sparse sample distributions augmented using expert advise.

The Parzen kernel smoothing procedure appears to work well in the univariate case, however, to be useful for whole farm simulation modelling the procedure must be applicable in the multivariate case. The next step is to extend Parzen kernal smoothed distributions to the multivariate case, as illustrated in the next section. For more details about kernel density methods, see Silverman (1986) and Scott (1992).

4. The multivariate kernel density simulation procedure

The general version of the multivariate empirical (MVE) distribution estimate procedure described by Richardson et al. (2000) is generalized for a smoothed multivariate distribution. The procedure uses a kernel density estimation (KDE) function to smooth the limited sample data of variables in a system individually, and then the dependencies present in the sample are

used to model the system.² The resulting stochastic procedure is called the multivariate kernel density estimate (MVKDE) of a random vector. The steps in specifying a MVKDE distribution are described below:

1. The starting point is the matrix of (often historical detrended) observations (often yields or prices):

$$\mathbf{X}_{i \times j} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1j} \\ X_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ X_{n1} & \cdots & \cdots & X_{nk} \end{bmatrix} \quad (1)$$

where $i = 1, \dots, n$ observations or states and $j = 1, \dots, k$ variables.

2. The estimated correlation matrix used to quantify the (linear) correlation among the variables is calculated using $\mathbf{X}_{i \times j}$ as:

$$\mathbf{P}_{k \times k} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1k} \\ & 1 & \cdots & \vdots \\ & & \ddots & r_{(k-1)k} \\ \text{sym.} & & & 1 \end{bmatrix} \quad (2)$$

where r_{ij} is the sample correlation coefficient between the vectors $X_{.j}$ and $X_{.j}$ for $j = 1, \dots, k$. The r_{ij} coefficients are typically product-moment correlations, but a similar procedure can be used for rank-based methods.

3. The correlation matrix is factored by the Cholesky decomposition:

$$\mathbf{R}_{k \times k} \text{ such that } \mathbf{P} = \mathbf{R}\mathbf{R}^T \quad (3)$$

² The model presented has been programmed in Excel using the Simetar Add-In (<http://www.simetar.com/>). This user-friendly software makes implementation easy. The simulation, risk analysis, and econometric capabilities of Simetar are described in Richardson (2004).

4. The minimum, $X_{Min,j}$, and the maximum, $X_{Max,j}$ bounds for each variable, k are determined. The cumulative probabilities for these are assumed to be $F(X_{Min,j})=0$ and $F(X_{Max,j})=1$.

5. For each variable, k , a new vector, X_{sj}^A , of dimension S ($s=2,\dots,S$) is created with given minimum $X_{1j}^A = X_{Min,j}$ (i.e. $s=1$ for the minimum observation) and maximum $X_{Sj}^A = X_{Max,j}$ by the formula:

$$X_{sj}^A = \left(\frac{1}{S-1}\right)(X_{Max,j} - X_{Min,j}) + X_{(s-1)j}^A \quad (4)$$

6. The smoothed percentiles for each X_{sj}^A between the extreme bounds $F(X_{Min,j})=0$ and $F(X_{Max,j})=1$ are based on a kernel density estimator (Silverman, 1986; Scott, 1992). For the j^{th} variable, the smoothed percentile is evaluated at a given point X_{sj}^A as:

$$\hat{F}(X_{sj}^A) = \frac{1}{nh_j} \sum_{i=1}^n K \left[\frac{(X_{sj}^A - X_{ij})}{h_j} \right] \quad (5)$$

where $K(\cdot)$ is the cumulative kernel function associated with a symmetric continuous kernel density $k(\cdot)$ such that $K(x) = \int_{-\infty}^x k(t)dt$, and h_j is the bandwidth.

The MVKDE of the random vector \tilde{X}_j is simulated as follows for each iteration or sample of the possible expanded states of nature:

1. Generate a k -vector, $\mathbf{z}_{k \times 1}$, of independent standard normal deviates (ISNDs).

2. Pre-multiply the vector of ISNDs by the factored correlation matrix to create a k -vector, $\mathbf{z}_{\mathbf{k}\times 1}^*$, of correlated standard normal deviates (CSNDs):

$$\mathbf{z}_{\mathbf{k}\times 1}^* = \mathbf{R}\mathbf{z} \quad (6)$$

3. Transform each of the j CSNDs to correlated uniform standard deviates (CUSDs) using the error function:

$$CUSD_j = ERF(z_j^*) \quad (7)$$

The error function, $ERF(z)$, is the integral of the standard normal distribution from negative infinity to z , which is the z -value from a standard normal table. The result of the function will be a value between zero and one.

4. The quantile from the j^{th} smoothed empirical distribution function is found by applying the inverse transform method. Given the $CUSD_j$ along with the respective vectors X_{sj}^A and $\hat{F}(X_{sj}^A)$, interpolate among the X_{sj}^A to calculate a random vector \tilde{X}_j . The procedure, as described by Richardson (2004), is as follows: First, a $CUSD_j$ vector is generated (from step 3 above). Second, the generated $CUSD_j$ is matched into its interval on the probability $\hat{F}(X_{sj}^A)$ scale (including $F(X_{Min,j})$ and $F(X_{Max,j})$) between the nearest lower $\hat{F}_L(X_{Lj}^A)$ and the nearest upper $\hat{F}_U(X_{Uj}^A)$. Third, match up the corresponding interval, between X_{Lj}^A and X_{Uj}^A (including the extremes $X_{Min,j}$ and $X_{Max,j}$). Fourth, generate one vector of simulated MVKDE variables using the formula:

$$\tilde{X}_j = X_{Lj}^A + (X_{Uj}^A - X_{Lj}^A) * \frac{(CUSD_j - \hat{F}_L(X_{Lj}^A))}{(\hat{F}_U(X_{Uj}^A) - \hat{F}_L(X_{Lj}^A))} \quad (8)$$

Three points should be noted. First, the first three steps that account for the dependency of the stochastic variables are the way random variates are generated from a normal (or Gaussian) copula (e.g., Cherubini et al., 2004). In agricultural economics, Richardson and Condra (1978) and King (1979) pioneered the normal copulas in their simulation models. Copulas other than the normal (e.g., t -copulas, skewed t -copulas) could have been used to deal with stochastic dependency. However, since the parameters in the copulas are estimated from the data set, the estimation procedures are problematic with sparse data. Second, our MVKDE procedure differs from conventional MVKDE procedures, since minimum and maximum bounds are included in the simulation procedure. Third, by choosing S in Eq. (4) large enough, the simulation procedure in step (4) will give so many points in the linear interpolation function that the approximation error will be minimal.

5. Application of MVKDE

The steps described to simulate a MVKDE distribution were applied to the sparse sampled yield data in Table 1. The correlation matrix for the three random variables was estimated without the subjective minimums and maximums. The MVKDE defined by the Parzen kernels for the augmented distributions and the correlation matrix was simulated for 500 iterations. Four statistical tests were performed on the simulated values for validation of the MVKDE procedure.

The first validation test is a series of Student- t tests, suggested by Vose (2000, p. 54), to test whether the simulated values are appropriately correlated. Adjusting the overall 95% confidence interval for multiple testing, the critical t -value is 2.40 at the individual 98.3% level and the four off-diagonal t -test statistics are all less than 1.5 (Table 2). This result indicates that the historical correlation coefficients observed for the three variables in MVKDE distribution were not statistically different from their respective correlation coefficients in the simulated data.

Table 2. Statistical validation tests to determine if the MVKDE procedure appropriately simulated the augmented yield distributions

Student-t Test of the Correlation Coefficients			
Confidence Level	98.3048%		
Critical Value	2.40		
	Yield 2	Yield 3	
Yield 1	0.51	1.45	
Yield 2		0.72	

Comparison of Sparse Data Multivariate Distribution to the Simulated Values for a MVKDE Distribution			
Confidence Level	95.0000%		
	Test Value	Critical Value	P-Value
Two Sample Hotelling T^2 Test	0.01	7.90	1.000
Box's M Test	6.51	12.59	0.368
Complete Homogeneity Test	10.33	16.92	0.324

The augmented mean vector of the three yield variables was not statistically different from the mean vector for the simulated data at the 95% level, according to the Two Sample Hotelling T^2 Test (Table 2) (Johnson and Wichern, 2002, pp. 210-220). The Box's M Test (Box, 1953) was used to test if the historical covariance matrix for the multivariate distribution was statically different from the covariance matrix observed for the simulated values. The null hypothesis that the two covariances are equal at the asymptotic 95% level is not rejected. The Complete Homogeneity Test simultaneously tested the historical mean vector and covariance to their respective statistics in the simulated data. The statistical test results fails to reject the hypothesis that the simulated means vector and covariance differ from their historical values at the 95% level (Table 2).

These four statistical tests indicate that the MVKDE procedure simulated the smoothed Parzen kernel distributions appropriately, in that it did not change the dependencies of the variables, did not change the means of the variables, and did not change the variance of the variables. Using the MVKDE procedure with a Parzen kernel distribution, it is easy to

simulate a sparse multivariate sample and ensure that historical correlation is maintained as long as the correlation matrix is not singular.

6. Summary, discussion and conclusion

It is always important to consider the reliability and the temporal and spatial relevance of data to be used in risk assessment, and to decide how to accommodate data imperfections. These matters are especially important when risk analyses must be based on sparse samples. Often the data are too few to provide a good basis for consistent probability distribution assessment. Several procedures for smoothing the CDF of a sparse sample are available in the literature; however, each procedure has its limitations. The use of kernels to smooth sparse samples is demonstrated in the paper. To expand the usefulness of kernel smoothing for whole-farm simulation modelling, a multivariate kernel smoothing procedure has also been illustrated.

The result provided for use of the multivariate kernel estimator to smooth out irregularities in sparse data show that this method is a useful methodological advance. The MVKDE method is more flexible than strictly parametric probability methods, and intuitively better than a simple linear interpolation of the empirical distribution.

The study also leads to a number of ideas for further research. The proposed smoothing method is only one of many alternatives for use in farm-level analysis. It would be useful to conduct a comprehensive study using Monte Carlo simulation to generate a synthetic, uncertain farm-level case in order to compare alternative smoothing methods. In the statistical field there is extensive discussion about choice of kernel function and bandwidth. Hence, for farm-level simulation models there is a need to explore the effect of choice of kernel function and bandwidth on model results. More generally, it seems that more work is needed to compare different smoothing techniques in combination with more advanced stochastic dependency modelling.

References

- Anderson, J.R., Dillon, J.L. and Hardaker, J.B. (1977). *Agricultural Decision Analysis*. Iowa State University Press, Ames.
- Box, G.E.P. (1953). Non-normality and tests on variances. *Biometrika* 40: 318-335.
- Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*. Wiley Finance, Chichester, UK.
- Hardaker J.B., Huirne R.B.M., Anderson J.R. and Lien G. (2004). *Coping with Risk in Agriculture, 2nd edn*, Wallingford: CABI Publishing.
- Hardaker, J.B. and Lien, G. (2005). Towards some principles of good practice for decision analysis in agriculture. Norwegian Agricultural Economics Research Institute, Oslo, Working paper 2005-1.
- Johnson, R.A. and D.W. Wichern. (2002). *Applied multivariate statistical analysis*. Prentice Hall, fifth edition, Upper Saddle River, NJ.
- Ker, A.P. and Coble, K. (2003). Modeling conditional densities. *American Journal of Agricultural Economics* 85: 291-304.
- King, R.P. (1979). Operational techniques for applied decision analysis under uncertainty. Ph.D. dissertation, Department of Agricultural Economics, Michigan State University.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33: 1065-1076.
- Reutlinger, S. (1970). Techniques for project appraisal under uncertainty. World Bank Occasional Paper No. 10. Baltimore: John Hopkins Press.
- Richardson, J.W. (2004). Simulation for Applied Risk Management. College Station, TX: Department of Agricultural Economics, Texas A&M University.
- Richardson, J.W. and Condra, G.D. (1978). A general procedure for correlating events in simulation models. Department of Agricultural Economics, Texas A&M University.
- Richardson, J.W. and Nixon, C.J. (1986). Description of FLIPSIM V: a general firm level policy simulation model. Texas Agricultural Experiment Station, Bulletin B-1528.
- Richardson, J.W., Klose, S. L. and Gray, A.W. (2000). An applied procedure for estimating and simulating multivariate empirical (MVE) probability distributions in farm-level risk assessment and policy analysis. *Journal of Agricultural and Applied Economics* 32: 299-315.
- Schlaifer, R. (1959). *Probability and Statistics for Business Decisions*. McGraw-Hill, New York.
- Scott, D.W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York.
- Silverman, B.W. (1985). Some aspects of the spline smoothing approaches to non-parametric regression curve fitting (with discussion). *Journal of Royal Statistical Society B* 47: 1-52.
- Silverman, B. W. (1986). *Density Estimation and Data Analysis*. Chapman and Hall, New York.
- Vose, D. (2000). *Risk Analysis: a Quantitative Guide*. John Wiley & Sons, Ltd., second edition, New York.
- Yatchew, A. (1998). Nonparametric regression techniques in economics. *Journal of Economic Literature* 36: 669-721.