PARAMETRIC AND NON-PARAMETRIC CROP YIELD DISTRIBUTIONS AND THEIR EFFECTS ON ALL-RISK CROP INSURANCE PREMIUMS

by

Calum Turvey*
and
Jinhua Zhao**

WORKING PAPER WP99/05

Department of Agricultural Economics and Business
University of Guelph
Guelph, Ontario

January 1999

*Calum Turvey is an associate professor in the Department of Agricultural Economics and Business at the University of Guelph.
**Jinhua Zhao is an assistant professor in the Department of Economics at Iowa State University.

THE DEPARTMENT OF AGRICULTURAL ECONOMICS AND BUSINESS

PARAMETRIC AND NON-PARAMETRIC CROP YIELD DISTRIBUTIONS AND THEIR EFFECTS ON ALL-RISK CROP INSURANCE PREMIUMS

by

Calum G. Turvey

and

Jinhua Zhao

Department of Agricultural Economics and Business

University of Guelph

Guelph, Ontario.

January, 1999

Calum Turvey is an associate professor in the Department of Agricultural Economics and Business at the University of Guelph and Jinhua Zhao is an assistant professor in the department of economics at Iowa State University.

PARAMETRIC AND NON-PARAMETRIC CROP YIELD DISTRIBUTIONS AND THEIR EFFECTS ON ALL-RISK CROP INSURANCE PREMIUMS

Abstract

Normal, gamma and beta distributions are applied to 609 crop yield histories of Ontario farmers to determine which, if any, best describe crop yields. In addition, a distribution free non-parametric kernel estimator was applied to the same data to determine its efficiency in premium estimation relative to the three parametric forms. Results showed that crop yields are most likely to be described by a beta distribution but only for 50% of those tested. In terms of efficiency in premium estimation, minimum error criteria supports use of a kernel estimator for premium setting. However, this gain in efficiency comes at the expense of added complexity.

Key words: crop insurance, crop yield distributions, kernel distributions, insurance premiums

PARAMETRIC AND NON-PARAMETRIC CROP YIELD DISTRIBUTIONS AND THEIR EFFECTS ON ALL-RISK CROP INSURANCE PREMIUMS

Actuarial soundness of publicly provided all-risk crop insurance is an important criterion in setting premiums because of its influence on farmers' participation in crop insurance and it's impact on government expenditures. A premium is said to be actuarially sound if it equals expected indemnities thereby implying an expected zero-profit position for both the insured and the insurer if transaction costs are zero. Actuarial soundness by definition requires assessment of the underlying probability distribution of crop yields.

Prior assumptions of underlying yield probabilities which differ from actual probabilities can lead to errors in premium setting which can lead to problems of adverse selection. Adverse selection is the result of asymmetry in information between insureds and the insurer. In the absence of perfect information, insurers are unable to completely define yield risks and consequently cannot fully appraise risk for premium setting. Consequently, crop insurers tend to offer all-risk crop insurance using pooled or aggregate yield data to estimate premiums but pay indemnities based on individual yield outcomes. All farmers growing a specific crop pay a common premium with individual yield and indemnity histories being used only to make adjustments to the base premiums over time. Thus many farmers with different yield risk profiles can purchase insurance for essentially the same price with high risk farmers receiving substantially more benefits than low risk farmers.

Mathematically, yield insurance premiums on an individual farm basis are computed using

$$PI(f) = P \int_{-\infty}^{z} (Z-Y) f(Y)\, dY \tag{1}$$

where Z is the coverage level, P the elected price level, Y farm yields, and f(Y) the assumed probability distribution. If the true distribution is defined by the probability distribution function

$$PI(f) - PI(g) = P \int_{-\infty}^{z} (Z-Y) (f(Y) - g(Y))\, dY. \tag{2}$$

g(Y), then the error in actuarial premiums is

If farmers are aware of their own probability of loss, g(Y), then adverse selection could take place with many high risk farmers participating when A(f) - A(g) < 0 and many low risk farmers opting out when A(f) - A(g) > 0.

In view of this, one of the more pressing problems facing insurance rate makers is the ambiguous nature of yield risks which differ across crop types and even within crop type but across different crop growing regions. Differences in yield risk can be manifested in different yield distributions. For example different combinations of soil type and climatic conditions can affect yield distributions in such a way that crops grown in one region may have a high probability of receiving a low yield outcome, while in other regions the same crop with the same mean yield may have a high probability of obtaining a high yield outcome. How different assumptions about crop yield probability distributions affect crop insurance premiums is a goal of this study. In particular the research attempts to assess yield distributions of Ontario cash crops in relation to common parametric forms (e.g. normal, beta, gamma); evaluate the stability of commodity yield distributions across different farms in different regions of Ontario; and to assess insurance premium differentials under alternative yield distributions.

The problem of determining probabilistic outcomes as they relate to actuarial science and insurer liabilities is one which is common across all types of insurance and contingent liabilities. Assessment of probabilities in relation to differing assumptions about the underlying probability distribution is however complex in most cases. Crop insurance is an exception, because historical yield observations are generally available. Moreover, with the exception of trends due to technology, dynamics do not affect state contingent outcomes. Thus, allowing outcomes to be temporally independent permits use of static distributions which are less complex than dynamic ones. None-the-less, true crop yield distributions are still unknown, and whether or not they follow one parametric type or another is usually assumed. However, it may be the case where none of the common parametric distributions adequately reflect the empirical distribution. To address this problem non-parametric techniques must be employed. Consequently, an important aspect of this research is the application of a non-parametric kernel estimating technique which is employed for comparative purposes as an alternative hypothesis to the parametric distributions. Crop insurance premiums are also derived from the kernel distribution. Since the kernel function is flexible it may be able to take on the basic characteristics of the two-parameter distributions listed above, and in this respect provide a minimal error approach to premium estimation.

The manuscript proceeds as follows: in the next section a review of literature pertaining to crop yield distributions is made; next the parametric and non-parametric distributions used in the study are presented and the data described; the following sections present results and conclude the paper.

BACKGROUND

Some parametric distributions have been used or tested in research related to crop yield. Day (1965) recommended the beta distribution for crop yield; Nelson and Preckel (1989) used the beta distribution in their production function estimation; Pope and Ziemer (1984) and Gallagher (1986) suggest a gamma distribution and Kenkel et al. (1991) tested the normal, beta, gamma, logistic, and extreme type A distributions (see Johnson and Kotz (1970a and 1970b) for their description). As with other research Kenkel et al. (1991) are unable to establish a distribution prior. Common distributions used in crop yield analyses include normal, beta and gamma distributions. These distributions, all of which are used in this study, evolve as a two-parameter family since entire distributions can be described by the mean and variance. These parametric distributions will be compared to a non-parametric kernel density function. The focus is solely on the nature of the distribution functions themselves and how differing assumptions of probability distribution affect actuarial crop insurance premiums. Each distribution is described in the following section. Detailed reviews of the parametric distributions can be found in Johnson and Kotz (1970 a,b), or Morgan and Henrion (1990).

Normal Distribution

The density function for the normal distribution with mean µ and standard deviation F is

$$f(Y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left[\frac{y-\mu}{2\sigma}\right]^{2}} \tag{3}$$

given by

The normal distribution is symmetric and unbounded between [4, 4]. In most studies of yield distribution however the likelihood of a predicted f() being less than zero is negligible although the probability of 0 yield is finite. Typically the normal distribution will be truncated at Y = 0.

Beta Distribution

$$f(Y) = \frac{1}{\beta(p,q)} \cdot \frac{(y-a)^{p-1}(b-y)^{q-1}}{(b-a)^{p+q-1}} \tag{4}$$

The beta distribution is given by

where a and b, (a # y # b), are scale parameters which define the boundary of the distribution; p

$$\beta(p,q) = \frac{\Gamma(p)\,\Gamma(Q)}{\Gamma(p+Q)} \tag{5}$$

and q, (p > 0, q > 0) are shape parameter; $(p,q) is the beta function,

and ((C) are gamma functions with

$$\Gamma(p) = \int_{o}^{\infty} X^{p-1} e^{-x} \, dX \tag{6}$$

and X = (y-a)/(b-a). Note that if a and b are known then $\$(p,q)$ can be derived from only 2 parameters. (In this study a = 0 and b equals maximum observed yield plus one standard

$$\tilde{p} = \frac{\mu^2 - \mu^3 - \sigma^2 \mu}{\sigma^2}, \tag{7}$$

deviation.) Estimates of p and q can be obtained from the sample mean, μ, and variance, F^2 with

$$\tilde{q} = \frac{\mu (1-\mu)^2 - \sigma^2 (1-\mu)}{\sigma^2} \tag{8}$$

and

The beta distribution is symmetric when $p = q$, positively skewed when $p < q$ and negatively skewed when $p > q$.

Gamma Distribution

$$f(Y) = \frac{Y^{a-1} e^{-y/b}}{b^a \Gamma(a)} \tag{9}$$

The form of Gamma distribution is for Y $ 0

where a and b are shape parameters (a, b > 0) and using the sample mean and variance $a = \mu^2/F^2$ and $b = \mu/F^2$.

Shape and scale parameters of the 3 distributions are defined by the sample mean and standard deviation using the methods of moments. Alternatively these parameters could be estimated through maximum likelihood techniques as in Antle (1983), Taylor (1984), and Nelson and Preckel (1989). Day (1965) comments that maximum likelihood techniques would provide little improvements over the method of moments, a conclusion which was affirmed by our pretest of the two techniques to a small sub-sample of farms.

Non-parametric Kernel Estimators

In contrast to the above parametric procedures an alternative non-parametric technique is available. Kernel estimators belong to the class of non-parametric estimators such as histogram which estimate density functions directly from the sample without assuming any distributional form. Given an ordered sample Y_1, Y_2 .... Y_n, the kernel estimator of the density function f(Y) with

$$\hat{f}(Y) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{Y-Y_i}{h}\right) \qquad \textbf{(10)}$$

kernel K is defined by (Silverman 1986)

where h is a smoothing parameter and K satisfies $_{-4}I^4$ k(Y)dY = 1. Note that the function K can be any valid probability distribution function and this distribution function is defined around each of the n elements of the empirical distribution. Further note the role of the smoothing parameter h; h is also called the window width because as it increases (decreases) the dispersion of outcomes around the discrete point Y_i decreases (increases). Hence choice of k(C) and h fully determine the kernel estimator. In this study k(C) is assumed to be a triangular distribution.[i] The smoothing parameter h was chosen to minimize the mean integrated square error between the estimate (Y) and the true empirical F(Y).[ii] (Exact procedures are beyond the scope of this text but can be found in Silverman 1986). The procedure assumes that the kernel estimates are at least

---

[i]. To be more precise, other kernels including the biweight, the normal and the Epanechivok, were also tried (see Silverman). Biweight and Triangular results were comparable, and normal and Epanechivok were comparable, but results from the former were more realistic than the latter. In addition, an adaptive kernel, in which h is constructed to differ at varying locations, was also tried but improvements were negligible.

[ii]. If (x) is an estimate of the true distribution f(x) then the mean square error (MSE) is $E[(x) - f(x)]^2$ and the mean integrated square error is MISE = IMSE (f(x))dx = E $I[(x) - f(x)]^2 dx$. The MSE measures error at a point whereas MISE is a global measure of efficiency.

asymptotically unbiased and consistent (Rao 1983).

Yield Distributions and Goodness of Fit

A statistical approach to evaluating the goodness of fit of the 3 parametric distribution types to the empirical distribution is the non-parametric Kolmogorov-Smirnov (K-S) test which has good properties for small sample data. Although it has been found not to perform well against a null hypothesis of normal distribution and the Shapiro-Wilk test (Shapiro and Ziemer 1968) it must be recognized that the K-S test is distribution free whereas the Shapiro-Wilk test applies only to tests of normality.

In particular if F(Y) is the true but unknown cumulative probability distribution and (Y) is the hypothesized CDF then the null hypothesis of the goodness of Fit test is H_0: F(Y) = (Y). The alternative hypothesis is a composite hypothesis which includes all distributions other than the hypothesized one (Birnbaum 1953).

The K-S test is based on a distance measure equal to D_n = Sup[F(Y)-(Y)] with D_n being distribution free for any continuous (Y) (Gibbons 1985). Being distribution free enables the comparison of different D_n's for different hypothesized distributions. If more than 1 distribution type is hypothesized it is possible that the K-S test could fail to reject all null hypotheses. However, a feature of the distance measure is that the hypothesized distribution with the lowest D_n value can be used with more confidence. Hence, while the K-S test provides credibility for selection of a particular distribution type there always exists a possibility of Type II error. Even though the K-S test is consistent against any alternative F(Y) à (Y) it may be asymptotically biased with a finite sample. Fortunately, a significant advantage of the K-S test statistic is that it has small (finite) sample properties, so that the test can be considered exact with small sample size (Bradley 1968). This provides a considerable advantage over other goodness-of-fit tests (e.g. Chi Square) when the sample size is small (Gibbons 1985). However, a major disadvantage of the K-S test is that confidence levels are based on the assumption that $F_0(x)$ is fully described by the population parameters. When mean and variance are estimated from the sample, the K-S test may be conservative (Gibbons 1985).

Premium Estimation

Premiums for the 3 parametric and 1 non-parametric distributions were computed according to equation (1). Since no closed form for the integration in (1) exists, integration was calculated by numerical approximation. The interval [O,Z] was divided into 1000 ordered sub-intervals of equal width. Each sub-interval was taken as a trapezoid and its area calculated accordingly. The value of the integration was the sum of the areas of all the sub-intervals multiplied by their respective values of P(Z-Y) where Y was taken as the lower point of the interval and P is the exogenously set price.

Data

Data were drawn selectively from over 96,000 individual farm yield observations provided by the Ontario Crop Insurance Commission. The selection was restricted to those farms with continuous yield observations over the maximum time horizon permitted by the data. Yield histories therefore ranged from a high of 19 years to a low of 13 years. Since these data were already adjusted for trend by the commission no further adjustments were made. A Wald-Wolfowitz test applied to a subsample of the data indicated that the null hypotheses of randomness could not be rejected, thus implying that attempts to adjust for trend beyond what was employed by the insurance commission was not required. In total a rich data set comprising 609 farms covering 5 crops, (spring grain, wheat, corn, soybeans, and white beans) and distributed across 10 Ontario counties (spanning the province) were used. Elected prices were established by the average 1991 crop price reported by each county.

Procedures Summary

In summary, a K-S test will be used to test whether or not each of the 609 crop yield series is described by either a normal, gamma or beta distribution. While it may be possible to accept the null hypotheses for all 3 distributions on a single crop for a given farm, the distribution with the lowest D_n value will be taken as the correct distribution. Accordingly, the correct distribution can very across crops and regions. The kernel density function will be tested against these `correct' distributions. Finally, insurance premiums at 80% coverage (actual coverage levels in Ontario range from 75-85%) will be estimated for each of the 4 distributions and compared.

RESULTS

This section presents the results. First the distribution form of the crop yields will be compared; second, the best candidate distribution from the parametric form will be compared to the kernel functions; and third the effects of the distributional assumptions on insurance rate making will be assessed.

Crop Yield Distributions

_______Crop yield distributions for spring grain, wheat, corn, soybeans and white beans were first assessed using the 3 parametric forms. The results are reported in Table I. For only 1 of the 609 distributions were all 3 candidate functions rejected at the 10% level by the K-S test. Of the remaining farms the gamma distribution was rejected for only 7 farms, while normal and beta distributions were rejected for none. Of those farms with more than 1 candidate functions as possibilities the one with the smallest distance measure D_n was considered `best'.

The results are displayed in Table I. Of 609 farm/crops, 28.08% were characterized by a normal distribution, 50.08% by a beta distribution and 21.84% by gamma. The results appear to support Days conclusion that the beta distribution should be used to describe crop yields. But the conclusion is not general; Beta distribution describes only 44.6% of wheat, 44% of grain, 55% of corn, 49% of soybeans and 39% white beans; Type II error would be approximately 50% for all crops on average, and nearly 60% for some crops such as white beans. None-the-less, the normal distribution is clearly not representative, nor is gamma.

Premium Estimation

Crop insurance premiums at 80% coverage levels were numerically estimated for all 3 parametric distributions and the kernel estimator. A coverage level of 80% was chosen because it lies midrange between typical Ontario levels of 75%, 80%, and 85%. The premiums and support prices are listed in Table II.

No discernable patterns appear in the premiums in Table II. What is striking is the range of premiums. For example spring grain premiums in Russel county range from $13.60/acre using a beta distribution to low of $8.33 for the normal distribution; a 63% difference. On average the % difference between the high and low is 40% for wheat, 40% for grain, 29% for corn, 33% for soybeans, and 26% for white beans. Clearly such variance due solely to assumption of probability distribution could lead to problems of adverse selection. The problem of course is determining means by which such differences may be reduced, and hence the focus on the kernel function.

Kernel Estimates

In Table II the listed numbers and premiums for normal gamma and beta distribution reflect the `best' choice distribution based first on a 10% K-S test and second on the least distance criteria (i.e. minimum D_n). In this section these premiums are to be compared to the premiums generated from the kernel estimate.

For all 609 farms, premiums were computed for each parametric distribution as well as the kernel and using D_n the `best' was recorded. Each was subtracted from the premium obtained from the best of the four distributions. Operating under a null hypothesis that the minimum error is obtained through use of the kernel estimates, the overall performance of the four estimators

$$SRSSE = \sqrt{\sum_{i=1}^{n} (deviation)^2}, \quad (11)$$

was evaluated by finding the minimum square root of the sum of the squared error (SRSSE); where `deviation' is the difference between the `best' crop premium and the other distributions, and n is the number of farms in each crop category. Four SRSSE's, were therefore computed for each crop/county category.

Selecting which of the four distributions would minimize error in premiums can be made by comparing mean deviations or the SRSSE. In terms of mean deviations, results show that beta and kernel distributions were consistently positively deviated while normal and gamma premiums were consistently negatively deviated (Table III). The implication is that if either the kernel or beta estimators are universally assumed, premiums would tend to be biased upward while if gamma or normal estimators are universally assumed, estimated premiums would tend to be biased downward. Low risk farmers would be discouraged from purchasing crop insurance under the first case while high risk farmers would be encouraged in the second case.

Interpretation of the results in Table III is given by the following examples: Suppose a normal distribution was assumed for all wheat crops. From previous discussions and Table I, of 47 wheat crops only 13 (27.6%) had distributions closely approximated by a normal distribution, implying 72.3% of wheat crops were best described by an alternative. Relative to these alternatives an assumption of normal distribution would result in average premiums being $.48/acre below the `best' actuarial values (Table III) with a relatively high standard deviation of $.77/acre. Using the average deviation as a selection criteria suggests that presumption of a kernel distribution for all crops and all farms would result in a minimum error (.45), with beta placing second (.99), normal third (-1.11) and gamma fourth (-1.18). Use of this measure assumes that the overall objective is to minimize the error across all farms, without much concern for error at the individual farm level.

An alternative to using mean deviations to select distribution type is the SRSSE. Since the SRSSE is the geometric mean of the premium deviations for all of the farm crops, it represents the total deviation regardless of whether the deviations were positive or negative. Use of the SRSSE therefore minimizes the deviation at the individual farm level. On average SRSSE is minimized for the beta distribution (56.8%) with the kernel estimates second (46.7). Gamma and normal distributions had the highest average SRSSE with 48.2 and 56.8, respectively.

Neither the mean deviation or SRSSE provide consistent assessments across crops. Except for soybeans the deviation is lowest for the kernel function, although for soybeans the difference between the minimum (for normal distribution) and the kernel function, are not statistically different from one another. In contrast, SRSSE indicates minimum error with a beta distribution for wheat and corn, a gamma distribution for grains and white beans and a normal distribution for soybeans.

The ambiguity in the rankings of the distributions in Table I and the clear differences in their premium estimates in Table II provide an added complication to the insurance problem. Clearly it would be impractical and costly for insurance agencies to first approximate each distribution and then test to find which provided the minimum error relative to the empirical data in order to compute actuarial premiums. It would, therefore, be advantageous for insurers to adopt a single distribution form which minimizes the error in premiums.

The results thus far show that either the beta distribution or the kernel estimator could be used for setting premiums. As shown in Table I the beta distribution was selected as best of the 3 parametric forms for 50.08% of all cases while the normal and gamma distributions were best for only 28.08%, and 21.84% of the cases, respectively. Based on the premium deviation criterion the kernel estimator could not be rejected as the `best' criterion in 305 of the 609 cases (50%).

In comparing the non-parametric distribution to the three parametric ones the kernel was rejected in 134 of 171 cases where the normal distribution was considered best; in 79 of 305 cases where beta was considered best; and 91 of 133 cases in which gamma was considered best. The kernel was <u>not</u> rejected in 268 of 438 cases in which the normal distribution was rejected; 79 of 304 cases in which beta was rejected; and 263 of 476 cases in which gamma was rejected.

What is interesting about these results is the flexibility which the kernel estimates possess: In the absence of unambiguously defined parametric distributions the kernel estimates are able to mimic all 3 parametric distribution types and in the process generate premiums with minimal error against the `best' of the 3 distributions in at least 50% of the cases. There is an advantage in this; by minimizing the error in distribution choice, with the added flexibility of a non-parametric form, the kernel estimate can more accurately reflect yield risk, thereby increasing the likelihood that premium estimates are correct and thus minimizing the problem of adverse selection.

CONCLUSIONS

The problem of adverse selection is one in which insureds have more information about the probabilities of risky outcomes than the insurers. In the context of crop insurance this study has illustrated the ambiguous nature of Ontario crop yield distributions. This ambiguity spanned both crop type and region. In the absence of perfect information it would appear, therefore, that adverse selection cannot be eliminated without substantial information costs assessing each crop's distribution.

The kernel function in this study has shown substantial flexibility in minimizing errors in premiums. That is, flexibility in the nonparametric form can approximate, in 50% of the cases, the appropriate level of down-side risk on which premiums are based. However, this implies an overall type II error of 50%. Nonetheless, in terms of mean error it was found that errors (measured relative to the empirical distribution) in premiums were generally low with the kernel function. Unfortunately, such gains in efficiency are obtained only with the added complexity of the kernel estimator.

Finally, although there is a substantial variance in the premiums across farms, the average range of premiums does not appear to be high on a per acre basis. However, crop insurance is considered a variable cost of production, and any errors, even moderately significant, can have an effect on farmers' budgeted returns. For example, an error of $5.27/acre for grain in Russel county could account for a decrease in per acre returns of greater than 10%. The extent to which farmers acknowledge possible errors depends upon their own perception of underlying down-side risk relative to that reflected in the offered premiums. The behaviourial response to perceived differences is the mitigating factor for adverse selection and moral hazard. The extent and costs of adverse selection and moral hazard is left for future study.

References

Antle, J.M. 1983. "Testing the Stochastic Structure of Production: A Flexible Moment-Based Approach." Journal of Business and Economic Statistics, 1:192-201.

Birnbaum, Z.W. 1953. "Distribution-Free Tests of Fit for Continuous Distribution Functions." Ann. Math Statist., 24:1-8.

Bradley, J.V. 1968. *Distribution-Free Statistical Tests*. Prentice-Hall, Inc.

Day, R.H. 1965. "Distribution of Field Crop Yields." American Journal of Agricultural Economics. 47:713-41.

Gallagher, P. 1986. "U.S. Corn Yield Capacity and Probability: Estimation and Forecasting with Non-symmetric Disturbances." North Central J. of Agricultural Economics. 8:109-122.

Johnson, N.L. and S. Kotz. 1970a. *Continuous Univariate Distributions-1*. Houghton Mifflin Company.

Johnson, N.L. and S. Kotz. 1970b. *Continuous Univariate Distributions-2*. Houghton Mifflin Company.

Kenkel, P.L., J.C. Buzby, and J.R. Skees. 1991. "A Comparison of Candidate Probability Distributions for Historical Yield Distributions." University of Kentucky, A.E.#9107.

Morgan, M.G. and M. Henrion. 1990. *Uncertainty: a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press.

Nelson, C.H. and P.V. Preckel. 1989. "The Conditional Beta Distributions as a Stochastic Production Function." American Journal of Agricultural Economics, 71:370-78.

Pope, R.D. and R.F. Ziemer. 1984. "Stochastic Efficiency, Normality, and Sampling Errors in Agricultural Risk Analysis." American Journal of Agricultural Economics, 66:31-40.

Rao, B.L.S.P. 1983. *Nonparametric Functional Estimation*. Academic Press.

Shapiro, R.D. and R.F. Ziemer. 1968. "A Comparative Study of Various Tests for Normality." J. Amer. Statist. Assoc., 63:1343-72.

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

Taylor, C.R. 1984. "A Flexible Method for Empirical Estimating Probability Functions." Western Journal of Agricultural Economics, 9:66-76.

Table I: Candidate Yield Distributions of Ontario Field Crops

County	Crops	Number of Observations	Number of Farms	Expected Yield	F	Number of Farms[a] Normal	Beta	Gamma
Prescott	Grain	18	23	53.85	23.39	15	7	1
	Corn	18	43	76.59	20.68	15	23	5
Russel	Grain	17	21	43.28	18.04	3	11	7
	Corn	18	33	79.35	21.08	7	21	5
Dundas	Corn	14	34	82.76	16.19	7	19	8
Ottawa	Grain	18	24	52.60	16.88	16	7	1
	Corn	18	44	84.78	12.81	7	29	8
Wellington	Grain	14	30	63.67	12.40	6	18	6
	Corn	15	25	82.01	12.48	8	5	12
Norfolk	Wheat	14	14	38.76	9.12	4	7	3
Perth	Wheat	15	16	64.25	7.55	6	7	3
	Grain	18	25	65.40	11.69	10	11	4
	Corn	19	32	99.39	16.14	8	17	7
	Soybeans	8	31	36.23	3.44	4	12	15
	White beans	19	28	23.88	5.85	14	11	3
Kent	Corn	18	19	115.71	13.40	5	9	5
	Soybeans	19	29	38.28	7.55	6	15	8
Essex	Corn	17	8	102.01	21.54	1	6	1
	Soybeans	19	36	35.42	7.48	7	19	10
Lambton	Wheat	18	5	65.96	7.98	0	4	1
	Corn	17	16	106.50	17.47	4	11	1
	Soybeans	19	21	33.71	5.27	6	12	3
Middlesex	Wheat	16	12	60.04	7.84	3	3	6
	Corn	18	21	103.20	12.95	5	12	4
	Soybeans	13	19	34.80	2.37	4	9	6
Summary	Wheat	-	47	-	-	13	21	13
	Grain	-	123	-	-	50	54	19
	Corn	-	275	-	-	67	152	56
	Soybeans	-	136	-	-	27	67	42
	White beans	-	28	-	-	14	11	3
Total			609	-	-	171	305	133
Percent			100	-	-	28.08	50.08	21.84

[a] The numbers in these 3 columns represent the frequency by which the `best' distribution was selected using the K-S test and the lowest $D_n$ value. For example, of the 23 farms in Prescott county growing grain, 15 farms had yields best characterized by a normal distribution, 2 a beta distribution and 2 a gamma distribution. For all farms and all crops 28.08% were `best' characterized by a normal, 50.08% beta, and 21.84% gamma distributions.

Table II: Mean Crop Insurance Premiums of Different Estimators

County	Crop	Elected Price ($/bu.)	Normal ($/acre)	Beta ($/acre)	Gamma ($/acre)	Kernel ($/acre)	Range[a] ($/acre)
Prescott	Grain	2.02	9.64	14.4	11.9	10.7	4.76
	Corn	2.89	15.2	19.4	15.7	18.7	4.20
Russel	Grain	2.02	8.33	13.6	11.4	10.5	5.27
	Corn	2.89	14.4	17.9	14.3	18.1	3.80
Dundas	Corn	2.89	8.64	10.8	8.40	10.3	2.40
Ottawa	Grain	2.02	8.58	11.2	9.19	9.33	.75
	Corn	2.89	8.95	11.2	8.64	11.4	2.76
Wellington	Grain	1.86	4.36	5.43	4.14	5.58	1.44
	Corn	2.67	6.28	7.60	5.77	7.01	1.83
Norfolk	Wheat	3.13	5.67	6.99	5.54	6.77	1.45
Perth	Wheat	3.81	2.38	3.39	2.10	3.34	1.29
	Grain	1.86	4.49	5.46	4.24	5.62	1.38
	Corn	2.67	6.86	8.31	6.11	8.78	2.67
	Soybeans	6.21	2.35	3.64	2.31	2.88	1.33
	White beans	9.12	11.9	14.4	11.4	12.3	3.00
Kent	Corn	2.57	4.78	6.06	4.23	6.02	1.79
	Soybeans	6.21	7.04	8.40	6.33	8.56	2.23
Essex	Corn	2.57	8.96	10.8	8.10	11.2	3.10
	Soybeans	6.21	8.20	9.75	7.55	9.73	2.20
Lambton	Wheat	3.13	1.43	2.14	1.21	2.38	1.17
	Corn	2.57	6.53	8.22	5.91	8.92	3.01
	Soybeans	6.21	5.02	6.12	4.46	6.41	1.95
Middlesex	Wheat	3.13	2.61	3.40	2.32	3.17	1.08
	Corn	2.57	5.23	6.56	4.65	6.94	2.29
	Soybeans	6.21	3.52	3.50	2.29	3.15	1.23
Average[a]	Wheat	-	3.32	4.33	3.09	4.22	1.24
	Grain	-	6.88	9.64	7.83	8.15	2.76
	Corn	-	9.33	11.7	9.05	11.7	2.65
	Soybeans	-	5.34	6.63	4.88	6.49	1.61
	White beans	-	11.90	14.40	11.40	12.30	3.00

[a] The range indicates the difference between the highest premium and the lowest premium. For example, grains premiums in Prescott county range from a low of $9.64/acre for the normal distribution, to a high of $14.40/acre for the beta distribution. The difference is $4.76/acre.

[b] The averages reported here are measured across all farms and counties.

Table III: Evaluating Different Estimators With Premium Deviation ($/acre).[a]

Crop	Measure	Normal	Beta	Gamma	Kernel
Wheat	Mean	-.48	.53	-.71	.41*
	Standard Deviation	.73	.55	.68	.69
	SRSSE	6.02	5.22*	6.75	5.50
Spring Grain	Mean	-1.38	1.39	-.42	-.10*
	Standard Deviation	2.43	1.93	1.73	2.08
	SRSSE	30.9	26.3	19.8*	23.0
Corn	Mean	-1.38	.96	-1.67	.93*
	Standard Deviation	2.30	1.40	1.61	1.93
	SRSSE	44.5	28.1*	38.5	35.6
Soybeans	Mean	-.57*	.73	-1.03	.58
	Standard Deviation	.92	.90	1.04	1.11
	SRSSE	12.7*	13.5	17.10	14.4
White Beans	Mean	-.88	1.62	-1.36	-.45*
	Standard Deviation	1.49	1.64	1.53	2.17
	SRSSE	9.16*	12.22	10.8	11.7
Average	Mean	-1.11	.99	-1.18	.54*
	Standard Deviation	2.02	1.43	1.55	1.81
	SRSSE	56.8	39.1*	48.2	46.7

[a] The means in this table are computed by subtracting the premiums under each category from the parametric distributions found to be `best' using the K-S, $D_n$ criteria. For example, if wheat yield distributions were assumed to be normal, the mean deviation across all wheat crop distributions from the `best' distribution is -.48. The standard deviation is the standard deviation of the differences.

The SRSSE = $\sqrt{\sum_{i=1}^{n} (\text{deviation})^2}$ as defined in equation (11). Values which are starred (*) indicate a minimum for the criterion indicated.

ENDNOTE