A Flexible Method for Empirically Estimating Probability Functions

C. Robert Taylor

This paper presents a hyperbolic trigonometric (HT) transformation procedure for empirically estimating a cumulative probability distribution function (cdf), from which the probability density function (pdf) can be obtained by differentiation. Maximum likelihood (ML) is the appropriate estimation technique, but a particularly appealing feature of the HT transformation as opposed to other zero-one transformations is that the transformed cdf can be fitted with ordinary least squares (OLS) regression. Although OLS estimates are biased and inconsistent, they are usually very close to ML estimates; thus use of OLS estimates as starting values greatly facilitates use of numerical search procedures to obtain ML estimates. ML estimates have desirable asymptotic properties. The procedure is no more difficult to use than unconstrained nonlinear regression.

Advantages of the procedure as compared to alternative procedures for fitting probability functions are discussed in the manuscript. Use of the conditional method is illustrated by application to two sets of yield response data.

Economists are increasingly aware of the need to formally incorporate risk and uncertainty into analyses of agricultural problems. Failure to account for uncertainty can result in imprecise if not blatantly incorrect empirical estimates (Anderson, 1982; Just and Pope).

Clearly, formal treatment of uncertainty is called for when risk averse behavior is anticipated. Furthermore, it is often imperative to explicitly consider uncertainty even under conditions of risk neutrality as the certainty equivalent requirements (Simon; Theil, 1957), which allow the replacement of random variables with their respective expected values (and thus use of a deterministic framework), are not satisfied in many risk neutral situations.

At one time it was considered acceptable to characterize random variables by their first two moments (mean and variance), but it is now recognized that such second-order approximations may not be adequate for many empirical studies. Although the first two moments are sufficient to characterize a normally distributed random variable, there appear to be few cases in agriculture where one can appeal to the Central Limit Theorem in order to theoretically justify a normal distribution.[1] The case for non-normality is supported by several empirical studies where third and even fourth moments of output have been found to be functions of input levels (Antle; Antle and Goodger; Anderson, 1974; Day).

In general, we cannot provide compelling theoretical arguments that, for example, the probability distributions of weather, crop yield, gross returns, or equipment failure follows one of the common theoretical distributions (e.g., lognor-

C. Robert Taylor is Professor of Agricultural Economics and Economics at Montana State University.

Development and presentation of the procedure proposed in this paper was enhanced considerably by comments from Oscar Burt. Bruce Beattie and Myles Watts provided valuable editorial suggestions.

[1] For a discussion of the applicability of the Central Limit Theorem to the distribution of total farm gross margins, see the exchange between Chen and Hazell.

Western Journal of Agricultural Economics, 9(1): 66–76

mal, Gamma, Beta, Poisson or normal). Also, the class of theoretical distributions which are analytically, statistically and computationally convenient is rather limited, especially for probability density functions (pdfs) with multiple modes or pdfs which are strongly skewed. Thus an important component of many risk analyses is empirical estimation of the form as well as parameters characterizing the pdf or cumulative distribution function (cdf). The problem of estimating the pdf or cdf is especially critical for safety-first and stochastic dominance considerations. The tail of the pdf is of crucial importance when safety-first decision rules are adopted, while the entire cdf is critical in stochastic dominance analyses.

This paper begins with a brief discussion of conventional methods for empirically fitting pdfs or cdfs—most if not all of which leave much to be desired for accurate, yet practical empirical application. Next, a new transformation procedure for estimating a cdf is presented; the pdf can be obtained by differentiation. The proposed approximation procedure uses a hyperbolic trigonometric (HT) transformation to constrain a polynomial function to the zero-one range. A polynomial function is suggested for many pdfs because of its foundation in approximation theory, and because it is linear in parameters. Parameters characterizing the HT transform should be estimated by maximum likelihood (ML).

An appealing feature of the proposed transform, as opposed to other zero-one transformations or constraints, is that ordinary least squares (OLS) regression can be used to obtain starting values for the ML search procedure thereby greatly facilitating ML estimation. Use of the conditional HT procedure is illustrated by application to two sets of data (Grissom and Spurgeon) on yield response to nitrogen that Day used in his application of the Pearson system of distributions. The focus of the paper is on conditional cdfs (and thus conditional pdfs) because of their importance in decision models; however, the procedure works equally well for unconditional cdfs.

Review of Existing Methods

This review of existing methods for empirically fitting cdfs or pdfs is divided into three parts—discrete approaches, simple continuous approaches, sophisticated continuous approaches, and recent approaches involving estimation of conditional density functions.

Discrete Representations

On the surface it might seem reasonable to use a discrete empirical representation of a pdf—a histogram. However, use of a discrete pdf in most economic models is not entirely satisfactory for three reasons. First, the statistical literature is quite vague as to how many intervals to use in constructing a histogram. With small samples, specifying a large number of intervals is akin to separate representation of each data point, while use of a small number of intervals generally gives a featureless picture of the pdf (Tarter and Kronmal). Second, it may be desirable to use intervals of the random variable which are smaller than those for which there are frequency data. In such cases, some kind of interpolation is required. Third, in situations where there is a logical basis for some order of continuity of the distribution, use of a histogram does not exploit the statistical leverage that can be achieved by introducing continuity. Just as it is often desirable to fit a production function to smooth out and nonlinearly interpolate data points, it is often desirable to fit a continuous pdf to smooth out and interpolate histogram data. We now turn to a survey of methods that have been proposed to fit continuous pdfs or cdfs.

Simple Continuous Representations

Relatively simple methods for fitting commonly used continuous pdfs or cdfs generally fall into three classes: (a) free-hand fitting of a cdf; (b) fitting a simple mathematical function; or (c) estimation of the moments of a pdf in the Pearson system of distributions. None of these approaches are satisfactory for most empirical studies. The simplest technique—free-hand fitting of a cdf—is not appealing for most applications[2] because of four drawbacks: (1) it is totally subjective and statistical properties such as bias and consistency cannot be determined; (2) it is extremely difficult to employ for fitting conditional cdfs while exploiting continuity of the conditional relationship; (3) it is sometimes difficult to incorporate into computerized models; and (4) the pdf cannot be obtained since the equation of the cdf is unknown.

Some researchers have resorted to fitting relatively simple functions such as an exponential (Dixon and Sonka), a polynomial (Held and Helmers), or a triangular pdf (Richardson and Condra). As noted by Dixon and Sonka, the simple exponential functions are quite restrictive. The disadvantages of using polynomials to fit cdfs are that they are not restricted to the zero-one range and they are not necessarily monotonic. Discontinuities in the triangular pdf are implausible for many stochastic processes and assuming this restrictive form may lead to serious approximation biases.

As noted previously, selection of a distribution from the Pearson system is not always satisfactory. The class is rather restrictive for stochastic processes that have pdfs with multiple modes or strongly skewed pdfs. In addition, the goodness-of-fit tests used to determine whether a data set was generated by a particular distribution lack power; that is, the type II errors associated with the tests may be quite large with small samples.

Sophisticated Continuous Representations

Several researchers have resorted to rather sophisticated methods for fitting cdfs or pdfs. Prominent classes of flexible methods which have been reported in statistical and mathematical literature are: (a) spline functions obtained by minimizing error sum of squares plus a prespecified roughness penalty (Craven and Wahba); and (b) Fourier series methods which minimize error sum of squares (Kronmal and Tarter); or (c) Fourier methods which maximize a likelihood function less a roughness penalty (Good and Gaskins). All of these methods are quite difficult to use.

Spline functions are difficult to estimate, especially when the location of the knots are also estimated. Zero-one restrictions must be placed on the cdf, and it is usually desirable to require continuity at least in the first derivatives of the cdf.

The author's experience with the Fourier series methods suggests that they are so flexible that the cdf essentially goes through all data points, resulting in an implausibly wavy pdf. To get a plausible pdf, the analyst must parametrically tighten the roughness penalty until reasonable results are obtained. (This problem is graphically illustrated in Tarter and Kronmal.) After repeatedly going through an extremely difficult and expensive computational procedure, the analyst is left with an estimated cdf that seems highly subjective (Parzen), and is difficult to incorporate into models because of its complicated trigonometric form, typically characterized by several hundred parameters. A more detailed review of these sophisticated pro-

[2] For purposes of this paper, fitting a pdf should be thought of in terms of fitting a cdf because we can directly relate sample observations to the height of a cdf but not to the height of a pdf.

cedures is found in Tarter and Kronmal; Fryer; and Wegman.

Recent Approaches Involving Conditional Density Functions

With many agricultural relationships an independent (decision) variable may influence the parameters of the pdf of a dependent (state) variable, but not influence the general functional form of the pdf. For example, the fertilization rate may influence the moments of a crop yield pdf, yet the same functional form may be appropriate for all fertilizer levels; that is, the parameters but not the form of the pdf may be conditional on the fertilization rate. Even if the same pdf functional form is not appropriate for all fertilization rates, we may expect a systematic or smooth relationship between pdfs for different rates. Separately fitting a pdf to each fertilization rate, as Day did, will not make efficient use of the data in either of the above situations. None of the flexible methods mentioned earlier were specifically designed to handle this problem, but they could be extended in a straightforward way to handle a conditional pdf. However, such modifications would make application of highly complicated procedures even more complicated and expensive.

Just and Pope recently suggested use of Harvey's heteroscedasticity correction procedure to account for the effect of an independent variable (X) on output (Y) variance. They suggested a three-stage nonlinear regression estimation procedure applied to a function of the form

$$Y = g(X) + [h(X)]^{1/2}\epsilon \qquad (1)$$

where g(X) and h(X) are functions to be estimated and ϵ is an error term with zero mean and unitary variance. The function g(X) accounts for the deterministic (i.e., expected) component of Y, while the function h(X) allows the variance of Y to change in a smooth manner with X. Although this procedure is easy to use relative to the highly sophisticated procedures and although it is more flexible than the simple procedures, it is unduly restrictive because an analyst seldom knows a priori that only the first two moments depend on an independent variable.[3] Moreover, in some decision models such as safety-first, accurate estimates of the tail of a distribution is more important than estimates of higher moments.

Just and Pope also suggested ML estimation of (1). Although the MLE procedure could be specified to allow higher moments to vary with X, the procedure requires the analyst to specify a priori the form of the distribution of ϵ. Hence, the procedure must be repeatedly applied for different assumed forms of the pdf of ϵ as well as different forms of g(X) and h(X).

Antle has proposed a moment-based method to represent stochastic conditional relationships. This procedure is an n^{th} degree approximation to a stochastic process, achieved by estimating an n^{th} degree polynomial whose coefficients are functions of the first n moments of the distribution. Coefficients of the polynomial model are estimated using generalized least squares (GLS), although it may be necessary to use nonlinear programming to incorporate non-negativity constraints on estimation of even moments. The estimators have desirable asymptotic statistical properties. Moreover, if the range of the random variable is finite, the set of moments uniquely define the pdf. Although this moment-based method is quite flexible and has a sound statistical foundation, it is not practical for use in empirical studies that directly require the

[3] Only the first two moments are allowed to vary because $\ln|\hat{\epsilon}_t|$ or $\hat{\epsilon}_t^2$ (where $\hat{\epsilon}_t$ is the set of residuals from a nonlinear regression of y on g(X)) are used in estimating h(X).

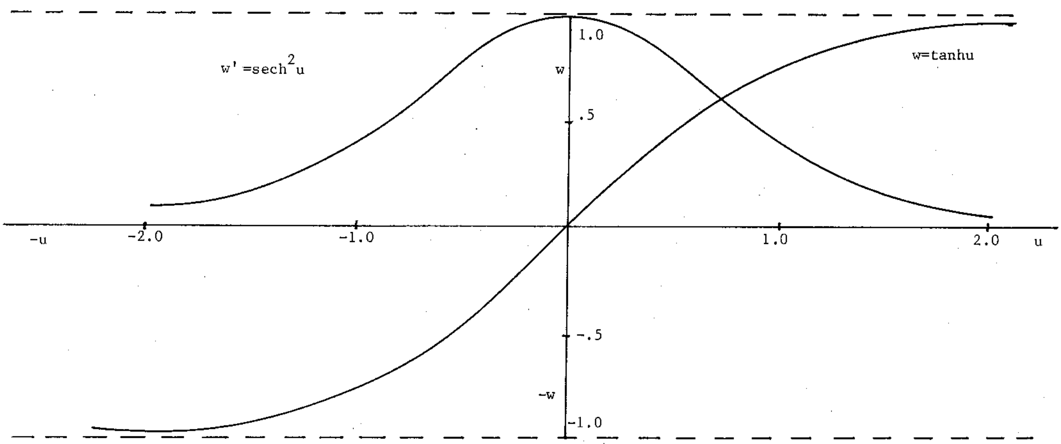

Figure 1. The Hyperbolic Tangent and Its Derivative, the Square of the Hyperbolic Secant.

equation of a pdf or cdf. That is, even though the moments uniquely define (in a theoretical sense) the underlying pdf, the analytical form of this pdf may be difficult if not impossible to obtain except in special cases. Thus, Antle's method may not be useful in safety-first and stochastic dominance analyses, although it may be practical when the pdf or cdf is not needed per se.

The preceding discussion suggests that existing methods of fitting cdfs are either operationally cumbersome or are unduly restrictive for some applications. Let us now consider the HT procedure which, while not a panacea, is appealing for many empirical studies. In ease-of-use and flexibility, the procedure lies between the simple functions and the complicated spline and Fourier techniques. The procedure is no more difficult to use than Just and Pope's procedure, and is as flexible as Antle's moment-based approach. Its advantage over the Antle approach is that it yields an explicit expression of the cdf and thus the pdf. Moreover the HT procedure approaches the problem of estimating a pdf head-on as a ML problem rather than in an ad hoc way as with the approaches advanced by Just and Pope, and Antle.

Transformation for Estimating a CDF

Consider a hyperbolic tangent

$$\tanh u = \frac{e^{u} - e^{-u}}{e^{u} + e^{-u}} \tag{2}$$

where $-\infty \leq u \leq \infty$ and $-1 \leq \tanh u \leq 1$. Figure 1 illustrates that the hyperbolic tangent has the right curvature properties for a unimodal cdf and that its derivative, the square of the hyperbolic secant, has the right properties for a pdf.

Now consider the transformation

$$F(Y|X) = .5 + .5 \tanh [P(Y,X)] \tag{3}$$

where $F(Y|X)$ is the cdf of Y conditional on X, and $P(Y,X)$ is a polynomial function of Y and X or a polynomial function of a transformation of Y and X such as ln Y and ln X.[4] For any value of $P(Y,X)$, trans-

[4] The function P(Y,X) need not be restricted to a polynomial specification; the polynomial specification is suggested because of its foundation in approximation theory and because it is linear in parameters. A polynomial in combinations of X and ln X, for example, can also be used. In a practical sense, flexibility of the procedure is limited only by the creativity of the analyst in specifying an appro-

formation (3) constrains $F(Y|X)$ to the interval zero-one. Since tanh u has one inflection point, this transformation allows for traditional bell-shaped pdfs. The function $P(Y,X)$ gives flexibility to the transformation, permits additional modes to the conditional pdf, and will allow for the pdf to be skewed in either direction, or to be symmetrical. Including X in $P(Y,X)$ allows for a systematic relationship between the pdfs associated with different values of X. Specification of $P(Y,X)$ to include interaction terms between Y and X will allow for substantial changes in the basic shape of the pdf for different X; without interaction terms, the moments but not the general shape of the pdf will vary with X.

Consider ML estimation of (3). Differentiation of (3) with respect to Y gives the conditional pdf.

$$f(Y|X) = .5P'(Y,X)\text{sech}^2[P(Y,X)] \tag{4}$$

where $f(Y|X)$ is the conditional pdf and $P'(Y,X) = \partial P(Y,X)/\partial Y$. On the basis of (4) we can form the likelihood function

$$L(\beta) = \prod_{i=1}^{n} .5P'(y_i,x_i)\text{sech}^2[P(y_i,x_i)] \tag{5}$$

where β is the vector of m parameter values characterizing $P(Y,X)$, y_i and x_i are paired observations, and n is the total number of observations.

Since maximizing the logarithm of the likelihood function is equivalent to maximizing the function itself, it is useful to replace (5) by

$$\ln L(\beta) = n \ln(.5) + \sum_{i=1}^{n} \ln[P'(y_i,x_i)] + 2\sum_{i=1}^{n} \ln[\text{sech}[P(y_i,x_i)] \tag{6}$$

Taking the partial derivatives of (6) with respect to the parameter vector, β, and setting to zero gives a set of m equations that can be simultaneously solved for m parameters.

$$\sum_{i=1}^{n}\left(\frac{\partial P'(y_i,x_i)}{\partial\beta_k}\right)\left(\frac{1}{P'(y_i,x_i)}\right) - 2\sum_{i=1}^{n}\left(\frac{\partial P(y_i,x_i)}{\partial\beta_k}\right)\tanh[P(y_i,x_i)] = 0 \tag{7}$$

for $k = 1, 2, \ldots, m$. The asymptotic covariance matrix of β is given by the negative of the expected value of the inverse of the matrix of second partial derivatives of (6) with respect to the parameter vector. Thus, traditional asymptotic t-tests can be used to determine the significance of individual polynomial terms in $P(Y,X)$. Also, a likelihood ratio test (Theil, 1971, pp. 396–97) can be used to test for significance of individual polynomial terms or groups of terms.

Analytical solution of the m equations in (7) for various $P(Y,X)$ is impossible except in trivial cases, so a numerical search procedure must be used to solve (7).[5] Re-

priate mathematical form for $P(Y,X)$. Appropriate specification of $P(Y,X)$ will allow for approximation of U, J and truncated pdfs in addition to continuous unimodal and multi-modal pdfs. Of course, in any application, statistical considerations should dictate which form of $P(Y,X)$ is selected to approximate the pdf.

[5] A wide variety of numerical search routines can be used to solve for ML estimates of β. In general, performance (computational efficiency, global convergence, etc.) of search routines is highly problem specific. It should be cautioned that the search routines that usually perform quite well for nonlinear regression and some types of simultaneous nonlinear equations do not appear to perform well on this particular problem. For example, the Marquardt algorithm (Kuester and Mize), which combines the steepest ascent method with the Gauss method, is often very slow to converge and often converges to local optima using OLS starting values. The Newton-Raphson technique is not appealing because it requires analytical specification of the Hessian associated with (6), which places a very heavy burden on the analyst and is also a possible source of error due to the complicated derivatives. Also, the Newton-Raphson technique sometimes breaks down because the Hessian is not positive definite in some iterations.

A routine that does work well in most if not all cases is the secant method proposed by Wolfe. A FØRTRAN program that tailors Wolfe's algorithm peated solution of (7) for various P(Y,X) can be quite onerous; however, a particularly attractive feature of the hyperbolic tangent transformation, (3), as opposed to other zero-one transformations, is that its inverse[6]

$$z = .5 \ln\left[\frac{F(Y|X)}{1 - F(Y|X)}\right] = P(Y,X) \qquad (8)$$

is a closed-form expression that is linear in parameters if P(Y,X) is linear in parameters. With linearity, OLS estimates of β can be obtained;[7] however, with time-series or experimental data, OLS estimates are biased because a nonstochastic variable—a transformation of F(Y|X) whose value is determined by the sample size—is treated as a dependent variable and the stochastic variable, which is treated as an independent variable, is correlated with the error term.

Although biased, OLS estimates of β can be used as initial guesses for the numerical search technique used to solve (7) for ML estimates. Since OLS estimates are often very close to ML estimates, one of the most difficult problems (having good starting values) in using ML procedures is overcome. And, as a practical matter, OLS can also be used to select a plausible set of polynomial terms to include in P(Y,X) for ML estimation.

An Application to Yield Response Data

To illustrate the flexibility and performance of the hyperbolic tangent transformation for conditional distributions, cdfs for cottom and corn yield conditional on nitrogen fertilization rate were estimated. Data were from a 37-year experiment by Grissom and Spurgeon and are the same data used by Day in his application of the Pearson system of distributions.

OLS and ML estimates of the cotton

to solving (7) for up to a third degree polynomial specification of P(Y,X) is available from the author. The program uses two routines in the IMSL Library, which is available on most mainframe computers. To use the program, the analyst need only provide observations on Y and X, specify which of the ten cubic terms to include in P(Y,X), and OLS (or other) starting values for β. An optional scaling feature is built into the program. Output from the program includes $\hat{\beta}$, asymptotic t-values for $\hat{\beta}$, the parameter correlation matrix, the asymptotic variance-covariance matrix of $\hat{\beta}$, the norm of the system of equations, (7), (i.e., the sum of squared deviations from the necessary conditions), and the value of the log-likelihood function, (6). The program costs only a dollar or two to run, in most cases.

[6] Given $w = \tanh(u)$, the inverse hyperbolic tangent is defined to be $u = \tanh^{-1}(w)$. To derive the logarithmic form of $\tanh^{-1}(w)$, consider $w = (e^u - e^{-u})/(e^u + e^{-u})$. Rearranging terms gives $e^{2u} = (1 + w)/(1 - w)$, and thus $u = .5 \ln((1 + w)/(1 - w)) = \tanh^{-1}(w)$. Using the logarithmic form of $\tanh^{-1}$, equation (8) can be derived from equation (2): rearranging terms in (2) we have $\tanh P(Y,X) = 2(F(Y|X) - .5)$; thus $P(Y,X) = \tanh^{-1}[2(F(Y|X) - .5)] = .5 \ln[(1+2(F(Y|X) - .5))/(1-2(F(Y|X) - .5))] = .5 \ln[F(Y|X)/(1 - F(Y|X))]$.

[7] In connection with OLS estimation of (8), two data cases must be considered. One case is where there are several observations on Y for each value of X. This case is typically encountered with experimental data such as on yield (Y) and fertilizer rate (X). The second case is one for which there is only one or at best a few values of Y associated with each value of X. In the latter case OLS estimation, *but not ML estimation*, will require grouping some of the values of X to obtain several values of Y for each X category, then using the mean of each X category as the observation on X. To simplify discussion of estimation of (8), we will consider only the former case and assume that we have the same number of observations on Y for each value of X. Given a sample of $n > 1$ observations on Y for the r^{th} value of X ($r = 1, 2, \ldots, R$), ranked from smallest to largest, $Y_{1r} < Y_{2r} \ldots < Y_{ir} \ldots < Y_{nr}$, we can assign to each a cumulative frequency $F(Y_{ir}|X_r) = i/n$. Then all $F(Y_{ir}|X_r)$ except $F(Y_{nr}|X_r)$ can be transformed by (8) to give a finite Z_{ir}; the problem with $F(Y_{nr}|X_r)$ is that Z_{nr} is infinite. As a practical cure, $F(Y_{nr}|X_r)$ can be slightly adjusted downward to give a finite value for Z_{nr}. If P(Y,X) is linear in parameters, OLS can be applied to the modified data set (Z_{ir}, Y_{ir}, X_r) to obtain an estimate of (3).

Carefully note that in the case where observations must be artificially grouped, the grouping is used only to obtain starting values (with OLS) for use in ML estimation of β. ML estimation does not require any such grouping.

TABLE 1. OLS and ML Estimates of the cdf for Cotton Yield (Y) Conditional on Nitrogen Application Rate (X).

Coefficient of:	OLS	ML
Intercept	−.4369 E + 1 (21.65)	−.5278 E + 1 (9.41)
Y	.6728 E − 2 (16.48)	.8209 E − 2 (7.99)
Y^2	−.2561 E − 5 (10.35)	−.3177 E − 5 (5.53)
Y^3	.4139 E − 9 (8.99)	.4649 E − 9 (4.54)
X	−.4840 E − 1 (22.54)	−.5326 E − 1 (5.81)
X^3	.6674 E − 5 (6.99)	.8356 E − 5 (2.04)
$\bar{R}^2$	.9591	

Notes: The functional form for the cdf is

$$F(Y|X) = .5 + .5 \tanh[\beta_1 + \beta_2 Y + \beta_3 Y^2 + \beta_4 Y^3 + \beta_5 X + \beta_6 X^3]$$

Asymptotic t-values are given in parentheses.

TABLE 2. OLS and ML Estimates of the cdf for Corn Yield (Y) Conditional on Nitrogen Application Rate (X).

Coefficient of:	OLS	ML
Intercept	−.2900 E + 1 (15.66)	−.2803 E + 1 (7.41)
Y	.2214 (11.96)	.2050 (6.64)
Y^2	−.5390 E − 2 (9.35)	−.4645 E − 2 (5.11)
Y^3	.5758 E − 4 (10.26)	.4594 E − 4 (4.97)
X	−.1027 (9.71)	−.1090 (4.73)
Y·X	.4177 E − 2 (7.43)	.4117 E − 2 (3.52)
X·Y^2	−.5973 E − 4 (8.45)	−.5219 E − 4 (3.54)
$\bar{R}^2$	.9339	

Notes: The functional form for the cdf is

$$F(Y|X) = .5 + .5 \tanh[\beta_1 + \beta_2 Y + \beta_3 Y^2 + \beta_4 Y^3 + \beta_5 X + \beta_6 YX + \beta_7 XY^2]$$

Asymptotic t-values are given in parentheses.

yield cdf are shown in Table 1. The likelihood ratio test mentioned previously was used to determine which polynomial terms to include in P(Y,X). Figure 2 shows ML estimates of the cotton yield pdf for three of the seven experimental nitrogen levels. Day also found distributions skewed toward higher yields; however, Day's results do not show a tail on the left side of the distribution, because the Pearson system is more restrictive than the method used in this paper.

OLS and ML results for the corn data are shown in Table 2, with the pdfs for three nitrogen rates shown in Figure 3. Interestingly, the data showed that a bimodal pdf exists at high, but not low, nitrogen rates. Day did not find this bi-modality. It is not the purpose of this paper to attempt to explain why bi-modality exists; however, it should be pointed out that the interaction terms (Table 2) which give rise to bi-modality are quite significant.

Note the substantially smaller t-ratios in the case of ML as contrasted to OLS estimates of individual terms in Tables 1 and 2. It is plausible that these differences are attributed to the OLS bias resulting from treating a stochastic variable as an independent variable.

Estimates of the Production Function

In many empirical decision analyses, it is necessary to have the expected value of Y given X, as well as the conditional pdf. Once the parameters of P(Y,X) have been obtained the conditional expectation of Y can be obtained by

$$E(Y|X) = \int_{-\infty}^{\infty} y[.5P'(y,X)\operatorname{sech}^2[P(y,X)]\, dy \qquad (9)$$

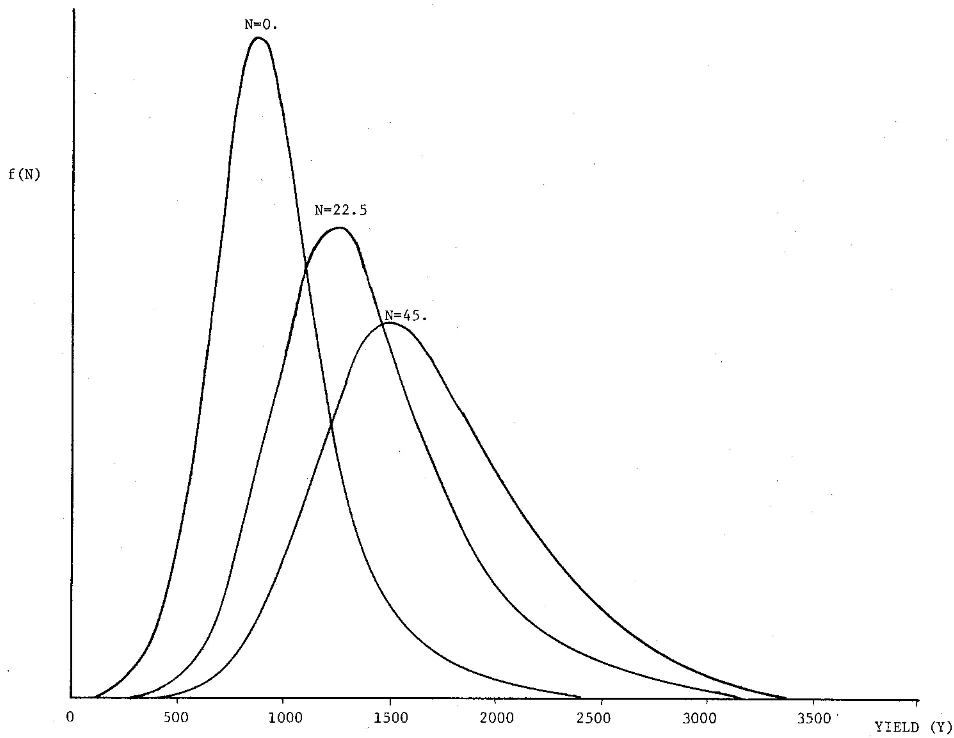

Figure 2. MLE Estimates of the Cotton Yield pdf for Different Nitrogen Fertilization Rates (N).

Unfortunately, the integration in (9) is not analytically possible except in rather trivial cases. However, numerical integration is feasible.[8]

As pointed out by Day, modal values of Y may also be important in decision analyses. Modal values can be obtained by solving the following equation for Y given X

$$0 = .5P''(Y,X)\text{sech}^2[P(Y,X)] - [P'(Y,X)]^2\text{sech}^2[P(Y,X)] \cdot \tanh[P(Y,X)] \quad (10)$$

Median values of Y for given X can be obtained by solving

$$0 = P(Y,X) \quad (11)$$

Analytical solution of (10) or (11) is not practical for most P(Y,X), but they can be solved numerically. For many specifications of P(Y,X), equations (10) and (11) are not well behaved enough for a derivative procedure for locating roots, such as Newton's method, to work well unless the initial guess is good. Consequently, the method of false position or a similar method is recommended for solution of (10) or (11).[9]

Concluding Remarks

Although the hyperbolic tangent procedure for empirically estimating a cdf is

[8] An inexpensive numerical integration routine that is widely available is DCADRE in the IMSL library. Alternatively, sample FORTRAN programs for numerical integration are given in Chapter 6 of Stark.

[9] A canned routine for the method of false position is ZFALSE in the IMSL library. Alternatively, a short FORTRAN program is given in Stark, pp. 92–95.

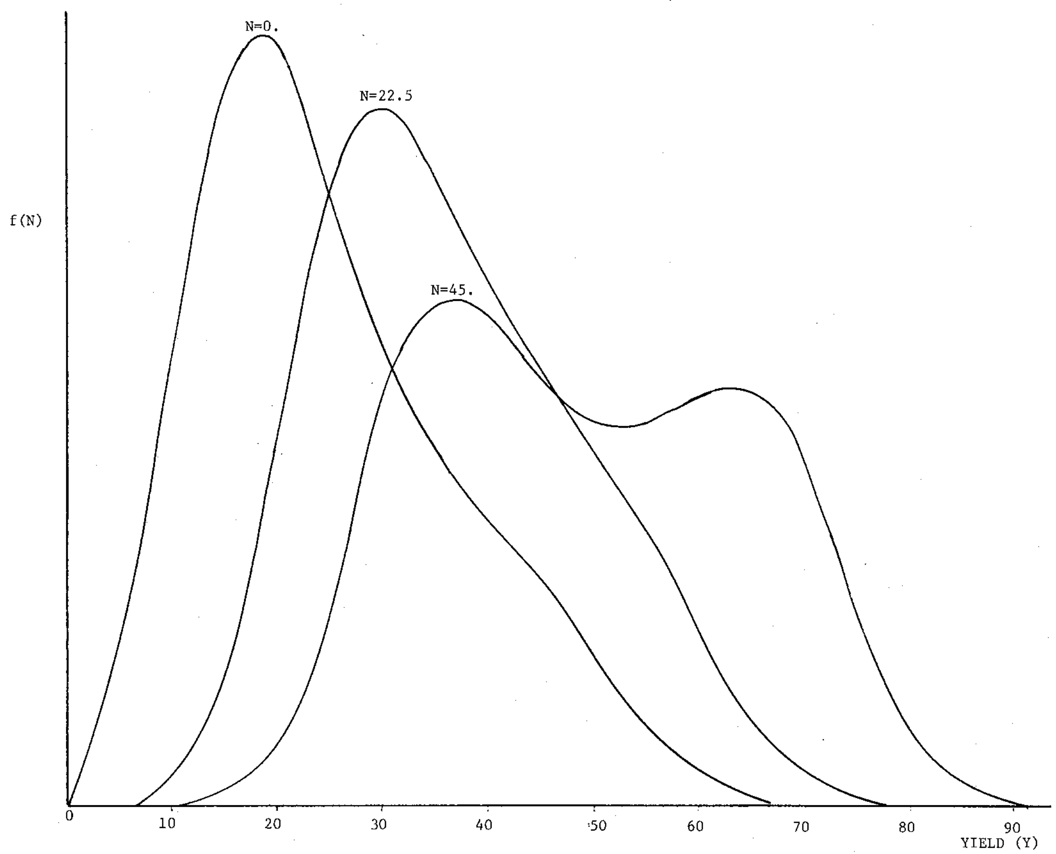

Figure 3. MLE Estimates of the Corn Yield pdf for Different Nitrogen Fertilization Rates (N).

an approximation, it does have several advantages over other available techniques. First, it is easy to use compared to procedures with equal flexibility. Secondly, the procedure has been shown to have the flexibility to closely approximate common theoretical probability distributions (Taylor), as well as fit data for many unconventional distributions. Thirdly, the procedure can be used to estimate conditional cdfs, which is not possible with most other procedures in their current state of development. Finally, with the ML approach, smoothing of data is controlled by traditional asymptotic statistical tests. Although the HT procedure may be viewed as somewhat subjective, its use is no more subjective than estimating any polynomial function where the degree of the polynomial is unknown.

References

Anderson, J. R. "Sparse Data, Estimational Reliability, and Risk-Efficient Decisions." *American Journal of Agricultural Economics,* 56(1974): 3 564–72.

Anderson, J. R. "Agricultural Economics Interdependence and Uncertainty." *Australian Journal of Agricultural Economics,* 26(1982): 2 89–97.

Antle, J. M. "Testing the Stochastic Structure of Production: A Flexible Moment-Based Approach." *Journal of Business and Economic Statistics,* 1(1983): 3 192–201.

Antle, J. M. and W. J. Goodger. "Measuring Stochastic Technology: The Case of Tulare Milk Production." Working Paper 83-4, University of California, Davis, Dept. of Agricultural Economics, 1982.

Chen, J. T. "A Linear Alternative to Quadratic and

Semivariance Programming for Farm Planning Under Uncertainty: Comment." *American Journal of Agricultural Economics,* 53(1971): 4 662–4.

Craven, P. and G. Wahba. "Smoothing Noisy Data with Spline Functions." *Numerische Mathematik,* 31(1979): 4 377–403.

Day, R. H. "Probability Distributions of Field Crop Yields." *Journal of Farm Economics,* 47(1965): 3 713–41.

Dixon, B. L. and S. T. Sonka. "A Note on the Use of Exponential Functions for Estimating Farm Size Distributions." *American Journal of Agricultural Economics,* 61(1979): 3 554–7.

Fryer, M. J. "A Review of Some Non-Parametric Methods of Density Estimation." *Journal of the Institute of Mathematics and its Applications,* 20(1977): 3 335–54.

Good, I. J. and R. A. Gaskins. "Density Estimation and Bump-Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data." *Journal of the American Statistical Association,* 75(1980): 369 42–56.

Grissom, P. and W. I. Spurgeon. *Fertility Practices for Cotton and Corn in the Yazoo-Mississippi Delta.* Agricultural Experiment Station Bulletin No. 614, State College, Miss.: Miss. State University, 1961.

Harvey, A. C. "Estimating Regression Models With Multiplicative Heteroscedasticity." *Econometrica,* 44(1976): 3 461–65.

Hazell, P. B. R. "A Linear Alternative to Quadratic and Semivariance Programming for Farm Planning Under Uncertainty: Comment: Reply." *American Journal of Agricultural Economics,* 53(1971): 4 664–5.

Held, L. J. and G. A. Helmers. "Estimation of a Survival Function for a Simulated Northern Plains Wheat Farm." *North Central Journal of Agricultural Economics,* 4(1982): 1 19–26.

International Mathematical and Statistical Library. *The IMSL Library,* 8th Ed., IMSL Inc., Houston, Texas.

Just, R. E. and R. D. Pope. "Production Function Estimation and Related Risk Consideration." *American Journal of Agricultural Economics,* 61(1979): 2 276–84.

Kronmal, R. and M. Tarter. "The Estimation of Probability Densities and Cumulatives by Fourier Series Methods." *Journal of the American Statistical Association,* 63(1968): 323 925–52.

Kuester, J. L. and J. H. Mize. *Optimization Techniques with Fortran.* McGraw-Hill Book Co., New York, 1973.

Parzen, E. "Comment: Density Estimation and Bump-Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data." *Journal of the American Statistical Association,* 75(1980): 369 56–59.

Richardson, J. W. and G. D. Condra. "Farm Size Evaluation in the El Paso Valley: A Survival/Success Approach." *American Journal of Agricultural Economics,* 63(1981): 3 430–7.

Simon, H. A. "Dynamic Programming Under Uncertainty With a Quadratic Criterion Function." *Econometrica,* 24(1956): 1 74–81.

Stark, P. A. *Introduction to Numerical Methods.* MacMillan Publishing Co., Inc., New York, 1970.

Tarter, M. E. and R. A. Kronmal. "An Introduction to the Implementation and Theory of Nonparametric Density Estimation." *The American Statistician,* 30(1976): 3 105–12.

Taylor, C. R. "A Simple Method for Estimating Empirical Probability Density Functions." Staff Paper 81-1. Department of Agricultural Economics and Economics, Montana State University, 1981.

Theil, H. "A Note on Certainty Equivalence in Dynamic Planning." *Econometrica,* 25(1957): 2 346–49.

Theil, H. *Principles of Econometrics.* John Wiley and Sons, Inc., 1971.

Wegman, E. J. "Nonparametric Probability Density Estimation: A Comparison of Density Estimation Methods." *Journal of Statistical Computation and Simulation,* 1(1972): 3 225–45.

Wolfe, P. "The Secant Method for Simultaneous Nonlinear Equations." *Communications of ACM,* 2(Dec. 1959): 12–13.