Journal of Agricultural and Resource Economics 38(1):64–77

A Nested Test for Common Yield Distributions with Application to U.S. Corn

Jesse Tack

We propose the use of maximum-entropy techniques to nest and test the functional form of commonly used yield densities. We demonstrate how common parametric yield models can be nested within a maximum-entropy framework and subsequently tested for using standard hypothesis tests. We include an empirical application that tests the beta distribution against a more general maximum-entropy alternative using county-level corn yield data for the U.S. Corn Belt and find evidence in favor of the alternative generalized maximum-entropy distribution.

Key words: beta distribution, corn, crop yield, distribution, maximum entropy

Introduction

This article proposes the use of maximum-entropy techniques to nest and test the functional form of commonly used yield probability density functions (pdf). This research contributes to the large body of literature examining the credibility of distributional assumptions for crop yields (see Claassen and Just, 2011, and references therein). We demonstrate how common parametric yield models can be nested within a maximum-entropy framework and subsequently tested for using standard hypothesis tests. We include an empirical application that tests the beta distribution against a more general maximum-entropy alternative using county-level corn yield data for the U.S. Corn Belt. The nested hypothesis tests favor the alternative generalized maximum-entropy distribution, and we find evidence that the beta distribution misrepresents the “fatness” of the yield distribution’s tails, the importance of which for accurately predicting climate change impacts has been recently stressed (Nordhaus, 2011; Pindyck, 2011; Weitzman, 2011).

Recent work in the parametric yield modeling literature has focused on nesting a wide range of skewness-kurtosis combinations within a single parametric model of crop yields. For example, Ramirez and McDonald (2006) use the Johnson family of distributions, which nests the normal distribution within a class of non-normal alternatives. However, comparing these distributions to the commonly used beta distribution—as in Lu et al. (2008)—can be difficult because the beta is not directly related to the Johnson family. Since this relationship is non-nested, researchers cannot use standard hypothesis testing frameworks such as the conventional Wald, Lagrange multiplier, and likelihood ratio tests; rather, they must rely on more *ad hoc* methods such as a simple ranking of (potentially penalized) maximized likelihood functions. This is not meant as a criticism of this and other research that use ranking approaches (e.g., Sherrick et al., 2004; Norwood, Roberts, and Lusk, 2004), but rather as an illustration that even the more flexible models do not permit nested tests for the beta distribution that has been used to model crop yields since at least Day (1965).

A maximum-entropy (ME) density can be obtained by maximizing Shannon’s information entropy measure subject to known moment constraints. This distribution is “uniquely determined

Jesse Tack is an assistant professor in the Department of Agricultural Economics at Mississippi State University.
We would like to thank Christopher McIntosh, two anonymous referees, Keith Coble, and Ardian Harri for their helpful comments and suggestions, and Ximing Wu for guidance in coding the maximum-entropy density estimation. All mistakes are our own.

Review coordinated by Christopher S. McIntosh.

as the one which is maximally noncommittal with regard to missing information, and that it agrees with what is known, but expresses maximum uncertainty with respect to all other matters" (Jaynes, 1957; Wu, 2003). The ME density has been used in econometrics (e.g., Golan, Judge, and Miller, 1996; Zellner, 1997; Zellner and Tobias, 2001; Shen and Perloff, 2001), finance (e.g., Buchen and Kelly, 1996; Hawkins, 1997), and the income distribution literature (e.g., Wu, 2003; Wu and Perloff, 2005). Within the agricultural economics literature, it has been used to study food demand (e.g., Golan, Perloff, and Shen, 2001), land use (e.g., Miller and Plantinga, 1999), and crop production (e.g., Lence and Miller, 1998; Zhang and Fan, 2001). Outside of Stohs and LaFrance (2004) and Tack, Harri, and Coble (2012), the ME density has not been widely used in the crop yield modeling literature despite several theoretical and empirical advantages.

This article provides a general framework for nesting and testing a candidate density against a large class of alternatives. A limitation of the proposed framework is that the candidate density must belong to the exponential family of distributions, which includes many commonly used distributions. Additionally, we provide what we believe to be the first nested test of the beta distribution in the yield modeling literature.

Maximum-Entropy Framework

The maximum-entropy principle states that among all the distributions that satisfy certain moment constraints, we should choose the one that maximizes Shannon's information entropy.[1] For a random variable X, the ME density is obtained by maximizing Shannon's (1948) entropy measure:

$$W = -\int f(x)\ln f(x)dx, \tag{1}$$

subject to the moment constraints:

$$E[\phi_j(x)] = \int \phi_j(x)f(x)dx = \mu_j, \quad j = 0, \ldots, J, \tag{2}$$

where μ_j are known values supplied by the researcher. In practice, the constraint for $j = 0$ is used to ensure that the density $f(\cdot)$ integrates to unity by setting $\phi_0(x) = \mu_0 = 1$ so that:

$$E[\phi_0(x)] = \int f(x)dx = 1. \tag{3}$$

The associated Lagrangian is:

$$L = -\int f(x)\ln f(x)dx - \left[\lambda_0 \int f(x)dx - 1\right] - \sum_{j=1}^{J} \lambda_j \left[\int \phi_j(x)f(x)dx - \mu_j\right], \tag{4}$$

where $\lambda_0, \ldots, \lambda_J$ are the Lagrange multipliers associated with the constraints in equation (2). The solution to this maximization problem, obtained by calculus of variation, is given by:

$$f(x, \boldsymbol{\lambda}) = \exp[-\lambda_0 - \sum_{j=1}^{J} \lambda_j \phi_j(x)], \tag{5}$$

where λ_0 is set to $\log\left(\int \exp\left(-\sum_{j=1}^{J} \lambda_j \phi_j(x)\right) dx\right)$ to ensure the density integrates to unity (see Zellner and Highfield, 1988; Golan, Judge, and Miller, 1996).

[1] This section follows discussions examining maximum entropy densities in Park and Bera (2009) and Stengos and Wu (2009).

The ME density is of the exponential family and can be completely characterized by the moments $E[\phi_j(x)]$, $j = 1, \ldots, J$, the sample counterparts of which are sufficient statistics of the density. These characterizing moments capture known prior information surrounding the random variable, and by using them a least biased distribution is achieved via the ME principle. Lastly, for a given set of characterizing moments defined by $[\phi_0, \phi_1, \ldots, \phi_J]$ and associated known values $[1, \mu_1, \ldots, \mu_J]$, the maximized entropy is just $W = \lambda_0 + \sum_{j=1}^{J} \lambda_j \mu_j$ because:

$$
\begin{aligned}
W &= -\int f(x, \boldsymbol{\lambda}) \ln f(x, \boldsymbol{\lambda}) dx \\
&= \int \sum_{j=0}^{J} \lambda_j \phi_j(x) \exp[-\sum_{j=0}^{J} \lambda_j \phi_j(x)] dx \\
&= \sum_{j=0}^{J} \lambda_j \mu_j.
\end{aligned} \tag{6}
$$

Common Maximum-Entropy Densities

Including certain characterizing moments as constraints in the maximum-entropy framework will generate specific, well-known distributions. For example, suppose we restrict the support to the interval $[a, b]$ and only require that the density integrates to unity. Since the ME density is always determined by equation (5), in this specific case we have:

$$
f(x, \lambda) = \exp(-\boldsymbol{\lambda}_0) = \exp\left[-\log\left(\int_a^b \exp(0) dx\right)\right] = (b-a)^{-1}, \tag{7}
$$

which is the pdf corresponding to the uniform distribution. If we additionally have information on the mean, so that ϕ_1 is the identity function $\phi_1(x) = x$, and we use the additional characterizing moment $E(x) = \int x f(x) dx = \mu$ and assume that the support covers the positive real line, then:

$$
f(x, \boldsymbol{\lambda}) = \exp(-\lambda_0 - \lambda_1 x) = \exp\left[-\log\left(\int_0^\infty \exp(-\lambda_1 x) dx\right) - \lambda_1 x\right] = \lambda_1 \exp(-\lambda_1 x), \tag{8}
$$

which is the pdf corresponding to the exponential distribution. Furthermore, if we model the support as the entire real line and let $\phi_0(x) = 1$, $\phi_1(x) = x$, and $\phi_2(x) = x^2$, then the solution is $f(x, \boldsymbol{\lambda}) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2)$, the pdf for the normal distribution. [2]

Park and Bera (2009) provide characterizing moments for a variety of other distributions, including the beta, log-normal, gamma, and Weibul. Furthermore, the Pearson family and its extensions described in Cobb, Koppstein, and Chen (1983) are all ME densities (Stengos and Wu, 2009).

Building Beta's Nest

We focus on a nested test for the beta distribution, but this approach generalizes to any member of the exponential family. We focus on the beta because most of the empirical literature in agricultural economics over the past decade has used the beta distribution to model crop yields (Lu et al., 2008; Babcock, Hart, and Hayes, 2004).

The beta pdf is the ME solution on the unit interval under characterizing moments defined by $\phi_1(x) = \ln(x)$ and $\phi_2(x) = \ln(1-x)$, so that the associated moment constraints are

[2] To see that this is the pdf for the normal, start with the standard specification $f(x, \mu, \sigma) = (2\pi\sigma^2)^{-1/2} \exp[-(x-\mu)^2/2\sigma^2]$, expand the quadratic term, and then collect terms according to powers of x. A last step in which $(\lambda_0, \lambda_1, \lambda_2)$ are defined in terms of (μ, σ) yields the expression in the text.

$E[\ln(x)] = \int \ln(x) f(x) dx = \mu_1$ and $E[\ln(1-x)] = \int \ln(1-x) f(x) dx = \mu_2$. The solution takes the form $f^{Beta} = \exp[-\lambda_0 - \lambda_1 \ln(x) - \lambda_2 \ln(1-x)]$ and the maximized entropy is $W^{Beta} = \lambda_0 + \lambda_1 \mu_1 + \lambda_2 \mu_2$.

Now consider as an alternative the model defined over the unit interval with characterizing moments defined by the k raw moments $\phi_i(x) = x^i$, $i = 1, \ldots, k$. Let the associated moments be constrained by $\theta = [1, \theta_1, \ldots, \theta_k]$ according to $E(x^i) = \int x^i f(x) dx = \theta_i$, $i = 0, 1, \ldots, k$, in which case the associated solution and maximized entropy are $f^{Alt} = \exp[-\lambda_0 - \sum_{i=1}^{k} \lambda_i x^i]$ and $W^{Alt} = \lambda_0 + \sum_{i=1}^{k} \lambda_i \theta_i$. This is a natural alternative to consider for several reasons. First, Wu (2003) successfully used it to model the U.S. income distribution, where it compared favorably to the log-normal and gamma distributions. Second, the raw moments of the *ex post* estimated density will match their respective sample counterparts, an intuitively appealing constraint. We refer to this alternative density as a rank k exponential from here forward.

Given a $T \times 1$ vector of sample data $\boldsymbol{x}$, define the parameter estimates for each of the models by $\hat{\boldsymbol{\lambda}}^{Beta}$ and $\hat{\boldsymbol{\lambda}}^{Alt}$, the corresponding densities by $\hat{f}^{Beta}$ and $\hat{f}^{Alt}$, and the associated maximized entropy measures by $\hat{W}^{Beta}$ and $\hat{W}^{Alt}$. In general, the entropy of the ME density provides a benchmark for comparing these distributions by measuring their entropy discrepancy, which is the simple difference between $\hat{W}^{Beta}$ and $\hat{W}^{Alt}$. A number of indices based on this discrepancy have been developed, including entropy power (Shannon, 1948) and its extensions (e.g., Vasicek, 1976), the entropy power variance ratio (Dudewicz and van der Meulen, 1981), and the entropy power fraction (Gokhale, 1983). More recently, Soofi, Ebrahimi, and Habibullah (1995) developed an information discrimination (ID) distinguishability index based on the Kullback-Leibler discrimination function. In this case, one would distinguish between $\hat{f}^{Beta}$ and $\hat{f}^{Alt}$ based on a measure of their relative entropy $K(\hat{f}^{Beta} : \hat{f}^{Alt}) = \int \hat{f}^{Beta}(x) \ln[\hat{f}^{Beta}(x) / \hat{f}^{Alt}(x)] dx$.

In addition to these information-theoretic measures of difference, likelihood functions could also be used to rank distributions. As discussed in Stengos and Wu (2009), ME is equivalent to maximum likelihood estimation (MLE) when the distribution is a member of the exponential family, since the maximized log-likelihood is proportional to the maximized entropy. That is, for an arbitrary estimated pdf $f(x, \hat{\boldsymbol{\lambda}})$, the maximized log-likelihood $\hat{l}$ can be rewritten as:

$$\hat{l} = \sum_{t=1}^{T} \ln f(x_t, \hat{\boldsymbol{\lambda}}) = T \sum_{j=0}^{J} \hat{\lambda}_j \mu_j = T\hat{W}. \tag{9}$$

Thus, the maximum-entropy approach can be used to rank yield models using likelihood functions as in Norwood, Roberts, and Lusk (2004), Lu et al. (2008), Sherrick et al. (2004), and Ramirez and McDonald (2006). In addition, one can easily construct additional model selection criteria for further model comparison, such as the Akaike and Bayesian Information Criterion measures.

As currently formulated, the proposed maximum-entropy framework does not provide a nested hypothesis test for the beta distribution. This is because there does not exist a subset of the parameter space of the rank k exponential density that generates the beta density. This shortcoming is easily remedied by building a hybrid model combining the moment constraints associated with beta and rank k exponential densities. This hybrid model (and thus the associated nesting procedure) is valid for any candidate and alternative densities provided they are members of the exponential family and have common support.

Specifically, let the first two characterizing moments of the hybrid model defined on the unit interval be $\phi_1(x) = \ln(x)$ and $\phi_2(x) = \ln(1-x)$ and the next k be defined by $\phi_i(x) = x^i$, $i = 3, \ldots, k+2$. Including the normalizing constraint that ensures the density will integrate to unity, there are a total of $J = k + 3$ moment constraints:

$$
\begin{aligned}
E[\phi_0(x)] &= \int f(x)dx = 1, \\
E[\phi_1(x)] &= \int \ln(x) f(x)dx = \mu_1, \\
E[\phi_2(x)] &= \int \ln(1-x) f(x)dx = \mu_2, \\
E[\phi_j(x)] &= \int x^{j-2} f(x)dx = \mu_k, \quad j = 3, \ldots, J-1.
\end{aligned} \tag{10}
$$

The associated solution for the hybrid model is:

$$
f^{Hyb} = \exp[-\lambda_0 - \lambda_1 \ln(x) - \lambda_2 \ln(1-x) - \sum_{j=1}^{k} \lambda_{j+2} x^j], \tag{11}
$$

and the maximized entropy is $W^{Hyb} = \lambda_0 + \sum_{j=1}^{J-1} \lambda_j \mu_j$. Equation (11) makes it clear that the restriction $\lambda_3 = \ldots = \lambda_{k+2} = 0$ generates the beta density and the restriction $\lambda_1 = \lambda_2 = 0$ generates a density of the rank k exponential form. Thus, both the beta and the proposed alternative are properly nested within the hybrid model.

In general, the hybrid model nests the candidate density and that of the alternative, so the resulting density is likely to appear different relative to the candidate density. However, the advantage of this approach is that the researcher can control how different the hybrid density is by selecting the alternative. For this particular example, one might be concerned that the proposed rank k exponential alternative is undesirable because the density does not necessarily decay to 0 at the upper bound of the unit interval support as the beta does. While this was not a concern for the empirical application discussed in this article, one could include one or more characterizing moments of the form $\phi_i(x) = \ln^i(1-x)$, $i \in \{2k+1; \forall k \in \mathbb{N}\}$ to ensure that the alternative density decays to 0 as x approaches 1.

As proposed in Stengos and Wu (2009), it is possible to exploit the equivalence between ME and MLE estimates to conduct a likelihood ratio (LR) test of the nested hypothesis. Considering the J-dimension parameter space $\boldsymbol{\Lambda}_j$, we want to test the restriction $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}_m$, a subspace of $\boldsymbol{\Lambda}_J$, $m \leq J$. Defining by W_m and W_J the maximized entropy values for the restricted and unrestricted models, the test statistic $R = -2T(W_m - W_J)$ is distributed χ^2 with $J - m$ degrees of freedom. Thus, given parameter estimates and the associated maximized entropy for the hybrid model $\hat{W}^{Hyb}$, we can test both the beta and the proposed alternative against the hybrid using the test statistics $R^{Beta} = -2T(\hat{W}^{Beta} - \hat{W}^{Hyb})$ and $R^{Alt} = -2T(\hat{W}^{Alt} - \hat{W}^{Hyb})$.

The proposed approach easily generalizes to other commonly used distributions in the yield literature. To build a nest for the log-normal density, start with the characterizing moments defined by $\phi_1(x) = \ln(x)$ and $\phi_2(x) = \ln^2(x)$; for the normal use $\phi_1(x) = x$ and $\phi_2(x) = \ln^2(x)$; for the gamma use $\phi_1(x) = x$ and $\phi_2(x) = \ln(x)$; and for the Weibull use $\phi_1(x) = x^a$ and $\phi_2(x) = \ln(x)$. For each case, if a rank k exponential model is chosen as the alternative then the implied hybrid density is defined as in equation (11), with $\phi_1(x)$ and $\phi_2(x)$ replacing $\ln(x)$ and $\ln(1-x)$, in which case the candidate and alternative densities are properly nested.[3] The alternative density and corresponding hybrid must always be defined over the support of the candidate density to ensure a properly nested structure. Furthermore, one is not restricted in using higher order raw moments to construct the alternative model, and an interesting avenue for future research would be the consideration of

[3] The resulting hybrid for some of these cases would contain redundancies regarding the included characterizing functions (e.g., the hybrid for the gamma would include $\phi(x) = x$ twice). In this situation, the second instance of the characterizing function can simply be eliminated from the hybrid density without invalidating the properly nested structure of the candidate and alternative densities.

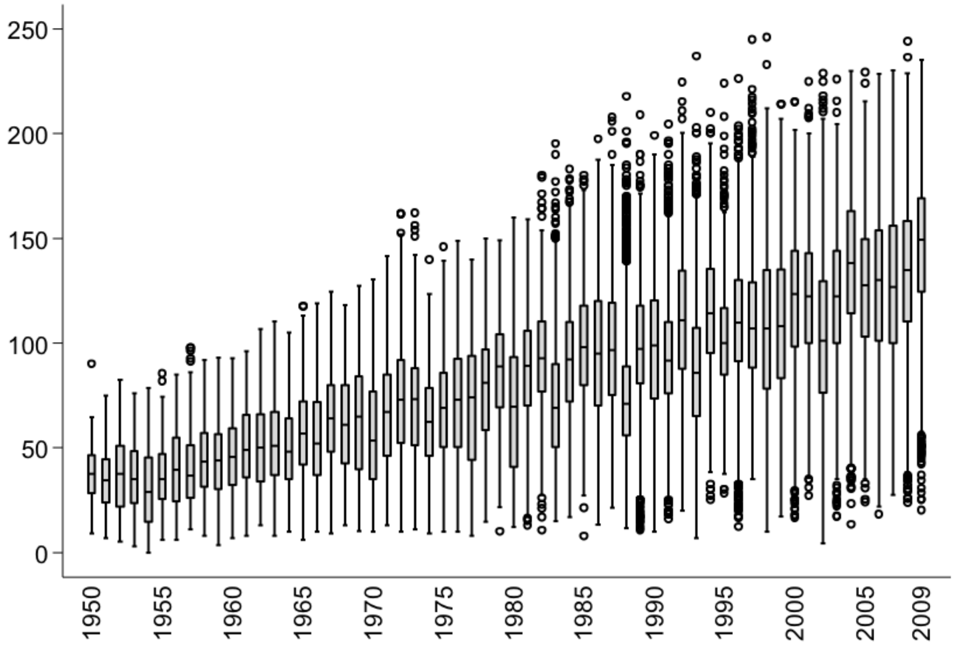

Figure 1. Annual Box Plots for the Raw Yield Data

alternative characterizing moments such as a sequence of logarithmic functions $\ln^i(x)$ for $i = 1, \ldots, k$, or trigonometric functions such as $\sin(x)$, $\cos(x)$, and $\tan^{-1}(x)$.[4]

Data

We focus on the U.S. Corn Belt and include any county in Illinois, Indiana, Iowa, Missouri, or Ohio that has reported a corn yield to the National Agricultural Statistics Service at any time between 1950 and 2009. We define yield as production over harvested acreage, which produces a dataset containing 496 counties and 29,057 observations. County-level yields are commonly used in the literature as they are the most disaggregate level where long time series are available (Tack, Harri, and Coble, 2012). From a policy perspective, these distributions are useful for analyzing county-triggered yield and revenue support crop insurance products. The raw yield data are presented in figure 1; there is a significant amount of inter- and intra-annual variation. The data also display a noticeable increase in both the mean and variance of corn yields over time.

We follow the standard practice of scaling both the mean and variance of yield outcomes to the final year in the dataset. We employ the fixed effects model:

$$y_{ist} = \alpha_i + \beta_{1s}t + \beta_{2s}t^2 + \varepsilon_{it}, \tag{12}$$

where the dependent variable y_{ist} denotes crop yield in county i in state s in period t, α_i captures county-specific time-invariant factors that influence yields, and $\beta_{1s}t + \beta_{2s}t^2$ controls for state-specific technological change. We use the proportional variance model discussed in Harri et al. (2011), $Var(\varepsilon_{it}) = \sigma^2[E(\hat{y}_{ist})]^2$, where the variance of yields moves in direct proportion to predicted (trending) yield $\hat{y}_{ist}$. We adjust residuals according to:

$$v_{it} = \hat{\varepsilon}_{it}\frac{\hat{y}_{isT}}{\hat{y}_{ist}}, \tag{13}$$

[4] Park and Bera (2009) provide a nice overview of how different characterizing moments can be used to capture particular distributional shapes such as thick tails, peakedness, and asymmetry.

Table 1. Sample Data, Detrended Normalized Yields 1950–2009

State	Mean	Std Dev	Min	Max	No. of Counties	N
Illinois	0.4186	0.0788	0.0765	0.667	102	6114
Indiana	0.4691	0.0729	0.1748	0.6669	92	5459
Iowa	0.4649	0.0799	0.0848	0.6672	99	5940
Missouri	0.3836	0.0976	0.0004	0.6677	115	6362
Ohio	0.4761	0.0713	0.1586	0.6673	88	5182

where $\hat{\varepsilon}_{it}$ are the residuals and $\hat{y}_{isT}$ is the predicted yield for the final year in the dataset T. Equation (12) is estimated for each state separately, and the yield data for each county are constructed as $\tilde{y}_{ist} = \hat{y}_{isT} + v_{it}$. A final normalizing step is performed by dividing all outcomes by the factor $1.5\max_t\{\tilde{y}_{ist}\}$.[5] Summary statistics for the detrended and normalized data are presented in table 1.

We implicitly assume that the quadratic trend correctly specifies technical innovations across time and that state-level pools of the normalized yield data are representative of county-level data. We conduct several robustness checks to evaluate the credibility of these assumptions for the empirical results that follow. The details and associated implications of these robustness checks are described in more detail below.

Empirical Results

We estimate the beta, rank k exponential, and hybrid density functions for each of the five Corn Belt states. Within each state, county-yield observations are pooled and the densities are estimated using Matlab.[6] We estimate five densities: the beta, three versions of the rank k exponential ($k = 3$, 5, and 7), and the hybrid.[7] As discussed earlier, the hybrid model combines the moments from the beta and rank k exponential ($k = 7$) densities. Newton's method can be problematic when a large number of moments are included as constraints, so we use the sequential updating method developed in Wu (2003) to estimate the parameters of each density.

Since the ME estimates are equivalent to MLE estimates for all of the densities considered here, we calculate the log-likelihood according to equation (9). Column 1 of table 2 reports these values for each of the estimated densities in each of the five states. Log-likelihood is nondecreasing as one moves down the list of estimated densities within each state. This is because additional moment constraints provide additional flexibility in the ME framework. Indeed, ME can provide the same amount of flexibility as kernel-density-type estimators, provided the number of moment constraints is sufficiently large.

Both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are easily calculated once the log-likelihood is known and can help detect over-fitting. The results are reported in columns 2 and 3 of table 2. The AIC favors the rank 5 density for all states except Ohio, for which the more flexible rank 7 density is favored. The BIC, which has a greater complexity penalty, favors the rank 5 density for all states.

Densities are commonly compared using their information discrepancies. Soofi, Ebrahimi, and Habibullah's (1995) ID index, denoted $ID(f : f^*)$, is a normalization of the Kullback-Leibler

[5] Since we are building a test for the two-parameter beta distribution, this normalization of the data to the unit interval is required. The approach easily generalizes to the three-parameter case, but the second characterizing moment would become $\phi_2 = \ln(c - y)$ where c is the upper bound of the support. We fully acknowledge that the definition of the upper bound used here is *ad hoc* and could potentially affect the empirical findings, as is common when working with the beta distribution in practice.

[6] Code available from the author upon request. The Matlab code for the rank k exponential is available on Ximing Wu's web page at http://agecon2.tamu.edu/people/faculty/wu-ximing/. Matlab code for the beta and hybrid models was written by the authors and available upon request.

[7] The first three raw moments are included in each of the rank k densities because yield distributions are likely skewed (Hennessy, 2009a,b).

Table 2. LR Tests for Estimated Densities

	(1) Log Likelihood	(2) Akaike Information Criterion	(3) Bayesian Information Criterion	(4) Information Discrimination Index	(5) Likelihood Ratio Test Statistic
Illinois					
Beta	6688	-2.1868	-2.1835	–	–
$k = 3$	7093	-2.3191	-2.3147	0.096	–
$k = 5$	7139	-2.3334	-2.3268	0.0074	91.71
$k = 7$	7140	-2.3330	-2.3242	0.00011	1.22
Hybrid	7140	-2.3323	-2.3213	0.0000022[a]	0.00[a]
				0.071[b]	903.65[b]
Indiana					
Beta	6501	-2.3807	-2.3771	–	–
$k = 3$	6717	-2.4595	-2.4547	0.062	–
$k = 5$	6777	-2.4806	-2.4733	0.011	119.1
$k = 7$	6778	-2.4805	-2.4708	0.00023	3.275
Hybrid	6778	-2.4797	-2.4676	0.0000025[a]	0.00[a]
				0.049[b]	554.63[b]
Iowa					
Beta	6423	-2.1616	-2.1582	–	–
$k = 3$	6972	-2.3461	-2.3415	0.15	–
$k = 5$	7101	-2.3890	-2.3822	0.022	258.9
$k = 7$	7103	-2.3889	-2.3799	0.00028	3.563
Hybrid	7104	-2.3886	-2.3774	0.00027[a]	2.38[a]
				0.11[b]	1362.64[b]
Missouri					
Beta	5382	-1.6909	-1.6878	–	–
$k = 3$	5963	-1.8734	-1.8692	0.096	–
$k = 5$	5981	-1.8784	-1.8720	0.0028	35.24
$k = 7$	5982	-1.8781	-1.8696	0.00021	2.672
Hybrid	5983	-1.8777	-1.8671	0.000085[a]	1.14[a]
				0.090[b]	1202.16[b]
Ohio					
Beta	6286	-2.4248	-2.4210	–	–
$k = 3$	6461	-2.4923	-2.4872	0.048	–
$k = 5$	6485	-2.5007	-2.4931	0.0045	47.67
$k = 7$	6491	-2.5023	-2.4922	0.0013	12.43
Hybrid	6491	-2.5015	-2.4889	0.0000064[a]	0.00[a]
				0.039[b]	411.45[b]

Notes: Likelihood Ratio Test Statistic is $f_{k+2}(x)$ versus $f_k(x)$ unless otherwise noted.
[a] Test for the Hybrid versus the $k = 7$ model.
[b] Test for the Hybrid versus the Beta model.

distance between two distributions f and f^*, with values closer to 0 indicating that the two densities are more similar. Column 4 of table 2 reports the indices, and the first three entries within each state correspond to $ID(f^{k=3} : f^{Beta})$, $ID(f^{k=5} : f^{k=3})$, and $ID(f^{k=7} : f^{k=5})$. The two values in the final entry correspond to $ID(f^{Hyb} : f^{k=7})$ and $ID(f^{Hyb} : f^{Beta})$. Each of the rank k exponential densities and the hybrid provide more information than the beta, and the information discrepancy decreases as more moment constraints are included.

Taken as a whole or individually, the measures provided in columns 1 through 4 provide information that can be used to select (or rank) the various models. When available, results from nested hypothesis testing can greatly complement these measures. We calculate LR test statistics from the maximized entropy values as described earlier in the article, which are provided in column 5 of table 2. The first two entries report ratios for $f^{k=5}$ versus $f^{k=3}$ and $f^{k=7}$ versus $f^{k=5}$, and the two values in the final entry are for f^{Hyb} versus $f^{k=7}$ and f^{Hyb} versus f^{Beta}. There are seven degrees of freedom for the hybrid-beta test, for all others there are two. The corresponding critical values for a 5% significance level are $\chi^2(7) = 14.07$ and $\chi^2 = 5.99$.

The nested hypothesis test results for the rank k exponential models make it clear that including more than just the first three moments is strongly supported by the data. Thus, while skewness is widely recognized in the yield literature as an important characteristic of yield distributions, it should not be considered a threshold beyond which additional moments do not matter. This illustrates a strength of the proposed approach, as the ME framework's provision of sequentially nested tests of more complicated models is especially important for yield modeling. Crop yield research is often limited by small datasets in which the number of observations per unit can be very small. Under our approach, a researcher can begin with a parsimonious model (e.g., the normal which is just the rank 2 exponential) and sequentially test whether the additional flexibility gained from including higher order moments is warranted.

The nested test of the hybrid density versus the rank 7 exponential suggests that the proposed restrictions are consistent with the data at a 5% level of significance. The LR test statistics are below the critical value of 5.99 for all states, providing robustness of this finding across the U.S. Corn Belt. Conversely, the nested test of the hybrid density versus the beta suggest that the proposed restrictions are not consistent with the data at the 5% level (critical value of 14.07), and again we see robustness across the included states. Taken together, these tests provide a nested framework for testing a candidate density against a proposed alternative, and the evidence strongly suggests rejection of the beta in favor of the rank 7 exponential.

The credibility of the results in table 2 relies on a correct specification of the yield trend and that the state-level pools of normalized yield data are representative of county-level data. For the former, we consider two additional technological change specifications, a one-knot and two-knot linear spline, which are consistent with the detrending approach currently used by the Risk Management Agency. We find that the overall pattern of results provided in table 2 does not change dramatically across either alternative model and thus conclude that these results are robust across alternative trend specifications.

To addresses the appropriateness of state-level pools, we demean and rescale the county-level normalized yield data as in Claassen and Just (2011). This approach ensures that the resulting observations are mean 0 and unit variance for each county, and thus poolable across counties within each state. Again, we find that the overall pattern of results provided in table 2 does not change dramatically under this alternative approach and thus conclude that results are representative of what one should expect at the county level. Results of both credibility checks are available from the authors upon request.

Figure 2 provides a histogram and kernel-density plot for the underlying normalized yield data overlaid with fitted densities for the beta and rank 7 models for Iowa.[8] Relative to the rank 7 exponential, the beta distribution misrepresents the density across nearly all relevant subsections of

[8] Graphs for the other states display similar results and are available from the authors upon request.

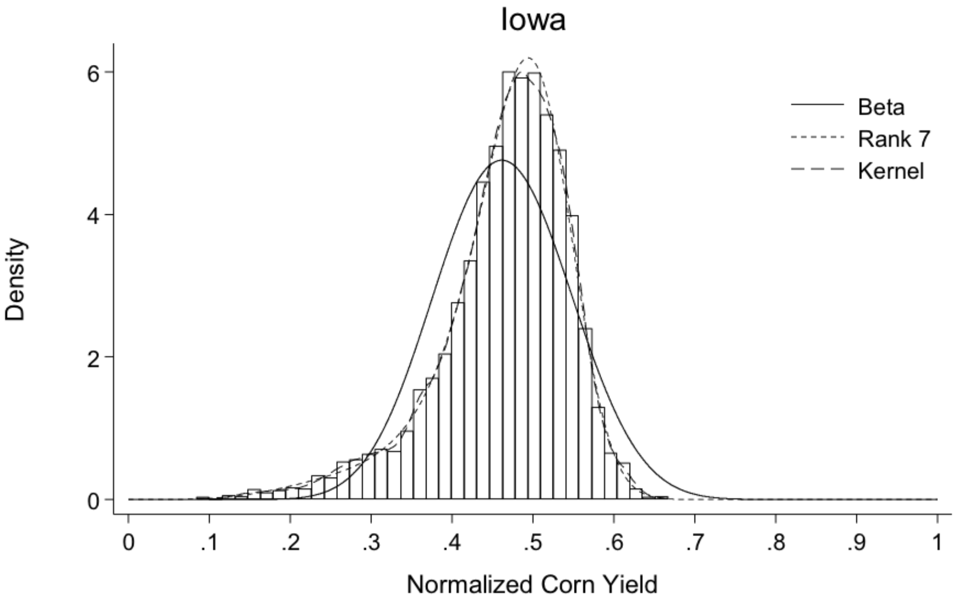

Figure 2. Histogram and Kernel Density for the Underlying Data, Overlaid with Fitted Densities for the Beta and Rank 7 Exponential Models

the support: it underestimates the lower tail (catastrophic losses), overestimates the density between the lower tail and the mean (shallow losses), and overestimates the upper tail (bumper crops). The fitted rank 7 exponential compares favorably with a kernel-density plot of the underlying data, which is not surprising given that the rank 7 model can be thought of as a seventh-order approximation of the log density.

Economic Significance

To measure the economic significance of the differences between the fitted beta and rank 7 exponential models, we compare implied premium rates for Group Risk Plan (GRP) crop insurance policies. The GRP program is triggered by county-level yield losses and has received much attention in the literature due to its potential to mitigate adverse selection and moral hazard problems often associated with farm-based policies (Harri et al., 2011). Given the fitted densities $f^{k=7}$ and f^{Beta}, the implied rates for coverage levels $cov \in [.5, .9]$ and models $m \in \{k=7, Beta\}$ are calculated as the ratio of expected indemnity over liability:

$$rate^{m}_{cov} = E(indemnity^{m}_{cov})/liability^{m}_{cov}, \tag{14}$$

where:

$$E(indemnity^{m}_{cov}) = \int_{0}^{y^{m}_{cov}} [y^{m}_{cov} - y)/cov] f^{m}(y)dy, \tag{15}$$

$$liability^{m}_{cov} \equiv y^{m}_{cov} = covE^{m}(y) = cov \int_{0}^{1} y f^{m}(y)dy. \tag{16}$$

The factor $1/cov$ in the expected indemnity calculation reflects the "disappearing deductible" built into GRP contracts (Barnett et al., 2005).

To facilitate rate comparisons, we report the ratio of calculated rates for each state, $ratio_{cov} = rate^{Beta}_{cov}/rate^{k=7}_{cov}$, in figure 3. The overall pattern of these ratios suggests that rates

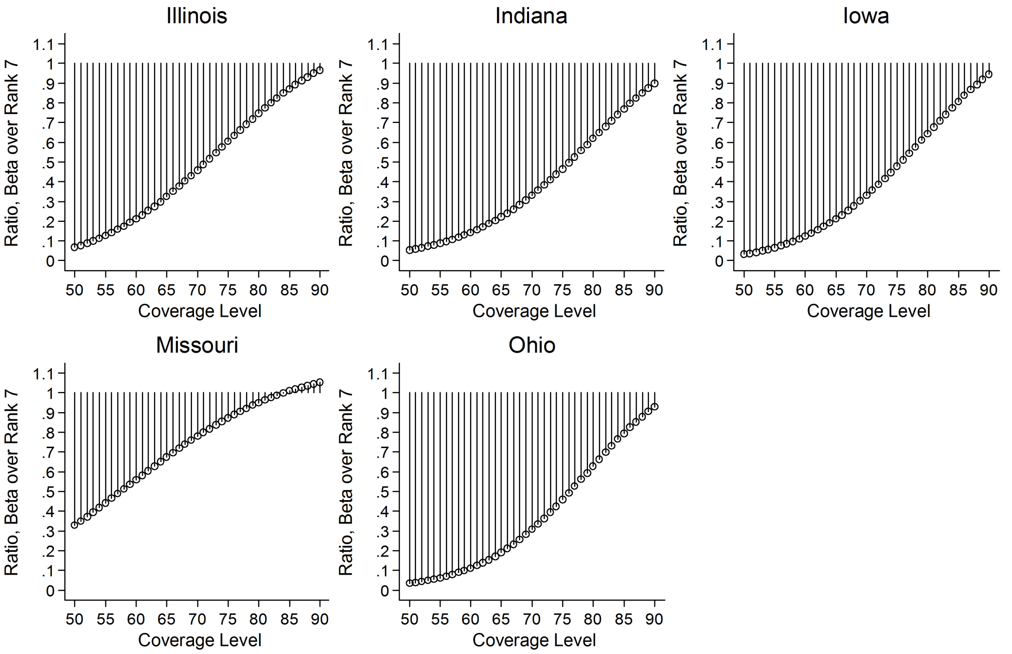

Figure 3. Pair Plots for GRP Rate Ratios under Beta and Rank 7 Exponential Models

are substantially smaller for the beta distribution across the lower range of coverage levels but approach the rate for the rank 7 model as the coverage level approaches 90%. This is consistent with the findings in figure 2, where the beta underestimates the density in the lower tail and then overestimates the density among shallow-loss outcomes in a compensating manner. Overall, the substantial differences in the GRP premium rates across models suggest that the differences in density estimates carry economic significance.

Conclusion

This article provides a general framework for nesting and testing a candidate density against a large class of alternatives, so long as the densities are members of the exponential family. This requirement is not especially limiting, as many commonly used densities such as the beta, normal, log-normal, gamma, and Weibull are all members of this family. We include an empirical application that focuses on testing the beta distribution using county-level corn yield data for the U.S. Corn Belt. Our findings suggest that the beta distribution be rejected in favor of our proposed alternative.

The testing framework and empirical findings presented here can provide useful guidance about the appropriateness of distributional assumptions in a variety of contexts. As discussed in the text, the beta distribution has been used extensively in research focusing on crop yields, and our findings suggest rejecting this distribution in favor of a parametric alternative that mimics the flexibility of commonly used kernel-density estimators. Thus, one might be concerned that previous applications based on the beta distribution employed an overly restrictive parametric assumption and that the findings might not be robust to the use of the more flexible alternative proposed here. This potential concern provides a rich area for future research.

Our empirical application includes an analysis of the economic significance of the distributional differences between the beta and the proposed alternative. The economic significance is established using a simple comparison of premium-rate differences for Group Risk Plan crop-insurance contracts; however, a more holistic comparison might include an out-of-sample exercise that examines whether private insurers can accrue economic rents through the reinsurance option provided by the Federal Crop Insurance Corporation's Standard Reinsurance Agreement (as in Harri et al., 2011). In this context, future research could consider whether the in-sample testing procedures presented here are a leading indicator of out-of-sample rating performance exercises in the spirit of Harri et al. (2011).

Since the proposed testing framework is general enough to include any exponential distribution, it could also prove useful for pricing livestock revenue insurance products based on the Asian basket option. As discussed in Hart, Babcock, and Hayes (2001), pricing these options requires "an analytic approximation to produce closed-form probability density functions for the price averages of the futures prices" (pg. 560), and they consider both the log-normal and the inverse gamma distributions as candidates. Given that both of these candidate distributions are members of the exponential family defined over the positive real line, one could define a hybrid model that included the characterizing moments for each distribution and then use the testing framework presented here. As the appropriateness of the log-normality assumption remains an active area of concern in the literature, future research using this framework seems warranted.

While the above discussion has focused mainly on applications in a crop insurance context, the testing procedure presented here can also provide useful guidance for empirical studies that use moment-models in the spirit of Antle (1983) and Just and Pope (1978). Given Tack, Harri, and Coble's (2012) framework for the Moment-Based Maximum-Entropy model, which essentially links moment-models with maximum-entropy density estimation, one can sequentially add additional equations to the system of moments and test the appropriateness of this inclusion via implied changes in the resulting maximum-entropy distribution. Given the broad use of these models across several academic disciplines, including agricultural economics, agronomy, and climate change, this could prove a rich area for future research.

[Received September 2012; final revision received February 2013.]

References

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

Babcock, B. A., C. E. Hart, and D. J. Hayes. "Actuarial Fairness of Crop Insurance Rates with Constant Rate Relativities." *American Journal of Agricultural Economics* 86(2004):563–575.

Barnett, B. J., J. R. Black, Y. Hu, and J. R. Skees. "Is Area-Yield Insurance Competitive with Farm-Yield Insurance?" *Journal of Agricultural and Resource Economics* 30(2005):285–301.

Buchen, P. W., and M. Kelly. "The Maximum Entropy Distribution of an Asset Inferred from Option Prices." *Journal of Financial and Quantitative Analysis* 31(1996):143–159.

Claassen, R., and R. E. Just. "Heterogeneity and Distributional Form of Farm-Level Yields." *American Journal of Agricultural Economics* 93(2011):144–160.

Cobb, L., P. Koppstein, and N. H. Chen. "Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family." *Journal of the American Statistical Association* 78(1983):124–130.

Day, R. H. "Probability Distributions of Field Crop Yields." *Journal of Farm Economics* 47(1965):713–741.

Dudewicz, E. J., and E. C. van der Meulen. "Entropy-Based Tests of Uniformity." *Journal of the American Statistical Association* 76(1981):967–974.

Gokhale, D. V. "On Entropy-Based Goodness-of-Fit Tests." *Computational Statistics & Data Analysis* 1(1983):157–165.

Golan, A., G. C. Judge, and D. Miller. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. New York: Wiley, 1996.

Golan, A., J. M. Perloff, and E. Z. Shen. "Estimating a Demand System with Nonnegativity Constraints: Mexican Meat Demand." *Review of Economics and Statistics* 83(2001):541–550.

Harri, A., K. H. Coble, A. P. Ker, and B. J. Goodwin. "Relaxing Heteroscedasticity Assumptions in Area-Yield Crop Insurance Rating." *American Journal of Agricultural Economics* 93(2011):703–713.

Hart, C. E., B. A. Babcock, and D. J. Hayes. "Livestock Revenue Insurance." *Journal of Futures Markets* 21(2001):553–580.

Hawkins, R. J. "Maximum Entropy and Derivative Securities." *Advances in Econometrics* 12(1997):277–301.

Hennessy, D. A. "Crop Yield Skewness and the Normal Distribution." *Journal of Agricultural and Resource Economics* 34(2009a):34–52.

———. "Crop Yield Skewness under Law of the Minimum Technology." *American Journal of Agricultural Economics* 91(2009b):197–208.

Jaynes, E. T. "Information Theory and Statistical Mechanics." *Physical Review* 106(1957):620–630.

Just, R. E., and R. D. Pope. "Stochastic Specification of Production Functions and Economic Implications." *Journal of Econometrics* 7(1978):67–86.

Lence, S. H., and D. J. Miller. "Recovering Output-Specific Inputs from Aggregate Input Data: A Generalized Cross-Entropy Approach." *American Journal of Agricultural Economics* 80(1998):852–867.

Lu, Y., O. A. Ramirez, R. M. Rejesus, T. O. Knight, and B. J. Sherrick. "Empirically Evaluating the Flexibility of the Johnson Family of Distributions: A Crop Insurance Application." *Agricultural and Resource Economics Review* 37(2008):79–91.

Miller, D. J., and A. J. Plantinga. "Modeling Land Use Decisions with Aggregate Data." *American Journal of Agricultural Economics* 81(1999):180–194.

Nordhaus, W. D. "The Economics of Tail Events with an Application to Climate Change." *Review of Environmental Economics and Policy* 5(2011):240–257.

Norwood, B., M. C. Roberts, and J. L. Lusk. "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions." *American Journal of Agricultural Economics* 86(2004):1032–1043.

Park, S. Y., and A. K. Bera. "Maximum Entropy Autoregressive Conditional Heteroskedastic Models." *Journal of Econometrics* 150(2009):219–230.

Pindyck, R. S. "Fat Tails, Thin Tails, and Climate Change Policy." *Review of Environmental Economics and Policy* 5(2011):258–274.

Ramirez, O. A., and T. U. McDonald. "The Expanded Johnson System: A Highly Flexible Crop Yield Distribution Model." In *Proceedings,* Long Beach, CA: American Agricultural Economics Association, 2006.

Shannon, C. E. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27(1948):623–656.

Shen, E. Z., and J. M. Perloff. "Maximum Entropy and Bayesian Approaches to the Ratio Problem." *Journal of Econometrics* 104(2001):289–313.

Sherrick, B. J., F. C. Zanini, G. D. Schnitkey, and S. H. Irwin. "Crop Insurance Valuation under Alternative Yield Distributions." *American Journal of Agricultural Economics* 86(2004):406–419.

Soofi, E. S., N. Ebrahimi, and M. Habibullah. "Information Distinguishability with Application to Analysis of Failure Data." *Journal of the American Statistical Association* 90(1995):657–668.

Stengos, T., and X. Wu. "Information-Theoretic Distribution Test with Application to Normality." *Econometric Reviews* 29(2009):307–329.

Stohs, S. M., and J. T. LaFrance. "A Learning Rule for Inferring Local Distributions over Space and Time." In J. P. LeSage and R. K. Pace, eds., *Spatial and Spatiotemporal Econometrics,* Advances in Econometrics. Amsterdam: Elsevier, 2004, 295–331.

Tack, J., A. Harri, and K. Coble. "More than Mean Effects: Modeling the Effect of Climate on the Higher Order Moments of Crop Yields." *American Journal of Agricultural Economics* 94(2012):1037–1054.

Vasicek, O. "A Test for Normality Based on Sample Entropy." *Journal of the Royal Statistical Society. Series B (Methodological)* 38(1976):54–59.

Weitzman, M. L. "Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change." *Review of Environmental Economics and Policy* 5(2011):275–292.

Wu, X. "Calculation of Maximum Entropy Densities with Application to Income Distribution." *Journal of Econometrics* 115(2003):347–354.

Wu, X., and J. M. Perloff. "China's Income Distribution, 1985-2001." *Review of Economics and Statistics* 87(2005):763–775.

Zellner, A. "The Bayesian Method of Moments (BMOM): Theory and Applications." *Advances in Econometrics* 12(1997):85–105.

Zellner, A., and R. A. Highfield. "Calculation of Maximum Entropy Distribution and Approximation of Marginal Posterior Distributions." *Journal of Econometrics* 37(1988):195–209.

Zellner, A., and J. Tobias. "Further Results on Bayesian Method of Moments Analysis of the Multiple Regression Model." *International Economic Review* 42(2001):121–139.

Zhang, X., and S. Fan. "Estimating Crop-Specific Production Technologies in Chinese Agriculture: A Generalized Maximum Entropy Approach." *American Journal of Agricultural Economics* 83(2001):378–388.