Are Crop Yield Distributions Negatively Skewed?

A Bayesian Examination

Xiaodong Du
Postdoc Research Associate
Center for Agricultural and Rural Development
Iowa State University
xdu@iastate.edu

David A. Hennessy
Professor
Department of Economics, and
Center for Agricultural and Rural Development
Iowa State University
hennessy@iastate.edu

Cindy Yu
Assistant Professor
Department of Statistics
Iowa State University
Cindyyu@iastate.edu

Poster prepared for presentation at the Agricultural & Applied Economics Association 2010 AAEA, CAES, & WAEA Joint Annual Meeting, Denver, CO, July 25-27, 2010.

Are Crop Yield Distributions Negatively Skewed?
A Bayesian Examination

Xiaodong Du
David A. Hennessy
Cindy Yu
Department of Economics
CARD Center for Agricultural and Rural Development
IOWA STATE UNIVERSITY

Introduction

RISKS ORIGINATED FROM the random nature of yield have significant impacts on farmers' production and marketing decisions. A better understanding of crop yield distributions is important both for crop producers and for the crop insurance industry, where contract payout patterns are sensitive to distribution tails. There is a long-established literature on how inputs affect mean and variance of crop yield distributions while controversy surrounds skewness attributes of typical yield distributions. In this study, we extend the stochastic production model in Just and Pope (1978) to accommodate skewness. The model is applied to several experimental yield datasets for corn and cotton at the small plot-level. We conduct inference within a Bayesian Monte Carlo Markov Chain (MCMC) framework (Gelman et al., 2004).

Stochastic Technology

LET EXPERIMENTAL PLOT yield be given by

$$y=f(N)+g(N)\varepsilon^{h(N)} \qquad (1)$$

where N is an input, e.g., nitrogen. Here ε is random while f, g, and h are functions defined to be consistent with the Just-Pope method. Let

$$f(N) = \alpha_0 N^{\alpha_1}; \qquad g(N) = \beta_0 N^{\beta_1}; \qquad h(N) = \gamma_0 N^{\gamma_1} \qquad (2)$$

The moments for crop yield are:

$$\mathbb{E}[y] = \alpha_0 N^{\alpha_1} + \beta_0 N^{\beta_1}\mathbb{E}[\varepsilon^{\gamma_0 N^{\gamma_1}}]; \qquad (3)$$

$$\mathbb{E}\left[(y - \mathbb{E}[y])^2\right] = \beta_0^2 N^{2\beta_1}\,\mathrm{Var}\,(\varepsilon^{\gamma_0 N^{\gamma_1}})$$

$$\mathrm{Skew}\,(y) = \mathbb{E}\left[(\varepsilon^{\gamma_0 N^{\gamma_1}} - \mathbb{E}\,[(\varepsilon^{\gamma_0 N^{\gamma_1}}])^3\right] / \left\{\mathbb{E}\left[(\varepsilon^{\gamma_0 N^{\gamma_1}} - \mathbb{E}\,[\,(\varepsilon^{\gamma_0 N^{\gamma_1}}]\,)^2\,\right]\right\}^{3/2}.$$

The stochastic production function in eqn. (1) is mean, variance, and skewness flexible.

Empirical Methodology

GIVEN THAT FIELD experiments are conducted on discrete and limited nitrogen application levels, we estimate the yield distribution for each nitrogen level individually. Following (1), for nitrogen level i ($i =$ 1,2..., I), yields are modeled as

$$y^i = a_0^i + \mathbf{X'}\,\boldsymbol{\beta} + b^i\varepsilon^{1/c^i},\ \varepsilon \sim D \cdot J(\alpha)$$
$$J(\alpha) = \mathrm{Beta}\,(\alpha, \alpha). \qquad (4)$$

where $y^i = [y_0^i, y_2^i, y_k^i]'$ denotes k yield observation and a_0^i denotes the constant term which varies with nitrogen input, $f(N^i)$. The matrix $\mathbf{X} = [x_1, x_2, \ldots x_L]'$ denotes the L controlled variables including, e.g., rotation effect, and location/time dummies with the corresponding coefficient vector $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots \beta_L]'$. Scale of crop yield is represented by $b^i = g(N^i)$.

We specify the yield random variations as ε^{1/c^i}, where ε is assumed to follow a symmetric beta distribution Beta (α, α) on the range (0, D), and skewness is represented by the parameter c^i. The beta distribution has long been popular in yield distribution models (Nelson and Preckel, 1989). Skewness is introduced into an originally symmetric distribution on ε through c^i. In doing so, it is possible to retain some well known properties of symmetric distributions and compare different classes of skewed distributions in a common framework. The corresponding yield skewness Skew (y^i), is

$$\frac{\dfrac{B(\alpha+3/c^i,\alpha)}{J(\alpha)} - \dfrac{3B(\alpha+2/c^i,\alpha)B(\alpha+1/c^i,\alpha)}{[J(\alpha)]^2} + 2\left(\dfrac{B(\alpha+1/c^i,\alpha)}{J(\alpha)}\right)^3}{\dfrac{B(\alpha+2/c^i,\alpha)}{J(\alpha)} - \left(\dfrac{B(\alpha+1/c^i,\alpha)}{J(\alpha)}\right)^2} \qquad (5)$$

In this study, inference is conducted within a Bayesian MCMC method, which has the advantage of easily incorporating inequality constraints on parameters into the estimation procedure. The inequality constraint is to ensure that parameter estimates are consistent with relationships implied by the underlying distribution asumption, which is

$$0 < \left[\frac{1}{D}(y^i - a_0^i + \mathbf{X'}\,\boldsymbol{\beta})/b^i\right]^{c^i} < 1.$$

Furthermore, the Bayesian procedure is particularly suitable as the model specified in (4) is highly nonlinear in the parameters. We employ the random-walk Metropolis-Hasting algorithms for updating posterior draws. Normal priors are adopted for all parameters. After convergence, draws by the Gibbs sampler are used to compute the mean and standard error of each parameter.

Data

THREE DATASETS ARE employed: (i) corn yield dataset A: experiments in Floyd County, Iowa from 1979 to 2003 with four nitrogen levels (lb./ac.), 0, 80, 160, and 240. Control variables include time trend and corn-after-corn (CC) rotation; (2) corn yield dataset B: experiments on four Iowa farms in 1986-1991 with ten nitrogen levels (lb./ac.): 0, 25, 50, 75, 100, 125, 150, 200, 250, and 300. Variables include dummies for four years and CC rotation; (3) cotton yield dataset C: experiments in three Texas counties, 1998-2002, with four nitrogen levels (lb./ac.), 0, 50, 100, and 150. Control variables are dummies for four years and two locations.

Results

THE GIBBS SAMPLER is coded in Matlab and ran for 10,000 cycles with the first 5000 as burn-in. The mean derived and standard error of the skewness in (5) for each nitrogen level are presented in the table below. The results indicate that the skewness parameter c^i generally increases with the levels of nitrogen inputs, while the corresponding derived distribution skewness decreases. In other words, the impact of the nitrogen input on the crop yield distribution is in general skewness-decreasing.

The Figure plots the posterior predictive yield distribution at mean parameter values of Dataset A.

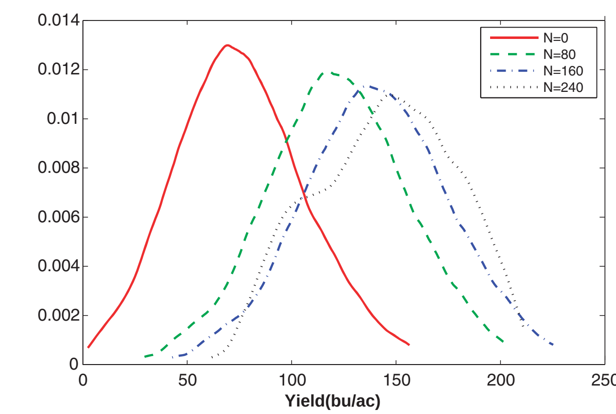
Estimated skewness and standard error (in parentheses)

Dataset A				
Nitrogen	0	80	160	240
Skewness	0.1576 (0.0628)	-0.1430 (0.0456)	-0.1566 (0.0536)	-0.1593 (0.0580)

Dataset B				
Nitrogen	0	25	50	75
Skewness	0.2566 (0.0528)	0.1658 (0.0686)	-0.0324 (0.0523)	-0.1720 (0.0425)
Nitrogen	100	125	150	200
Skewness	-0.1741 (0.0417)	-0.2146 (0.0422)	-0.2100 (0.0422)	-0.2199 (0.0431)
Nitrogen	250	300		
Skewness	-0.1933 (0.0444)	-0.2218 (0.0406)		

Dataset C				
Nitrogen	0	50	100	150
Skewness	0.0592 (0.0697)	-0.1728 (0.0504)	-0.1464 (0.0437)	-0.1529 (0.0440)

Posterior Predictive Yield Distributions for nitrogen levels (N) of 0, 80, 160, 240 lb./ac.



References

Gelman, A., J.B. Carlin, H.S. Stern, & D.B. Rubin. 2004. Bayesian Data Analysis. 2nd edition. Chapman & Hall/CRC.

Just, R. & R.D. Pope. 1978. Stochastic Specification of Production Function and Economic Implication. *Journal of Econometrics* 7:67-86.

Nelson, C.H. and P.V. Preckel. 1989. The Conditional Beta Distribution As a Stochastic Production Function. *American Journal of Agricultural Economics* 71: 370-378.