

Modeling Multivariate Crop Yield Densities with Frequent Extreme Events

Shu-Ling Chen

*Graduate Research Associate, Department of Agricultural, Environmental and
Development Economics. The Ohio State University*

and

Mario J. Miranda

*Professor and Chair of Graduate Studies, Department of Agricultural, Environmental
and Development Economics. The Ohio State University*

*Selected Paper prepared for presentation at the American Agricultural Economics
Association Annual Meeting, Denver, Colorado, August 1-4, 2004*

*Copyright 2004 by Shu-Ling Chen and Mario J. Miranda. All rights reserved. Readers
may make verbatim copies of this document for non-commercial purposes by any means,
provided that this copyright notice appears on all such copies.*

Modeling Multivariate Crop Yield Densities with Frequent Extreme Events

Shu-Ling Chen

and

Mario J. Miranda

Abstract

Measuring the lower tail of a crop yield distribution is important for managing agricultural production risk and rating crop insurance. Common parametric techniques encounter difficulties when attempting to model extreme yield events. We evaluate and compare alternative models based on our candidate distributions for high risk counties.

Introduction

Modeling of yield distributions continues to receive much attention in the crop insurance and agricultural risk management literature. The importance of properly modeling yield distributions stems in part from the growing number of public and private crop yield and revenue insurance products that have been introduced in recent years under the Federal Crop Insurance program. Accurate assessment of yield distributions, particularly their lower tails, is necessary for precise computation of crop insurance premium rates. Inaccurate rates can lead to adverse selection problems and poor actuarial performance of the crop insurance program. Accurate assessment of co-variance structures is also paramount for precise rating of the federal government's Standard Reinsurance Agreement, which provides reinsurance for a variety of crop insurance products across the USA.

Which statistical distribution can best explain the behavior of yields remains an unsettled question. An extensive literature has highlighted the challenges associated with modeling yields for the rating of crop insurance. Several well-known parametric models, such as the beta distribution (e.g. Nelson and Preckel) or the lognormal distribution (e.g. Goowin, Roberts and Coble), are widely applied in this research area. However, Goodwin and Ker, and Ker and Coble argue that beta distribution for crop yields are inadequate and propose, instead, nonparametric and semi-parametric methods. Just and Weninger argue that the normal distribution, which has lost favor in recent years for modeling yields, remains a reasonable candidate for yield densities because of misspecification and data limitation problems. Other yield distributions have been proposed in the empirical literature for modeling yields, including the Weibull

distribution, variants of the Burr distribution and the standard nonparametric kernel method.

The various distributions that have been proposed each have their own relative merits for the modeling of yields. Most of the candidate distributions are continuous in nature and thus encounter problems when attempting to model yield distributions for high risk farms or counties that may experience complete crop failure, implying that probability mass is stacked on zero. Unbounded distributions, such as the normal distribution, imply the potential for negative yield realizations. Bounded distributions, such as the beta, also have problems in representing very high risk distributions. The beta distribution, for example, tends to flatten and then take on unreasonable “U-shapes” when the variance of the distribution rises. Such shapes for yield distributions are not in accordance with our agronomic expectations. Other problems also exist for many parametric distributions commonly used to model yield risks. For example, some distributions are undefined for certain values of the distribution parameters. This is true for the normal and Burr distributions.

In this paper, we compare and contrast various candidate distributions for the modeling of crop yields, particularly with regard to their ability to predict extreme-events in agricultural regions where complete crop failures are relatively common. We focus on models based on Burr, beta and nonparametric distributions, with applications to selected counties in Texas where extremely low yields occurs with more frequency than is experienced in other parts of the country.

Modeling Crop Yields Distribution

A. The Univariate Case

Our goal is to estimate conditional yield densities in circumstances in which extreme events, such as complete crop failures, are relatively common. We derive maximum likelihood estimates of the parameters of alternative candidate distributions for county level yield data and evaluate our results in terms of the credibility of certain distributions. Two measures of goodness of fit, Chi-Square test and Anderson-Darling (AD) test, are employed.

We employ a two-step estimation process in the modeling yields. First, in order to control for technical progress in crop production, trend yields are estimated using ordinary least-squares assuming that trend yields follow a second-order polynomial in time. In particular, we assume that

$$y = \delta_0 + \delta_1 t + \delta_2 t^2 + \varepsilon,$$

where y is the yield per planted acre, t denotes time, and ε is error term. All yields are then converted to 1982 equivalents. Alternative parametric distributions are used to explain variations in detrended yields. The following distributional forms are considered:

Normal distribution

The probability density function of X is

$$p_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (\sigma > 0)$$

where μ is mean parameter, and σ is scale parameter.

Weibull distribution

The unconditional probability density function of X is

$$p_x(x) = c\sigma^{-1}\{(x-\theta)/\sigma\}^{c-1} \exp[-\{(x-\theta)/\sigma\}^c], \quad (\theta < x)$$

where θ is known lower threshold, σ is scale parameter, and c is shape parameter.

Beta distribution

The unconditional probability density function of X is

$$p_x(x) = \frac{1}{B(\alpha, \beta)} \frac{(x-\theta)^{\alpha-1} (\theta + \sigma - x)^{\beta-1}}{\sigma^{\alpha+\beta-1}} \quad \theta < x < \theta + \sigma$$
$$= 0 \quad x \geq \theta + \sigma \text{ or } x \leq \theta$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

where θ is known lower threshold parameter, σ is scale parameter ($\sigma > 0$), α is the first shape parameter ($\alpha > 0$), and β is the second shape parameter ($\beta > 0$).

Burr distribution

The cumulative density function is

$$F(x) = 1 - \frac{1}{(1+x^c)^k} \quad x \geq 0$$
$$= 0 \quad x < 0$$

where $c, k \geq 1$ are real numbers.

The probability density function

$$\begin{aligned}
 F'(x) &= f(x) \\
 &= \frac{kcx^{c-1}}{(1+x^c)^{k+1}}
 \end{aligned}$$

is unimodal at $x = \left(\frac{c-1}{ck+1}\right)^{1/c}$ if $c > 1$, and L-shaped if $c = 1$.

Standard nonparametric kernel methods

The kernel density estimator places a bump or individual kernel at each sample realization from the density of interest. The estimate of the density at any given point in the support is simply the sum of the individual kernels at that point.

The kernel estimate of a density function can be represented as a convolution of the sample distribution function with the chosen kernel and thus

$$\hat{f}(x) = \int K_h(x-u) dF_n(u)$$

where h is the bandwidth or smoothing parameter, $K_h(u) = 1/h K(u/h)$, K is the kernel function, and $F_n(u)$ is the sample distribution function. K is assumed to be a square integrable symmetric probability density function with a finite second moment and compact support. Denoting $\mu_2(K) = \int u^2 K(u) du$ and $R(K) = \int K(u)^2 du$ while letting f be the unknown density of interest, standard properties for second order kernels are

$$\begin{aligned}
 E\hat{f}(x) - f(x) &= \int K(u)[f(x-hu) - f(x)] du \\
 &= 1/2 h^2 \mu_2(K) f''(x) + O(h^4) \\
 Var[\hat{f}(x)] &= (nh)^{-1} f(x) R(K) + o[(nh)^{-1}],
 \end{aligned}$$

and thus

$$\begin{aligned}
MSE\{\hat{f}(x)\} &= (nh)^{-1} f(x)R(K) + 1/4 h^4 [\mu_2(K)]^2 [f''(x)]^2 + o[(nh)^{-1} + h^4] \\
MISE\{\hat{f}\} &= (nh)^{-1} R(K) + 1/4 h^4 [\mu_2(K)]^2 R[f''(x)] + o[(nh)^{-1} + h^4].
\end{aligned}$$

B. Multivariate Case: Modeling of the Dependence Structure

The problem that we encounter in modeling multivariate crop yield distributions is that crop yield may have different suitable distribution for each county. As argued by Barry and Ker:

“strong spatial dependence, an empirical stylized fact, negates appealing to central limit theorems (CLTs) for dependent processes when considering mean yields. These theorems require that spatial dependence dies off at a sufficiently quick rate or that spatial dependence disappears after some finite distance. While this is certainly true for yield data, it is almost never true for the spatial region”

Thus, to calculate correlation of crop yield failures across counties, a flexible density function is required. We will posit that the multi-dimensional vector of yield random variables denoted $Y = \{y_1, y_2, \dots, y_k\}$ possesses correlation matrix of Y is a $k \times k$ matrix, V , which is defined as

$$V = \begin{bmatrix} 1 & \rho_{12} & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{21} & 1 & \cdots & & \rho_{2k} \\ \vdots & & \ddots & & \vdots \\ \rho_{k1} & \cdots & & \rho_{kk-1} & 1 \end{bmatrix}.$$

Data

Our analysis utilizes NASS county-level yield data collected from the late 1950s for major U.S. field crops: corn, soybeans, wheat and cotton. Let us take upland cotton as example. The major state produced upland cotton in U.S. is Texas, which accounts for 23.74% of the total planted yield during the 1956-1997. The existence of extreme event yield is defined if the realized yields fall below 60% coverage level of predicted planted crop yield. Most counties with high percentage of extreme-event yield are located in the South Texas district. There are 15 counties included in this district. To avoid the inefficient statistical issue in the paper, all counties with time length less than 30 years located in the South Texas district are dropped out. Only 8 counties, including Brooks, Dimmit, Duval, Frio, Jim Hogg, Jim Wells, Live Oak and Zavala, are studied.

Empirical Evidence

The sample moments of upland cotton planted yield in Table 1 shows that all county-level yields exhibit positive skewness in the South Texas District of Texas, except Frio County. The coefficient of skewness varies from -0.021 to 1.815, which is consistent with our argument about extreme-event yield county. When a county has high percentage of extreme-event yield, mass yield tends to stack on left tail of the yield density. In the previous literature, county-level crop yield are considered to be “fatter tailed” than standard normal distribution. However, our data summary indicates that the coefficients of kurtosis among all extreme-event yield counties except Duval and Live Oak County are near zero or even negative. For most non-Gaussian random variable, the coefficient of kurtosis is nonzero. In the statistical literature, a random variable that has

negative kurtosis is called sub-Gaussian and that has positive kurtosis is referred as super-Gaussian. Sub-Gaussian typically specifies a “flat” or “less peaked” probability density function. A reasonable explanation for the contradictory may be addressed by the feature of extreme-event yield.

The goodness-of-fit tests for alternative distributions are provided in Table 2. Surprisingly, Anderson-Darling test rejects the distributions more frequently than Chi-Square test, especially for the beta distribution. The upland cotton planted yield beta distribution for all counties but Dimmit County are rejected by Anderson-Darling test at 5% significant level while only 3 counties’ yield beta distributions are rejected by Chi-square test. In the panel A and B of Table 2, the Weibull distribution fits the data best compared to the normal distribution and the beta distribution. The Weibull yield distribution is rejected for only one county, Live Oak County, which may be explained by its discontinuous yield data. Also, the test results point out that it is more likely to reject the normal distribution as a county has relatively high percentage of extreme-event planted yield. That demonstrates the previous finding of positive skewness of yields. It may be appropriate to argue that the percentage of extreme-events in planted yield play an essential role in testing the goodness-of-fit for the normal distribution.

Figure 1-3 illustrate all county-level extreme-event upland cotton yield densities. The histogram of yield data exemplifies that high percentage of extreme-event yield has relatively higher amount yield stack on left tail of the density function. The larger percentage of extreme-event yield, the higher positive value of skewness of yield density.

Table 1: Summary Statistics of Upland Cotton Planted Yield: South Texas District, Texas, 1956-1997

County	Code	Observation	Mean	Std Dev	Skewness	Kurtosis
Brooks	047	42	524.76	299.65	0.656	- 0.527
Dimmit	127	39	542.52	215.87	0.578	0.513
Duval	131	42	338.29	249.83	1.815	3.873
Frio	163	40	501.35	171.69	- 0.021	- 0.467
Jim Hogg	247	39	520.86	286.36	0.722	- 0.366
Jim Wells	249	42	306.31	109.79	0.323	0.819
Live Oak	297	42	345.68	185.89	1.456	2.901
Zavala	507	42	671.93	158.58	0.427	- 0.420

Table 2: Goodness-of-Fit Tests of Candidate Distributions: Upland Cotton Planted Yield, 1956-1997

<i>Panel A.</i>		Normal		Weibull		Beta	
County	% of extreme-event planted yield	Chi-Sq	P-value	Chi-Sq	P-value	Chi-Sq	P-value
Brooks	35.71	9.849	0.020	4.363	0.225	6.954	0.073
Dimmit	15.38	7.169	0.127	6.091	0.192	8.101	0.088
Duval	47.62	38.505	< 0.001	8.349	0.080	17.701	0.001
Frio	20.00	1.816	0.611	1.742	0.628	4.766	0.190
Jim Hogg	43.59	3.659	0.301	1.422	0.700	4.043	0.257
Jim Wells	16.67	6.545	0.088	6.312	0.097	9.861	0.020
Live Oak	30.95	11.952	0.008	7.813	0.050	11.304	0.010
Zavala	2.38	1.528	0.676	2.091	0.554	4.669	0.198

<i>Panel B.</i>		Normal		Weibull		Beta	
County	% of extreme-event planted yield	AD	P-value	AD	P-value	AD	P-value
Brooks	35.71	0.858	0.025	0.315	> 0.250	1.140	0.008
Dimmit	15.38	0.316	> 0.250	0.270	> 0.250	0.471	> 0.250
Duval	47.62	1.995	< 0.005	0.721	0.056	2.755	< 0.001
Frio	20.00	0.236	> 0.250	0.250	> 0.250	1.134	0.008
Jim Hogg	43.59	0.963	0.015	0.516	0.197	1.404	0.002
Jim Wells	16.67	0.376	> 0.250	0.447	> 0.250	1.256	0.003
Live Oak	30.95	1.302	< 0.005	0.788	0.038	2.307	< 0.001
Zavala	2.38	0.306	> 0.250	0.443	> 0.250	1.115	0.007

IV. Conclusion

Although modeling yield densities has been a popular subject in crop insurance, the issue of extreme-events has not been extensively explored in the agricultural economics literature. The contribution of this paper is to compare the performance of several popular yield densities in circumstances in which extremely yield events are relatively common and to provide a method to formally test whether these correlations deviate from what would be expected under multivariate extreme-event crop yield densities. Moreover, implications for the modeling of yield risks, including the rating of crop insurance contracts, will be offered.

Figure 1. County-level Upland Cotton Normal Yield Densities

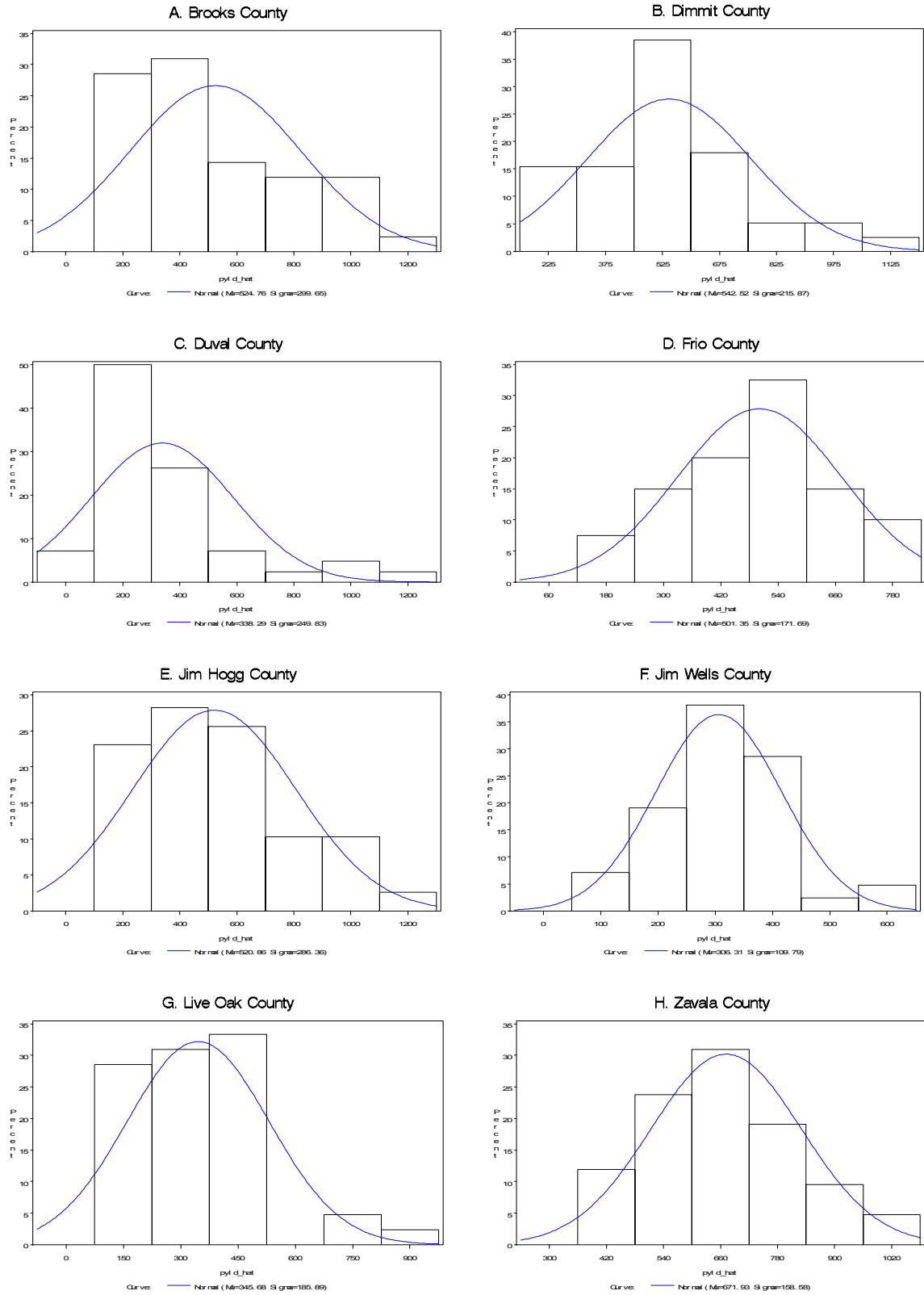


Figure 2. County-level Upland Cotton Weibull Yield Densities

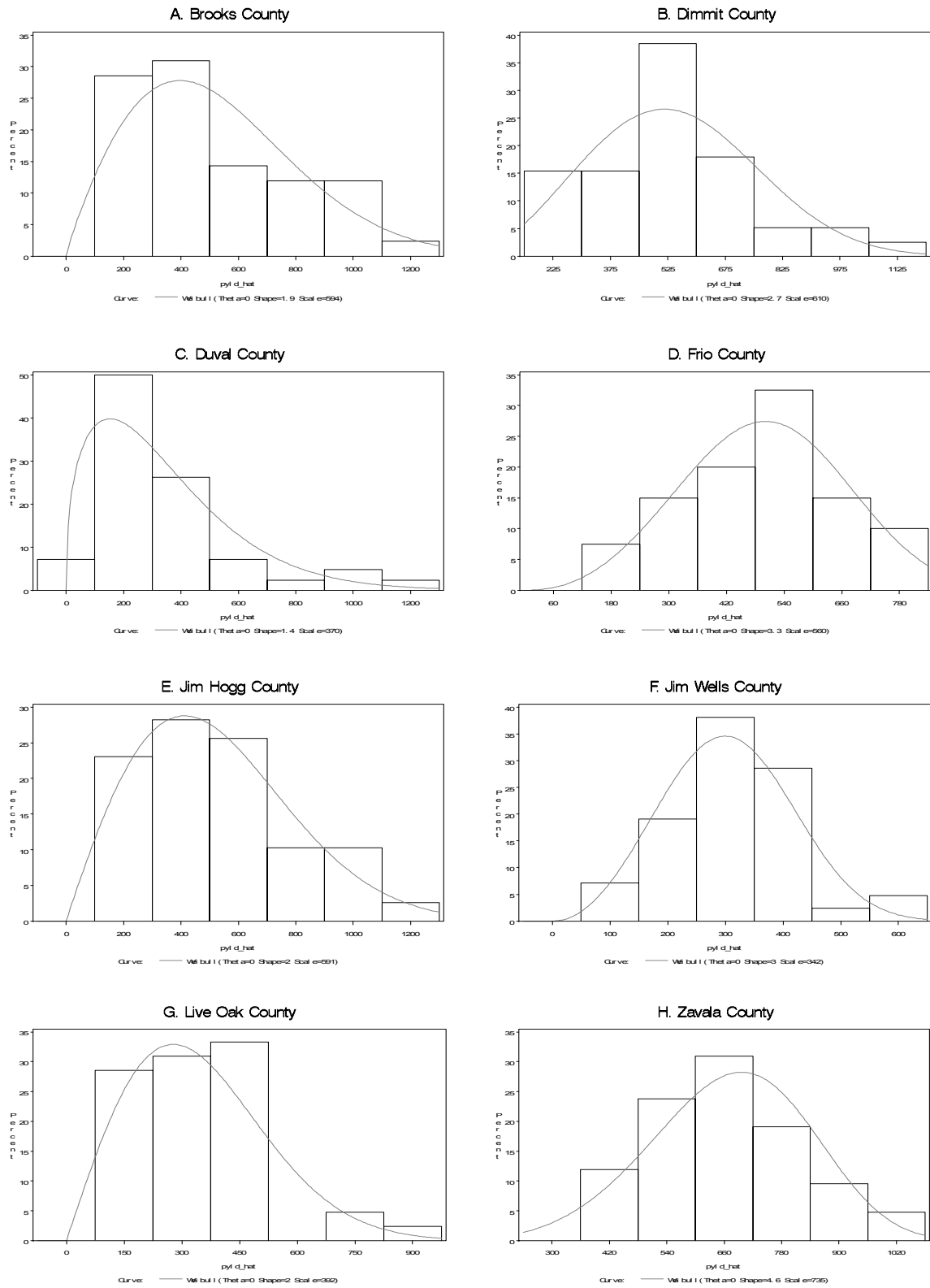
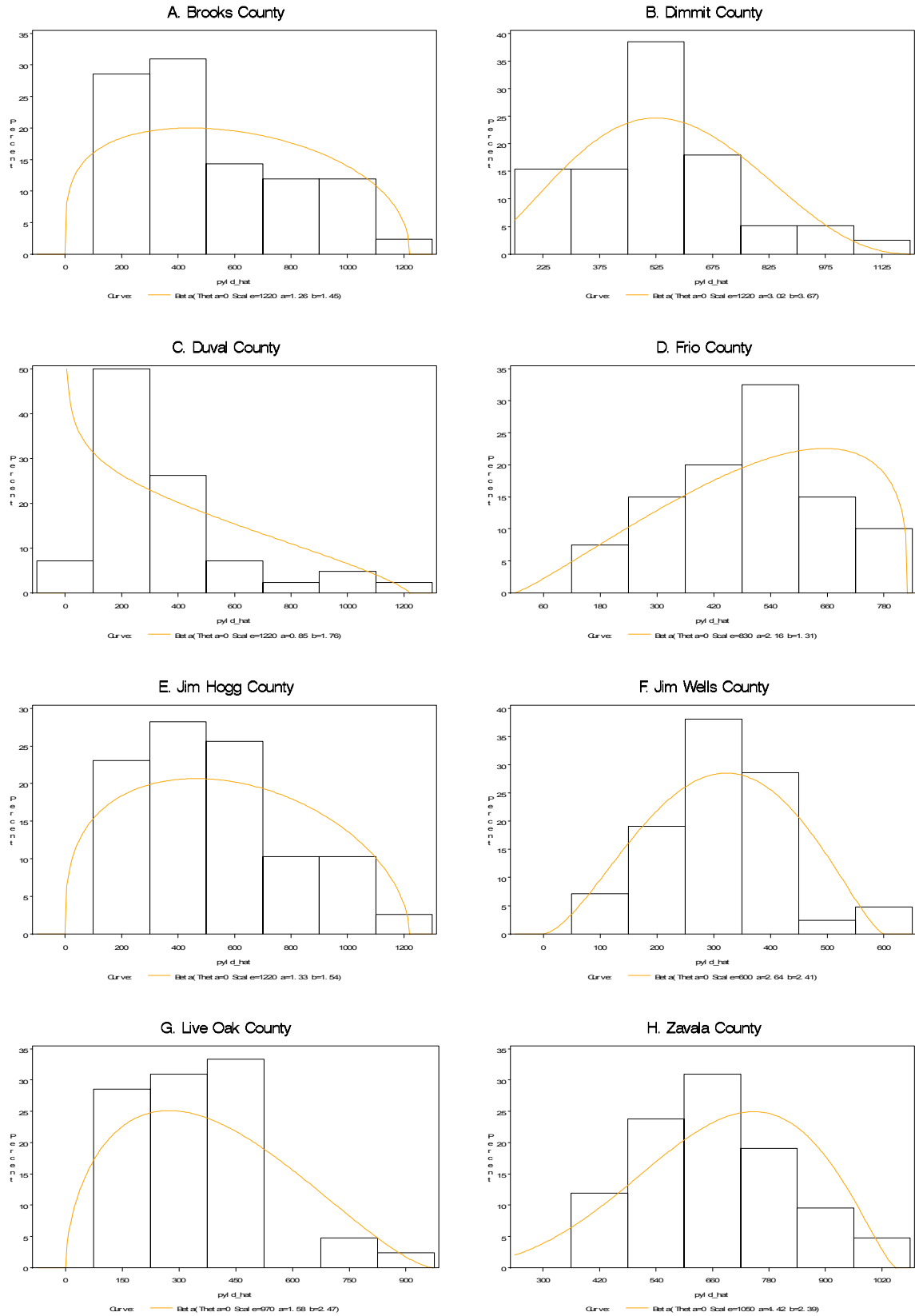


Figure 3. County-level Upland Cotton Beta Yield Densities



Reference

- Barry K. Goodwin. "Problems with Market Insurance in Agriculture." *Amer. J. Agr. Econ.* 83 (August 2001): 643-649.
- Barry K. Goodwin, and Alan P. Ker. "Modeling Price and Yield Risk." *A Comprehensive Assessment of the Role of Risk in U.S. Agriculture*. Kluwer Academic Publisher. 2002.
- Duncan, John, and Robert J. Myers. "Crop Insurance under Catastrophic Risk." *Amer. J. Agr. Econ.* 82 (November 2000): 842-855.
- Hartmann, P., S. Straetmans, and C. G. De Vries. "Market Linkages in Crisis Periods." *ECB. Working Paper.* 71 (July 2001)
- Johnson, Norman L. and Samuel Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*. New York: John Wiley & Sons, Inc. 1972.
- Longin, François M. "Value at Risk and Extreme Values." *IFAC Computation in Economics, Finance and Engineering: Economic Systems*. Cambridge, UK (1998): 45-49.
- Longin, François, and Bruno Solnik. "Extreme Correlation of International Equity Markets." *The Journal of Finance*. (April 2001): 649-676.
- Mahul, Olivier, and Brian D. Wright. "Designing Optimal Crop Revenue Insurance." *Amer. J. Agr. Econ.* 85 (August 2003): 580-589.
- Mahul, Olivier. "Managing Catastrophic Risk through Insurance and Securitization." *Amer. J. Agr. Econ.* 83 (August 2001): 656-661.
- Miranda, Mario, and Dmitry V. Vedenov. "Innovations in Agricultural and Natural Disaster Insurance." *Amer. J. Agr. Econ.* 83 (August 2001): 650-655.
- Moss, C. B., and J. S. Shonkwiler. "Estimation Yield Distributions with a Stochastic Trend and Nonnormal Errors." *Amer. J. Agr. Econ.* 75 (November 1993): 1056-62.
- Ramaswami, Bharat, and Terry L. Roe. "Aggregation in Area-Yield Crop Insurance: The Linear Additive Model." *Amer. J. Agr. Econ.* 86 (May 2004): 420-431.
- Sherrick, Bruce J., Fabio C. Zanini, Gary D. Schnitkey, and Scott H. Irwin. "Crop Insurance Valuation Under Alternative Yield Distributions." *Amer. J. Agr. Econ.* 86 (May 2004): 406-419.