Journal of Quantitative Economics, Vol. 8 No.1, January 2010

NORMALITY OF INDIAN CROP YIELDS: APPLICATION OF PANEL ANALYSIS

SALEEM SHAIK[1]

Abstract

This paper has two-fold contribution. First, normality of the crop yield residuals are examined using panel statistical procedures accounting for trend, autocorrelation and heteroskedasticity. Second, to evaluate the importance of accounting for spatial and temporal variation on the normality of crop yield residuals, the changes in the skewness, kurtosis and D'Agostino-Pearson (K^2) *omnibus normality test across panel and time-series models are examined. These are examined with an empirical application to district level data from 1950-2002 for 15 crops and 14 states in India composed of a total of 3143 individual reporting districts in India. Results indicate crop yield residuals were normally distributed in 65 percent and 73 percent of districts, respectively based on skewness and kurtosis statistics. Accounting for spatial and temporal variation seems to change the distribution of crop yield residuals in 20 percent, 14 percent and 17 percent of districts based on skewness, kurtosis and omnibus tests respectively.*

Keywords: Crop yield normality; Skewness, Kurtosis, and Omnibus Test; Indian crop yields; District level data from 1956-2002; panel and time-series procedures.

JEL Classification: C23, Q1, G22, O53

1. Introduction

The current focus in Indian crop insurance is on the development of new products related to weather derivatives along with or in competition with conventional yield insurance products. An essential component of any new product or improvement in existing products is to understand, estimate and identify the distributions of yield, revenue, or loss-cost for crop insurance, the distributions of temperature and precipitation for weather derivatives, and finally the interaction of yield/revenue/loss-cost and temperature/precipitation distributions. Recognizing the importance of identifying distributions, many authors have estimated and examined the normality of crop yield residuals using short and/or long term data. This is important for public and private insurance companies and risk management specialists as it provides information to develop actuarially sound premium rates, to differentiate low versus high risk farms, districts or states, and avoids asymmetric issues like adverse selection, moral hazard and fraud.

[1] 504 Richard Barry Hall, Dept of Agribusiness and Applied Economics, NDSU, Fargo, ND 58108-6050, E-mail: saleem.shaik@ndsu.edu

Several problems are commonly encountered when attempting to estimate the distributions using farm level or aggregate crop yield data. With farm level data a major issue is N (number of cross-sections) is greater than T (time-series) and correlated yields due to periodic area-wide events. Use of longer-term aggregate data is complicated due to technological, yield trends and spatial variability. It is more likely the longer-term yield data would accurately reflect the likelihood of temporal-spatial random and systematic events. Previous studies identified evolving technological by modeling the distributions accounting for trends, heteroskedasticity or autocorrelation using time-series techniques.Recent literature focusing on crop yield distributionhave tested for normality using disaggregate farm-level and aggregate county-level data from the U.S (Day, 1965; Gallagher, 1986 and 1987). There is hardly any research examining the normality assumption of developing countries crop yields.

Even with available cross-section time-series data, panel statistical procedures were seldom used in estimation of the crop yield distributions. Panel statistical procedures have several advantages over conventional cross-section or time-series statistical methods (see Hsiao, 2000). These advantage include reduction in collinearity among exogenous variables, allow more complicated models that are not possible in cross-sectional or time-series data, parsimonious, and finally accounting for temporal and spatial variation will lead to remaining errors or residuals used to examine normality that are truly random compared to the traditional time-series procedures.

Given these advantages, first there is need to revisit and test normality of crop yields accounting for temporal and spatial variation using panel techniques.Second, changes in the normality results of crop yield residuals estimated from panel and time-series statistical procedures are quantified and categorized into four groups. If the crop yield residuals are normal in both panel and time-series model it is categorized as normal to normal (N-N). If residuals are non-normal in both models it is categorized as non-normal to non-normal (NN-NN). The remaining two groups of interest include changes from normal to non-normal (N-NN) and non-normal to normal (NN-N). The last two categories are of importance as they would indicate the extent of divergence in the normality results.

In this paper, normality of the crop yield residuals are examined for food grains (rice and wheat), millets (bajra, barley, jowar, maize and ragi), pulses (gram and tur), oilseeds (groundnut, rape and mustard seed, sesamum, soybean and sunflower) and fiber (cotton) crops using panel statistical procedures. The next section develops panel statistical procedures to model crop yield distributions for de-trending, autocorrelation and heteroskedasticity correction as well as skewness, kurtosis, and omnibus normality tests. In the third section an empirical application to district data from 14 states over the period 1956 to 2002 examinesnormality of crop yield residuals. Changes in the normality results of crop yield residuals estimated from panel and time-series statistical techniques are discussed in this section. Finally, I conclude with future research issues.

2. Modeling Crop Yields

Time series procedures to model crop yield distributions useparametric, non-parametric and semi-parametric methods for short and/or long term data are available in the existing literature (Atwood et al, 2002 and 2003). Previous procedures proposed de-trending crop yields by identifying and removing any systematic variation by estimating the degree of polynomial trends, correcting for autocorrelation, heteroskedasticity and use the remaining random errors to conduct the skewness, kurtosis, and the omnibus normality tests. These procedures could be easily extended to panel statistical techniques that account for spatial and temporal variation leading to errors or residuals that are truly random to test for normality.

Panel Statistical Procedures

Statistical procedures to estimate and test crop yield residuals for normality involve detrending and accounting for autocorrelation and heteroskedasticity. The time-series model can be represented as:

$$Y = \alpha + \beta^{\gamma}\tau^{\gamma} + \varepsilon \qquad \text{… (1)}$$

where Y represents a 1×NT matrix; τ represents a K×NT matrix of exogenous time trend variables with N and T representing the spatial (cross-section) and temporal (time series) dimension; α is the intercept, β^{γ} is the associated parameters of linear (τ), quadratic (τ^2) and cubic (τ^3) time trend variables to be estimated for each cross-section (districts in this case); and ε represents a 1×NT matrix of pure random error.

Here the focus is on the additive errors of the one-way random effects model as it allows the estimation, testing the degree of polynomial trend and accounts for spatial and temporal variation. Second, one-way was chosen over two-way random effects model as two-way random effects model cannot be estimated due to the use of time trend polynomials as exogenous. This can be represented in vector form as

$$\varepsilon = Z_u u + Z_w w \qquad \text{… (2)}$$

where $\begin{matrix} Z_u = (I_N \otimes \iota_T); & u' = (u_1, u_2, \ldots\ldots, u_N) \\ Z_w = (I_N \otimes I_T); & w' = (w_1, w_2, \ldots\ldots, w_{NT}) \end{matrix}$, and

I_N and I_T (ι_N and ι_T) represent an identity matrix (vector of one's) of dimensions N and T respectively; and represent the random error components with zero means and covariance matrix:

$$E\begin{pmatrix} u \\ w \end{pmatrix}(u'\ w') = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_w^2 \end{pmatrix} \qquad \text{… (3)}$$

Equation (1) along with equation (2) and (3) can be written as one-way random effects panel model:

$$Y = \alpha + \beta^{\gamma}\mathbf{T}^{\gamma} + u + w \qquad \text{… (4)}$$

where u represents a 1×*N* matrix of temporally invariant spatial random error, and w represents a *1×NT* matrix of pure random error; and α is the intercept, β^{γ} is the associated parameters of time trend variables to be estimated for each state.

Since each panel consists of all districts within a state, the degree of polynomial trend is estimated for each state using equation (4). The degree of polynomial for each state is determined based on the F-test at the 5% significance level. Once the degree of polynomial is estimated, the panel statistical procedure with pre-determined polynomial trend is estimated and the residuals from the model are tested for autocorrelation and heteroskedasticity. If the residuals from the panel model are found to be autocorrelated, a first order autocorrelation structure is imposed on the error, w or $w_{i,t}$ as

$$w_{i,t} = \rho w_{i,t-1} + \varepsilon_{i,t} \qquad \text{… (5)}$$

and the model is re-estimated with the appropriate degree of polynomial.

Next, the square of the error (w) from equation 4 with first order autocorrelation structure imposed (equation 5) is regressed on the time polynomial to check for the presence of heteroskedasticity. This can be represented as:

$$w^2 = \alpha + \beta^{\gamma} \mathbf{T}^{\gamma} + u + \omega \qquad \text{… (6)}$$

where w^2 represents a *1×NT* matrix; $\mathbf{T}$ represents a *K×NT* matrix of exogenous time trend variables with *N* and *T* representing the spatial (cross-section) and temporal (time series) dimension; α' is the intercept, β^{γ} is the associated parameters of linear (τ), quadratic (τ^2) and cubic (τ^3) time trend variables to be estimated for each state; where u represents a 1 × *N* matrix of temporally invariant spatial random error, and ω represents a *1×NT* matrix of pure random error.

The residuals or error terms are made variance stationary in the presence of heteroskedasticity by dividing the residuals, w by the standard deviation of the predicted residuals, $\hat{\sigma}_w$. This can be represented as:

$$\hat{\hat{w}} = \frac{w}{\hat{\sigma}_w} \qquad \text{… (7)}$$

Once the residuals are corrected for trend, autocorrelation and heteroskedasticity, the skewness, kurtosis and omnibus tests are conducted to examine the normality of crop yields.

Source of difference between panel and time-series statistical procedures

In addition to the rational provided above for the use of panel statistical procedures, the main difference is the accounting of spatial errors in panel procedures. To illustrate the distinction, let us take the difference between panel model, Y^{Panel} defined in equation 4 and the time-series model, $Y^{\text{Time-series}}$ defined in equation 1. This can be represented as:

$$\begin{aligned} Y^{\text{Panel}} - Y^{\text{Time-series}} &= \left(\alpha + \beta^{\gamma}\mathbf{T}^{\gamma} + u + w\right) - \left(\alpha + \beta^{\gamma}\mathbf{T}^{\gamma} + \varepsilon\right) \\ &= \left(\alpha + \beta^{\gamma}\mathbf{T}^{\gamma}\right) - \left(\alpha + \beta^{\gamma}\mathbf{T}^{\gamma}\right) \quad \text{......systematic component} \\ &\quad \text{and} \\ &= (u + w) - (\varepsilon) \quad \text{......random component} \end{aligned} \qquad \text{… (8)}$$

The difference between the models stems from two sources. The first difference is reflected in the systematic component and echoed in the degree of polynomial and heteroskedasticity trend estimated by the time-series and panel models. Unlike panel model, the time series model does not take into account information across cross-sections within a state and over time in the estimation of systematic component, $\left(\alpha+\beta^{\gamma}T^{\gamma}\right)$. The panel model estimates a common degree of polynomial or heteroskedasticity trend for all the cross-sections within a state. In contrast the time series model estimates the degree of polynomial or heteroskedasticity trend for each cross-section within a state. The second difference is the random component, i.e., the errors or residuals used to test for normality of crop yield residuals. Panel model accounts for temporally invariant spatial errors (u) and the remaining error or residual, w is used to test for normality of crop yields. Unlike panel model, the time series model does not take into account common variation across cross-sections within a state in the estimation of systematic components leading to the random component that is totally driven by individual cross-section variation over time.

These systematic and random component differences across the panel and time-series statistical procedures are quantified in the form of degree of polynomial and heteroskedasticity trend and its implications on the normality tests and changes in the normality results between the two models.

Skewness, kurtosis and omnibus tests

The normalized residuals from equation (7) are tested for normality following D'Agostino, Belanger, and D'Agostino, Jr. (1990) two directional statistics. These two directional statistics for skewness $(\sqrt{\beta_1})$ and kurtosis (β_2) are given by

$$\sqrt{\beta_1}=\frac{E(X-\mu)^3}{\sigma^3} \quad \text{and} \quad \beta_2=\frac{E(X-\mu)^4}{\sigma^4} \qquad \dots (8)$$

where E, μ and σ are the expected value operators, mean, and standard deviation, respectively of normalized residuals. Negative values for skewness indicate data are skewed to the left. Positive values for skewness indicate data are skewed to the right. The kurtosis for a standard normal distribution is three and for this reason kurtosis is defined as $\left(\beta_2-3\right)$. A positive value for kurtosis indicates a peaked distribution, i.e., leptokurtic or thick tailed $\left((\beta_2-3)>0\right)$ and a negative value of kurtosis indicates a flat distribution, i.e., platykurtic or thin tailed $\left((\beta_2-3)<0\right)$.

The D'Agostino-Pearson (K^2) omnibus normality test is based on the joint D'Agostino skewness test $(\sqrt{\beta_1})$ and Anscombe-Glynn kurtosis test (β_2) moment and represented as

$$K^2=Z^2(\sqrt{\beta_1})+Z^2(\beta_2) \qquad \dots (9)$$

Where $Z(\sqrt{\beta_1})$ and $Z(\beta_2)$ are the standard normal deviates equivalent to observing $(\sqrt{\beta_1})$ and (β_2) statistics (Armitage and Colton, 1998). The (K^2) statistic has approximately a Chi-squared distribution with two degrees of freedom when the population is normally distributed.

3. Indian District Data

The district is the smallest administrative unit in India for which data on crops are available. This study covers a total of 3,143 districts across 15 crops and 14 states in India for the period 1956-2002.This is a unique data set as it provides historical information on food grains (rice and wheat), millets (bajra, barley, jowar, maize and ragi), pulses (gram and tur), oilseeds (groundnut, rape and mustard seed, sesamum, soybean and sunflower) and fiber (cotton) district level yields that has never been collected and used in the estimation. District level data is available from four publications - Area and Production of Principal Crops in India; Agricultural Situation in India; Statistical Abstracts of India; and Crop and Season Reports of the various States.

4. Empirical Application and Normality Test Results

Normality assumption of the crop yields is examined for food grains (rice and wheat), millets (bajra, barley, jowar, maize and ragi), pulses (gram and tur), oilseeds (groundnut, rape and mustard seed, sesamum, soybean and sunflower) and fiber (cotton) crops using district data from 14 states over the period, 1956-2002.

The degree of polynomial estimated for trend and heteroskedasticity correction is presented in Table 1. The percentage of districts in which crop yield residuals are normally distributed based on skewness, kurtosis and the D'Agostino-Pearson omnibus test are presented in tables 2 for each crop. Figure 1 and 2 present the histogram of skewness and kurtosis values of each district for each crop. Changes in the normality results are discussed by comparing the skewness, kurtosis and omnibus values of crop yield residuals estimated from panel and time-series statistical procedures. The result of the changes in the normality between panel and time-series model is presented in table 3.

Panel model results from table 1 indicate 19 percent, 28 percent and 44 percent of the districts have a linear, quadratic and cubic trend respectively in their crop yields. The remaining 8 percent indicate no trend in their crop yields. Time series model indicate 35 percent, 26 percent and 11 percent of the districts had a linear, quadratic and cubic trend respectively in their crop yields. For heteroskedasticity, 60 percent and 71 percent of the crop yield residuals indicated homoskedasticity based on panel and time series model respectively. The remaining 40 percent and 39 percent of the districts based on panel and time series model indicated heteroskedasticity as the time trend is significant. This finding suggests the need for "variance-stabilization" of the crop yield residuals to correct for heteroskedasticity. With panel model, if a particular state had a linear trend all the districts in the state were treated as linear trend. This was done so that we could compare between the time series and panel models.

Normality of crop yield residuals is discussed using skewness, kurtosis and the D'Agostino-Pearson omnibus test. Percentages of districts in which crop yield residuals are normally distributed based on tests are presented in table 2 for each crop.

Panel model results from table 2 and figure 1 show crop yield residuals in 65 percent of districts were normally distributed and in the remaining 9 and 26 percent of the districts crop yield residuals were negatively and positively skewed, respectively. Similarly, based on kurtosis statistics and figure 2, crop yield residuals in 73 percent of the districts indicate normality of crop yield residuals and the remaining 27 percent of districts indicate leptokurtic or thick tailed $((\beta_2 - 3) > 0)$ distributions. Based on skewness tests, crop yield residuals in 59 percent, 73 percent, 69 percent, 67 percent and 56 percent of the districts growing food grains (rice and wheat), millets (bajra, barley, jowar, maize and ragi), pulses (gram and tur), oilseeds (groundnut, rape and mustard seed, sesamum, soybean and sunflower) and fiber (cotton) are normally distributed. Kurtosis test results indicate crop yield residuals in 61 percent, 75 percent, 75 percent, 73 percent and 46 percent of food grains (rice and wheat), millets (bajra, barley, jowar, maize and ragi), pulses (gram and tur), oilseeds (groundnut, rape and mustard seed, sesamum, soybean and sunflower) and fiber (cotton) growing districts are normally distributed. The omnibus test suggest crop yield residuals in 53 percent, 70 percent, 68 percent, 66 percent and 38 percent of food grains (rice and wheat), millets (bajra, barley, jowar, maize and ragi), pulses (gram and tur), oilseeds (groundnut, rape and mustard seed, sesamum, soybean and sunflower) and fiber (cotton) growing districts are normally distributed.

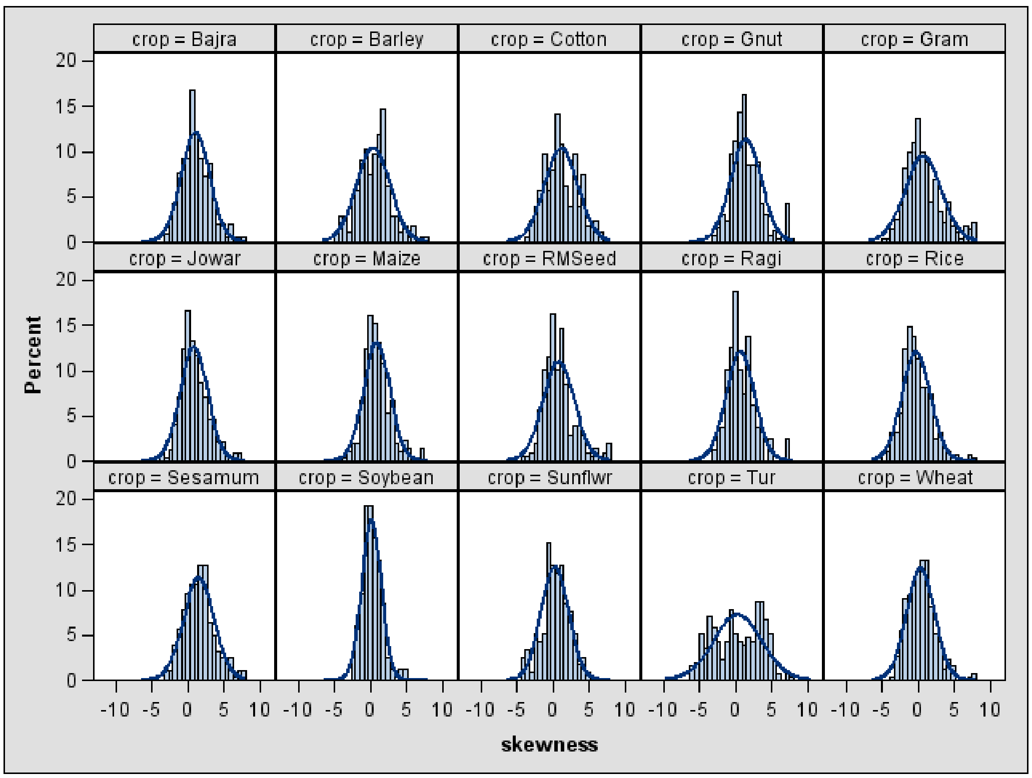

Figure 1. Histogram of district level skewness values for panel model, 1956-2002

Results of time-series model presented in table 2 and figure 1 indicate crop yield residuals in 68 percent of districts were normally distributed and the remaining 9 and 23 percent of districts indicated crop yield residuals were negatively and positively skewed respectively. Similarly, based on kurtosis statistics and figure 2, crop yield residuals in 69 percent of districts indicate normal distribution and the remaining 31 percent of districts indicated leptokurtic or thick tailed $((\beta_2 - 3) > 0)$ distributions. Time-series model results show 3.5 percent of additional districts have realized normally distributed crop yield residuals based on skewness statistics compared to panel model. In contrast kurtosis statistics indicate 1.6 percent of additional districts have realized non-normally distributed crop yield residuals with time-series model. The omnibus test indicated less than 1 percent change in the normality results between the time-series and panel models. Comparison of the normality result at the aggregate level masks the true difference in the changes in normality results across time-series and panel models.

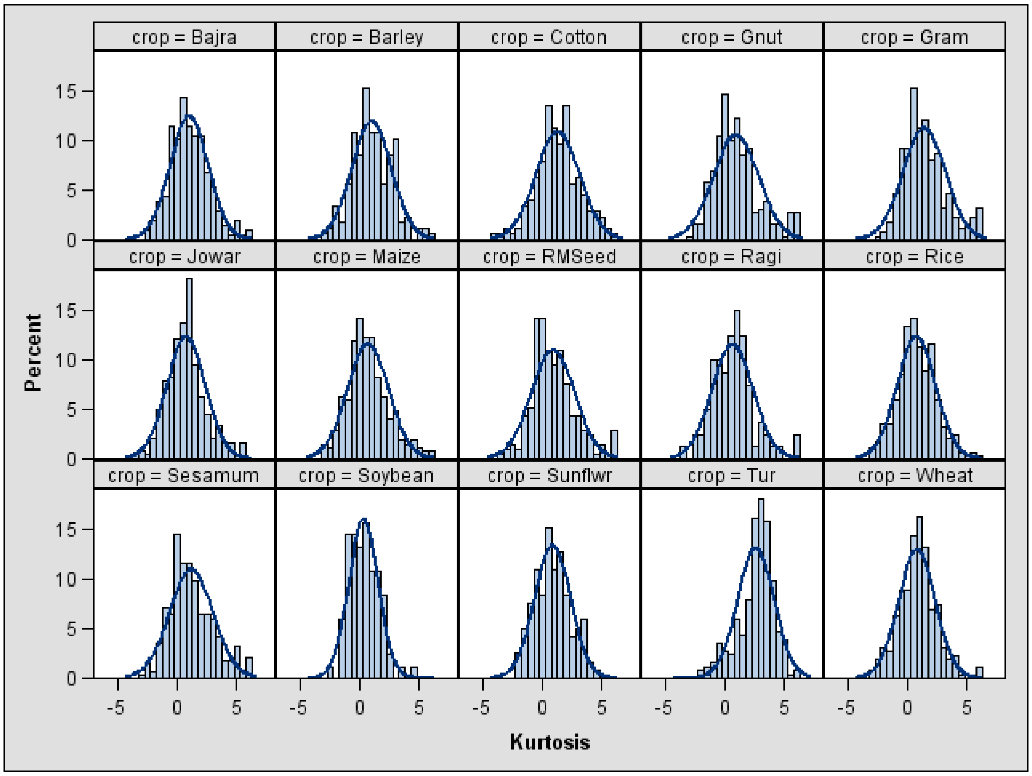

Figure 2. Histogram of district level kurtosis values for panel model, 1956-2002

Changes in the normality results are discussed by comparing the skewness, kurtosis and omnibus values of crop yield residuals estimated from panel and time-series statistical procedures for each observation. If the crop yield residuals are normal in panel and time-series models, yields are categorized as normal to normal (N-N). If the residuals are non-normal in both models it is categorized as non-normal to non-normal (NN-NN). The remaining two groups are of interest as the normality results change from normal to non-normal (N-NN) and from non-normal to normal (NN-N). The percentage of districts in each of the four categories by crop is presented in table 3.

Next, changes in normality results of crop yield residuals estimated from panel and time-series statistical techniques are categorized into four groups - normal in both models (N-N), non-normal in both models (NN-NN), changed from normal to non-normal (N-NN) and changed from non-normal to normal (NN-N).Based on skewness, kurtosis and omnibus tests the percentage of districts in each of the four categories by crop is presented in table 6. Changes in the normality results are reported and discussed by crop based on skewness, kurtosis and omnibus tests.

Table 1. Degree of Polynomial for Panel Model by Crop, 1956-2002

Crop	*Number of States*	*Degree of Polynomial for trend estimation*				*Degree of Polynomial for Heteroskedasticity estimation*			
		Zero	*One*	*Two*	*Three*	*Zero*	*One*	*Two*	*Three*
Time Series Model									
Bajra	12	1	26	127	54	106	80	21	
Barely	9		32	65	79	129	11	36	
Cotton	11	7	31	16	122	84	31	40	21
Gram	14	76	65	54	64	214	18		26
Gram	13	27	14	114	118	201	37	35	
Jowar	12		76	73	92	147	60		34
Maize	13		23	66	179	196	21	33	18
Ragi	8		24	41	15	66	10	4	
Rice	14		44	74	164	163	83	15	21
RMseed	13	3	66	37	104	102	100		7
Sesamum	14	65	69	121	27	201	56	25	
Soybean	6	4	62		17	83			
Tur	8	28	18	16	56	10	107		
Wheat	13	57	43	18	136	46	19	126	63
Panel Model									
Bajra	12	39	96	60	13	143	60	4	1
Barely	9	26	56	52	42	110	51	13	2
Cotton	11	47	55	32	42	102	42	8	24
Gram	14	118	66	65	10	211	37	2	9
Gram	13	67	106	84	16	213	42	11	7
Jowar	12	65	111	44	21	170	58	8	5
Maize	13	77	113	55	23	220	34	6	8
Ragi	8	38	26	12	4	71	3	3	3
Rice	14	42	111	90	39	238	30	5	9
RMseed	13	59	86	44	21	153	37	9	11
Sesamum	14	123	68	66	25	203	60	11	8
Soybean	6	33	43	1	6	71	9	1	2
Tur	8	45	40	25	8	80	27	4	7
Wheat	13	88	43	90	33	55	125	58	16

Table 2. Districts with Normal/non-normal Distributed Crop Yield Residuals across Panel and Time-series Models

State	*Districts*	*Skewness*		*Kurtosis*		*Omnibus*	
		Normal	*Non-Normal*	*Normal*	*Non-Normal*	*Normal*	*Non-Normal*
PANEL							
Bajra	200	127	73	151	49	130	70
Barley	176	115	61	122	54	108	68
Cotton	176	104	72	107	69	93	83
Groundnut	258	163	95	190	68	165	93
Gram	273	172	101	183	90	162	111
Jowar	237	170	67	184	53	167	70
Maize	268	195	73	202	66	191	77
Ragi	78	57	21	60	18	55	23
Rice	282	206	76	207	75	195	87
Rape-Mustard Seed	209	155	54	151	58	145	64
Sesamum	282	167	115	190	92	166	116
Soybean	81	70	11	74	7	71	10
Sunflower	118	80	38	91	27	78	40
Tur	248	72	176	58	190	38	210
Wheat	257	185	72	196	61	182	75
Total	3143	2038	1105	2166	977	1946	1197
TIME SERIES							
Bajra	200	141	59	139	61	127	73
Barley	176	123	53	114	62	110	66
Cotton	176	104	72	101	75	91	85
Groundnut	258	179	79	198	60	176	82
Gram	273	181	92	173	100	163	110
Jowar	237	168	69	182	55	164	73
Maize	268	197	71	204	64	180	88
Ragi	78	59	19	61	17	59	19
Rice	282	188	94	207	75	182	100
Rape-Mustard Seed	209	152	57	151	58	140	69
Sesamum	282	182	100	198	84	173	109
Soybean	81	77	4	73	8	74	7
Sunflower	118	89	29	91	27	87	31
Tur	248	123	125	34	214	40	208
Wheat	257	169	88	188	69	162	95
Total	3143	2132	1011	2114	1029	1928	1215

Table 3. Percentage (%) Changes in the Skewness, Kurtosis and D'Agostino-Pearson (K^2) **Omnibus Normality Test across Panel and Time-series Models**

State	*N-N*	*NN-NN*	*N-NN*	*NN-N*	*N-N*	*NN-NN*	*N-NN*	*NN-N*
	Skewness				*Kurtosis*			
Bajra	60	23	5	12	66	19	10	4
Barley	53	18	13	17	60	26	10	5
Cotton	47	28	13	13	49	31	12	9
Groundnut	62	29	1	7	71	20	3	6
Gram	55	26	8	11	58	28	9	5
Jowar	63	20	9	8	71	16	7	6
Maize	66	20	7	7	70	18	6	6
Ragi	66	16	8	10	71	15	6	8
Rice	58	18	15	9	65	18	8	8
Rape-Mustard Seed	69	21	6	4	66	21	7	7
Sesamum	53	29	6	11	62	24	5	8
Soybean	87	5	0	8	86	4	6	5
Sunflower	63	19	5	13	72	18	5	5
Tur	20	39	11	31	14	73	11	2
Wheat	56	19	16	9	67	17	10	7
Total	57	23	8	11	61	25	8	6
Omnibus								
Bajra	57	26	9	8				
Barley	52	28	10	11				
Cotton	39	35	14	13				
Groundnut	61	29	3	7				
Gram	51	32	9	9				
Jowar	63	22	8	7				
Maize	62	24	9	5				
Ragi	64	16	8	13				
Rice	55	22	14	9				
Rape-Mustard Seed	62	25	8	5				
Sesamum	51	31	8	10				
Soybean	83	4	5	8				
Sunflower	61	21	5	13				
Tur	13	78	4	5				
Wheat	55	21	16	8				
Total	53	30	9	8				

N-N = normal to normal; NN-NN = non-normal to non-normal;
N-NN = normal to non-normal; and NN-N = non-normal to normal

Based on skewness results across all crops analyzed, 57 percent and 23 percent of districts indicate the crop yield residuals are normally distributed and non-normally distributed respectively by panel and time-series models. With the exception of barely, cotton, sesamum and tur other crops exhibited higher percentage of districts with normally distributed crop yield residuals in panel and time-series models. Bajra, groundnut, gram, sesamum and tur crops exhibited higher percentage of districts with non-normally distributed crop yield residuals in panel and time-series models. Accounting for spatial and temporal variation seems to change the distribution of crop yield residuals from normal (time-series model) to non-normal (panel model) in 9 percent of districts. Barley, cotton, rice, tur and wheat crops exhibited higher percentage of districts changed crop yield residual distribution from normal to non-normal. For the remaining11 percent of districts, accounting for spatial and temporal variation seems to change the distribution of crop yield residuals from non-normal (time-series model) to normal (panel model) distribution.With the exception of groundnut, jowar, maize, rice, rape and mustard seed, soybean and wheat crops exhibited higher percentage of districts changed crop yield residual distribution from non-normal to normal.

Kurtosis results across all crops analyzed, indicated 61 percent of districts with normally distributed crop yield residuals and 25 percent of districts with non-normally distributed crop yield residuals in panel and time-series models. With the exception of cotton and tur, other crops exhibited higher percentage of districts with normally distributed crop yield residuals in panel and time-series models. Barley, cotton, groundnut, gram, rape and mustard seed, sesamum and tur exhibited higher percentage of districts with non-normally distributed crop yield residuals in panel and time-series models. Accounting for spatial and temporal variation seems to change the distribution of crop yield residuals from normal (time-series model) to non-normal (panel model) in 8 percent of districts. Bajra, barley, cotton, tur and wheat exhibited higher percentage of districts changed crop yield residual distribution from normal to non-normal compared to overall average across all crops. For the remaining 6 percent of districts, accounting for spatial and temporal variation seems to change the distribution of crop yield residuals from non-normal (time-series model) to normal (panel model) distribution. With the exception of cotton, ragi, rice, rape and mustard seed, sesamum and wheat, other crops exhibited lower percentage of districts changed crop yield residual distribution from non-normal to normal.

Based on omnibus results across all crops analyzed, 53 percent and 30 percent of districts indicate the crop yield residuals are normally distributed and non-normally distributed respectively by panel and time-series models. With the exception of cottonand tur, other crops exhibited higher percentage of districts with normally distributed crop yield residuals in panel and time-series models. Cotton, gram, sesamum and tur exhibited higher percentage of districts with non-normally distributed crop yield residuals in panel and time-series models. Accounting for spatial and temporal variation seems to change the distribution of crop yield residuals from normal (time-series model) to non-normal (panel model) in 9 percent of districts. Barley, cotton, rice and wheat exhibited higher percentage of districts changed crop yield residual distribution from normal to non-normal. For the remaining 8 percent of districts, accounting for spatial and temporal variation seems to change the distribution of crop yield residuals from non-normal (time-series model) to normal (panel model) distribution. With the exception of bajra, groundnut, jowar, maize, rape and mustard seed, soybean, tur and wheat, other crops exhibited higher percentage of districts changed crop yield residual distribution from non-normal to normal.

5. Summary and Conclusions

This paper provides a two-fold contribution to the literature. First, the paper examines the normality of the food grains (rice and wheat), millets (bajra, barley, jowar, maize and ragi), pulses (gram and tur), oilseeds (groundnut, rape and mustard seed, sesamum, soybean and sunflower) and fiber (cotton) crop yield residuals using panel statistical procedures. The panel statistical procedures model crop yield distributions accounting for trend, autocorrelation and heteroskedasticity. Second, the normality of crop yields is examined using skewness, kurtosis and omnibus tests under alternative representations of the data, a panel structure and using a traditional time-series formulation. These are examined with an empirical application to district level data from 1956-2002 for 15 crops and 14 states in India composed of a total of 3143 individual reporting districts in India.

To summarize in 65 percent of districts crop yield residuals were normally distributed and in the remaining 9 and 26 percent of districts residuals were either negatively or positively skewed, respectively. Similarly, based on kurtosis statistics, 73 percent of districts indicated normality of crop yield residual and the remaining 27 percent of districts indicated leptokurtic or thick tailed $((\beta_2 - 3) > 0)$ distributions. Based on an omnibus test crop yield residuals in 53 percent, 70 percent, 68 percent, 66 percent and 38 percent of food grains (rice and wheat), millets (bajra, barley, jowar, maize and ragi), pulses (gram and tur), oilseeds (groundnut, rape and mustard seed, sesamum, soybean and sunflower) and fiber (cotton) growing districts were normally distributed.

Normality results suggest accounting for spatial and temporal variation in panel statistical techniques lead to changes in the results of normality testing of crop yield residuals. Based on skewness, kurtosis and omnibus results 57 percent, 61 percent and 53 percent of districts respectively indicated the crop yield residuals are normally distributed by panel and time-series models. Accounting or not accounting for spatial and temporal variation seems to change the distribution of crop yield residuals in 20 percent, 14 percent and 17 percent of districts based on skewness, kurtosis and omnibus tests respectively.

The outcome of this research has potential implications on the crop insurance program in India. If the crop insurance program in India assumes normality and estimates the premium rate would lead to serious implications in 23 percent, 25 percent and 30 percent of the district based on skewness, kurtosis and omnibus test respectively. This would be a future research area to purse based on historical indemnity and premiums paid rather than simulated indemnity and premiums from historical yields to develop actuarially sound premium rates. Second the outcome of the results are important for public and private insurance companies and risk management specialist as it would provide information to differentiate low versus high risk farms, districts or states, and avoid asymmetric issues like adverse selection, moral hazard and fraud.

Reference

Armitage P., and Colton T. (1998), *Encyclopedia of Biostatistics*, Wiley. New York Vol:4, pp: 3075-3079.

Atwood, Joseph A., S Shaik and Myles J. Watts (2002), “Can Normality of Yields Be Assumed for Crop Insurance?” *Canadian Journal of Agricultural Economics*. 50: 171-184.

Atwood, J., S. Shaik, and M. Watts (2003), “Are Crop Yields Normally Distributed? A Reexamination.” *American Journal of Agricultural Economics* 85: 888-901.

D’Agostino, R.B., A. Belanger, and R.B. D’Agostino, Jr (1990),“A Suggestion for Using Powerful and Informative Tests of Normality.”*The American Statistican*. 44: 316-321.

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

Gallagher, P (1986), “U.S. Corn Yield Capacity and Probability: Estimation and Forecasting with Nonsymmetric Disturbances.”*North Central Journal of Agricultural Economics*. 8: 109-122.

Gallagher, P (1987), “U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Distributions.” *American Journal of Agricultural Economics*. 69 : 796-803.

Hsiao, C., (2003), *Analysis of Panel Data*, 2nd edition, Cambridge: Cambridge University Press (Econometric Society monograph no. 34).

♦♦♦