Journal of Economics and Behavioral Studies
Vol. 3, No. 6, pp. 395-401, Dec 2011 (ISSN: 2220-6140)

A Regional Analysis of Corn Yield Models: Comparing Quadratic versus Cubic Trends

*Francis Annan[1], Henry De-Graft Acquah[2]
[1]Mississippi State University, USA
[2]University of Cape Coast, Cape Coast, Ghana
*fa131@msstate.edu

Abstract: This study investigates county-level corn yield trend models using quadratic and cubic trend estimations. The study empirically revealed that the cubic trend is more appropriate for yield data from the West, Midwest and South regions. The linear and the quadratic trend models were respectively found to be more appropriate for yield data from the Plains and Atlantic regions. Results suggest that the data should be allowed to determine the appropriate trend relationships to avoid trend misspecification. Additionally, the yield trends are found to be inconsistent across all the regions. Different locations tend to exhibit different yield trends. It is therefore recommended that differences between regions be recognized when we conduct yield trend tests and not to generalize our results to other regions.

Keywords: *Corn yield, regional modeling, quadratic model, cubic model, linear model*

1. Introduction

Accurate modeling of yield distributions in agricultural and applied economics is imperative because it fundamentally form a basis for economic decisions and crop insurance. The importance of properly modeling yield distributions stems in part from the dramatic growth in participation in the U.S. crop insurance program and the introduction of a broader range of new crop insurance products after the enactment of the 2000 Agricultural Risk Protection Act (Goodwin, Vandever and Deal, 2004; Glauber, 2004). Several studies have been carried out to empirically determine the distributional model that best characterizes crop yields. When crop yields are used to fit models, the trends component is estimated before assessing the distribution of the yield data-generally in an ad-hoc manner (Zhu et al. 2008). Since crop yields tend to be upward trending, it is important to remove the trend effect of technology before proceeding with the estimation of the actual yield distribution. Typically, these studies involve modeling yield distributions that specifically are based on conventional approach of estimating the trend components of the time series of the yield data. Consequently, after fitting the trend to the data the detrended data is modeled. Numerous studies have used the quadratic trend for trend estimation (Zhu et al, 2008; Sherrick et al, 2000).

According to Ker and Goodwin (2000) and Ker and Coble (2003), exogenous trends in yields that occur due to technical change pose a challenge for the modeling of yield distributions. Trend estimation therefore requires a great deal of attention. However, very little attention has been given to the estimation of trends in relation to spatial and trend type. In this paper, statistical analysis was undertaken to estimate and analyze corn yield trends across the different National Agricultural Statistics Services (NASS) regions. Specifically, the study seeks to: (1) investigate whether a cubic model is more appropriate than a quadratic model and vice versa; (2) investigate whether the findings from (1) are consistent across NASS production regions. The remainder of the paper is organized as follows. Section two provides some background and reviews related previous literature. Section three exposits the data used for the study and the data analysis techniques. Empirical results are presented and discussed in section five. The summary and conclusion of the findings drawn from the study are finally presented.

2. Literature Review

This section is devoted to offer some background on estimation of yield trends and modeling yield distribution in agricultural production. Several modeling of yield distributions and trend estimation have been introduced from previous literature. These include the Just and Weninger (1999) study, hereafter referred to as JW. According to JW, analyzing random yield distributions require isolating the truly random

components of the yield distributions. JW indicated that most analyses proceed by assuming the deterministic trend can be approximated by a lower-order polynomial function. JW demonstrated how misspecification of the deterministic component (trend) of yields, represented by differences between the true and assumed trend specifications, causes non-stationarity of yield deviations and incorrect assessment of kurtosis, and consequently normality in crop yields. JW opted for a flexible polynomial trend where the polynomial is determined by the data. However, their study did not compare these trends cross all the regions, specifically NASS production regions. Furthermore, a study by Harri et al (2009) explored many alternative trend-modeling procedures. Both the deterministic and stochastic trend models were investigated. However, they found a limited support for stochastic trends in crop yields. This study focuses on the deterministic trend approach, seeking to investigate the quadratic trends of corn yield and concurrently allow for cubic trend estimation, and comparing these trends across the NASS production regions for consistence.

3. Methodology

This section presents the conceptual framework, data and econometric techniques employed in the study. Specifically, econometric techniques such as polynomial modeling and nested dummy variable regression are emphasized.

Conceptual Framework: Determining the correct functional form or model to describe technological change is crucial (Harri et al. 2008), for modeling of yield distributions. Said differently, yield data typically span a significant time period and thus requires the removal of the upward trend in the yield observations that is primarily driven by technological change. This is to ensure that we can compare these yields through time. A deterministic trend, instead of stochastic trend is usually used to capture development of yields. Zhu et al. (2008) pointed out that the trend component of yields should be controlled before their distributions are estimated. Usually, to remove the trend component in yields, a detrending regression is used to fit a quadratic trend model as shown in the equation below:

$$y_t = \alpha + \beta_1 t + \beta_2 t^2 + e_t \qquad (1)$$

Where y_t is the observed crop yield data in year t, and t = 1, 2, 55 stands for $year$ = 1955, 1956, 2009. After regressing the yields on a quadratic (or linear) time trend, the residuals $\hat{e}_t = y_t - \hat{y}_t$ and the predicted yield can be calculated, where $\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 t + \hat{\beta}_2 t^2$. The residuals (detrended county-yield series) $\hat{\varepsilon}_{it}$ must satisfy the distributional assumptions of the classical regression (i.e. $\hat{\varepsilon}_{it} \sim N(0, \sigma^2)$). That is, it should have a zero expectation, should be normally distributed, should be homoskedastic, should not be auto correlated and should be uncorrelated with the explanatory variables.
As pointed out by Harri et al (2009), Enders (1995) specified the deterministic trend of the series as a polynomial function of time: $D_t = \psi_o + \psi_1 t + \psi_2 t^2 + ,\ldots$ where $\psi_0, \psi_1, \psi_2, \ldots$ are estimated parameters. Sequential t-tests are used to determine the appropriate polynomial degrees. The original data series are made trend stationary by subtracting the deterministic component. Polynomial modeling was greatly considered by JW in their study. This study investigated the quadratic and cubic trend models and compared these trend models across all five NASS regions to examine consistency.

Data and Methods: County level yield data for corn was obtained from National Agricultural Statistics Services of the United States Department of Agriculture (NASS). The time span for the data is 1955 to 2009 (i.e. 55 years of historical data). The target crop utilized for this study was irrigated corn-grain. A representative county was selected from all the five NASS regions to represent each region. The representative county selection decision was based on first choosing from each region the state with the largest number of planted acres for 2009. Within the identified state, the county with the largest planted acres for 2009 was selected. All counties without 55 continuous years of historical yield data are excluded. The yield variables for the selected representative counties were investigated.

***Model specification*:** Two trend models, the quadratic and the cubic trend polynomial models were examined. The quadratic model is nested in the cubic model as follows:

$$y_t = \alpha + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + e_t \quad (2)$$

***Hypothesis testing*:** Based on the objectives of the study, the hypotheses that the quadratic trend is more appropriate than the cubic trend (H_0: $\beta_3 = 0$) and the trend components are the same across the NASS production regions were investigated. More generally, the dummy variable regression is specified as follows:

$$y_t = \sum_{i=1}^{5} \alpha_i + 5\sum_{i=1}^{5} t\beta_{1i} I_i + \sum_{i=1}^{5} t^2 \beta_{2i} I_i + \sum_{i=1}^{5} t^3 \beta_{3i} I_i + e_t \quad (3)$$

where: i = 1, 2, 3, 4 and 5 representing the regional dummies for West, Midwest, Plains, South and Atlantic regions respectively. In order to test for the consistency across all the five NASS production regions, a restriction was placed on the general model specified above indicating that the coefficients for the trend variables for each individual region or state is the same, however allowing for different intercepts for each production region. Sequentially, an F test is used to estimate the consistency of the restricted model with the unrestricted or general model, defined as follows:

$$F[J, n-K] = \frac{(R^2_{UR} - R^2_R)/J}{(1-R^2_R)/(n-K)}$$

Where J is the number of imposed parameter restrictions; n is the number of observation; K is the number of parameters' in the unrestricted or general model; RSS_R is the residual sum of squares from the regression of the restricted model; RSS_{UR} is the residual sum of squares from the regression of the unrestricted model. Heteroskedasticity was tested for the yield data sets used for this study using the White's test. The White's test looks for evidence of an association between the variance of the disturbance term and the regression without assuming any specific relationship. The squared residuals were used as a proxy for the disturbance term since the disturbance terms for the observations are unobserved. The squared residuals were regressed on the explanatory variables in the model, their squared terms and their cross products, omitting any duplicative variable. A number of univariate tests for normality of the residuals (yield distributions) were explored. These include Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling tests. The results are presented in table 1.

4. Results and Discussion

The results of the polynomial trends for the five selected counties are presented in table 2. In total, five counties were selected to represent the five NASS production regions. These include Yuma County (Colorado), Kossuth County (Iowa), Custer County (Nebraska), Yazoo County (Mississippi), Union County (Kentucky) for the West, Midwest, Plains, South and Atlantic regions respectively. The results in table 2 demonstrate whether the cubic model is more appropriate than the quadratic model in the respective regions. We observe from table 2 that the cubic trend is deemed an appropriate model for three regions namely West, Midwest, and South. In which case, the null hypothesis ($H_o: \beta_3 = 0$) is rejected in these regions consequently. On the other hand, in the Plains and Atlantic regions, the cubic trend was not statistically different from zero at 5% level. Thus we fail to reject the null hypothesis ($H_o: \beta_3 = 0$) in these two regions. A further investigation about the Plains and Atlantic regions was carried out. When the quadratic model was tested separately (unreported) for both regions (Plains and Atlantic), it was revealed that the linear trend is appropriate in the Plains. On the other hand, the quadratic trend is the appropriate model in the Atlantic region.

The results generally suggested that the appropriate trend model varies across regions and hence it is important to give attention to the type of detrending regression to fit for empirical yield modeling. As pointed out by Just and Weninger (1999), allowing the data to determine the trend model will prevent the possibility of misspecification of the deterministic component (trend) of yield data in empirical work. Additionally, as evidenced in the F test results presented (Pr > F = <0.0001), the deterministic trend models are not the same across all the five NASS production regions. A p-value of <0.0001 suggested that the null hypothesis that the trend models were consistent or same across all the NASS production regions should be rejected as a result at one percent level of significance. This finding further indicates that, a generalization of the deterministic

trends of yields across these NASS production regions may lead to misspecification and should be avoided. The examination of degree of the polynomial trend must be given an important attention. Again, the yield data should be allowed to determine the degree of the deterministic polynomial trend to use for empirical analysis. The results on white's test for heteroskedasticity and normality test with the polynomial detrending of the yield data for a significance level of 0.05 is presented in table 1. The presence of heteroskedasticity was rejected in all the data sets considered for this study since the P-values were all greater than 0.05 alpha levels. The results for Shapiro-Wilk (W) normality test was only reported because the other tests show similar estimates and statistical inference. As shown in table 1, the yield distributions for Yuma and Kossuth counties representing the West and Midwest regions are not normally distributed. On contrast, Union, Yazoo and Custer counties representing the Atlantic, South and Plains had normally distributed yields. The rejection of normality in Yuma and Kossuth counties was due to the obvious large negatively skewed yields, reported in table 1.The absence of multicollinearity in the data sets was assumed for this study.

Table 1: White's Test for Heteroskedasticity and Univariate Normality tests for Corn

County	Heteroskedasticity test		Normality test		
	Chi-square value (χ^2_5)	P value	W-statistic	P value	Skewness
Yuma	6.64	0.2488	0.927*	0.0058	-0.926
Kossuth	5.52	0.2488	0.895*	0.0005	-1.601
Union	3.46	0.2488	0.954*	0.0649	-0.772
Yazoo	9.23	0.2488	0.986	0.8226	0.203
Custer	5.98	0.2488	0.955	0.0699	-0.454

An "*" indicates statistical significance at the α = 0.05

Table 2: Polynomial regression results for Corn

Variable	Parameter Estimate	Standard Error	P value
	a)Corn Yield-Yuma County		
Intercept	25.74*	8.61	0.0031
t	9.89*	1.54	0.0001
t^2	-0.27*	0.074	0.0003
t^3	0.003*	0.001	0.0034
	b)Corn Yield-Kossuth County		
Intercept	77.94*	8.61	<0.0001
t	4.47*	1.54	0.0033
t^2	-0.16*	0.074	0.03
t^3	0.0025*	0.001	0.015
	c) Corn Yield-Union County		
Intercept	72.43*	8.61	<0.0001
t	2.26	1.54	0.143
t^2	-0.037	0.074	0.615
t^3	0.0006	0.001	0.532
	d) Corn Yield-Yazoo County		
Intercept	35.99*	8.61	<0.0001
t	-1.031	1.54	0.502
t^2	0.165*	0.074	0.027
t^3	-.0021*	0.001	0.041
	e) Corn Yield-Custer County		
Intercept	41.84*	8.61	<0.0001
t	4.54*	1.54	0.004
t^2	-0.081	0.074	0.276
t^3	0.0009	0.001	0.391

An "*" indicates statistical significance at the α = 0.05

(ii) Consistency of trend across NASS production regions:

The study also seeks to investigate whether these estimated trends are consistent across all the 5 NASS regions. The result revealed in this study is presented as below:
F value = 7.30; Pr > F = <.0001**. (An "**" indicates statistical significance at the $\alpha = 0.01$).

5. Conclusion

This paper examines county-level corn yield trend models using quadratic and cubic trend estimations. Five counties were selected to represent the NASS regions. A demonstration of whether the cubic or quadratic trend is more appropriate for each of the NASS production regions was investigated. The consistency of these trends across all the regions was also examined. As a priori, the results show that Corn yields are normally distributed in the Atlantic, South and Plains but not normally distributed in the West and Midwest regions. Normality of yield distributions therefore varies from region to region for Corn. Homoskedasticity was found in all corn yields across all regions.

The empirical results of this paper showed that, the cubic trend is more appropriate for yield data in the West, Midwest and South regions. On the other hand, the linear and the quadratic trend models are more appropriate for yield data from the Plain and Atlantic regions respectively. Thus, the data should be allowed to determine these trend relationships to avoid trend misspecification. These results appear to have implications for examining agricultural decision making under uncertainty(Just and Weninger, 1999), sine yield trend misspecification have dramatic effects on insurance analysis. By testing for the consistency of the trends across all the production regions, it was found that these yield trend models are inconsistent across all the regions. Different locations tend to exhibit different yield trend. Thus, it is important to recognize the differences between regions when we conduct yield trend tests and not to generalize our results to other regions.

Endnotes: [1]See also Just and Weninger, Harri et al. [2]Selection of more than one county to represent a region may generate efficient and interesting results.

References

Enders, W. (1995). Applied Econometric Time Series. New York: Wiley.

Glauber, J. W. (2004). Crop Insurance Reconsidered. *American Journal of Agricultural Economics*, 86, 1179 - 1195.

Goodwin, B. K., Vandever, L. M. & Deal, L. J. (2004). An Empirical Analysis of Acreage Effects of Participation in the Federal Crop Insurance Program. *American Journal of Agricultural Economics*, 86(4), 1058-1077.

Harri, A., Coble, H. K., Erdem, C. & Knight, O. T. (2009). Crop Yield Distributions: A Reconciliation of Previous Research and Statistical Tests for Normality. *Review of Agricultural Economics*, 31, 163-182.

Just, E. R. & Weninger, Q. (1999). Are Crop Yields Normally Distributed? *American Journal of Agricultural Economics*, 81, 287-304.

Ker, A. P. & Goodwin, K. B. (2000). Nonparametric Estimation of Crop Insurance Rates Revisited. *American Journal of Agricultural Economics*, 83, 463-478.

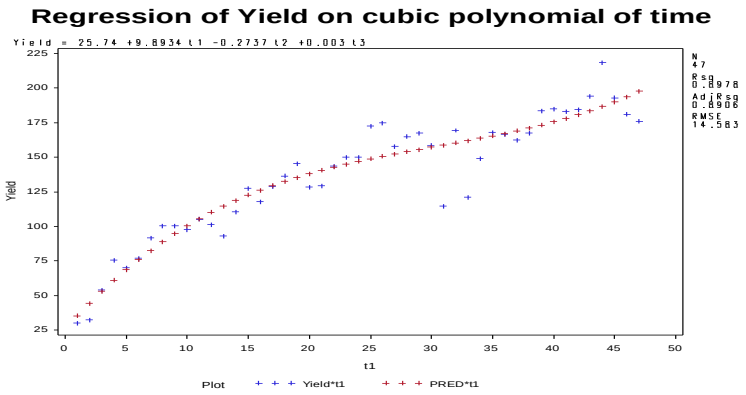
Ker, A. P. & Coble, H. K. (2003). Modeling Conditional Yield Distributions. *American Journal of Agricultural Economics*, 85, 291-304.

Sherrick, B. J., Zanini, C. F., Schnitkey, D. G. & S. Irwin, H. (2000). Crop Insurance Valuation under Alternative Yield Distributions. *American Journal of Agricultural Economics*, 86, 653-659.

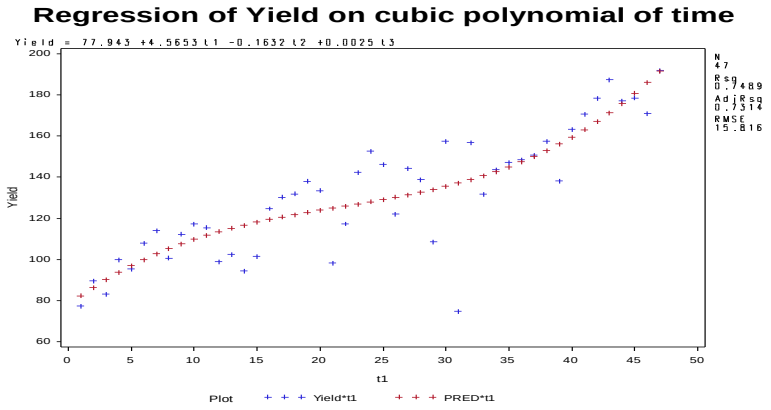
Zhu, Y., Goodwin, K. B. & Ghosh, S. (2008). Time-varying Yield Distributions and the Implications for Crop Insurance Pricing. NC State Agricultural Economics Workshop, Raleigh, NC.

APPENDIX A: GRAPHS: Fitted Trends

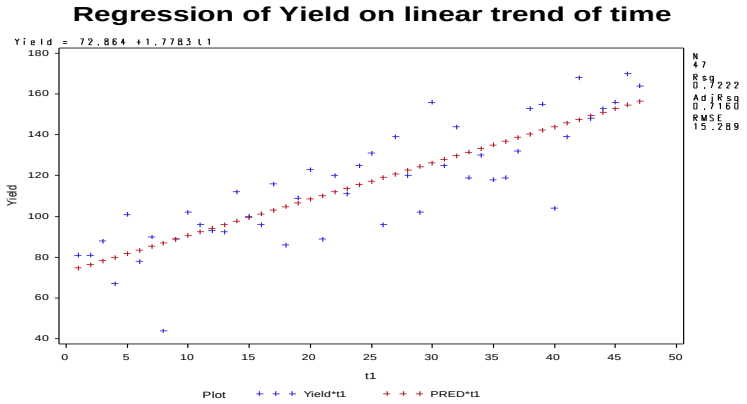
1) Colorado State:



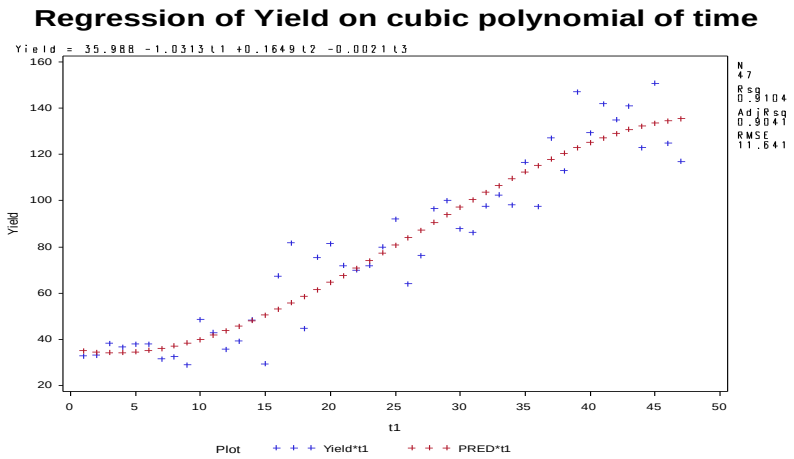
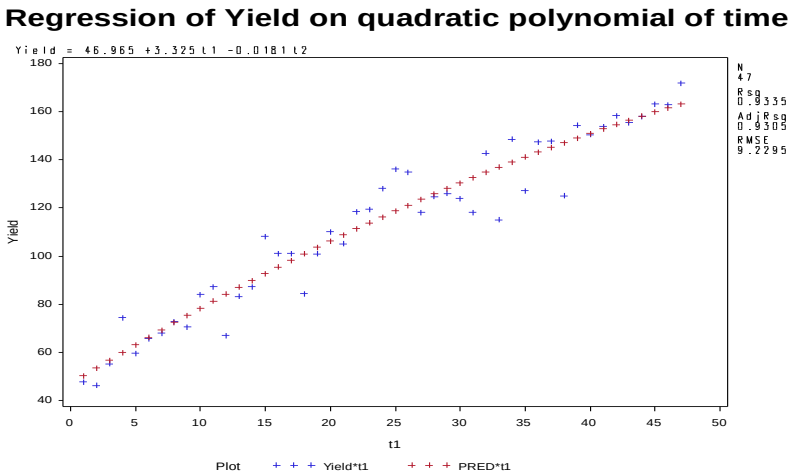
2) Iowa State:



3) Kentucky



4) Mississippi state



5) Nebraska state