Time-varying Yield Distributions and the U.S. Crop Insurance Program

Ying Zhu
SAS Institute, Inc
ying.zhu@sas.com

Barry K. Goodwin
Department of Agricultural and Resource Economics
North Carolina State University

Sujit K. Ghosh
Department of Statistics
North Carolina State University

Selected Paper prepared for presentation at the Agricultural & Applied Economics Associations 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24- 26, 2011.

Abstract

The objective of this study is to evaluate and model the yield risk associated with major agricultural commodities in the U.S. We are particularly concerned with the nonstationary nature of the yield distribution, which primarily arises because of technological progress and changing environmental conditions. Precise risk assessment depends on the accuracy of modeling this distribution. This problem becomes more challenging as the yield distribution changes over time, a condition that holds for nearly all major crops. A common approach to this problem is based on a two-stage method in which the yield is first detrended and then the estimated residuals are treated as observed data and modeled using various parametric or nonparametric methods. We propose an alternative parametric model that allows the moments of the yield distributions to change with time. Several model selection techniques suggest that the proposed time-varying model outperforms more conventional models in terms of in-sample goodness-of-fit, out-of-sample predictive power and the prediction accuracy of insurance premium rates.

Key Words: Crop Insurance, Model Comparison, Time-Varying Distribution

Time-varying Yield Distributions and the U.S. Crop Insurance Program

The Federal Crop Insurance program represents an important component of U.S. agricultural policy and is intended to protect farmers from yield and revenue risk. Accurate modeling of crop yield distributions is essential for the proper design of crop insurance contracts and to the maintenance of an actuarially sound insurance program. Historical agricultural yield data suggest a strong upward trend in crop yields (figure 1(a)). Advances in technology, germplasm, breeding techniques, the development of new hybrids and changes in environmental factors may significantly affect the distributions of crop yields. These changes can complicate efforts to accurately model yield distributions using data observed over time.

Many studies have attempted to determine the distributional model and estimation methods that best characterize crop yield distributions. Modeling approaches in the current literature range from non-parametric (Goodwin and Ker, 1998) to parametric methods based on the assumption that crop yields are independently and identically distributed. The parametric approach of modeling yields usually involves selection and specification of candidate distribution families, parameter estimation and goodness-of-fit assessments. Among others, the Beta distribution is popularly used in practice due to its flexibility and ability to represent the skewness typically associated with crop yield distributions. The notion of a conditional Beta distribution for yields was introduced by Nelson and Preckel (1989). Other popular candidates used in the literature include the lognormal distribution (Day, 1965), the Normal distribution (Just and Weninger, 1999), the Weibull distribution (Chen and Miranda, 2004) and the Logistic distribution (Sherrick et al., 2004). Evidence of non-normal yields has been presented by a number of authors, including Taylor (1990), Ramirez (1997) and Ramirez, Misra, and Field (2003).

In many cases, agricultural yields display a strong upward trend over time and the deviations from trend (residuals) frequently display heteroscedasticity (see figure 1(a)) and thus violating the assumption that yields are identically distributed. A very common approach to modeling yield risk using time-series data has been to first detrend the time series data and then estimate the yield distribution using the detrended yield data, thereby treating the estimated, detrended yields as "observed" data. These approaches are often referred to as "two stage" methods; the first stage fits a trend model to the data while the second stage uses the detrended data to model the distribution. Examples of such two-stage detrending procedures can be found in, among others, Miranda and Glauber (1997), Swinton and King (1991), and Atwood, Shaik, and Watts (2003).

In this two-stage method, it is crucial to determine the correct functional form of the regression representing trend in the first stage and then to establish the correct distributional properties of the detrended data, including such characteristics as skewness, kurtosis, and heteroscedasticity. However, it has been recognized that the resulting estimated residuals, representing the detrended yields, are subject to the estimation uncertainty associated with sampling variability in the first stage estimates of trend and thus may not necessarily provide an accurate representation of the actual yield distribution. Although any biases induced at the first-stage asymptotically approach zero when the correct functional form is used in the regression and errors are homoscedastic, the uncertainty induced at the first stage, if not accounted for in the second stage estimates of the yield distribution, will lead to inaccurate estimation of the variance in the final estimates. The magnitude of this effect can be large especially when the errors are heteroscedastic (Robinson (1987)) and thus can potentially induce significant adverse selection into an insurance program if ignored.

This standard two-stage method has been one of the most popular approaches to removing time trends and modeling the distribution of crop yields. A similar two-stage method is used to rate the Group Risk (GRP) and Gross Revenue Insurance (GRIP) programs, though this method does address the potential for heteroscedasticity. However, it is possible to account for the uncertainties associated with the first stage estimates and adequately represent characteristics of the yield distribution (such as deterministic trends and heteroscedasticity) by applying an alternative simultaneous estimation method. We propose a likelihood based estimation method that simultaneously estimates the trend (conditional mean) and higher order conditional moments of the yield density by using a flexible class of parametric distributions. We also provide a set of model validation tools that enables a researcher to test the validity of the proposed class of distributions in approximating the true underlying data generation mechanism.

The method, along with its validation measures proposed here, allows one to measure conditional yield risk in a dynamic setting and thereby calculate premium rates for crop insurance contracts in a more accurate and systematic way. Our method essentially models the first four conditional moments of the distribution simultaneously by allowing location, scale, skewness and kutosis parameters of the specific distributional family to evolve over time, whereas the more common two-stage method usually allows one to model only the location (conditional mean) and sometimes the scale (conditional variance) to reflect changes over time. A more complete and coherent picture of technological progress and the consequential changes in yield risk can be provided by simultaneously modeling the time trend and the distributional parameters.

A Conventional Two-Stage Estimation Framework

In most empirical analyses, a deterministic trend is used to capture the dynamics of the expected yields and thus to represent the variation of yields around this expected level.[1] The trend component is usually controlled for before assessing the distribution of yields–generally using a homoscedastic parametric or nonparametric regression model. Popular regression models include a log-linear specification based on polynomials, kernel regression, smoothing splines, and partial linear models (Gyorfi et al. (2002)). We illustrate this idea by using a quadratic trend as well as a nonparametric trend model.[2]

Consider the following trend model:

$$y_t = m(\mathbf{x}_t) + \varepsilon_t \tag{1}$$

where y_t is the observed crop yield in year t, $(t = 1, \ldots, T)$, $m(x)$ denotes the regression function $E(Y_t|X_t = x)$, x_t represents linear or nonlinear time indexes representing trend, and ε_t represent residuals that are assumed to be independently distributed with mean zero. The regression function $m(\cdot)$ can be estimated nonparametrically using kernel methods or smoothing spline methods. Alternatively, if we assume a parametric functional form for $m(\cdot)$, then the regression coefficients can be obtained using ordinary least squares (OLS).[3] In either case, the residuals are obtained as $\hat{\varepsilon}_t = y_t - \hat{m}(x_t)$. We considered both

[1]The main justification for using a deterministic component is that, if crop yield variables evolve slowly through time, then approximation of a deterministic component may be sufficient to model the yield distribution (Just and Weninger, 1999).

[2]The selection of these two trend models is intended to provide a benchmark for comparison purposes. There are other detrending methods such as log-linear regression. Since the focus of this study is to compare the two-stage approach and the time-varying method that we propose as an alternative, we use representative methods to illustrate the concepts. A comprehensive survey of all possible detrending methods is beyond the scope of this study.

[3]We assume that $m(x_t) = m_0(x_t, \beta)$, where m_0 is a known functional form up to some finite dimensional regression coefficient vector β.

quadratic and nonparametric trend models. The Kolmogorov-Smirnov (K-S) 2-sample goodness of fit (GOF) test suggests that the two residual distributions are not significantly different between the nonparametric and parametric models based on the data in this study. On the basis of this test, the quadratic detrending method is used as a benchmark.

Our empirical analyses presented in this paper are based on applications to USDA's National Agricultural Statistics Services (NASS) county-level average yields.[4] Figure 1(b) presents the nonparametric residual plot of annual corn yields in Iowa, which shows that the deviations from trend tend to be proportional to the level of the yields. To account for this temporal heteroscedasticity effect, a rescaled form of the deviations from a trend-based, forecasting equation is often suggested. This approach, though ad hoc, is commonly used in practice (see, for example, Miranda and Glauber (1997), Atwood, Shaik, and Watts (2003)). By dividing each error by its associated forecast, the residuals can be scaled to the year T equivalent predicted yield.

We use a goodness of fit (GOF) specification test to determine the appropriate distribution for the detrended yield $\tilde{y}_t$. A Q-Q plot based on the residuals $\hat{\varepsilon}_t$ (figure 1(b)) suggests that the residuals are more negatively skewed than what would be implied by the normal distribution, which suggests that a Beta distribution may be a viable candidate. A GOF test for the Beta distribution (based on a Chi-square statistic) confirms that a Beta distribution provides a reasonable fit for the normalized county-level yields typically applied in this two-stage approach. For example, the GOF test yields a p-value of 0.51 for Kossuth County and 0.62 for Adair County Iowa all-practice corn yields. We use $Beta(\alpha, \beta, \theta, \delta)$ to denote a Beta distribution with location parameter $\theta \geq 0$, scale parameter $\delta > 0$, and shape

[4] The data are available at the NASS website at http://www.nass.usda.gov.

parameters $\alpha, \beta > 0$.[5] This implies that the yields follow a Beta distribution with constant shape parameters and time-varying location and scale parameters, i.e., $y_t \sim Beta(\alpha, \beta, \tilde{\theta}_t, \tilde{\delta}_t)$, with $\tilde{\theta}_t = \hat{\rho}_t\theta, \tilde{\delta}_t = \hat{\rho}_t\delta$ and $\hat{\rho}_t = \frac{\hat{y}_t}{\hat{y}_T}$. The log-likelihood function of a general Beta distribution based on the detrended data $\tilde{y}_t$ with two shape parameters α, β and location θ and scale δ parameters, is given by,

$$
\begin{aligned}
\text{LLF}(\alpha, \beta, \theta, \delta | \tilde{y}_t, t = 1, \ldots, T) \quad = \quad & \sum_{t=1}^{T} (\alpha - 1) \log(\tilde{y}_t - \theta)^+ + \sum_{t=1}^{T} (\beta - 1) \log(\delta + \theta - \tilde{y}_t)^+ \\
& - \sum_{t=1}^{T} \log(B(\alpha, \beta)) - \sum_{t=1}^{T} (\alpha + \beta - 1) \log(\delta) \qquad (2)
\end{aligned}
$$

where $\log(B(\alpha, \beta)) = \log(\Gamma(\alpha)) + \log(\Gamma(\beta)) - \log(\Gamma(\alpha + \beta))$ and $\log a^+ = \log a$ if $a > 0$ and $\log a^+ = 0$ otherwise, which ensures that $\theta \leq \tilde{y}_t \leq \theta + \delta \ \forall t$, for any $\theta, \delta > 0$.

We obtain the parameter estimates $(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\delta})$ by maximizing the $\text{LLF}(\alpha, \beta, \theta, \delta)$ based on the normalized values of $\tilde{y}_t$. The results are presented in table 1. The predicted mean yield can be calculated from the detrended model as:

$$
\hat{\hat{y}}_t = \frac{\hat{y}_T}{\hat{y}_t} \hat{\tilde{y}}_t = \frac{\hat{m}(x_T)}{\hat{m}(x_t)} (\hat{\theta} + \hat{\delta} \frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}}) \qquad (3)
$$

As we have noted, using a first stage estimation to detrend yield data and then treating the resulting detrended yields as if they were observed without error may not be appropriate because the first stage estimation error is ignored (e.g., $\hat{\varepsilon}_t$'s are assumed known for the LLF in equation 2.) A more systematic inferential method may be needed to accurately capture trend effects and model conditional yield risk.

[5] In other words, $\frac{\tilde{y}_t - \theta}{\delta} \sim Beta(\alpha, \beta)$, where $Beta(\alpha, \beta)$ represents a standard Beta distribution defined on $(0, 1)$ with shape parameters $\alpha, \beta > 0$.

A Time-Varying Yield Distribution Model

In this section, a flexible class of parametric models is proposed which allows us to simultaneously and coherently specify the first four moments using suitable polynomials of time and the coefficients of the polynomials are estimated simultaneously by maximizing the resulting likelihood function. Several alternative models are examined to measure conditional yield risk. For instance, instead of using polynomials to models the first four moments of the proposed distribution, one may use knot-based splines. In contrast to typical methods, the time-varying model accounts for parameter uncertainty by maximizing the time-varying likelihood function, which includes time-trend parameters and the distributional parameters in one step. The results of this proposed model are compared to those based on the conventional two-stage approach described in the previous section for several important crops and counties drawn from U.S. county-level data.

The basic assumption of the time varying model is that the parameters of the distribution follow a specific temporal pattern, such that the whole temporal changes of the yield distribution can be captured by the time-varying shape and scale parameters. The resulting parameter estimates are consistently estimated if the likelihood function is appropriately specified.

These time varying parameters evolve according to an exponential form. This particular functional form ensures that the Beta shape, scale, and location parameters are positive at every observation. We evaluated two different time trend structures for the parameters of the yield distributions—a standard linear trend and a quadratic trend model. However our method is not restricted to these chosen functional forms.[6] The log-likelihood function of the time-varying Beta distribution

[6]Of course, other functional forms including quadratic specifications could be used to ensure positive parameters. For instance, quite generally we can model any of these Beta parameters as $\exp\{\sum_{j=1}^{J} \psi_j(t) b_j\}$, where $\psi_j(\cdot)$'s may represent members of collection of J basis functions

is identical to that of the constant Beta distribution (equation 2) with the notable exception that the shape and scale parameters are allowed to vary with time and thus appear as $\alpha_t, \beta_t, \delta_t$, and θ_t.

Because the quadratic specification nests the linear trend, a standard likelihood ratio test can be used to evaluate the statistical significance of the quadratic terms and thus to select the optimal trend specification. Note that the Beta distribution is characterized by four parameters (α, β, θ, and δ). For simplicity and numerical stability of the maximum likelihood approach, we fix the minimum possible yield to be equal to zero in each case (i.e., by setting $\theta_t = 0$ for all t). We allow each parameter of the Beta distribution to vary over time through a functional relationship of the form (e.g., $\alpha = exp(f(\mathbf{b}, t))$ where $f(\cdot)$ is a linear or quadratic function of time). Such a specification allows us to use an unconstrained maximization of the likelihood function. As our results demonstrate below, the quadratic terms were not found to be statistically significant for the data sets that we have analyzed and thus our final representation of the conditional moments use a standard linear trend.

The predicted value $\hat{y}_t$ from the time-varying model is given by

$$\hat{y}_t = \hat{\delta}_t \frac{\hat{\alpha}_t}{\hat{\alpha}_t + \hat{\beta}_t} \tag{4}$$

where $\alpha_t = \alpha(t, \hat{\mathbf{b}}), \beta_t = \beta(t, \hat{\mathbf{b}})$, and $\delta_t = \delta(t, \hat{\mathbf{b}})$.

(e.g., choosing $\psi_j(t) = t^{j-1}$ we obtain polynomials while choosing $\psi_j(t) = (t - t_j)^3_+$ we obtain cubic polynomials with knots t_j's). Alternatively, one may also specify functional form using the first four moments of the Beta distribution, which may require a constrained optimization of the likelihood function.

Empirical Application

The time-varying model not only addresses the dynamic characteristics of yield distributions, but also provides a more flexible specification of heteroscedasticity and higher order moments (e.g., skewness and kurtosis). We implement the time-varying model by applying the methods to the top 10 counties in the major producing states for corn, soybeans, cotton. These county/crop combinations include the following: Iowa all-practice corn from Kossuth, Sioux, Pottawattamie, Plymouth, Webster, Pocohontas, Hardin, Franklin, Clinton and Woodbury counties; Iowa soybeans from Kossuth, Sioux, Pattawattamie, Plymouth, Webster, Woodbury, Benton, Grundy, Crawford and Tama counties; Texas upland cotton from Gaines, Lubbock, Hockley, Lynn, Dawson, Hale, Terry, Crosby, Floyd and Martin counties.[7]

It is widely recognized that the rate of technological progress varies considerably across different crops. Our results are presented in figure 2 and demonstrate that Iowa corn and soybean yields are skewed, kurtotic and exhibit strong time trend effects and varying degrees of heteroscedasticity through time. In contrast, Texas cotton yields appear to have a more modest time trend, though strong evidence of temporal heteroscedasticity is exhibited.

The maximum likelihood estimates of this time-varying Beta distribution with a linear time trend in the exponent and a quadratic time trend structure are shown in table 1. A likelihood ratio test statistic of the two alternative models has a value of 4.12, which does not reject the null hypothesis that the quadratic trend parameters are equal to zero and thus supports the adequacy of the linear specification.

[7] Although our choice of counties encompasses a significant proportion of the overall production of each crop in the relevant states (and further reflects a significant amount of the GRP crop insurance liability and premium), we also considered analysis for a much wider range of all counties (for which data existed) in each state evaluated. The results were very consistent with what is presented below. In order to conserve space, we only present results for the top ten counties in prominent states for each crop. However, detailed results for other counties are available from the authors on request. In addition, analysis of shorter series of yield data were also considered and found to yield similar conclusions. These results are also available on request.

The MLE estimates can be used to evaluate the time-varying Beta density for any given year. Figures 2(d), 2(e) and 2(f) illustrate the dynamic evolution of the densities that are estimated by each time-varying model for corn, soybeans and cotton yields. Various moments of the distributions appear to evolve over time. The density plots of these estimated time-varying distributions suggest different means, skewness coefficients, and maximum values of corn yields for each year. In figures 2(a), 2(b) and 2(c), we present estimated densities for both the time-varying model and the more conventional detrended model. In every case, the time-varying densities show a smaller degree of leptokurtosis than is the case for standard, two-stage detrended yield data.

Table 2 presents log-likelihood values for the two alternative models for a number of counties. In almost every case, the time-varying model provides a superior fit to the data than the conventional model, even after adjustments (within the context of alternative information criteria) for the number of parameters. This is also illustrated in figure 3, which contains a side-by-side bar plot of the LLF values for all major county/crop combinations considered in our analysis.[8]

Model Performance and Specification Tests

We considered a number of specification tests and evaluations of forecasting performance of the alternative models. Vuong's nonnested specification test (Vuong (1989)) is a likelihood-based test for model selection. Vuong's test statistic is given by:

$$v = \frac{n^{\frac{1}{2}} LR_n(\hat{\theta}_n, \tilde{\theta}_n)}{\hat{\omega}_n} \tag{5}$$

[8]MLEs for these other counties are available upon request from the authors.

where $LR_n(\hat{\theta}_n, \tilde{\theta}_n) = L_n^f(\hat{\theta}_n) - L_n^g(\tilde{\theta}_n)$, $L_n^f(\hat{\theta}_n)$ is the maximum likelihood function of the time-varying model and $L_n^g(\tilde{\theta}_n)$ is the maximum likelihood function of the two-step model. $\hat{\omega}_n$ is defined as:

$$\hat{\omega}_n^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\log\frac{f(Y_t|X_t;\hat{\theta}_n)}{g(Y_t|X_t;\tilde{\theta}_n)}\right)^2 - \left(\frac{1}{n}\sum_{t=1}^{n}\log\frac{f(Y_t|X_t;\hat{\theta}_n)}{g(Y_t|X_t;\tilde{\theta}_n)}\right)^2$$

The test statistic v is approximately distributed as a standard normal random variable. As specified, if $v > c$, where c is the critical value[9], we reject null that the models are the same in favor of the alternative time-varying model F_θ. Alternatively, if $v \leq -c$, we would reject the null in favor of the detrended model G_θ. Vuong's test statistics v are presented in table 2 and in a majority of cases (87%) support the time-varying specification.

Table 2 also presents goodness-of-fit comparisons for conventional models (model I) and time-varying models (model II) based on the Akaike Information Criterion (AIC) (Akaike (1974)) and Schwarz's Bayesian Information Criterion (BIC) (Schwarz (1978)). Smaller values of the AIC or BIC indicate a better fit. Both figure 3 and table 2 show that the time-varying Beta has the lowest AIC and BIC for most if not all counties, which indicates that it is the most parsimonious and optimal model that we have considered in this article. Moreover, $\Delta AIC(\Delta AIC = AIC - \min(AIC))$ and $\Delta BIC(\Delta BIC = BIC - \min(BIC))$ in table 2 are significantly large for the conventional detrended Beta model,[10] which also offers evidence in support of the time-varying model (see Burnham and Anderson, 2003).

Table 3 presents the results of comparisons of ten-year, out-of-sample forecasts,

[9]We can choose a critical value c from the standard normal distribution that corresponds to the desired level of significance (e.g. for $c = 1.96$; $Pr(z \geq |\pm c|) = 0:05$).

[10]As an example, $\Delta AIC = 88.16, \Delta BIC = 88.25$ for detrended model for Webster county soybean yields in Iowa.

two-step-ahead forecasts and a cross validation (leave-one-out) test. The out-of-sample forecast method essentially evaluates which method is better at forecasting the first moment of yields. This, of course, has direct relevance for the estimation of crop yield distributions and the subsequent rating of crop insurance contracts. Note, however, that these tests only compare models in one aspect of the yield distribution—the first moment (the mean). Thus, likelihood based specification tests may provide more information about goodness of fit for the entire distribution.

The cross-validation method ranks competing models based on their out-of-sample forecasting performance with some observations being randomly left out. For example, the "leave-one-out cross-validation test" is conducted for all counties considered for Iowa all-practice corn for the 82 years of county-level annual yields from 1926 to 2007. We drop each observation from the sample, fit the model, and use the estimates to forecast the omitted observation. The predicted and actual yields are compared to get the cross-validation Root Mean Squared Error (RMSE) in each period.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_{(i)})^2}$$

where $\hat{Y}_{(i)}$ is the prediction for Y_i obtained by fitting the model with observation i omitted.

We sum the cross-validation errors and obtain the RMSE for the two competing models. Results (table 3) indicate that the time-varying Beta distribution model out-performs the constant Beta model in most of the major agricultural production counties. Specifically, eight of the ten top Iowa corn production counties, nine of ten Iowa top soybean production counties, and six of seven Texas top cotton production counties exhibit a better cross-validation performance in the time-varying model. The resulting RMSEs of the time-varying model for these yield data are smaller than that of the conventional model. The differences of the RMSE between the two competing models are bigger for corn and cotton than soybeans. This is consistent with what we have observed in the practice of genetic improvement and biotechnological progress in agriculture. There have been less biotech innovations for soybeans than for corn and cotton and therefore the yield distribution of soybeans is less affected. As a result, the two competing methods do not make a big difference in the out-of-sample predictive power for soybeans yields. In addition to computing RMSEs, one may also compute the Spearman's correlation between the Y_i's and $\hat{Y}_{(i)}$'s or generate a Q-Q plot to check other distributional characteristics between the observed and (leave-one-out) predicted values.

In the current group risk crop insurance programs in the U.S., yields are forecast two years into the future. These forecasts are then used to establish insurance guarantees. In light of this, we considered an additional out-of-sample forecast evaluation intended to provide an analog to the forecasts used in these area-wide programs. In this approach, models are ranked based on their out-of-sample forecasting performance for a series of two-year ahead and ten-year ahead forward forecasts. For example, to predict 1993's yield, the estimates are based on the sample from 1926 to 1991; to predict 1994's yield, the estimates are based on the sample from 1926 to 1992, etc.

Another out-of-sample test is conducted by partitioning the entire sample into two parts and estimating parameters based on the first part of the data for the period 1926 to 1997 (the first 72 observations), then the estimated parameters are used to compute the expected (mean) yields for the out-of-sample period spanning 1998 to 2007 (the second part of the data). The mean of the squared difference between the predicted value and the actual yield value is calculated as a "leave-ten-out" forecast error $RMSE_{10}$.

The out-of-sample measures are computed for selected major crop/county combinations in the U.S. and such predictive measures again provide comprehensive evidence that the time-varying approach represents an improvement across all criteria considered. Table 3 shows that time-varying model has smaller values of both $RMSE_2$ and $RMSE_{10}$ in most cases. Having noted this, we must point out that the out-of-sample comparison test is only based on the accuracy of first moment mean prediction, which is not an overall evaluation of the entire yield distribution. Since the time-varying model is an alternative to the conventional two-stage model to estimate the yield distribution and to forecast the mean, these two models may display different out-of-sample performance based on different yield data in terms of mean prediction. Recall that strong evidence, as presented in table 2, supports the time-varying model's performance in estimating the entire yield distribution in terms of likelihood based tests and nonnested model distribution tests.

Table 4 presents alternative methods to comparing the two competing models. By using a regression method, we can consider which model's predicted values better explain the variation of the actual yields. To this end, we regress actual yields on each of the alternative predictions. The results indicate that only the coefficient on predicted yields from the time-varying model is significantly different from zero, which suggests the time-varying model yields a better prediction of the actual yield. Further, the intercept term is also not significantly different from zero, indicating that the chosen model has no systematic bias. Likewise, the coefficient on the time-varying model prediction is not significantly different from one, suggesting that the chosen model has no scale bias.

Simulation of a Group Risk Insurance Program

Yield–based insurance policies in the federal crop insurance program include the individual, farm-level multiple peril crop insurance (MPCI) and the county-level Group Risk Plan (GRP), which is based upon county-average yields from NASS. An important policy parameter in the GRP program is the premium rate, which is based on the county-average yield distribution. In this section, we evaluate the economic impacts of adopting rates based on the time-varying distribution methods. If the yield distributions change over time, premium rates should be adjusted accordingly. The premium rates from the proposed time-varying approach are illustrated with simulated data and a rate cross-validation test is conducted to compare the predictive accuracy of the premium rates from the time-varying approach with those of the conventional two-stage approach. Standard crop yield insurance pays an indemnity at a predetermined price to replace yield losses. Under the GRP, insured farmers collect an indemnity if the county average yield falls beneath a guarantee, regardless of the farmers' actual yields. Loss probabilities correspond to the likelihood that yields y below some threshold will be observed, which is given by the area under the density curve to the left of the guaranteed yield. Consider an insurance contract that insures some proportion ($\lambda \in (0,1)$) of the expected crop yield (y^e). If $y < \lambda y^e$, the insurer will pay $(\lambda y^e - y)p$ as an indemnity, where p is a predetermined price. An actuarially fair premium is defined by the expected loss of this contract, which takes the form of

$$\text{E(Loss)} = E[(\lambda y^e - y)I(y \leq \lambda y^e)]p = E[(\lambda y^e - y)^+]p \tag{6}$$

where $a^+ = \max(0, a)$ for a number $a \in \mathbb{R}$. In the preceding discussion, y denotes the observed annual county level yield and y^e represents the predicted (guaranteed) yield. Calculation of expected loss requires estimation of the distribution of yields. We compare the conventional two-stage estimation method to the proposed time-varying distribution in terms of expected loss and premium rates.

In our simulation, one million yields are generated from the time-varying Beta distributions. The probability of yield loss, the expected yield loss, and the actuarially fair premium rate associated with a contract that guarantees 75 percent of the expected yield is calculated for each year. As shown in figure 4, the premium rates range from 0.83 percent in 1985 to 0.36 percent in 2006 for the case in which the yields are from the time-varying model. The rates change as the moments of the time-varying distribution evolve. In contrast, the premium rates calculated from a conventional two-stage Beta distribution model (model I) indicate a constant premium rate around 1.88 percent from 1927 to 2006 (figure 4). For crop insurance in 2006, the premium rate from the detrended Beta model is 1.52 percentage points higher than the premium from the time-varying Beta model (0.36 percent versus 1.88 percent). Thus, the conventional model tends to significantly over-price the same level of coverage.

Rate cross-validation is proposed to measure the predictive accuracy of premium rates of one model when the alternative model is true. Rate cross-validation can be tested as follows:

Step 1: Assume one of the alternative yield distribution models, denoted by j, is true and simulate a set of actuarially fair premium rates (denoted as $r_{true_j,t}$).

Step 2: Simulate 1000 sets of 80 pseudo-observations of corn yields from the corresponding true yield distribution.

Step 3: Obtain 1000 sets of MLEs based on these pseudo-observations; then calculate the pseudo actuarially fair premium rates (denoted as $r_{j',t}$) based on the MLEs.

Then we can compare the pseudo premium rates with the true rates and obtain the Mean Percentage Error (MPE) and the Root Mean Squared Error(RMSE).

Cross-validation demonstrates a smaller MPE and RMSE for the time-varying model. As is shown in table 4, when the true rate is derived from the conventional model (with an average rate equal to 0.0188), the mean squared error (MSE) of predicted rates of the time-varying model is 0.0087, which is 9.58% lower than the MSE (0.0097) obtained from the conventional model when the alternative (the average premium rate implied by the time-varying model is 0.0058) is true. In addition, the MPE is 0.45 for the time-varying model and 1.66 for the detrended model. Smaller MPE and MSE values indicate that the time-varying model is more accurate, flexible, and robust in terms of premium rate prediction. This prediction error can also be expressed in economic terms. For example, for a crop insurance contract with $1000 liability per acre, the rate cross-validation error of the premium is $8.68 for the time-varying model. The rate cross-validation error of the premium is $9.60 for the conventional model. Therefore, the predicted premium error of the time-varying model is $0.92 less than the detrended model per unit of insurance ($1, 000 of total liability in this example). In light of the fact that the total premium in the federal crop insurance program in 2009 was nearly $80 billion, pricing errors can result in substantial aggregate losses. Consequently, the accuracy of insurance rates is improved by applying the time-varying yield distribution model.

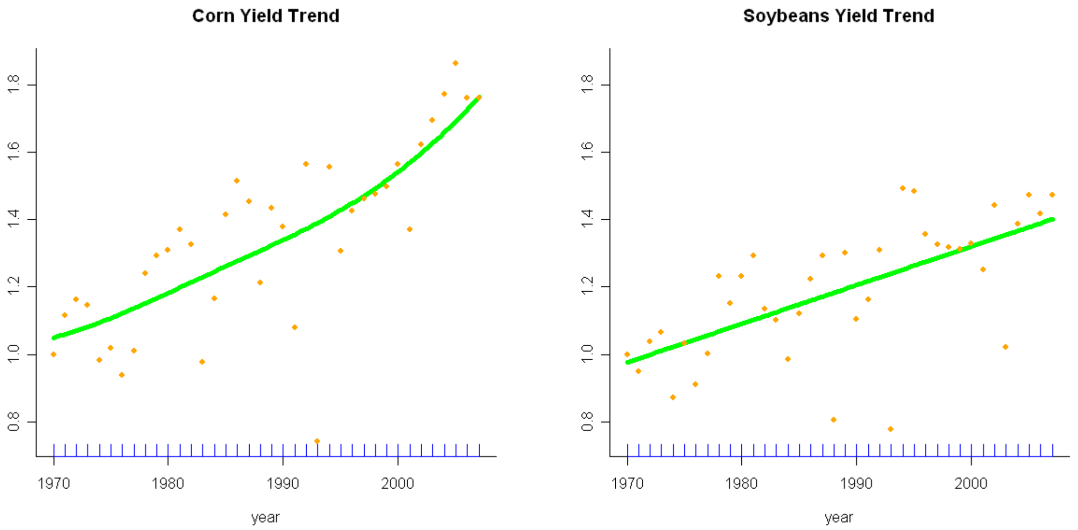
Conclusions

This study has examined the accuracy of alternative methods for measuring conditional yield risk under technological change. We propose a method for incorporating trends in the yield distribution that may offer a more accurate and consistent method for estimating the distribution of crop yields than other approaches commonly used in the literature. This method involves simultaneously estimating the time trend effects and the parameters of the yield distribution and therefore overcomes possible shortcomings associated with the more common approach of treating the detrended yields as "observed" data rather than data estimated from a previous detrending model.

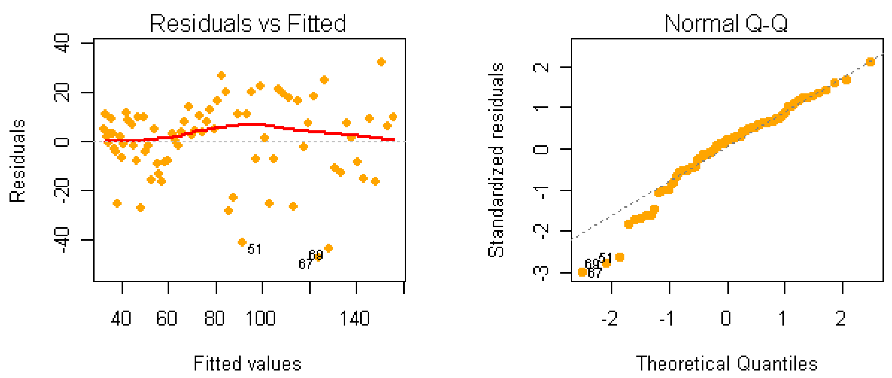
Several model selection tools are used to compare the in-sample goodness of fit and out-of-sample predictive power of the alternative models. The results show that the proposed time-varying model is superior to the conventional two-stage model in terms of providing a better fit (in terms of lower AIC and BIC criteria) and stronger out-of-sample predictive power for most of the major county/crop combinations. The results of out-of-sample prediction tests are consistent with prior expectations based on technological progress and biotechnology. In particular, multiple biotech traits and genetic improvements have occurred for corn and, to a lesser degree, for cotton. Much of the biotech innovations for soybeans have mainly involved herbicide tolerance. The proposed time-varying method therefore appears to offer greater improvement for corn and cotton than is the case for soybeans.

In a rate simulation exercise, the premium rate derived from the time-varying model showed significantly decreasing premium rates (from 0.83 percent in 1985 to 0.36 percent in 2006) over time, while the conventional model implied a constant rate (1.88 percent). A method of "rate cross-validation" demonstrated that the time-varying distribution model may offer significant advantages, even when the underlying yield trend process is misspecified.

Overall, this analysis reveals a dynamic evolution of yield distributions under technological change for major U.S. crop yields. In our data, which represents county-level yields for important crops in major growing areas, we find that the time-varying model provides a superior fit to the data. This study has policy implications that relate to improving the accuracy of assessing yield distributions in cases where parameters of the distribution evolve over time. When the distributions change, premium rates can be adjusted to represent the most recent information. This offers the potential to improve the accuracy of models used in rating crop insurance contracts and thus may improve risk management mechanisms to protect producers from risk. The improved time-varying method has practical implications for the GRP and GRIP programs as well as the design of other insurance contracts. Our applications assume a Beta distribution for each year. Future research may benefit from relaxing this assumption by using more flexible models such as a mixture of Beta distributions and nonparametric methods.

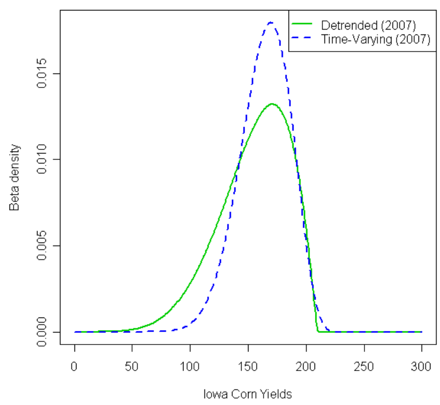


(a) Yield Trend of Different Crops (1970-2007)

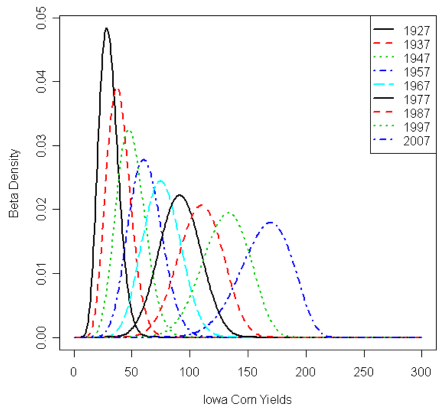


(b) Residual Plot of Annual Corn Yield, Adair County, Iowa

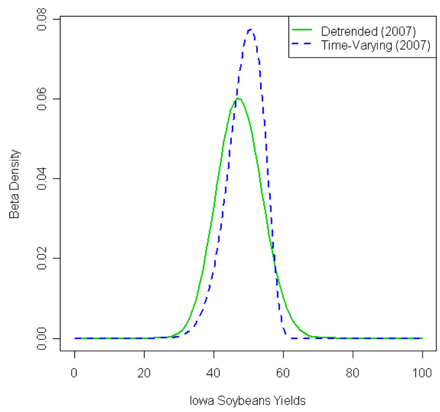
Figure 1: Scatter Plot and Residual Analysis



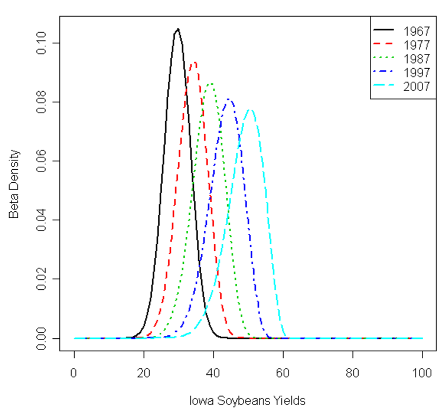
(a) Corn Yield Distribution of 2006: Detrended Beta vs. Time-Varying Beta



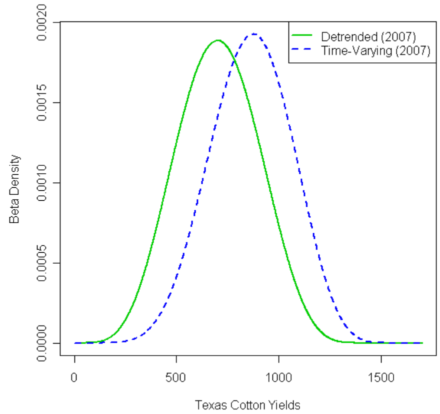
(d) 10-year Overlay Beta Density Plot for Corn



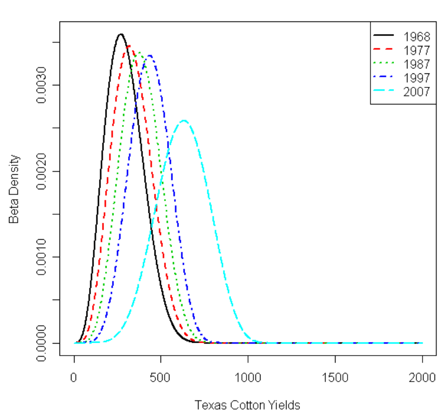
(b) Soybeans Yield Distribution of 2007: Detrended Beta vs. Time-Varying Beta



(e) 5-year Overlay Beta Density Plot for Soybeans

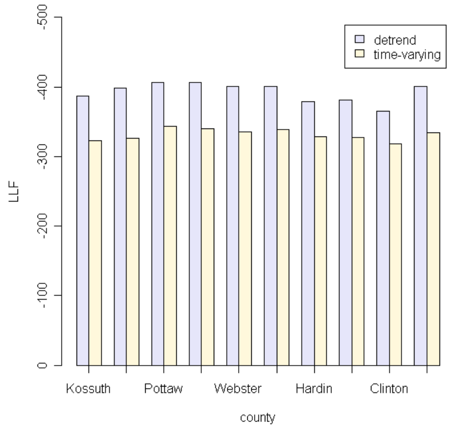


(c) Cotton Yield Distribution of 2007: Detrended Beta vs. Time-Varying Beta

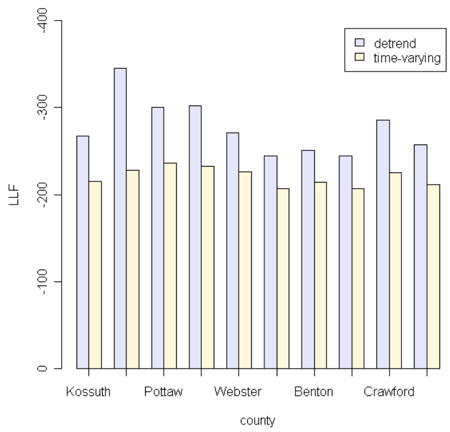


(f) 5-year Overlay Beta Density Plot for Cotton

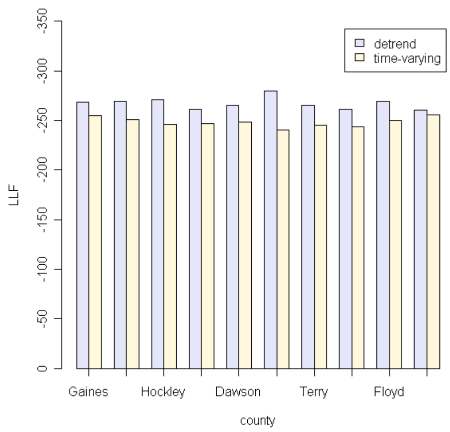
Figure 2: Estimated Time-Varying Beta Densities, Major Crop Yields in the U.S.



(a) LLF Comparison—Iowa Corn Yields



(b) LLF Comparison—Iowa Soybeans Yields



(c) LLF Comparison—Texas Cotton Yields

Figure 3: In-Sample Goodness-of-Fit Comparison of the Two Competing Models: LLF

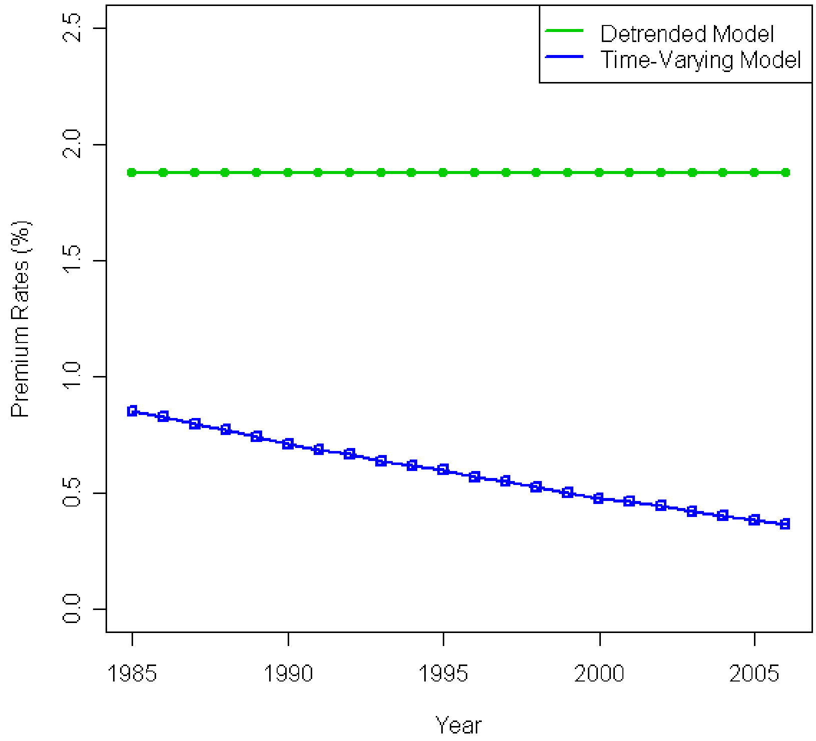


Figure 4: Premium Rates (for a 75% Coverage Level Crop Insurance Contract) for Time-Varying Model and Detrended Model (1985-2006)

Table 1: Maximum-Likelihood Parameter Estimates and Summary Statistics for Two-Stage Model and Time-Varying Models: Example for Adair County Corn Yields[c]

Two-Stage Model Based on Detrended Yield Data

Four-Parameter Beta (LLF = -378.69)			Three-Parameter Beta (LLF = -380.67)		
Parameters	**Estimates**	**Std. Error**	**Parameters**	**Estimates**	**Std. Error**
shape1(α)	5.99	0.21*	shape1(α)	5.99	0.19*
shape2(β)	2.10	0.23*	shape2(β)	2.07	0.23*
location(θ)	0.97	7.85	-	-	-
scale(δ)	203.43	1.04*	scale(δ)	204.13	1.07*

Time-Varying Models Based on Actual Yield Data

Linear Trend Structure[a] (LLF = -328.68)			Quadratic Trend Structure[b] (LLF = -326.62)		
Parameters	**Estimates**	**Std. Error**	**Parameters**	**Estimates**	**Std. Error**
b_1	2.38	0.32*	b_1	2.55	0.10*
b_2	0.43	0.75	b_2	0.16	0.40
b_3	—	—	b_3	-0.29	0.50
b_4	4.02	0.32*	b_4	2.95	0.10*
b_5	-2.71	1.29*	b_5	-1.63	0.30*
b_6	—	—	b_6	-2.61	4.8
b_7	7.47	14.99	b_7	12.26	117.70
b_8	-7.50	18.14	b_8	-15.27	138.15
b_9	—	—	b_9	-13.72	90.03

Time-Varying Models: **LLF(L): L1:** -328.68 **L2:** -326.62
LRT Statistics: $-2(L1-L2) = 4.12$ χ^2_4 p-value = 0.39

Notes: An asterisk * denotes statistical significance at the $\alpha = 0.05$ or smaller level
[a] the Time-Varying Beta Model with a linear trend structure is defined as: $y_t \sim (\alpha_t, \beta_t, 0, \delta_t)$ $\alpha_t = exp(b_1 + b_2\tilde{t})$; $\beta_t = exp(b_4 + b_5\tilde{t})$ $\delta_t = exp(b_7 + b_8\tilde{t})$
[b] the Time-Varying Beta Model with a quadratic trend structure is defined as: $y_t \sim (\alpha_t, \beta_t, 0, \delta_t)$ $\alpha_t = exp(b_1 + b_2\tilde{t} + b_3\tilde{t}^2)$; $\beta_t = exp(b_4 + b_5\tilde{t} + b_6\tilde{t}^2)$; $\delta_t = exp(b_7 + b_8\tilde{t} + b_9\tilde{t}^2)$
[c] Examples for other crops and counties are available from the author on request.

Table 2: Model Comparison Using In-Sample Goodness-of-fit Test and Non-nested Vuong's Test for Major Agricultural Yields

	Detrending Model–Model I				Time-Varying Model–Model II				
County	**K**	**LLF**	AIC/ΔAIC	BIC/ΔBIC	**K**	**LLF**	AIC	BIC	v^a
Iowa All-Practice Corn									
Kossuth	6	-386.90	785.80/129.17	785.28/129.17	6	-322.32	656.63	656.11	11.41*
Sioux	6	-398.085	808.17/143.62	807.65/143.62	6	-326.27	664.55	664.03	7.93*
Pottawattamie	6	-406.25	824.50/125.06	823.98/125.06	6	-343.72	699.44	698.92	12.61*
Plymouth	6	-406.47	824.94/133.44	824.43/133.44	6	-339.751	691.50	690.98	14.03*
Webster	6	-400.63	813.25/130.30	812.73/130.30	6	-335.48	682.95	690.98	13.15*
Pocohontas	6	-401.21	814.42/125.83	813.91/125.83	6	-338.30	688.60	688.08	12.95*
Hardin	6	-379.19	770.39/102.23	769.87/102.23	6	-328.08	668.16	667.64	10.76*
Franklin	6	-381.39	774.79/108.04	774.27/108.04	6	-327.37	666.75	666.23	8.55*
Clinton	6	-364.70	741.39/94.04	740.87/94.04	6	-317.67	647.35	646.83	8.99*
Woodbury	6	-401.22	814.45/133.34	813.93/133.34	6	-334.557	681.11	680.60	14.81*
Iowa Soybeans									
Kossuth	5	-267.22	544.44/102.06	544.01/102.15	6	-215.19	442.38	441.86	8.73*
Sioux	5	-345.65	701.3/233.88	700.87/233.97	6	-227.71	467.42	466.90	5.36*
Pottawattamie	5	-300.17	610.34/126.22	609.91/126.31	6	-236.06	484.12	483.60	8.27*
Plymouth	5	-302.13	614.26/136.72	613.83/136.81	6	-232.77	477.54	477.02	8.72*
Webster	5	-271.26	552.52/88.16	552.09/88.25	6	-226.18	464.36	463.84	8.79*
Woodbury	5	-244.77	499.54/73.28	499.11/73.37	6	-207.13	426.26	425.74	9.01*
Benton	5	-251.3	512.60/73.02	512.17/73.11	6	-213.79	439.58	439.06	9.57*
Grundy	5	-244.77	499.54/73.28	499.11/73.37	6	-207.13	426.26	425.74	9.01*
Crawford	5	-285.52	581.04/118.08	580.61/118.17	6	-225.48	462.96	462.44	6.07*
Tama	5	-257.38	524.76/89.44	524.33/89.53	6	-211.66	435.32	434.80	9.77*
Texas Upland Cotton									
Gaines	6	-268.23	548.46/27.38	547.94/27.38	6	-254.54	521.08	520.56	1.84
Lubbock	6	-269.56	551.12/38.22	550.60/38.22	6	-250.45	512.90	512.38	5.11*
Hockley	6	-270.87	553.74/50.14	553.22/50.14	6	-245.8	503.60	503.08	8.70*
Lynn	6	-261.55	535.1/29.56	534.58/29.56	6	-246.77	505.54	505.02	6.67*
Dawson	6	-264.87	541.74/32.74	541.22/32.74	6	-248.5	509	508.48	8.48*
Hale	6	-279.65	571.3/77.94	570.78/77.94	6	-240.68	493.36	492.84	2.33*
Terry	6	-264.96	541.92/40.34	541.40/40.34	6	-244.79	501.58	501.06	1.42
Crosby	6	-261.55	535.1/35.86	534.58/35.86	6	-243.62	499.24	498.72	1.29
Floyd	6	-268.92	549.84/37.26	549.32/37.26	6	-250.29	512.58	512.06	1.27
Martin	6	-260.08	532.16/8.86	531.64/8.86	6	-255.65	523.30	522.78	2.00*

Notes: An asterisk * denotes statistical significance at the $\alpha = 0.05$ or smaller level. K is the number of parameters in a model. "a" is the Vuong's test statistics for time-varying model vs. detrending model.

Table 3: Out-of-Sample Performance

County	Detrending Model–Model I			Time-Varying Model–Model II		
	$RMSE$	$RMSE_2$	$RMSE_{10}$	$RMSE$	$RMSE_2$	$RMSE_{10}$
Iowa All-Practice Corn						
Kossuth	14.54	22.21	22.62	14.52*	14.72*	10.26*
Sioux	13.52*	16.02	23.55	13.74	15.27*	14.56*
Pottawattamie	16.44	22.45	15.28*	16.05*	21.85*	19.13
Plymouth	15.57	19.45*	19.54*	15.56*	22.25	19.67
Webster	19.49	20.73	12.66	15.89*	17.72*	11.86*
Pocohontas	16.52*	22.13*	26.32*	16.58	22.19	29.18
Hardin	15.09	21.62	20.12	14.87*	19.13*	16.41*
Franklin	14.95	23.48*	18.51	14.50*	23.68	10.95*
Clinton	15.86	19.31*	24.17*	15.51*	19.57	26.63
Woodbury	14.76*	22.97	27.81	14.79	18.51*	16.12*
Iowa Soybeans						
Kossuth	4.12*	7.47	7.63	4.14	7.44*	7.60*
Sioux	4.18	5.13	5.67*	4.13*	4.82*	6.37
Pottawattamie	4.75	6.28	6.09*	4.73*	5.95*	6.27
Plymouth	4.74	4.73	5.06*	4.64*	4.06*	6.43
Webster	4.38	6.53*	6.36*	4.36*	6.61	6.92
Woodbury	3.74	6.09	5.82	3.69*	5.98*	5.71*
Benton	4.38	6.82	6.47	4.07*	5.61*	6.23*
Grundy	3.74	6.09	5.82	3.69*	5.98*	5.71*
Crawford	4.60	6.29	6.32	4.50*	6.12*	6.23*
Tama	3.99	6.67	6.20	3.96*	6.56*	6.18*
Texas Upland Cotton						
Gaines	130.72*	217.85	307.04*	130.90	217.23*	307.96
Lubbock	143.34	157.72	185.53*	128.04*	182.61*	196.43
Hockley	116.39	143.07*	194.23	100.13*	159.50	192.15*
Lynn	118.05	153.87	180.65	116.46*	136.78*	171.73*
Dawson	105.05*	96.13	84.69*	108.61	84.38*	163.53
Hale	155.32	187.42	239.46	113.24*	130.92*	116.32*
Terry	112.48*	174.85	277.63	129.23	133.25*	150.56*
Crosby	127.25	144.48*	153.71*	114.38*	165.91	161.32
Floyd	181.32	187.05	234.51	130.37*	158.82*	150.56*
Martin	146.23*	163.43	150.37*	148.57	153.27*	155.54

Note: an "*" indicates a smaller out-of-sample predicted error in the two competing models.

Table 4: Other Model Comparison Methods

Compared by Regression Method		
Variable	Parameter Estimate	p Value[a]
Intercept	-0.125	0.970
γ_1:Coefficient of Prediction Value of Detrended Beta	-0.065	0.890
γ_2:Coefficient of Prediction Value of Time-Varying Beta	1.068	0.034*

Rate Cross-Validation		
	Mean of True Rates from Conventional Model (0.01887)	Mean of True Rates from Time-varying Model (0.0058)
Root Mean Squared Error		
Conventional Predicted Rate (RMSE)	0	0.098
Time-varying Predicted Rate (RMSE)	0.093	0
Mean Percentage Error		
Conventional Predicted Rate (MPE)	0	1.66
Time-varying Predicted Rate (MPE)	0.45	0

Note: a: an "*" indicates statistical significance at the $\alpha = .10$ or smaller level.