Adaptive local parametric estimation of crop yields: Implication for crop insurance ratemaking

Zhiwei Shen[1*] Martin Odening[1] Ostap Okhrin [2]

1 Department of Agricultural Economics, Humboldt Universität zu Berlin, Berlin, Germany
2 Chair of Statistics and Econometrics esp. Transportation, Dresden University of Technology, Dresden, Germany

[*]Corresponding author: zhiwei.shen@agrar.hu-berlin.de

Paper prepared for presentation at the 156th EAAE Seminar

'Prospects for agricultural insurance in Europe'

Wageningen, Netherlands, October 3-4, 2016

Adaptive local parametric estimation of crop yields: Implications for crop insurance ratemaking

Zhiwei Shen, Martin Odening, Ostap Okhrin

Abstract

A rigorous estimation model of crop yields which ensures accurate and actuarially sound insurance premiums is of utmost importance to maintain sustainable and viable risk management solutions for producers, insurers, and governments. A major challenge in estimating crop yield models arises from non-stationarity of the data generating process due to technological change and climate change. In this paper, we introduce a local adaptive parametric approach to deal with the non-stationarity of crop yields and to estimate the time-varying parameters of crop yield models. Results from an empirical application to major crops in the U.S. indicate that the proposed model precisely captures the evolution of crop yield risks: yield risks for corn and cotton are decreasing, but are increasing for winter wheat. In terms of forecasting performance, the adaptive local parametric model, in general, outperforms the linear spline model that is commonly used in the current rating methodology. A rating analysis suggests that the proposed model has the potential to obtain more accurate rates and that most current insurance premium rates are overestimated for corn and cotton, but are underestimated for winter wheat.

Keywords: crop yields; adaptive estimation; local parametric approach; crop insurance pricing.

JEL codes: C14; Q19

1 Background

Since worldwide crop insurance programs have become increasingly expensive, it is not surprising that considerable attention has been paid to recover accurate premium rates of crop insurance contracts (Ker and Goodwin, 2000; Norwood et al., 2004; Ozaki et al., 2008; Annan et al., 2014; Ker et al., 2016). The estimation of yield distributions is of paramount importance for designing and rating crop insurance. A major challenge in estimating crop yield models arises from non-stationarity of the data generating process, i.e., the evolution of crop yield distributions over time cannot be adequately described by a simple model with constant parameters. For instance, an important feature of agricultural crop yields is that they usually show an upward trend over time and that deviations from the trend (residuals) frequently exhibit heteroscedasticity (see Figure 1 (a)). Major causes of non-stationarity are climate change and technological change. McCarl et al. (2008), for example, find that the mean and variance of key climate variables change over time. This, in turn, has a significant effect on average crop yields and yield variability. Tolhurst and Ker (2015) investigate the impact of technological developments on crop yields and conclude that technological change not only shifts the mean and variance of yields, but also affects other moments of the yield distribution. These changes make it difficult to determine the data generating process and to accurately model yield distributions using observed time series data. Hence, historical crop yield distributions need to be regularly updated. Otherwise, insurance losses and insurance premiums derived from these yield models will likely be biased.

The literature on crop yield modelling offers various approaches to deal with non-stationarity. Typically, a two-stage estimation procedure is applied that in a first step removes a trend component and heteroscedasticity from the data. Afterwards, a parametric or non-parametric distribution is fitted to the detrended data. The dynamics of average yields are captured by either a deterministic or a stochastic trend. Deterministic time trend models are dominant in the literature and consist of a simple linear trend, polynomial trend (Just and Weninger, 1999), and spline functions (Harri et al. 2011) (see Figure 1(a)). Trends using stochastic approaches have been estimated by the Kalman filter

(Kaylen and Koroma, 1991) or autoregressive integrated moving average (ARIMA) process (Goodwin and Ker, 1998). Harri et al. (2009) provide an empirical comparison of the deterministic and stochastic time trend models and find limited support for stochastic trends in crop yields. A time trend model only captures mean shifts of the crop yield distribution, but a growing body of empirical evidence has shown that higher-order moments also vary over time due to climatic change or technological change (Yu and Barcock, 2010; Edgerton et al., 2012). To account for the adjustment of heteroscedasticity in historical yields, Skees et al. (1997) assume proportional heteroscedasticity over time, i.e., that the standard deviations of the residuals increase proportionally with increases in yields. Harri et al. (2011) doubt the universal validity of the assumption and find that arbitrarily imposing a specific form of heteroscedasticity in insurance rate calculations limits actuarial soundness. To capture the non-stationarity of other higher moments, Zhu et al., (2011) propose a time-varying yield distribution model by allowing location, scale, skewness, and kurtosis parameters to evolve over time, while Tolhurst and Ker (2015) suggest using a mixture of Normals with embedded trend functions to account for potentially different rates of technological change in different components (e.g., mean, variance, and skewness) of the yield distribution.

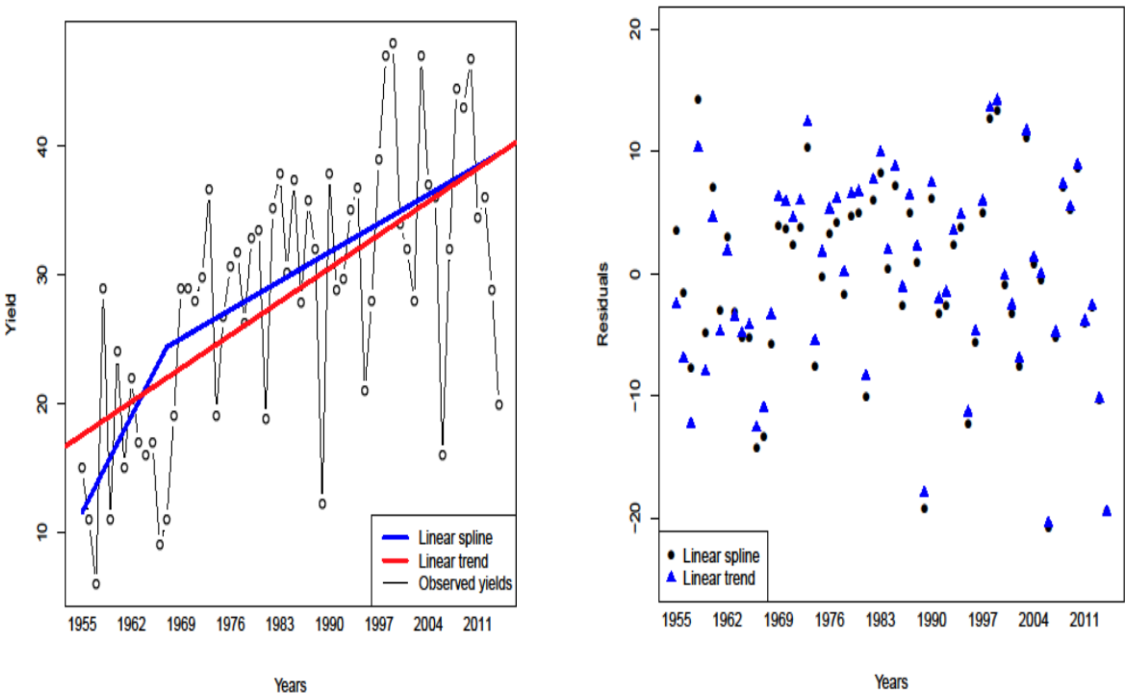
A further question related to the non-stationarity of the data generating process concerns the appropriate length of the sample period to be used for the calculation of crop insurance rates. In case of time-varying parameters, a shorter interval of historical data might be appropriate for estimation purposes. In the U.S., for example, the Risk Management Agency (RMA) uses less than 20 years of loss observations to calculate crop insurance premiums, arguing that yield losses from more than 20 years ago may not be representative for current agricultural risk despite data normalization (Ker et al., 2016). On the other hand, researchers have casted doubt on the use of short samples of available data since it may not be sufficient to properly determine crop yield loss distributions from short historic data, such as 30 years of data (Coble et al., 2010; Smith and Goodwin, 2010). The lack of historical data may lead to incorrect estimations of time trend and yield distribution and introduces a new model risk into the insurer's decision problem (Courbage and Liedtke, 2003). Indeed, different sample period selections, such as 20 or 30 years, are very likely to lead to different mean forecasts and risk assessments for crop yields (see Figure 1 (b) and Figure A1 (b)). Thus, insurers face a trade-off when choosing the appropriate sample length: in general, shorter intervals will result in larger variations compared to longer intervals, while longer intervals lead to increases in the modelling bias. The difference in using different sample lengths seems more prevalent in detrending corn yields (Figure A1) than detrending wheat yields (Figure 1). Despite the importance of this issue, studies that investigate the impacts of sample length selection in crop insurance pricing are rare. Woodard (2014) compares weather distributions based on 30 year data and 115 year data and finds little efficiency gains in ratemaking from using the longer period. On the other hand, Ker et al. (2016) report significant differences in the rating performance depending on the length of sample period of crop yield data.

A feature of the aforementioned approaches dealing with non-stationarity of crop yields is that they allow for some or all of model parameters to vary over time based on all of the observed historical data. An identification of these models requires either structural assumptions about the transition process over time or presumes that the parameters follow smooth functions of time. For instance, the linear spline models commonly used by the RMA require knowledge about the number and time of structural breaks. Meanwhile, answers to the question of optimal sample period length have remained inconclusive and the assumption of an arbitrary length of a rolling estimation window for all time points is rather restrictive. To address these limiting issues, we pursue an alternative flexible approach that is based on the local parametric assumption, i.e., an arbitrary nonstationary process can be well approximated by a simple time-homogeneous model within a given time interval (c.f. Spokoiny 2009). Such an approach makes sense, if the variability of yields is high compared with the variability of the underlying model parameters, so that the latter can be estimated from more recent data. A cornerstone of the adaptive local estimation procedure is the detection of change points at which the homogeneity assumption of the model parameters does no longer hold. The estimation procedure embeds a multiple change point test that rests on a small modelling bias condition. The small modelling bias condition means that the distance between the true and the estimated model parameter is bounded by a small constant with a high probability (in Section 2 we provide a formal definition of the small modelling bias condition). Based on this condition, the adaptive local estimation procedure identifies the largest data set for which a homogeneous parametric model can be assumed. In a way, this approach addresses the bias-variance trade-off in statistics: the longer the estimation period, the lower the variance of the parameter estimate, but increasing the data set entails the risk that the data no longer follow the same parametric model. The small modelling bias condition provides a well-defined solution for dealing with this trade-off. The theoretical properties distinguish this method from other heuristic procedures that also target the bias-variance trade-off, such as rolling window estimation or exponential smoothing. Recently, this local parametric approach has been successfully applied in many research fields, such as localizing temperature risk (Härdle et al., 2016), yield curve term structure modelling (Chen and Niu, 2014), localized realized volatility forecasting (Chen et al., 2010), and time-varying GARCH modelling (Čížek et al., 2009).

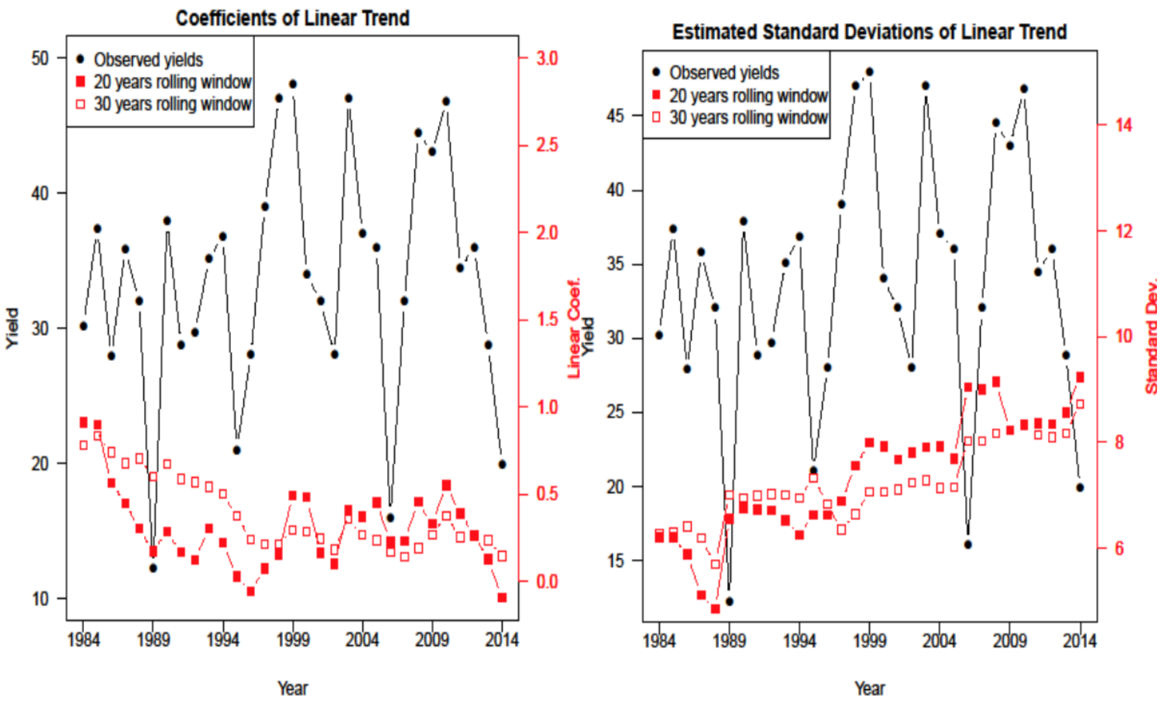
The objective of this study is to present a flexible and parsimonious local model to capturing the non-stationary nature of crop yields and to adaptively estimate crop yields and insurance rates. To be specific, we propose the adaptive local parametric approach to model county-level crop yield data in the U.S. We begin with a simple linear time trend model as the underlying local parametric model. The time-varying parameters of the local model are determined via adaptive data-driven statistical techniques. The idea is to find for each time point an optimal longest past time estimation interval for which the assumption of a (local) parametric model with constant parameters holds. The selection of the optimal longest interval is crucial to the procedure and is accomplished by a backward sequential testing procedure for each considered interval candidate. To examine the performance of the proposed model, we conduct an out-of-sample forecast as well as an insurance rating game that mimics the rating procedure applied by the RMA in the U.S. federal crop insurance program.

The contribution of this article to the existing literature is threefold. First, to the best of our knowledge, this is the first time to introduce a local adaptive procedure to deal with the non-stationarity of crop yields and to estimate the time-varying parameters of crop yield models. The results indicate that this approach has the potential to improve the quantification of crop yield risks and the estimation of crop insurance premiums. Second, the study contributes to the longstanding debate on the selection of the optimal sample period for crop yield data. Unlike previous ad hoc analyses, this paper offers a novel data-driven perspective to adaptively determine the appropriate sample length via a backward sequential testing procedure. Third, our empirical results contribute to the "increasing-risk-hypothesis" that has been raised in the climate change literature and has immediate implications for the pricing of crop insurance (Carriquiry and Osgood, 2012; Wang et al., 2013).

The rest of the paper is structured as follows. Section 2 describes the local adaptive crop yield model framework. We provide a brief overview of the appealing statistical properties of this approach and then describe the local parametric estimation procedure in detail, particularly the local change point test and the adaptive estimation. In Section 3, the model is applied to crop yield data such as winter wheat, corn, soybean, and cotton in the U.S. We present the empirical models for two representative counties for each crop, followed by a discussion of their estimation results. We also presents two assessments of forecast performances of different models in terms of crop yields and insurance rates, respectively. The final section provides a conclusion and discussion on the potential of local adaptive crop yield models and offers suggestions for further research.



(a) Winter Wheat in Ellis County, Kansas



(b) Linear time trend estimation based on rolling window data samples

Figure 1. Scatter plot and time horizon selection

2 Adaptive Local Parametric Approach

To capture the dynamics of crop yields, we consider a local perspective and develop an adaptive parametric crop yield model. The underlying idea is to find an optimal longest estimation window length (so-called *interval of homogeneity*) over which one fits a local parametric model (such as a simple linear trend) with constant parameters. For each time point, the interval of homogeneity is determined in a backward sequential testing procedure. At the beginning of the procedure, a small past-time interval of each point is taken to estimate the (locally) constant parameters. Then, the interval is iteratively extended and tested for local homogeneity of estimated parameters. The procedure ends when local homogeneity is rejected, i.e., a significant difference is detected in the values of the estimated parameters between current and previous intervals. At the end of the procedure, the longest local estimation window is chosen and the local parametric model is estimated for each time point. The resulting window size allows us to determine the current rate of technological change and mitigate the potential bias caused by historical crop yield data from decades ago.

While this approach has been applied to many complex models in the financial literature (e.g., Chen et al., 2010; Chen and Niu, 2014; Härdle et al., 2016), we incorporate the approach into the standard two-stage estimation framework for modelling crop yield risk. At the first stage, a simple linear time trend model is adopted as the underlying local parametric model, which is reasonable for short time intervals, such as 10 years. We assume that the crop yield y_t, at time $t \in (1, \dots, T)$ is as follows:

$$y_t = \alpha_t + \beta_t t + \epsilon_t, \tag{1}$$

where α_t, β_t denote the time varying local linear trend coefficient at time t and ϵ_t is a mean-zero random error term over a fixed interval $I = [t-m, t]$ of $(m+1)$ $(m \leq t)$ observations. To mitigate the effect of outliers on trend estimation, a robust iterative reweighted least square Huber M-estimator is employed (Finger, 2010; Annan et al, 2014). In the second step, the detrended yields $\tilde{y}_t$ are assumed to follow a normal distribution[1] with mean μ_t and time-varying variance σ_t^2. The distribution parameters $\theta_t = (\mu_t, \sigma_t)$ are time dependent and can be estimated by a (quasi) maximum likelihood estimation over a fixed interval I of detrended yields $\tilde{y}_t$. For time point t_0 the (quasi) maximum likelihood estimator $\tilde{\theta}(t_0)$ of $(m+1)$ detrended yields observations is defined as:

$$\tilde{\theta}(t_0) = \underset{\theta \in \Theta}{\operatorname{argmax}} L(\tilde{y}; I, \theta), \tag{2}$$

where $\Theta = R^+ \times R^+$ denotes the parameter space and $L(\tilde{y}; I_s, \theta)$ is the local log-likelihood function. We refer to $\tilde{\theta}(t_0)$ as the local maximum likelihood estimator. Note that the only difference with the conventional two-stage estimation is the fact that the local parametric approach depends on optimally selected time intervals *I*. The following sections explain the statistical technique of selecting the optimal interval based on which local homogeneity of estimated parameters hold. Since time varying trend coefficients and distribution parameters are both determined by the same selected intervals, we refer to $\theta_t = (\alpha_t, \beta_t, \mu_t, \sigma_t)$ from now on.

2.1 Small Modelling Bias Condition

A longer data period reduces the variance of the parameter estimate, but increasing the data set entails the risk of introducing a bias. To assess the trade-off between variability and bias of the parameter estimate and to determine the homogenous interval I, one has to quantify the quality of approximating the true (unknown) process y_t over an interval I by a parametric model $y_t(\theta)$ with constant parameter θ. The quality of the approximation is measured by the Kullback-Leibler divergence. For every interval I and every parameter $\theta \in \Theta$, this measure is defined by:

$$\Delta_I(\theta) = \sum_{t \in I} \mathcal{K}\{f(y_t(\theta)), f(y_t)\} = \mathrm{E}_\theta \log \frac{f(y_t(\theta))}{f(y_t)}, \tag{3}$$

where $f(\cdot)$ denotes a density distribution. The value $\Delta_I(\theta)$ measures the discrepancy between the underlying "true" process and the parametric model and thus enables us to control the modelling bias. To select the optimal choice interval, Čížek et al. (2009) introduce the small modelling bias

[1] Though the assumption of normal distributions for modelling crop yields has been criticized, most studies focus on fitting a global model and this assumption has not been investigated in the context of local parametric models. Previous studies on local parametric models have implied that normality is more likely to exist in a local window rather than in long time series data (Andriyashin et al., 2006; Wang et al., 2013; Härdle et al., 2016). Therefore, it is reasonable to assume that this assumption might be valid.

(SMB) condition, i.e., for some $\theta \in \Theta$, $\Delta_I(\theta)$ is bounded by a small constant with a high probability. Formally, for some $\theta \in \Theta$ and $\Delta > 0$,

$$\mathrm{E}_\theta \Delta_I(\theta) \leq \Delta. \tag{4}$$

Thus, the "true" model can well approximate on the interval I with parameter θ while keeping the modelling bias small according to Equation (4). The best parametric estimation on interval I can be defined by minimizing $\mathrm{E}\Delta_I(\theta)$ over $\theta \in \Theta$.

The risk arising in the estimation of locally constant modelling under SMB is bounded. Under the SMB condition, Polzehl and Spokoiny (2006) and Čížek et al. (2009) show that the estimation loss normalized by parametric risk bound $\mathfrak{R}_r(\theta)$ is stochastically bounded. In the case of a quasi-MLE estimation with loss functions ($L\big(I, \tilde{\theta}_I, \theta\big) = |L\big(I, \tilde{\theta}_I\big) - L(I, \theta)|$), Čížek et al. (2009) show that if they let $\tilde{\theta}_I$ be MLE estimators on an interval I and if the SMB holds for some I and $\theta \in \Theta$, then

$$\mathrm{E}_\theta \left\{ \log\left[1 + \frac{\left|L(I,\tilde{\theta}_I) - L(I,\theta)\right|^r}{\mathfrak{R}_r(\theta)}\right] \right\} \leq 1 + \Delta, \tag{5}$$

where $\mathfrak{R}_r(\theta)$ is an upper bound satisfying the following condition given the true parameter θ^*:

$$\mathrm{E}_{\theta^*} \left| L\big(I, \tilde{\theta}_I\big) - L(I, \theta^*) \right|^r \leq \mathfrak{R}_r(\theta^*). \tag{6}$$

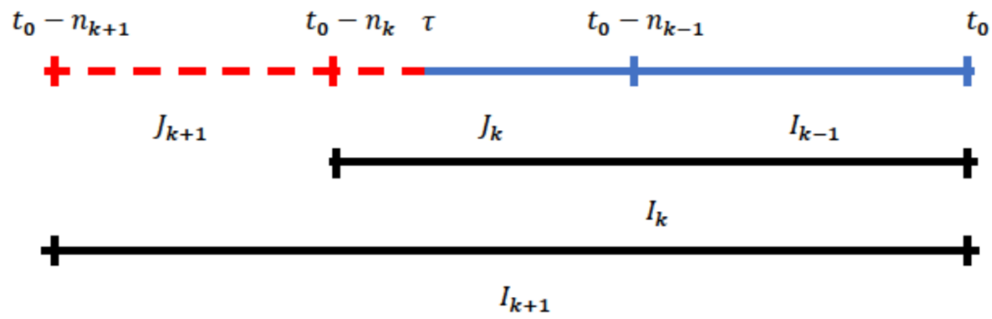
Equation (6) is called a "propagation condition". The bound given in equation (5) indicates that the risk in an estimated local constant model (under SMB) differs from the risk in the true constant model by a constant proportional to e^Δ. The risk bound in Equation (6) allows us to define the likelihood based confidence sets that can be used to determine critical values in the local homogeneity tests in sections 2.2 and 2.4. For more details, refer to Čížek et al. (2009) and Spokoiny (2009).

2.2 Local Change Point Detection Test

The local parametric approach crucially rests on the sequencing test of local time-homogeneity to search for an interval of homogeneity among the considered intervals I_k $(k = 0, 1, \ldots, K)$ at a fixed time point t_0. Here, we follow Härdle et al. (2015) and Čížek et al. (2009) and adopt the local change point detection test, in which the null hypothesis on parameter homogeneity for the intervals up to I_k is tested against the alternative hypothesis that a change point at unknown location τ within interval I_k exists. Assuming that the homogeneity assumption of interval I_{k-1} has not been rejected, the test statistic for testing possible change points in interval I_k is defined via the corresponding fitted log-likelihood $L(y; I, \theta)$ by:

$$T_k = \sup_{\tau \in J_k} \{ L_{A_{k,\tau}}(y, A_{k,\tau}, \tilde{\theta}_{A_{k,\tau}}) + L_{B_{k,\tau}}\big(y, B_{k,\tau}, \tilde{\theta}_{B_{k,\tau}}\big) - L_{I_{k+1}}(y, I_{k+1}, \tilde{\theta}_{I_{k+1}})\}, \tag{7}$$

where $J_k = I_k \backslash I_{k-1}$, $A_{k,\tau} = [t_0 - n_{k+1}, \tau]$ and $B_{k,\tau} = (\tau, t_0]$ represents two parts of observations in interval I_{k+1}. Since the change-point location τ is generally unknown, the test statistic is defined to consider the supremum of the log-likelihood ratio statistics over $\tau \in J_k$. Figure 2 visualizes the underlying construction of the test statistic. Suppose that at a fixed time point t_0 the parameter homogeneity within I_{k-1} has not been rejected. To test if the homogeneity of the interval should be extended to I_k, we examine all possible breakpoint τ within the data extension $J_k = I_k \backslash I_{k-1}$ by calculating the supremum of two log-likelihood values over the intervals $A_{k,\tau}$ and $B_{k,\tau}$ for any $\tau \in J_k$ and then compare it to the log-likelihood value estimated over I_{k+1}. The decision rule of the test requires the comparison of T_k with the corresponding critical values ξ_k. We reject the null hypothesis of parameter homogeneity if $T_k > \xi_k$.



The red dotted and blue line interval represent $A_{k,\tau}$ and $B_{k,\tau}$, respectively.

Figure 2. Graphical illustration of the construction of test statistics T_k.

2.3 Adaptive Estimation

In this section, we introduce adaptive estimation procedures. At a fixed time point t_0, we use historical observed data $Y_t, t \le t_0$ to estimate the unknown parameters $\theta(t_0)$ and repeat the procedure for each newly included time point t_0. The objective is to select the longest interval of homogeneity of Y_t over which the homogeneity assumption of the parametric model holds. Since the number of possible interval candidates can be large, we consider only a finite set of intervals, e.g., K+1 increasingly nested intervals $I_0 \subset I_1 \subset \cdots \subset I_K$. For each interval, the corresponding quasi-ML estimators $\tilde{\theta}_{I_0}(t_0), \tilde{\theta}_{I_1}(t_0), \ldots, \tilde{\theta}_{I_K}(t_0)$ can be determined for a fixed time point t_0. From now on, we ignore the index t_0 and describe the procedure for an arbitrary fixed time point. Let $\hat{\theta}_{I_k}$ refer to the accepted adaptive estimator in the interval I_k, and let $\hat{\theta}$ denote the optimal estimator based on the longest homogeneous interval $\hat{I}$.

The selection algorithm is built on a sequential testing procedure. It starts from the shortest interval I_0 over which the local homogeneity holds by assumption and the maximum likelihood estimator $\tilde{\theta}_{I_0}$is accepted, i.e., $\hat{\theta}_{I_0}$=$\tilde{\theta}_{I_0}$. Then, we iteratively extend to next longer intervals I_k over which the local change point test is conducted using Equation (7) to test the hypothesis of local homogeneity provided that the null hypothesis has not been rejected over I_{k-1}. The selected interval $\hat{I}$ corresponds to the longest accepted interval $I_{\hat{k}}$, such that:

$$T_k \le \xi_k,\ k \le \hat{k} \text{ and } T_{\hat{k}+1} > \xi_{\hat{k}+1}, \tag{8}$$

where ξ_k is the critical value at step k. The derivation of critical values is described in Section 2.4. Equation (8) indicates that a change point is detected over interval $I_{\hat{k}+1}$, i.e., parameter homogeneity of the interval $I_{\hat{k}+1}$ is rejected and extending the interval to $I_{\hat{k}+1}$ will introduce significant bias. As a last step, the longest accepted interval is $\hat{I} = I_{\hat{k}}$, resulting in the adaptive optimal estimator $\hat{\theta} = \hat{\theta}_{I_{\hat{k}}}$ for the fixed time point t_0. In summary, the procedure for a fixed time point t_0 is provided as follows:

1) Start with the smallest interval, $\hat{I} = I_0$, $\hat{\theta}_{I_0}$=$\tilde{\theta}_{I_0}$.
2) For $k = 1$, we test the interval I_1 for local homogeneity assumption. Select intervals I_2, I_1, and $J_1 = I_1 \backslash I_0$. If $T_1 \le \xi_1$, $\tilde{\theta}_{I_1}$ is accepted then $\hat{\theta}_{I_1}$=$\tilde{\theta}_{I_1}$. Otherwise, $\hat{\theta}_{I_1}$=$\hat{\theta}_{I_0}$, we accept the parameter estimator from the smallest interval as the optimal estimator for t_0.
3) For $k \ge 2$, select intervals I_{k+1}, I_k, and $J_k = I_k \backslash I_{k-1}$. $\tilde{\theta}_{I_k}$ is accepted and $\hat{\theta}_{I_k}$=$\tilde{\theta}_{I_k}$ if $T_k \le \xi_k$ and $\tilde{\theta}_{I_{k-1}}$ has not been rejected. Otherwise, $\hat{\theta}_{I_k}$=$\hat{\theta}_{I_{k-1}}$, where $\hat{\theta}_{I_k}$ is accepted after k steps.
4) The final estimate is $\hat{I} = I_{\hat{k}}$ and $\hat{\theta} = \hat{\theta}_{I_{\hat{k}}}$.

2.4 Calculation of Critical Values

Since the true distribution of the test statistic is unknown, the critical values have to be determined by simulation using the general approach of testing theory: to provide a prescribed performance of the procedure under the null hypothesis (Čížek et al., 2009; Chen and Niu, 2014; Härdle et al., 2015).

To be specific, we simulate 1,000 global homogeneous processes, i.e., a linear time trend model with constant parameters θ^* in model (1). The simulated data ensure homogeneity for all of the considered intervals. Under the null hypothesis of parameter homogeneity, the correct choice in the pure parametric situation is the largest considered interval I_K, over which the estimation loss of the ML estimator fulfill the risk bound Equation (6). If the adaptive procedure stops earlier at I_k with $k < K$, instead of $\tilde{\theta}_{I_K}$ we select the adaptive estimator $\hat{\theta}_{I_k} = \tilde{\theta}_{I_k}$, which can be interpreted as a "false alarm". The loss associated with such a false alarm is defined by $L_{I_K} = L(I_K, \tilde{\theta}_{I_K}) - L(I_K, \hat{\theta}_{I_k})$ and the corresponding risk bound given by Equation (6) due to the adaptive estimation changes to:

$$\mathrm{E}_{\theta^*}\left|L(I_K, \tilde{\theta}_{I_K}) - L(I_K, \hat{\theta}_{I_k})\right|^r \leq \rho \mathfrak{R}_r(\theta^*). \tag{9}$$

This implies that critical values ensures that the loss associated with a "false alarm" is at most a ρ-fraction of the parametric risk bound by the "optimal" or "oracle" estimate $\tilde{\theta}_{I_K}$ (Härdle et al., 2015). We select minimal critical values to ensure a small probability of such a false alarm. Similarly, at each step of the adaptive procedure, the estimate $\hat{\theta}_{I_k}$ after the k steps should satisfy:

$$\mathrm{E}_{\theta^*}\left|L(I_k, \tilde{\theta}_{I_k}) - L(I_k, \hat{\theta}_{I_k})\right|^r \leq \rho_k \mathfrak{R}_r(\theta^*),\ k = 1, \dots, K, \tag{10}$$

where $\rho_k = \rho k/K$ and $\mathfrak{R}_r(\theta^*) = \max\limits_k \left|L(I, \tilde{\theta}_{I_k}) - L(I, \theta^*)\right|^r$. The parameter ρ is the level of significance and influences the sensitivity of the procedure to homogeneity. Čížek et al. (2009) show that large values of ρ lead to smaller critical values.

Given that the sequential testing procedure is used in a local change point test, the critical values are computed through the following two steps:

Step 1. Consider first ξ_1 and let $\xi_2 = \xi_3 = \dots = \xi_K = \infty$. This leads to the estimates $\hat{\theta}_{I_k}(\xi_1)$ and the value ξ_1 is selected as the minimal one for which

$$\sup \mathrm{E}_{\theta^*}\left|L(I_k, \tilde{\theta}_{I_k}) - L(I_k, \hat{\theta}_{I_k}(\xi_1))\right|^r \leq \rho \mathfrak{R}_r(\theta^*)/K,\ k = 2, \dots, K, \tag{11}$$

Step 2. Suppose that $\xi_1, \xi_2, \dots, \xi_{l-1}$ is fixed from previous steps, and set $\xi_l = \dots = \xi_K = \infty$. With estimate $\hat{\theta}_{I_k}(\xi_1, \dots, \xi_l)$ for $k = l+1, \dots, K$, we find ξ_l as the minimal value which fulfills

$$\sup \mathrm{E}_{\theta^*}\left|L(I_k, \tilde{\theta}_{I_k}) - L(I_k, \hat{\theta}_{I_k}(\xi_1, \dots, \xi_l))\right|^r \leq \rho k \mathfrak{R}_r(\theta^*)/K,\ k = l+1, \dots, K. \tag{12}$$

3 Application to Crop Yield Insurance

3.1. Data and Model Specification

Due to data availability and quality, we utilize annual U.S. county-level crop yield data from the National Agricultural Statistical Service (NASS) to investigate the performance of the local parametric crop yield model. The considered crops are winter wheat, corn, soybean, and cotton. We choose two states for each crop according to their important role in the national production of each crop (see Table 1). County-level crop yield data cover the period from 1955 to 2014 with the exception of winter wheat in Texas, where data from 1968 to 2014 are available. After excluding counties without continuous yield records, we have data from 106, 178, 187, and 34 counties for winter wheat, corn, soybean and cotton, respectively. The broad coverage of the data set allows us to explore the importance of non-stationarity for different crops and different graphical locations. Practical relevance arises from the fact these data form the basis of area yield insurance programs, such as the group risk plan implemented by the RMA[2]. Summary statistics of yield data are provided in Table 1.

[2]All data and codes will be made available by the authors upon request.

Table 1. Summary Statistics of Yield Data

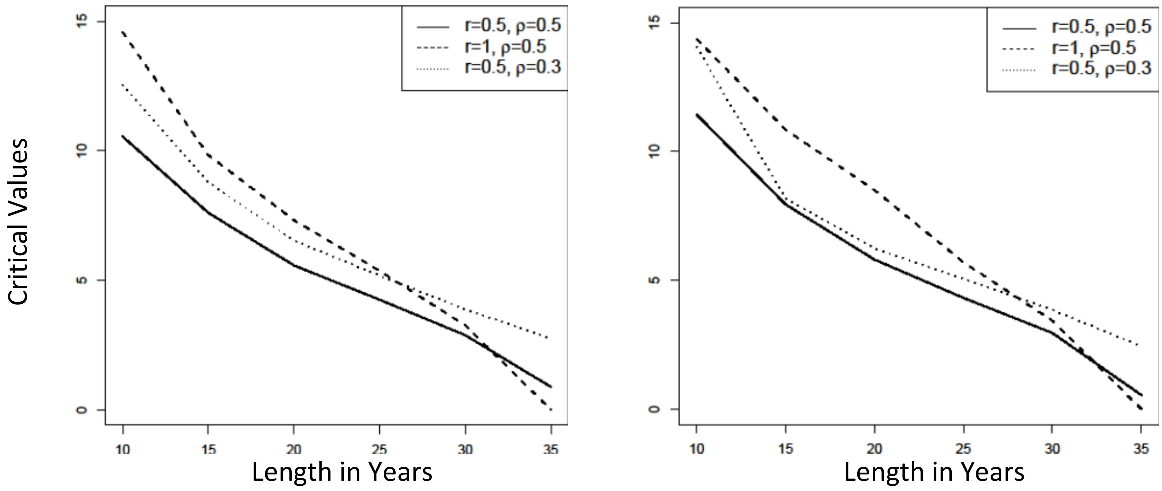
State	Number of Counties	Min.	Mean	Median	Max.	Std. Dev.	Coef. of Variation
			Winter Wheat				
Kansas	61	5.0	32.7	32.8	80.0	10.3	0.31
Texas	45	6.4	26.5	25.5	64.6	9.5	0.36
			Corn				
Illinois	79	19.0	114.5	114.0	236.0	39.2	0.34
Iowa	99	18.3	115.0	114.7	206.6	38.9	0.34
			Soybean				
Illinois	89	9.5	36.5	36.0	69.3	9.7	0.27
Iowa	98	7.3	37.3	36.9	64.0	9.6	0.26
			Cotton				
Georgia	20	127.0	589.1	564.0	1264.0	212.3	0.36
Mississippi	14	237.0	726.9	711.0	1482.0	208.8	0.29

The proposed adaptive local parametric approach depends on a set of parameters, namely, the considered interval candidates, the power index r, and the significance level value ρ. In the following, we justify the specification of these parameters. First, motivated by empirical applications in the literature we select $(K+1) = 8$ nested intervals from 5 to 40 historical observations, i.e., $\{5,10,15,20,25,30,35,40\}$. Here, we assume that the shortest interval (5 years) is always homogeneous and test if the homogeneity assumption applies to longer intervals via the local change point test. The longest interval includes 40 observations. This choice allows us to use the remaining data to evaluate the out-of-sample performance of the proposed method. Second, the hyper-parameters r and ρ are crucial to the calibration of risk bounds and critical values ξ_k. It has been shown that higher values of r lead to the acceptance of longer intervals of homogeneity and thus a higher modelling bias. Increasing ρ generally leads to an overall decrease of critical values ξ_k. We follow Chen and Niu (2014) and Härdle et al. (2013) and consider $r = 0.5$ and $\rho = 0.5$. The robustness and sensitivity of empirical results to different hype-parameters will be discussed in Section 3. Third, critical values also depend on the "true" parameters θ^*used in the Monte Carlo simulation. Fortunately, the previous studies document that the results are, in general, robust with respect to the selection of hypothetical parameters θ^*(Chen and Niu, 2014; Härdle et al., 2016). In our application, θ^* is specified as parameter estimates over the longest interval (i.e., 40 years starting from 1955).

3.2 Empirical Results

In this section we illustrate the implementation of the proposed local parametric crop yield model for eight representative combinations of crops and counties. The crucial step of the proposed procedure is to obtain critical values for the local change point test, for which we need to calculate the risk bound first. The simulated risk bounds $\mathfrak{R}_r(\theta^*)$ for each selected combination of crop and county are presented in Table A1 in the Appendix. We find risk bounds for different crop yields to be rather similar. This similarity might be due to the fact that the estimated θ^* based on a 40-year sample do not vary much across crops and regions, which is in line with the finding in Tolhurst and Ker (2015). As expected, larger values of risk power parameter r lead to larger values for risk bounds.

With simulated risk bounds at hand, critical values can be derived from Equation (11) and (12). These values for winter wheat and corn in two representative counties are displayed in Figure 3 for different significance levels ρ. As expected, critical values decrease with the length of data intervals, because the variance of the parameter estimate for short intervals is larger than for longer intervals. The effect of parameter r on the critical values is the same as it is on the risk bounds, i.e., an increase of r leads to an increase in critical values, particularly for shorter intervals. For longer intervals, differences in critical values with respect to the choice of r are moderate. Figure 3 further shows that decreasing ρ generally results in an increase in critical values, especially for the longer intervals. Again, the increase is moderate. That is, critical values are relatively robust to the choice of hyper-parameters ρ and r. We conjecture that this also holds for the final parameter estimates of the crop yield model. This is consistent with the results reported in financial or meteorological applications (Chen et al., 2010; Härdle et al., 2016). Similar findings apply to all combinations of crops and counties (see Figure A4).



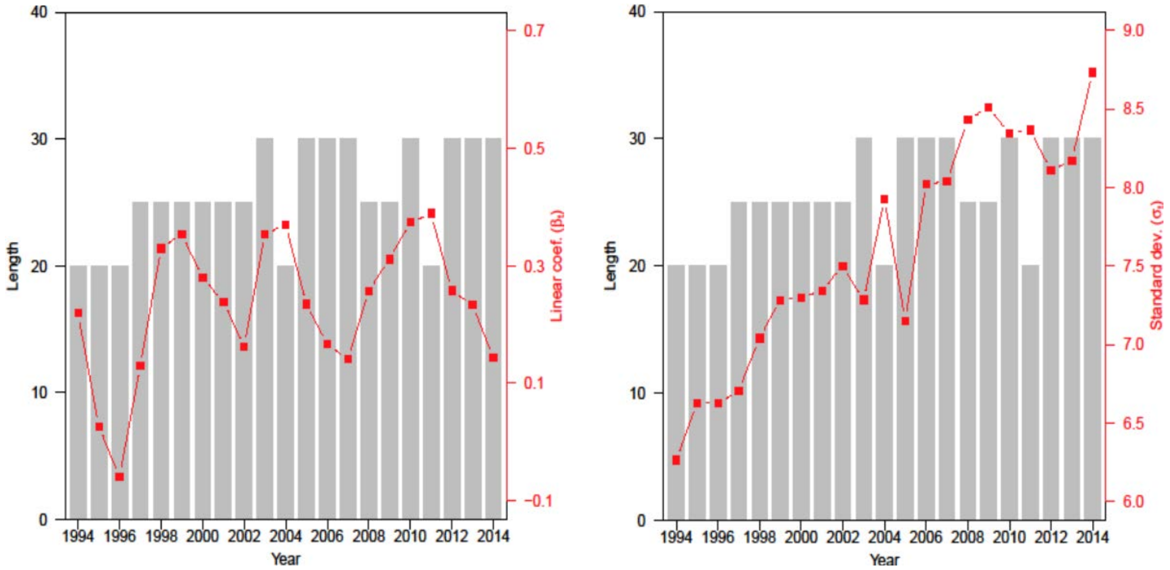
Wheat in Ellis County, Kansas **Corn in Story County, Iowa**

Figure 3. Simulated critical values for different values of parameters r and ρ

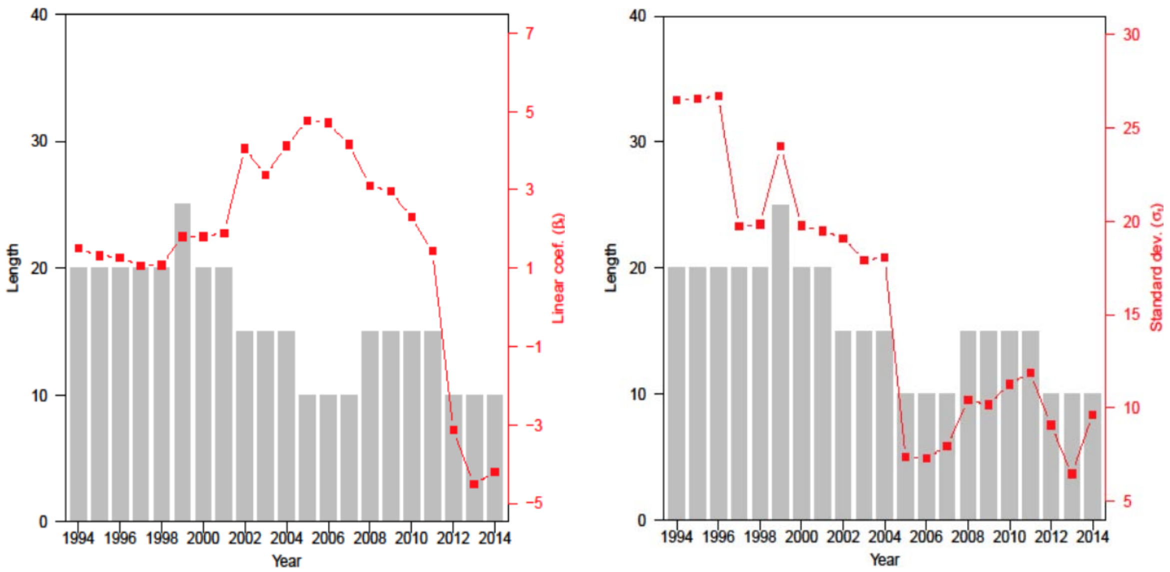
We can now apply the adaptive estimation procedures to select the longest interval for which the parameter homogeneity assumption is not violated. The estimation results based on adaptively selected optimal intervals for two representative counties are shown in Figure 4. Since the considered maximum estimation interval is 40, we estimate the time-varying parameters $\boldsymbol{\theta}$ via models (1) and (2) for 1994 based on 40 observations from 1955-1994. Then, the proposed local parametric approach will determine the optimal interval of homogeneity over which we can estimate the time-varying parameters for 1994 and then predict the crop yield for 1995. The estimation procedure is repeated for each time point from 1994 to 2014. The resulting optimal interval for wheat in Ellis in 1994, for example, is 20. The estimation of the trend model (1) should be based on observations from 1975 to 1994. This implies that a local change point (a structural break of the model parameters) has been detected between 1969 and 1974. In fact, this finding is in line with the visual inspection of observed wheat yields in Figure 1 (a). As the targeting year moves forward, optimal interval lengths for 1995, …, 2014 change, but mainly range from 20 to 30 years.

Based on the selected intervals of homogeneity, the estimated local parameters β_t and σ_t change considerably over time. The slope coefficient β_t and standard deviation σ_t represent the time-varying annual change and dispersion of crop yields, respectively, resulting from technological change and climate change over time. Unlike in a global linear trend model, the effect of severe crop shortfalls has more influence on the estimatedβ_t and $\sigma_t$in a local parametric model. This may reflect how quickly the advances in seed technology respond to changing climate conditions and could allow us to better capture current crop yield risks. For wheat in Ellis County, Kansas, the slope coefficients β_t for the period from 1994 to 2014 are mostly positive, but exhibit a decreasing pattern in recent years. It indicates that the rate of crop yield growth is decreasing. However, the magnitudes of the changes are rather moderate and range between 0.1 and 0.4 (bushels per acre). On the other hand, the standard deviation σ_t is constantly increasing over years, suggesting that wheat yield risk has become much higher in recent years compared to 20 years ago. For corn in Story County, Iowa, the trend coefficients β_t in recent years are also decreasing similar to that of wheat yields in Ellis County, Kansas, but the magnitudes of trend coefficients are considerably higher for corn. Interestingly, in contrast to the increasing dispersions of wheat yields, the estimated standard deviations σ_t for corn yields in the past ten years in the predicted dataset appear to be smaller than those of ten years before. This finding, which is also reported by Woodard et al. (2011), may reflect a higher resistance of corn against weather stress, especially drought (Yu and Babcock, 2009). In this situation, a global trend model estimated with historical corn data would overestimate standard deviations and corn yield risks. In turn, the corresponding crop insurance would be overpriced. The similar finding can be found for estimated standard deviations over time of cotton. For soybean, however, changes in the estimated trend coefficients and standard deviations over time appear to be moderate, suggesting that the assumption of homogeneous parameters likely holds over this time period. The results for soybean and cotton are presented in Figure A5 in the Appendix.

Since optimal intervals rely on the calibration of critical values based on the assumed hype-parameters, we investigate the robustness of the estimation results with respect to alternative hype-parameters. Figure A6 presents the adaptive estimation results based on alternative hyper-parameters for wheat in Ellis County, Kansas. The results in Figure 4 (a) and Figure A6, in general, confirm the robustness of the adaptive technique and low sensitivity to the choice of hype-parameters. The results suggest that in the data-driven adaptive procedure, changes in the simulated critical values derived from different choices of hype-parameters are rather moderate compared to changes in the test statistics due to the break point or structural change in empirical observations. Thus, similar optimal intervals are determined by the local change point test.



(a) Wheat in Ellis County, Kansas



(b) Corn in Story County, Iowa

Figure 4. Adaptively estimated trend β_t (left) and volatility σ_t (right) with optimal time intervals ($r = 0.5$ and $\rho = 0.5$)

3.3 Forecast Performance

In this section, we analyse the forecasting performance of the proposed model and compare it to alternative models. To evaluate the forecasting accuracy, we calculate one-step point forecasts of crop yields for each time point in the out-of-sample period, i.e., from 1995 to 2014 (and from 2008 to 2014 for wheat in Texas). The forecast accuracy is measured by the root mean squared error (RMSE). We compare the local parametric approach (LPA) with two benchmark models: The first model is a one-knot robust linear spline model (Spline) and the second is a linear trend model with a 40-year rolling window (RW40). The one-knot robust linear spline model reflects the trend-adjusting procedures in the current RMA's rating methodology (Annan et al., 2014; Harri et al., 2011). It is estimated with all available historical data. The 40-year rolling window estimation can be considered as a special case of the local parametric approach when the homogeneity assumption of the parametric model holds for all historical crop yield data and the longest possible interval (40 years) is always selected.

Figure 6 depicts the observed and predicted crop yields as well as the RMSE for all models. At first glance, the predicted crop yields from the LPA follow the crop yield dynamics more closely than the predictions of the other two models. The LPA also shows the smallest RMSE among representative crop-counties combinations with the exception of corn in County Mercer and soybean. To better understand under what conditions the LPA outperforms the benchmark models, we take a closer look at the results for different crops. We can see that the yield data follow different patterns. Wheat and cotton yield data exhibit structural changes in either their time trend or yield dispersion (cf. Figures 1 and A3), implying that data from 20 or 30 years ago are not useful for predicting current yields and probably introduce parameter bias for current years. In other words, the short data intervals are appropriate in this case. The LPA is able to detect these structural breaks and selects reasonable intervals of parameter homogeneity and thus provides better forecasts. The superiority of the LPA is significant for cotton (156.57 versus 170.06 for the RW40 and 182.26 for the Spline) and moderate for wheat (9.96 versus 10.3 for the RW40 and 10.55 for the Spline). In contrast, soybean yields are characterized by single erratic shortfalls rather than by structural breaks (cf. Figure A2). Thus, if one considers these shortfalls as outliers, the assumption of homogeneous parameters likely holds over the entire time period. The local change point test recognizes these single yield shortfalls as an indicator of a parameter change and rejects the homogeneity assumption for the longer time intervals. This results in higher variance of the parameter estimates and higher forecast errors (4.91) compared to the Spline model (4.86) and the rolling window model (4.88). However, the difference in RMSE is moderate and the plots of predicted soybean yield are rather identical across all of the models. Actually, a Diebold Mariano (DM) test with LPA as a benchmark shows that most of the aforementioned differences in the out-of-sample forecasts are not significant except for the case of cotton (see Table 2), which is likely due to the rather short sample period. In contrast, a Mincer-Zarnowitz (MZ) test[3] finds significant differences for wheat in Ellis and cotton in all counties.

Table 2 Forecast accuracy evaluation

Model	Wheat		Corn		Soybean		Cotton	
	Ellis	Coryell	Mercer	Story	Henry	Cedar	Dooly	Coahoma
				Diebold-Mariano test				
LPA				Benchmark				
Spline	-0.981	-0.071	0.214	-0.156	0.015	0.224	-0.945	-1.774*
RW40	-0.626	-0.527	-0.013	-0.067	-0.264	0.186	-0.588	0.073
				Mincer-Zarnowitz test				
LPA	1.541	13.528***	2.761*	7.456***	0.008	0.148	0.050	2.004
Spline	3.020*	4.692*	1.499	3.756**	0.201	0.734	6.744***	8.226***
RW40	3.281*	3.693	1.109	2.678*	0.228	0.678	3.508**	3.986**

Notes: In a DM test, the negative sign implies that the benchmark's loss is lower than that implied by other models. ***, **, and * represent the significance at the 1%, 5% and 10% level, respectively.
A higher MZ test statistic indicates that the null hypothesis is more likely to be rejected, and hence the considered model is less favorable.

[3] The MZ test is based on the idea that the error of an efficient forecast has to be unbiased and uncorrelated with the forecast itself according to the Mincer-Zarnowitz regression:

$$y_t = a + by_t^f + \epsilon_t$$

where y_t and y_t^f are observed and forecast values. The null hypothesis of an efficient forecast is: $H_0: a = 0, b = 1$.

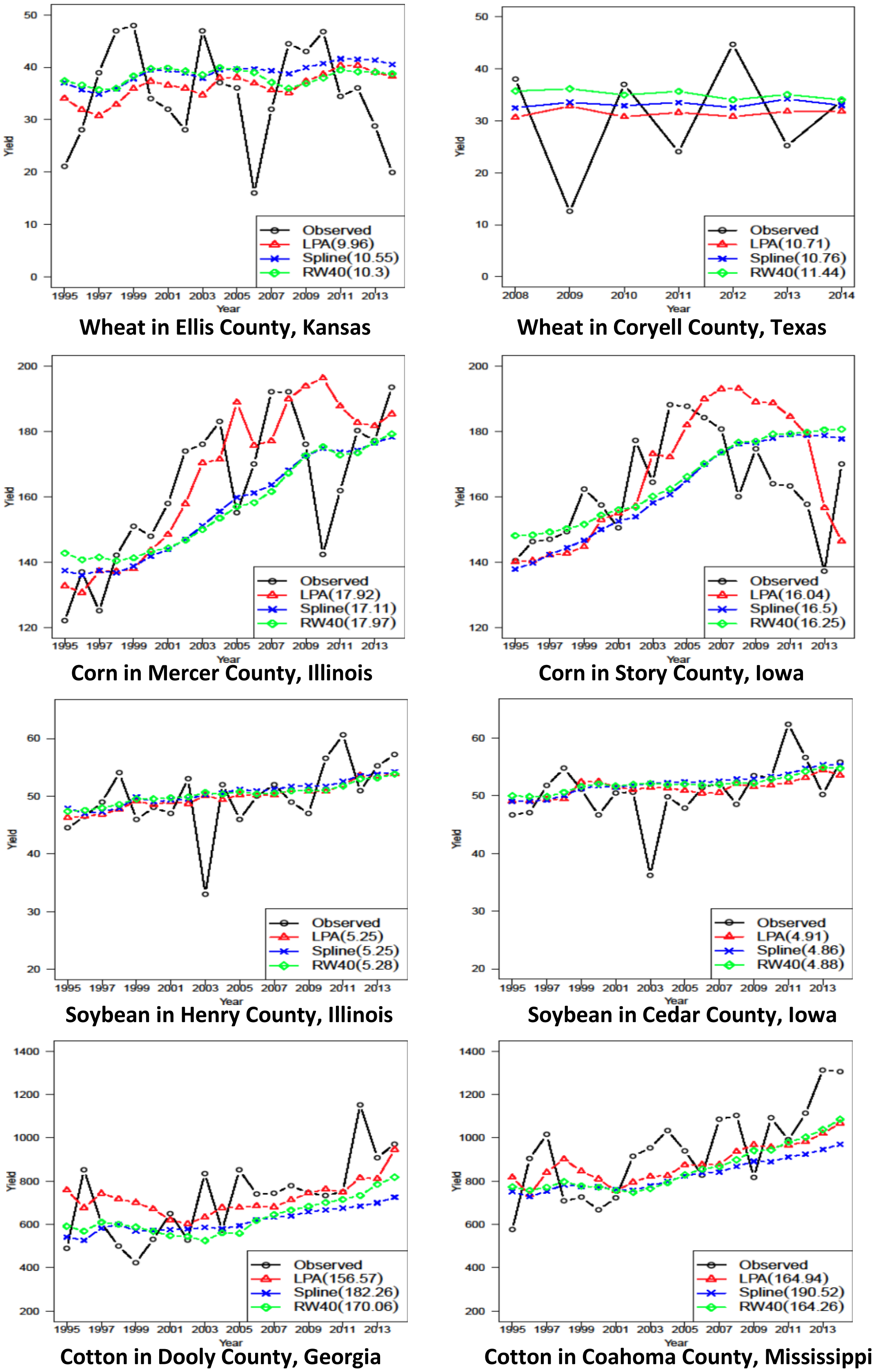


Figure 6. One-step ahead forecasts of crop yields under different models and RMSE in the parentheses

3.4 Out-of-Sample Rating Game

The previous analyses shed light on the forecasting accuracy of the local parametric approach. However, mean forecast errors may not provide full insight into economic implications and the value of the LPA for practical applications in crop insurance pricing. Insurance premium rates mainly reflects the current risk of crop yield and rely on variance estimation. To explore the potential of our proposed method for improving the accuracy of estimated crop insurance rates, we conduct an out-of-sample rating game which has been commonly used in the crop insurance literature (Harri et al., 2011; Annan et al., 2014; Tolhurst and Ker, 2015). The rating game mimics features of the U.S. crop insurance program, that is private insurance companies can choose to participate in the insurance programs and sell insurance contracts to farmers with premium rates set by the RMA of the government. Therefore, private insurance companies will likely re-estimate the premiums and compare their premiums with those set by the RMA. If their premiums are higher than the RMA's premiums, they cede the policies as they believe the RMA's premiums to be underpriced. Conversely, if their premiums are lower than the RMA's premiums, they retain the policies as they believe the RMA's premiums to be overpriced. Here, we compare the RMA premiums with the one derived from the LPA. Then, given actual realized crop yields, we calculate the loss ratios of retained policies and ceded policies for each county and period combination. We focus on an area yield insurance contract since county-level yield data are used in our study. An actuarially fair premium π for an area yield insurance policy with a coverage level c of (unconditional) expected yield y^e is given by:

$$\pi(c) = \Pr(y_t \le cy^e)\{cy^e - \mathbb{E}[y_t | y_t < cy^e]\} = \int_0^{cy^e} (cy^e - y_t)\, f(y_t|\mathcal{F}_t) dy_t \ , \qquad (23)$$

where y_t is the random crop yield at time t. The expectation is based on the conditional yield density $f(y_t|\mathcal{F}_t)$ with the available information set $\mathcal{F}_t$ at time t. Following usual practice, we select a 90% coverage level, i.e., $c = 0.9$ (e.g., Annan et al., 2014). In the simulated rating game, the yield guarantee cy^e for an insurance policy is determined by the RMA. Thus, the difference between private insurance premiums and the RMA's premiums depends solely on their estimations of the conditional yield density $f(y_t|\mathcal{F}_t)$.

For the out-of-sample period from 1995 to 2014, we repeat the rating game to derive premiums and loss ratios for each period and county combination using data from 1955-1994, …, 1955-2013. Note that for the local parametric approach, only the selected longest data interval of homogeneity will be used for insurance rating. According to Annan et al. (2014), the RMA models crop yields with a one-knot linear spline and then adjusts the residuals for heteroscedasticity. Based on the detrended crop yields, the empirical premium is estimated as the RMA's premium (π^g). In other words, this technique uses a burn analysis rating, whereas the LPA premium (π^p) is based on simulated predicted crop yields given the time-varying parameter estimates. The cede-retain decisions for each insurance policy are then made according to the comparison between π^g and π^p. Following usual convention, loss ratios are aggregated on a state basis using planted acreage for the final year 2014 as weights. If loss ratios for retained polices are smaller than those for ceded polices, we conclude that the LPA is better at estimating the premium rate than the RMA method. Statistical significance of differences in loss ratios can be tested by a randomization test.[4]

The results of the out-of-sample insurance rating game between the LPA model and the RMA method are presented in Table 3. The percentages of retained policies are quite high for all crop policies except for winter wheat, suggesting that most of the LPA premium rates are lower than the RMA rates. This observation is in line with our finding in the analysis in Section 3.2. that the one-

[4] With the percentage of retained contracts at hand, we randomly choose the same percentage of insurance contacts from the entire period-county contract pool and calculate the retained loss ratio. This step is repeated 5,000 times generating 5,000 loss ratios. P-values are calculated as the percentage of loss ratios that are smaller than the retained loss ratio. This procedure has been adopted by Harri et al. (2011), Tolhurst and Ker (2015) and Ker et al. (2016).

knot-spline model based on all historical data tends to overestimate current yield volatility. The fact that π^g frequently exceeds π^p results in a high number of retained policies, which, in turn, renders the randomization test less meaningful. In other words, the out-of-sample rating assessment may fail if the considered rating methodology leads to systematically lower or higher premiums than the RMA method. Nevertheless, we observe that in four of the eight state-crop combinations, insurance companies could gain a significant economic rent using the LPA rating method. Particularly, the results for winter wheat and cotton are in favor of the LPA methodology since it captures tail yield risks better than the RMA approach. Given its high ratio of policy payouts, potential economic rent of private insurance companies using the LPA could be much more considerable. For the other four state-crop combinations, the results show that either the RMA methodology leads to lower loss ratios or that the differences are not statistically significant.

Though the high number of retained policies renders the results of the out-of-sample rating assessment useless, it does not imply that a private insurance company will not be able to make profit or that the premiums derived from the proposed LPA method are underpriced. If the RMA premiums exceed actual average losses, profits can be made by retaining all policies. To capture this aspect, we also display the realized loss ratios of all policies in the last two columns of Table 3. If the realized loss ratio is smaller (larger) than 1, insurance premiums are supposed to be overpriced (underpriced). For winter wheat and cotton, the loss ratios based on the LPA premiums are closer to the expected long run value of 1 than for the RMA's premiums. Actually, the RMA's premiums are underpriced for winter wheat and overpriced for cotton. Interestingly, the loss ratio using LPA's premiums is also closer to one for corn, though the results for corn in Iowa in the out-of-sample rating assessment are not in favour of the LPA method.

Table 3. Out-of-Sample Insurance Rating Game

Crop-State	Number of Counties	Retained Policies (%)	Loss Ratio of Retained Policies	Loss Ratio of Ceded Polices	*p*-value	Payout Policies (%)	Loss Ratio of all Policies (π^g)	Loss Ratio of all Policies (π^p)
Winter Wheat								
Kansas	61	55.4	1.212	2.173	0.000	37.9	1.668	1.469
Texas	45	48.6	1.612	2.588	0.003	48.3	2.163	1.836
Corn								
Illinois	79	96.9	0.568	0.981	0.131	23.5	0.577	1.435
Iowa	99	92.4	0.301	0.101	0.985	14.8	0.289	0.845
Soybean								
Illinois	89	92.6	0.642	0.778	0.283	19.6	0.649	1.413
Iowa	98	89.5	0.822	0.738	0.630	20.1	0.815	1.507
Cotton								
Georgia	20	79.5	0.356	1.040	0.000	29.3	0.449	0.782
Mississippi	14	88.6	0.382	0.872	0.079	16.4	0.431	1.017

*Note: A *p*-value close to 0.00 indicates that the proposed method outperforms the RMA method. On the other hand, if a *p*-value is close to 1.00, the RMA method is better than the proposed method. The loss ratio of all policies in the last column is calculated based on the LPA's premiums.

4 Discussion and Conclusions

This article was motivated by the challenge of considering non-stationarity in the estimation of crop yield models, which is a building block for the pricing of crop insurance. This non-stationary can be a result of technological change and/or climate change. To deal with non-stationarity, various approaches have been proposed to allow some or all model parameters to vary over time. An identification of these models in the current literature requires either structural assumptions about the transition process over time or presumes that the parameters follow smooth functions of time. In this paper, we develop an alternative data-driven approach that is based on the local parametric assumption. To be specific, the idea of the local parametric approach is to find an optimal *interval of homogeneity* over which one can fit a local parametric model with constant parameters. The selection of interval of homogeneity is determined in a backward sequential testing procedure with an embedded local change point test. The advantage of adaptively and promptly detecting structural change makes the proposed local parametric approach more flexible and less restrictive for modelling time-varying parameters compared to previous approaches. In addition, the proposed approach enabled us to contribute to the longstanding debate over sample period selection of crop yields from a sound statistical perspective. The backward selected sample period allows us to more accurately determine the current rate of technological change and the current risk of crop yields, and therefore to mitigate the potential bias caused by historical crop yield data from more than four decades ago.

We apply the proposed local parametric approach to county-level winter wheat, corn, soybean, and cotton yields in a large number of counties in the U.S. Our empirical results demonstrate that the proposed local parametric approach selects reasonable intervals of parameter homogeneity, mainly ranging from 20 to 30 years before the current period. In contrast to earlier work on this issue, we relax the assumption of a fixed sample period over the entire dataset. In fact, a change of estimated local parameters, such as β_t and σ_t, over time allows us to capture, for example, the deceleration of crop yield growth, the decrease in corn and cotton yield variability, or the increase of wheat yield risk that has been found in the literature (e.g., Yu and Babcock, 2009). In terms of the forecasting accuracy of crop yields, the results show that the local parametric approach, in general, leads to the smaller forecast errors of crop yields across crops compared to traditional alternatives. This is particularly true when yield data exhibit structural changes or regime switches. A simulation exercise further documents that the local parametric approach has the potential to improve the pricing of insurance contracts. However, each method comes attached with cautions and limitations. A shortcoming of the proposed local parametric approach is that the inability of the local change point test to distinguish between unusual shortfalls and actual structural changes leads to the rejection of the longer interval of homogeneity and to more volatile parameter estimates. This problem could be mitigated by the exclusion or devaluation of outliers, which is also recommended in the average loss cost ratio approach in the RMA rating methodology (Coble et al., 2010).

This study provides the first empirical application of an adaptive local parametric model to crop yields. However, a number of potential extensions may further improve the performance of this model and are suggested for future research. First, one may apply alternative homogeneity tests that can detect the structural change, but are less sensitive to occasional catastrophes. A simple likelihood ratio test is an example (Härdle et al., 2016). Second, the underlying crop yield model, which is a simple linear trend mode in our case, could be refined. More sophisticated crop yield models include a mixture Normals with embedded trend functions (Tolhurst and Ker, 2015) or a model that takes into account extreme weather events through exogenous weather variables. Finally, the incorporation of heavy-tailed distributions such as the Weilbull distribution (Woodard, 2014) or Beta distribution (Zhu et al., 2011) might further improve the model results since a lot of empirical evidence suggest that area yields are not normally distributed.

References

Andriyashin, A., Benko, M., Härdle, W., Timofeev, R., Ziegenhagen, U. (2006). Color harmonization in car manufacturing processes, Applied Stochastic. Models in Business and Industry, 22:519–532

Annan, F., Tack, J., Harri, A., Coble, K. (2013). Spatial Pattern of Yield Distributions: Implications for Crop Insurance. American Journal of Agricultural Economics 96(1): 252-268.

Atwood, J., Shaik, S., Watts, M. (2003). Are crop yields normally distributed? A reexamination. American Journal of Agricultural Economics, 85(4), 888-901.

Carriquiry, M. A., Osgood, D. E. (2012). Index insurance, probabilistic climate forecasts, and production. Journal of Risk and Insurance, 79(1), 287-300.

Chen, Y., Niu, L. (2014). Adaptive dynamic Nelson–Siegel term structure model with applications. Journal of Econometrics, 180(1), 98-115.

Chen, Y., Härdle, W., Pigorsch, U.(2010). Localized Realized Volatility Modeling. Journal of the American Statistical Association, 105(492), 1376-1393.

Čížek, P., Härdle, W., Spokoiny, V. (2009). Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models. The Econometrics Journal, 12(2), 248-271.

Coble K.H., Knight T.O., Goodwin B.K., Miller M.F., Rejesus R.M. A Comprehensive Review of the RMA APH and COMBO Rating Methodology: Final Report 2010. www.rma.usda.gov/pubs/2009 /comprehensivereview.pdf (accessed 01.09.2015).

Edgerton, M. D., Fridgen, J., Anderson Jr, J. R., Ahlgrim, J., Criswell, M., Dhungana, P., Stark, S. B. (2012). Transgenic insect resistance traits increase corn yield and yield stability. Nature biotechnology, 30(6), 493-496.

Goodwin, B. K., Ker, A. P. (1998). Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts. American Journal of Agricultural Economics, 80(1), 139-153.

Härdle, W. K., Hautsch, N., Mihoci, A. (2015). Local adaptive multiplicative error models for high‐frequency forecasts. Journal of Applied Econometrics, 30(4), 529-550.

Härdle, W., Lopez, B., Okhrin, O. Wang, W. (2016) Localizing temperature risk. Journal of the American Statistical Association. http://dx.doi.org/10.1080/01621459.2016.1180985

Härdle, W. K., Okhrin, O., Okhrin, Y. (2013). Dynamic structured copula models. Statistics & Risk Modeling, 30(4), 361-388.

Harri, A., Coble, K.H., Ker, A.P., Goodwin, B.K. (2011). Relaxing heteroscedasticity assumptions in area-yield crop insurance rating. American Journal of Agricultural Economics, 93(3): 707-717.

Harri, A., Erdem, C., Coble, K. H., Knight, T. O. (2009). Crop yield distributions: A reconciliation of previous research and statistical tests for normality. Applied Economic Perspectives and Policy, 31(1), 163-182.

Just, R. E., Weninger, Q. (1999). Are crop yields normally distributed?. American Journal of Agricultural Economics, 81(2): 287-304.

Kaylen, M. S., Koroma, S. S. (1991). Trend, weather variables, and the distribution of US corn yields. Review of Agricultural Economics, 13(2), 249-258.

Ker, A. P., Coble, K. (2003). Modelling conditional yield densities. American Journal of Agricultural Economics, 85(2), 291-304.

Ker, A. P., Goodwin, B. K. (2000). Nonparametric estimation of crop insurance rates revisited. American Journal of Agricultural Economics, 82(2), 463-478.

Ker, A.P., Tolhurst, T.N., Liu, Y. (2016): Bayesian Estimation of Possibly Similar Yield Densities: Implications for Rating Crop Insurance Contracts. American Journal of Agricultural Economics 98(2): 360-382.

McCarl, B. A., Villavicencio, X., Wu, X. (2008). Climate change and future analysis: is stationarity dying?. American Journal of Agricultural Economics, 90(5), 1241-1247.

Norwood, B., Roberts, M. C., Lusk, J. L. (2004). Ranking crop yield models using out-of-sample likelihood functions. American Journal of Agricultural Economics, 86(4), 1032-1043.

Ozaki, V. A., Ghosh, S. K., Goodwin, B. K., Shirota, R. (2008). Spatio-temporal modeling of agricultural yield data with an application to pricing crop insurance contracts. American journal of agricultural economics, 90(4), 951-961.

Polzehl, J., Spokoiny, V. (2006). Propagation-separation approach for local likelihood estimation. Probability Theory and Related Fields, 135(3), 335-362.

Skees, J. R., Black, J. R., Barnett, B. J. (1997). Designing and rating an area yield crop insurance contract. American journal of agricultural economics, 79(2), 430-438.

Smith, V.H., Goodwin, B. K. (2010). Private and public roles in providing agricultural insurance in the United States. In Public Insurance and Private Market, edited by J. R. Brown, pp.173-209. Washington, DC.

Spokoiny, V. (2009). Multiscale local change point detection with applications to value-at-risk. *The Annals of Statistics*, 1405-1436.

Tolhurst, T.N., Ker, A.P. (2015). On technological change in crop yields. American Journal of Agricultural Economics, 97(1), 137-158.

Wang, W., Bobojonov, I., Härdle, W. K., Odening, M. (2013). Testing for increasing weather risk. Stochastic environmental research and risk assessment, 27(7), 1565-1574.

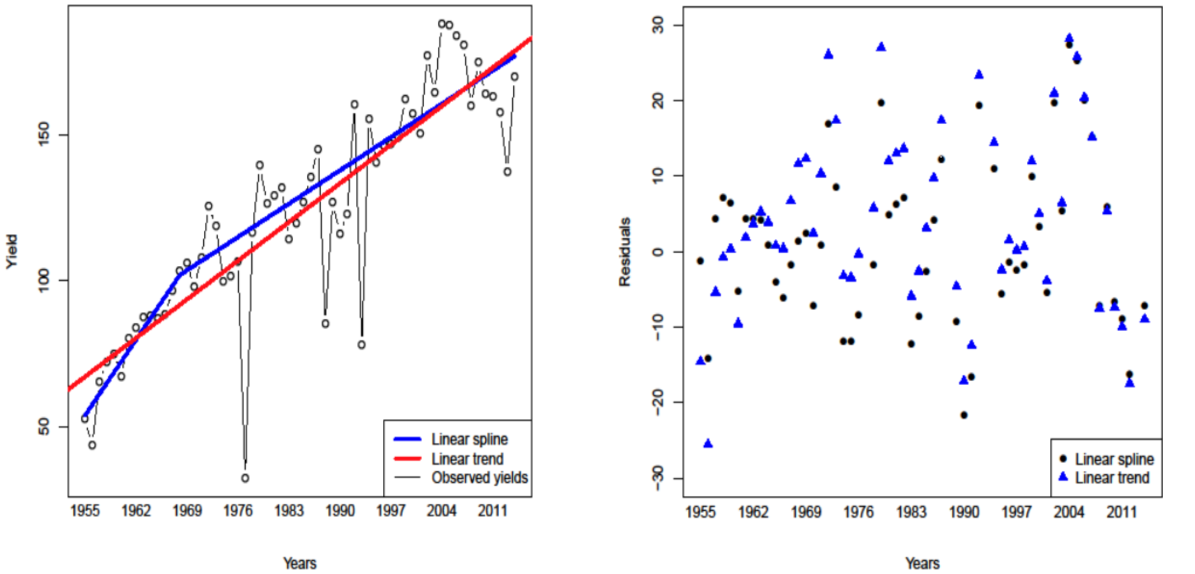
Woodard, J. D. (2014). Impacts of Weather and Time Horizon Selection on Crop Insurance Ratemaking: A Conditional Distribution Approach. North American Actuarial Journal, 18(2), 279-293.

Woodard, J. D., Sherrick, B. J., Schnitkey, G. D. (2011). Actuarial impacts of loss cost ratio ratemaking in US crop insurance programs. Journal of Agricultural and Resource Economics, 211-228.

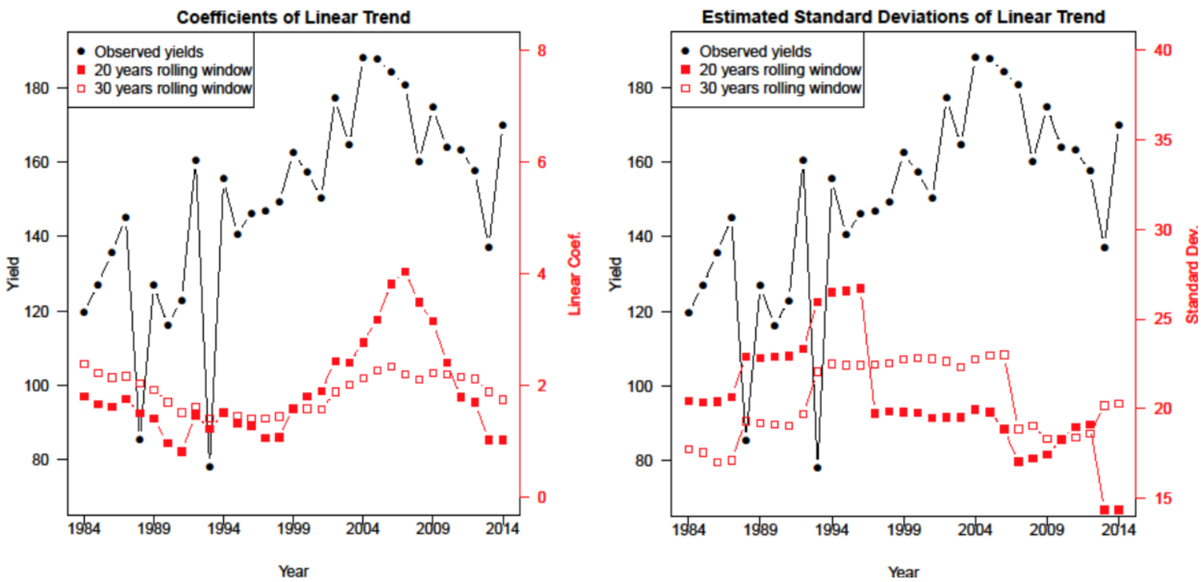
Yu, T., Babcock, B. A. (2010). Are US corn and soybeans becoming more drought tolerant? American Journal of Agricultural Economics, 92(5), 1310-1323.

Zhu, Y., Goodwin, B. K., Ghosh, S. K. (2011). Modeling yield risk under technological change: Dynamic yield distributions and the US crop insurance program. Journal of Agricultural and Resource Economics, 192-210.

Appendix

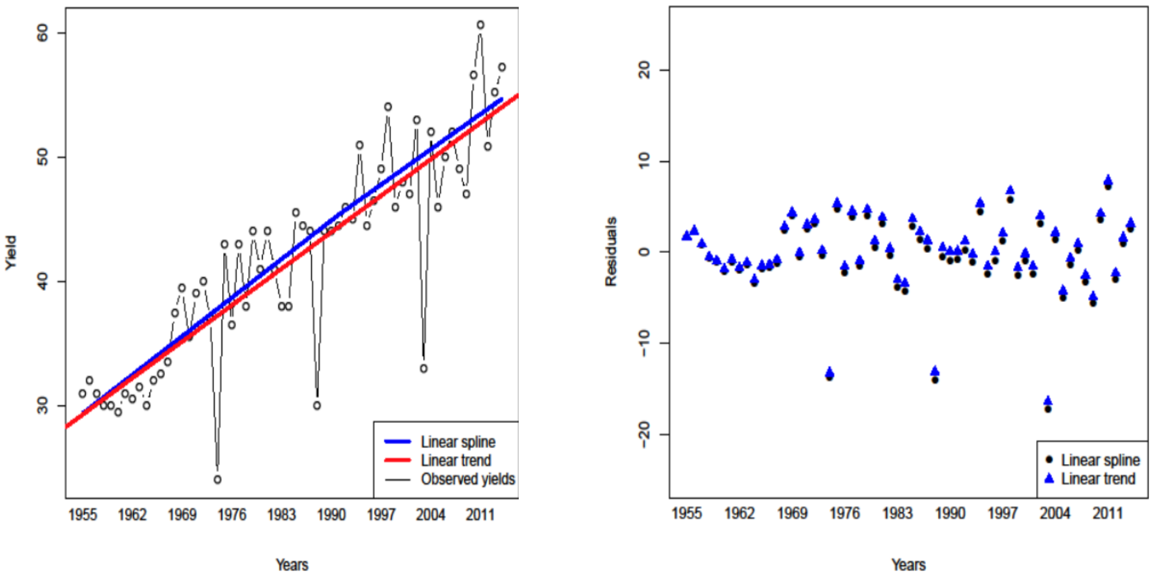


(a) Corn in Story County, Iowa

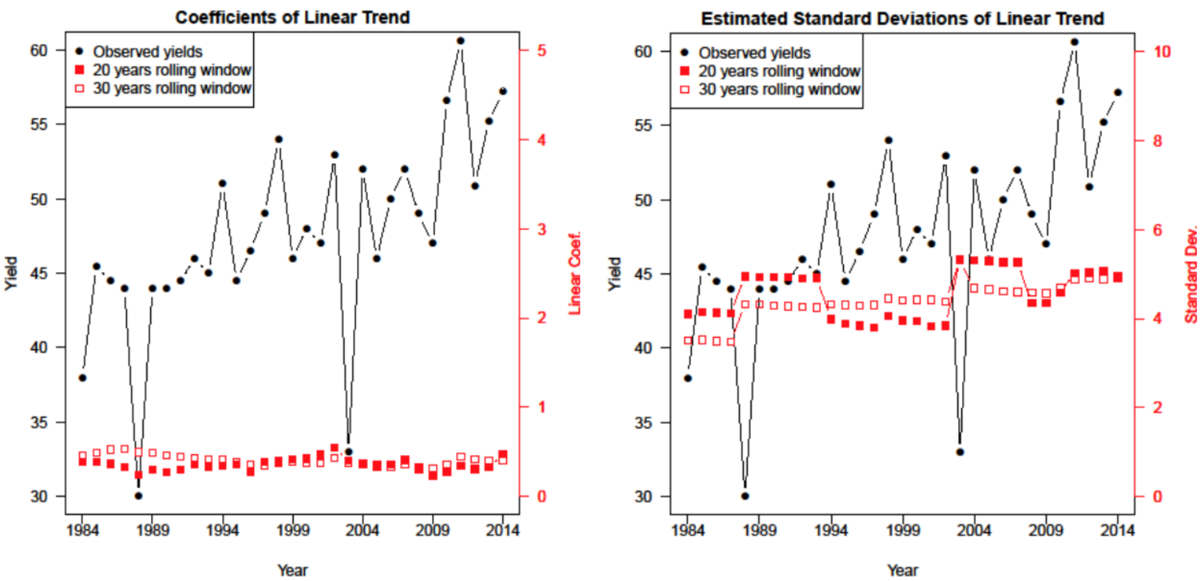


(b) Linear Time Trend Estimation based on Rolling Window Data Samples

Figure A1. Scatter plot and time horizon selection for corn

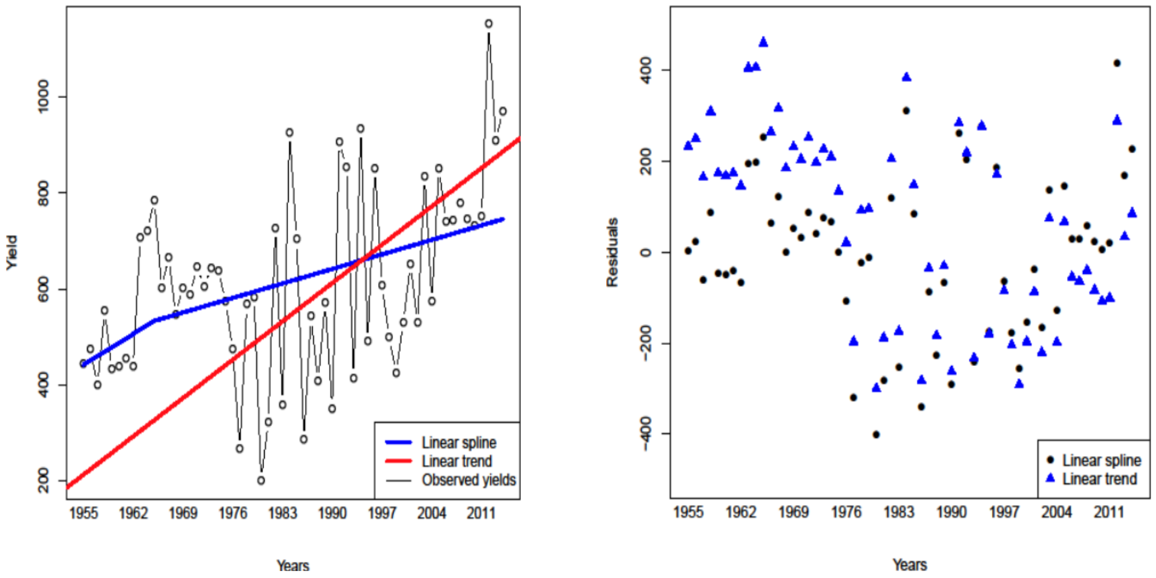


(a) Soybean in Henry County, Illinois

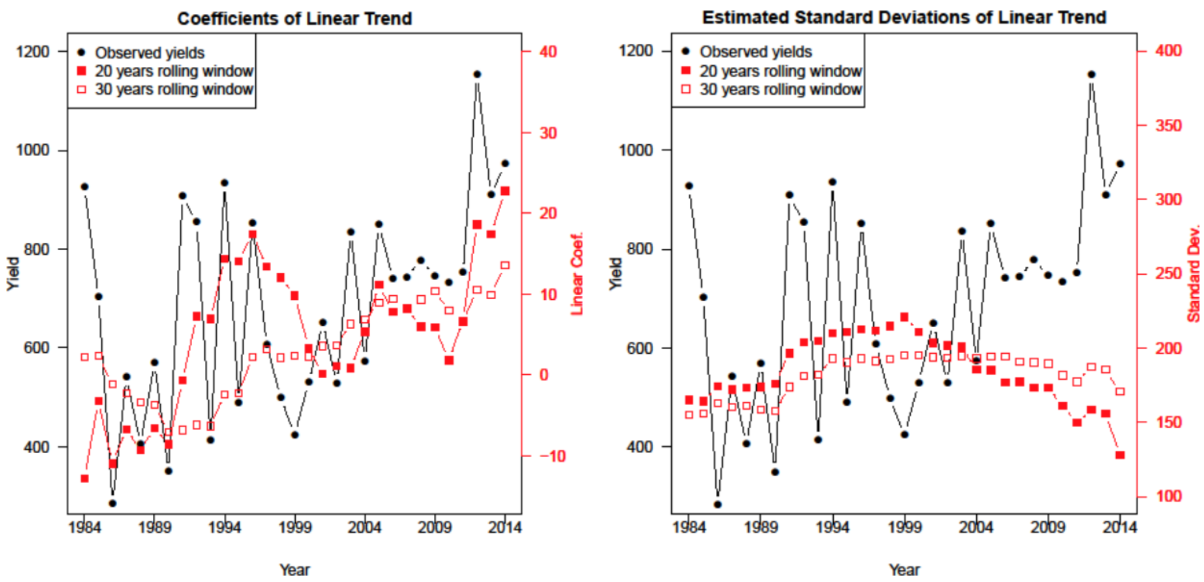


(b) Linear Time Trend Estimation based on Rolling Window Data Samples

Figure A2. Scatter plot and time horizon selection for soybean



(a) Cotton in Dooly County, Georgia



(b) Linear Time Trend Estimation based on Rolling Window Data Samples

Figure A3. Scatter plot and time horizon selection for cotton

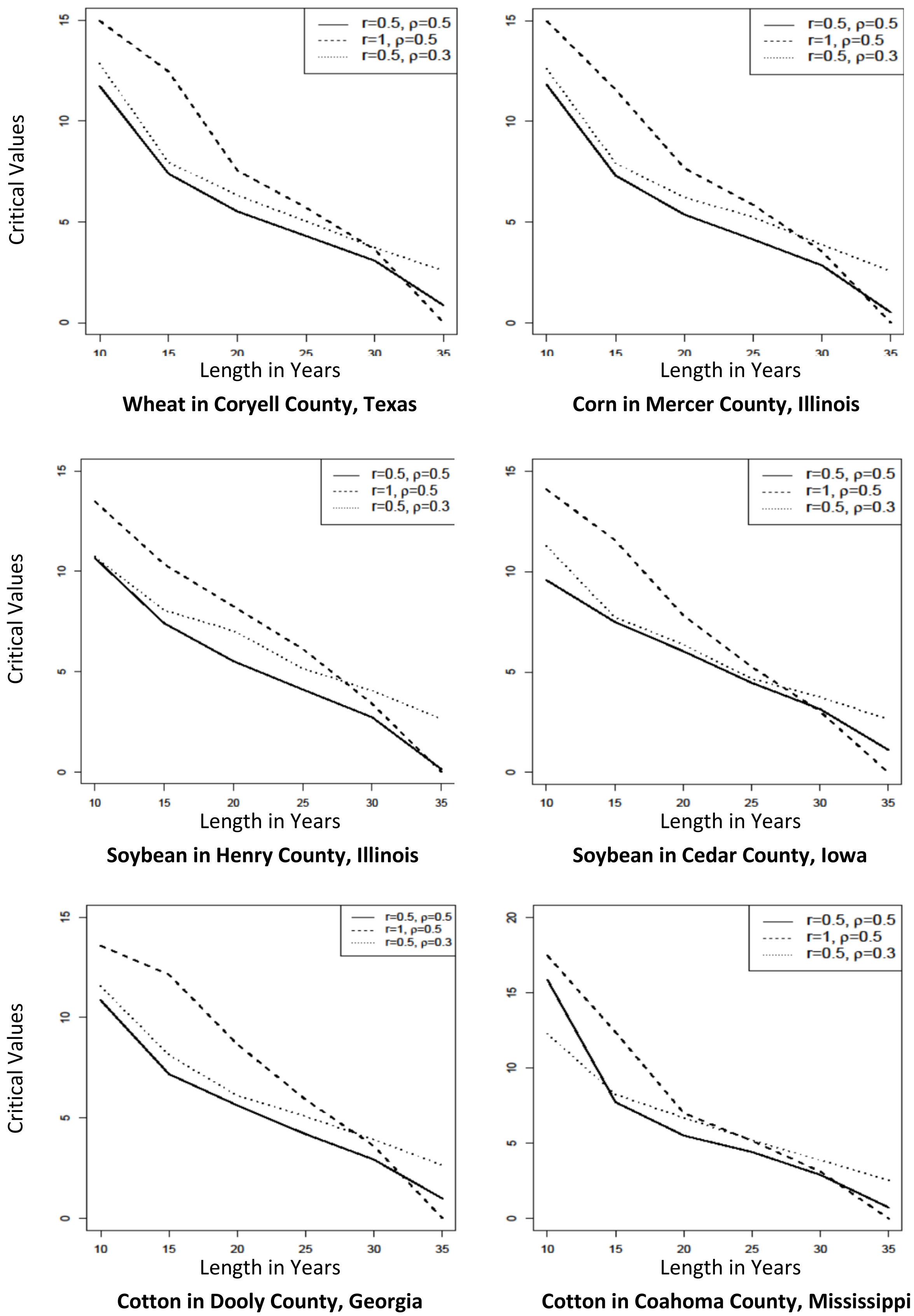
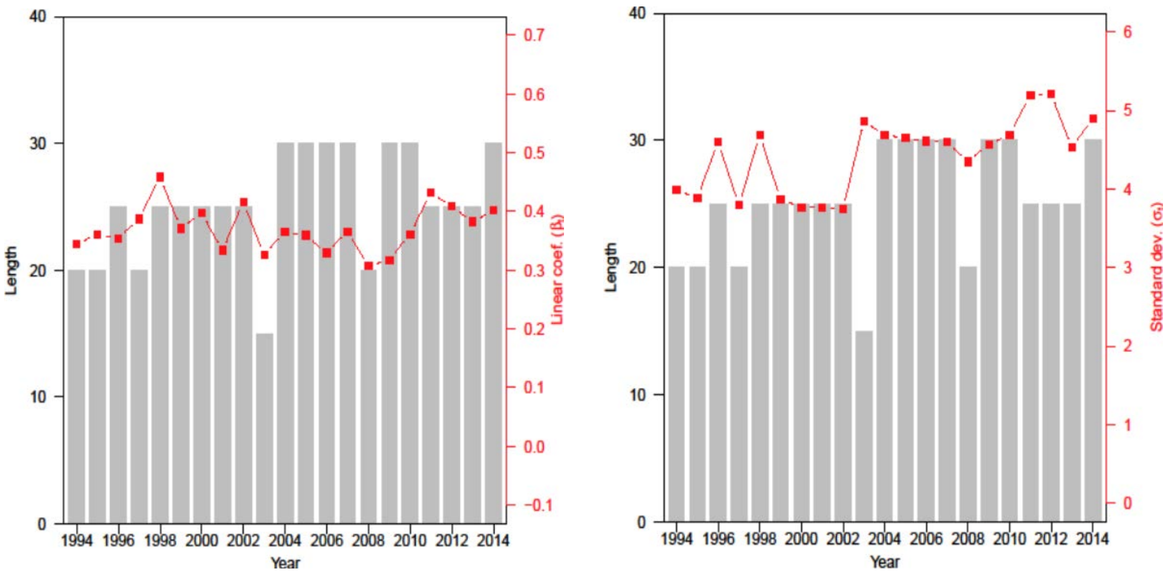
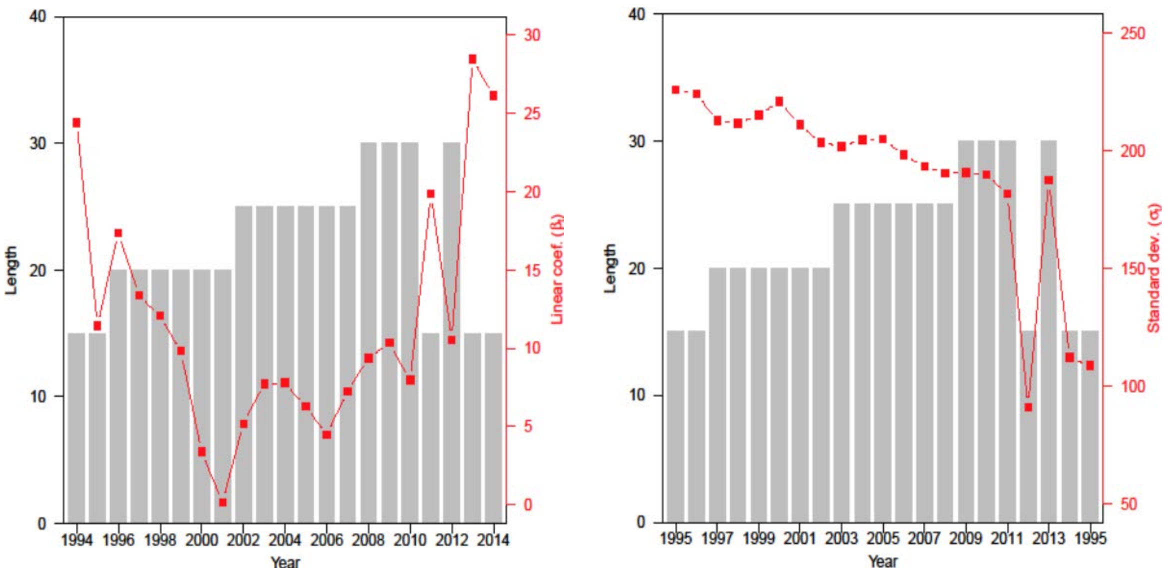


Figure A4. Simulated critical values for different values of parameters r and ρ

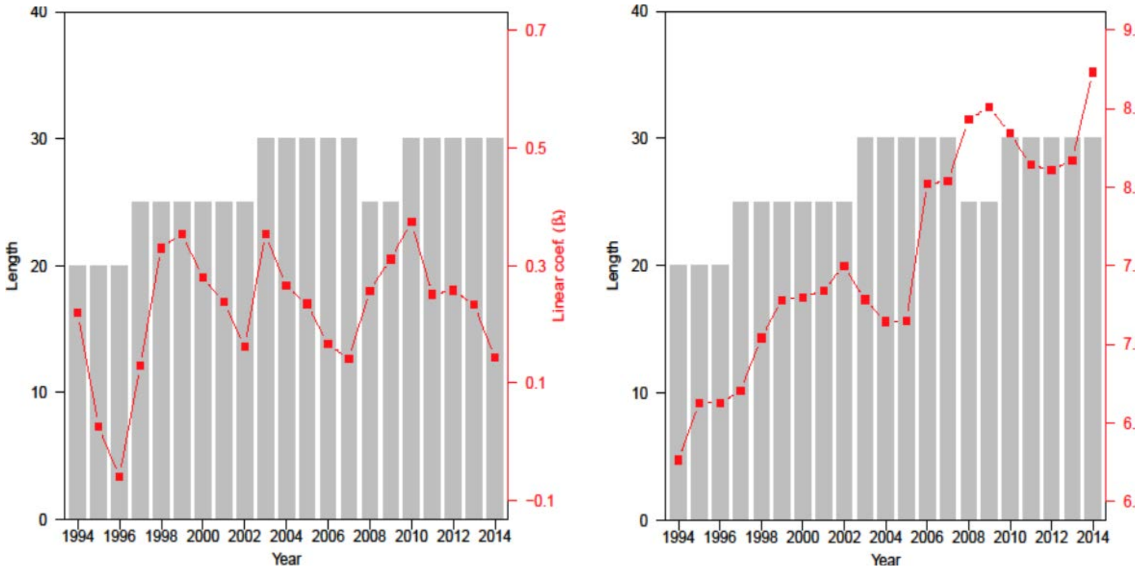


(a) Soybean in Henry County, Illinois

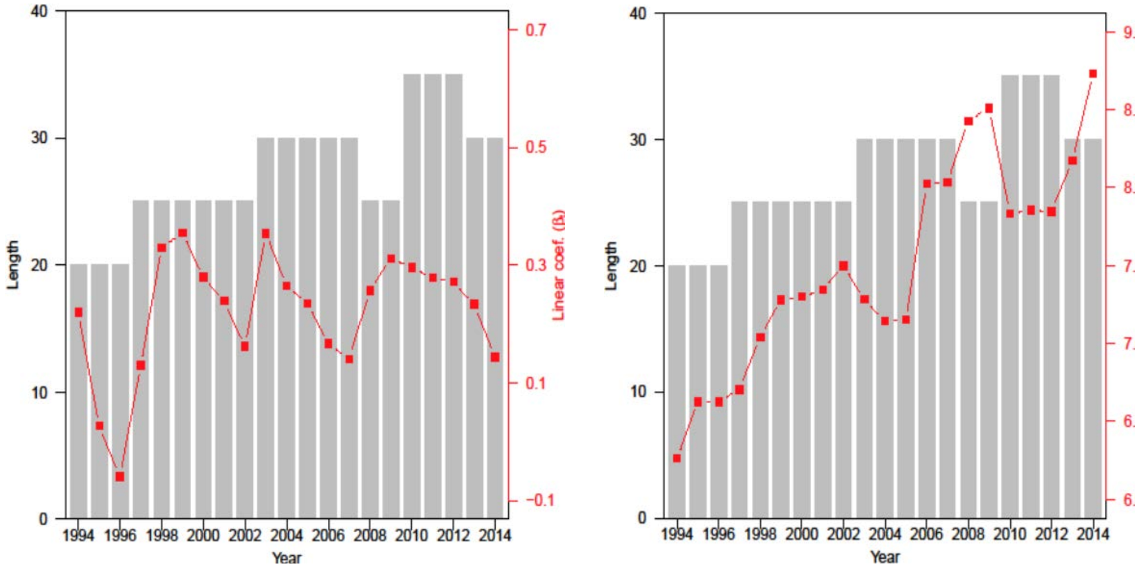


(b) Cotton in Dooly County, Georgia

Figure A5. Adaptive estimated β_t (left) and σ_t (right) with optimal intervals based on the scenario ($r = 0.5$ and $\rho = 0.5$) for soybean (Henry, Illinois) and cotton (Dooly, Georgia)



(a) Wheat in Ellis County, Kansas ($r = 1$ and $\rho = 0.5$)



(b) Wheat in Ellis County, Kansas ($r = 0.5$ and $\rho = 0.3$)

Figure A6. Adaptively estimated trend β_t (left) and volatility σ_t (right) and time intervals with alternative hyper-parameters

Table A1. Simulated risk bound $\mathfrak{R}_r(\boldsymbol{\theta}^*)$

Crop	State	County	$r = 0.5$ (baseline)	$r = 1$
Wheat	Kansas	Ellis	1.201	1.787
	Texas	Coryell	1.206	1.755
Corn	Illinois	Mercer	1.289	1.881
	Iowa	Story	1.328	2.156
Soybean	Illinois	Henry	1.276	1.848
	Iowa	Cedar	1.207	1.789
Cotton	Georgia	Dooly	1.191	1.711
	Mississippi	Coahoma	1.237	1.880