# Conditional Distributions of Crop Yields: A Bayesian Approach for Characterizing Technological Change

*A. Ramsey;*

*Virginia Polytechnic Institute and State University, Agricultural and Applied Economics, United States of America*

*Corresponding author email: aframsey@vt.edu*

*Abstract:*

*What changes in the distribution of crop yields occur as a result of technological innovation? Viewing observed yields as random variables, estimation of the yield distribution conditional on time provides one approach for characterizing distributional transformation. Yields are also affected by weather and other covariates, spatial correlation, and a paucity of data in any one location. Common parametric and nonparametric methods rarely consider these aspects in a unified manner. Comprehensive solutions for describing the distribution of yields can be considered ideal. We implement a Bayesian spatial quantile regression model for the conditional distribution of yields that is distribution-free, includes weather (covariate) effects, smooths across space, and models the complete quantile process. Results provide insight into the temporal and spatial evolution of crop yields with implications for the measurement of technological change.*

**JEL Codes:** Q18, G22

#1472

# Conditional Distributions of Crop Yields: A Bayesian Approach for Characterizing Technological Change

**Abstract**

What changes in the distribution of crop yields occur as a result of technological innovation? Viewing observed yields as random variables, estimation of the yield distribution conditional on time provides one approach for characterizing distributional transformation. Yields are also affected by weather and other covariates, spatial correlation, and a paucity of data in any one location. Common parametric and nonparametric methods rarely consider these aspects in a unified manner. Comprehensive solutions for describing the distribution of yields can be considered ideal. We implement a Bayesian spatial quantile regression model for the conditional distribution of yields that is distribution-free, includes weather (covariate) effects, smooths across space, and models the complete quantile process. Results provide insight into the temporal and spatial evolution of crop yields with implications for the measurement of technological change.

Quantities and prices are the fundamental units of economic analysis. In agriculture, quantities are often given in terms of crop yield: the amount of agricultural output produced per unit of land in production. Much of the world has seen dramatic increases in crop yields over the past century or so. Yield gains have allowed the agricultural sector to feed more people with both less land and less variable inputs. Consequent impacts on economic growth, food security, and environmental quality have been substantial. Given significant historical developments in crop yields, continued population growth and potential shifts in climate have spurred interest in the prediction and analysis of future innovations. The measurement of change in the distribution of yields has taken on increased significance, and it is this change that we quantify through Bayesian spatial quantile regression.

Observed yields can be viewed as random variables, having been drawn from a probability distribution. Randonmess in yields results from variation in weather over the growing period, invasions of pests, changes in production practices, technological change, and myriad other factors. If all possible covariates could be accounted for, observed crop yields would be deterministic. Precision technologies offer some hope that production agriculture may be moving towards this scenario. But at present, econometric models for yields are subject to a number of errors. These errors range from measurement error (which may include aggregation error) to model specification error. For the purposes of prediction and inference, models for the yield distribution should minimize error and facilitate economic interpretation.

Inference on the distribution of crop yields may be pursued in its own right or as an input to further economic analysis. Changes in the distribution over time can be ascribed partly to technological change; consequent impacts on economic decision-making arise (Tolhurst and Ker, 2014). The rate of adoption of a new technology hinges on the technology's effect on profitability, which is a function of its effect on yields (Griliches, 1957; Dixon, 1980). Improved yields result from both private and public investments in the research and development of new crops and production practices. Public funds spent on agricultural research, and institutional rules that stimulate private investment, have direct policy relevance and major impacts on agricultural productivity (Alston et al., 2010; Hurley, Rao, and Pardey, 2014; Sparger et al., 2013).

Estimated yield distributions are also a key component of the federal crop insurance program; this program is the linchpin of contemporary agricultural policy in the United States. Assumptions about both temporal variation and the parametric form of the conditional distribution of yields have

significant effects on premium rates and losses in the program (Ker and Goodwin, 2000; Ker and Coble, 2003; Zhu, Goodwin, and Ghosh, 2011; Sherrick et al., 2014). Because the program is subsidized by taxpayers, a priori specifications have implications for funding liabilities and realized expenditures. A central argument for the existence of government backed insurance is that systemic risk in agriculture is too large for private insurers to bear. The validity of this argument can be assessed by careful measurement of both the yield distribution and dependence across members of the insurance pool.

Complementary to the relevance of the conditional yield distribution to technological change and agricultural policy, measurement of the relationship between yields and weather facilitates evaluation of the potential impacts of climate change. Climate models can be used to project empirically determined relationships forward (Schlenker, Hanemann, and Fisher, 2006). Because farmers respond to changes in their environment, climate and technology are interrelated. Lobell et al. (2014) suggested that plant breeding targets yields under ideal growing conditions and may cause crops to become more sensitive to adverse weather conditions. If this hypothesis is correct, the rate of yield growth should be greater at high quantiles of the yield distribution. Testing the hypothesis necessarily requires an estimate of the conditional yield distribution. The importance of accounting for weather extends to agricultural policy more broadly. Models of weather and yields can be used to improve yield loss estimates in crop insurance by marrying long time series of weather data to shorter time series of insurance loss costs (Rejesus et al., 2015).

We implement Bayesian spatial quantile regression to analyze changes in the distribution of yields over time, the effects of covariates on the yield distribution, and spatial association between yields in different locations. The importance of integrating all of these components in a single model is evidenced through a straightforward example. Figure 1 shows all-practice corn yields in Adams County, Illinois from 1925 to 2015. The simplest way to capture the apparent upward trend in yields is to regress yields on time. Ordinary least squares assumes homoskedasticity of the residuals and a constant variance for the error distribution, which in this context is the same as the conditional yield distribution minus a location adjustment. Constant variance can be seen in the conditional density plots that have been overlaid on the regression line. The coefficient on time suggests that yields increase by roughly 1.5 bushels per acre per year. Of particular interest are a number of outlying observations that occur more than two to three standard deviations away from the mean of

the conditional distribution.

The regression in figure 1 is a restricted - and admittedly simplistic - view of the yield distri-bution. Weighted least squares could be used to impose a heteroskedasticity structure or feasible generalized least squares could be used to estimate the structure from the data. Variance would be allowed to vary, but the form of the conditional distribution would remain Gaussian. A more flexi-ble analysis of temporal distributional change is given by the quantile regressions in figure 2. The spread of the quantile regressions over time suggests heteroskedasticity, in line with the observance of a number of outliers after 1980 in figure 1 . While there are positive trends in yields at all five estimated quantiles, differences remain in the magnitudes of those coefficients associated with time. Figure 2 indicates that yields have increased more rapidly at upper quantiles of the distribution. All of the time coefficients are statistically different from zero, though coefficients at lower quantiles are measured with a high degree of uncertainty.

The quantile regressions are more compelling than ordinary linear regression precisely because they allow the conditional distribution to vary flexibly over time. This variation is not limited to the location or variance and no assumption is made as to parametric form. They provide enough structure that the effects of covariates - in this case the effect of time - are obvious. This is in contrast to fully nonparametric regression which usually says little about the conditional distribution of the dependent variable. Nonetheless, there are several considerations relevant for estimation of the conditional yield density that are not addressed through standard frequentist quantile regression.

Especially at extreme quantiles, a common problem is the lack of adequate yield data in many locations. We have the luxury of being able to strategically choose counties with long yield histories, but in many practical applications, similar conveniences are not available. One solution to the lack of available yield data is to borrow strength from data in neighboring locations. Because yields are affected by covariates that are spatially correlated, yields themselves will be spatially correlated. Models that borrow strength across space are desirable because they result in more accurate em-pirical estimates and predictions. Recognizing this additional feature of the problem, our proposed approach smooths across both space and quantile levels. Bayesian tests of the rate of technological change at different quantiles are easily implemented in this framework. Unlike other semipara-metric and nonparametric methods that have been applied to this problem, change can be precisely measured and interpreted.

# 1 The Conditional Yield Distribution

Crop yields, and yield distributions, are characterized by change over time and variation with the the weather. Yields in locations that are close to one another tend to be similar. Yield data in any one location is relatively sparse. Scarcity of data results in inferential difficulties, but the spatial correlation of yields suggests possible solutions. These solutions lead to improved inference on the effects of technology on the behavior of crop yields. Therefore, holistic approaches for modeling yields should consider all of these aspects. To borrow a rhetorical device from Deaton and Muellbauer (1980), an ideal approach for the modeling of yields is comprehensive in all of these dimensions.

A robust literature has dealt with yield distributions conditional on time. From a modeling perspective, the estimation is usually divided into two components: detrending of yields and estimation of the conditional yield density. Models for deterministic trends have included linear functions, polynomials, and splines. The residuals from the deterministic trend are used to construct normalized yields which are then treated as observed yields for the purposes of density estimation. Parametric examples include the normal distribution by Botts and Boles (1958), the beta distribution by Nelson and Preckel (1989), and the gamma distribution by Gallagher (1987). Nonparametric techniques were used in Goodwin and Ker (1998) and Ker and Goodwin (2000). As a result of short samples, available yield data may not provide conclusive results with respect to the choice of an optimal distribution. Nonparametric estimators typically have slow rates of convergence. Ker and Coble (2003) proposed a semiparametric method as a compromise between parametric efficiency and nonparametric flexibility.

Outside of the two-stage methodology for estimation of the yield density, several authors have examined the use of distributions with embedded trend functions. Zhu, Goodwin, and Ghosh (2011) embedded trend functions within three and four parameter beta distributions. Tolhurst and Ker (2014) used a mixture of two normal distributions with components corresponding to worst case and average growing conditions. These models naturally allow the distribution to change with time but may require the imposition of ad-hoc constraints in estimation. The interpretation of two normal components as representing catastrophic and average conditions breaks down when the components cross or when the intercept terms do not have proper ordering. Interpretation of the parameters of highly parameterized beta distributions can be challenging when measures of central tendency

and dispersion are functionals of these parameters. Methods using time-trending distributions have much in common with quantile regression, but unlike the constraints for a valid quantile process, the constraints do not always arise from statistical theory.

A somewhat distinct line of research has dealt with the effects of weather, soil type, and other covariates on crop yields. Much of the recent motivation for these studies has come from increased scientific understanding of climate change and the impacts that changing temperatures and extreme weather could have on American agriculture. Schlenker and Roberts (2006) considered the nonlinear effects of weather in determining yields. Their results were extended to additional covariates in Roberts, Schlenker, and Eyer (2013). Tack and Ubilava (2013) and Tack and Ubilava (2015) examined the ways that specific weather phenomena - the El Niño Southern Oscillation - can affect yields and drive losses in the federal crop insurance program. Antle (2010) used partial-moment functions as a means of capturing asymmetries in the effects of inputs on crop production. His results, along with those of Tack, Harri, and Coble (2012), highlight the necessity of accounting for the effects of weather and other covariates on higher moments of the yield distribution. Higher moments are often used as inputs for the economic study of decision-making under uncertainty.

Recognizing that biological constraints result in scarce yield data at specific locations, econometric methods have been designed to model correlation of yields across space and time. Ozaki et al. (2008) used a Bayesian conditional autoregressive model to capture spatial dependence across regions of Brazil. Ker, Tolhurst, and Liu (2015) developed a method for combining similar densities using Bayesian model averaging and applied this to a number of crops and counties. They found that private insurers could extract rents from government insurance programs using averages of independent county-level models. Tack and Holt (2016) showed that spatial correlation in corn yields tends to be stronger during cases of extreme weather; this phenomenon is otherwise known as state dependence. More fundamentally, spatial correlation affects the relevance of the central limit theorem for aggregate yields as well as the degree of systemic risk across insured units in crop insurance programs (Ker and Goodwin, 2000).

We have taken some liberties in categorizing previous work in crop yield modeling. The breadth of scholarship on, and amount of attention to, the topic should stand as testament to the yield as a fundamental quantity of economic analysis. A number of studies have addressed several of the relevant features of yields, and thus could be considered almost ideal under our simple definition. One

example is Ker, Tolhurst, and Liu (2015) who used time-varying mixtures of normals and Bayesian model averaging to combine mixture models from different locations. Their intent was to demonstrate the usefulness of pooling for crop insurance ratemaking and they did not explicitly consider technological change. The only study using quantile regression appears to be Barnwhal and Kotani (2013). Their focus was on the heterogeneous impacts of weather across the yield distribution and they used a relatively short (34 years) time series of yields. Yield data is likely to be particularly short in developing countries. Providing support for a quantile regression approach, they found significant effects of weather and trend variables across independently estimated quantiles

In this paper, we tackle features of yields in one comprehensive - and almost ideal - model. The model is capable of controlling for covariates and yields readily interpretable covariate effects. No assumptions are made on the form of the conditional distribution of yields, although a sensible centering distribution is specified. The model is also coherent in the sense that it is derived from a single density process. Furthermore, it smooths across spatial locations through a Gaussian process prior. Bayesian inference facilitates easy visual inspection, quantitative summaries, and statistical tests, of the effects of technological change across different quantiles of the response.

## 2   Bayesian Spatial Quantile Regression

In the case of ordinary linear regression, the conditional distribution of the dependent variable is normal with constant variance. Covariates shift the location of the conditional density, but otherwise have no effect on other moments of the distribution. Regressing crop yields on time, the implicit assumption is that technological change only affects the mean yield and not the shape of the distribution. Empirical evidence has repeatedly shown that the distribution of yields is skewed or otherwise characterized by non-normal features. The use of ordinary linear regression - and other more flexible econometric techniques for trend analysis - is not well-suited to inference when distributional change is of interest.

An alternative to strictly parametric methods are fully nonparametric approaches. Although nonparametric density estimators are completely flexible, they converge to the true distribution at a slow rate. It is also more difficult to incorporate and interpret covariate effects in nonparametric frameworks. While nonparametric strategies are suitable for estimating a stationary yield density,

we are interested in making inference on the way the conditional distribution of yields has changed over time. The incorporation of covariate effects is necessary for crop yields because the effects of technology and weather are easily confounded (Lusk, Tack, and Hendricks, 2017). Weather variables can also have important interactions among one another (Roberts, Schlenker, and Eyer, 2013).

As a compromise between parametric and nonparametric approaches, Koenker and Bassett (1978) proposed quantile regression. Quantile regression has the advantage of linearity in parameters - providing easily interpreted covariate effects - with a flexible conditional distribution. Suppose there is a simple random sample of size $N$ and that we wish to model the response $y_i$ as a function of $K$ covariates $\mathbf{X}_i$. The standard quantile regression model is

$$q(\tau|\mathbf{X}_i) = \mathbf{X}_i' \boldsymbol{\beta}(\tau) \tag{1}$$

where $q(\cdot)$ is the quantile function defined as $P(y_i < q(\tau|\mathbf{X}_i)) = \tau \in [0,1]$. The coefficients $\boldsymbol{\beta}$ vary with the quantile $\tau$ resulting in different covariate effects at different quantiles of the conditional distribution. Quantile regression thus provides a distribution-free approach to the estimation of the conditional density.

An estimate of equation 1 for several quantiles can be obtained by maximum likelihood. The problem is couched as one of optimization with the regression quantile given by the solution to

$$\min_{\boldsymbol{\beta} \in R^K} \left[ \sum_{i|y_i \geq \mathbf{X}_i'\boldsymbol{\beta}} \tau|y_i - \mathbf{X}_i'\boldsymbol{\beta}| + \sum_{i|y_i < \mathbf{X}_i'\boldsymbol{\beta}} (1-\tau)|y_i - \mathbf{X}_i'\boldsymbol{\beta}| \right] \tag{2}$$

Classical quantile regression has since been applied in many economic studies where the conditional distribution of the response is of primary interest (Buchinsky, 1994; Jalan and Ravallion, 2001; Sanglestsawai, Rejesus, and Yorobe, 2014). The quantile curves are usually estimated independently so estimates at extreme quantiles do not borrow information from estimates at quantiles near the mass of the distribution. Because of their independent estimation, quantile curves can cross, rendering the distribution of the dependent variables invalid. Methods for the simultaneous estimation of quantile curves have been proposed and empirical applications indicate that more precise coefficient estimates result (Bondell, Reich, and Wang, 2011; Tokdar and Kadane, 2012).

Bayesian methods for quantile regression extend to a number of domains including limited in-

dependent variables, instrumental variables, spatial data, and regularization (Hallin, Lu, and Yu, 2009; Lancaster and Jun, 2010; Li, Xi, and Lin, 2010; Benoit and Poel, 2012). Bayesian models are able to incorporate both temporal and spatial information into estimation of the quantiles while providing benefits in terms of inference and the construction of conditional densities. Consider an extension of the hypothetical random sample for equation 1 where the response $y_i$ is observed at space/time location $(s,t)_i$. The time and location of the $i$th observation are given by $s_i$ and $t_i$. The model of equation 1 can be extended to

$$q(\tau|\mathbf{X}_i, s_i) = \mathbf{X}_i' \boldsymbol{\beta}(\tau, s_i) \tag{3}$$

where the effects of the covariates now vary by both quantile level and location. A valid quantile process will be nondecreasing in the quantile for all values of the predictors. The advantage of a model and estimation procedure for the quantile process, as opposed to a single quantile curve, is that information is shared across quantile levels. Formulating such a model is clearly more complicated than classical quantile regression because parameter effects must be specified over a theoretically infinite number of quantile levels. To construct the process, let $M$ be a number of basis functions with $B_m(\tau)$ a basis function of $\tau$. Then the quantile process can be modeled as

$$\beta_j(\tau, \mathbf{s}) = \sum_{m=1}^{M} B_m(\tau) \alpha_{jm}(\mathbf{s}) \tag{4}$$

where $j = 1, \ldots, K$ and the $\alpha_{jm}(\mathbf{s})$ are basis coefficients that determine the shape of the quantile process. This semiparametric approach was first proposed by Reich, Fuentes, and Dunson (2011) who used quantile regression to study trends in summer ozone levels. Several extensions of the model have since been implemented and include Reich (2012) who accounts for residual correlation in predictors, Reich and Smith (2013) who considered censored data, and Smith et al. (2015) who developed a hierarchical variant of the model.

To obtain estimates of the quantile process, a suitable basis must be identified such that the monotonicity constraint on the quantile process can be imposed. Selection can be made among a number of different basis functions including parametric quantile functions (Reich, 2012), bernstein polynomials (Reich, Fuentes, and Dunson, 2011), or splines (Smith et al., 2015). The R package

BSquare can be used to implement non-spatial versions of Bayesian simulateneous quantile regression (Reich, 2013). The form of the basis having been selected, the number of basis functions must also be specified. If Bernstein polynomials are used, then

$$B_m(\tau) = \binom{M}{m} \tau^m (1-\tau)^{M-m} \tag{5}$$

which is the Bernstein basis polynomial and equation 4 is specifically a Bernstein polynomial of degree $M$. Bernstein polynomials have proven to be popular for statistical and economic problems where shape constraints must be imposed (Ryu and Slottje, 1996; Chak, Madras, and Smith, 2005; Wang and Ghosh, 2014). As the degree of the Bernstein polynomial goes to infinity, it converges uniformly to any arbitrary continuous function on the interval $[0, 1]$.

The relevant constraint is that the process in equation 4 must be monotonic increasing in $\tau$. Using the Bernstein polynomial basis, Reich, Fuentes, and Dunson (2011) show that this reduces to $\gamma_{jm} = \alpha_{jm} - \alpha_{jm-1} \geq 0$ for $m = 2, \ldots, M$. The constraints are ensured by a latent unconstrained variable where $\alpha_{jm}(\mathbf{s}) = \sum_{l=1}^m \gamma_{jl}(\mathbf{s})$ and

$$\gamma_{jm}(\mathbf{s}) = \begin{cases} \gamma_{jm}^*(\mathbf{s}) & \gamma_{1m}^*(\mathbf{s}) + \sum_{j=2}^K I(\gamma_{jm}^*(\mathbf{s}) < 0)\gamma_{jm}^*(\mathbf{s}) \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Specifying the $\gamma_{jm}^*(\mathbf{s})$ as a spatial Gaussian process results in smoothing across spatial locations. The processes have spatial covariance $\text{Cov}(\gamma_{jm}(\mathbf{s}), \gamma_{jm}(\mathbf{s}')) = \sigma_j^2 \exp(-||\mathbf{s}-\mathbf{s}'||/\rho_j)$ and mean $\bar{\gamma}_{jm}(\Lambda)$. The spatial covariance depends on $\sigma_j^2$ which is the variance of $\sigma_{jm}^*(\mathbf{s})$ and $\rho_j$ which is a parameter controlling the range of spatial correlation. $\Lambda$ is the hyperparameter set for the centering distribution.

Two approaches may be taken to estimate the model. In cases of small and moderate samples, the full model can be estimated with an MCMC sampling scheme. Reich, Fuentes, and Dunson (2011) develop an approximate method that first fits separate quantile regressions over a number of quantiles and then uses the Bayesian spatial model to construct the quantile process over these initial estimates. In terms of mean squared error, coverage, and power, the approximate and full models perform better than classical quantile regression. The MCMC algorithms are given in appendices to Reich, Fuentes, and Dunson (2011) and Reich (2012) and are not restated here.

The model ultimately provides a means of capturing changes in a conditional distribution while allowing a high degree of flexibility in terms of covariate effects. Covariates affect higher moments, and more generally the shape, of the distribution. These effects vary across locations and across quantile levels. The second point is particularly useful because of the increasing evidence of the nonlinear effects of weather on crop yields. This model allows for nonlinearities through quantile-varying covariates. Although the quantile process is linear in the covariates, interactions and other transformations of variables can be included. A modicum of spatial smoothing is induced through the Gaussian process prior.

## 3   Analysis of Illinois Corn Yields

We apply the approximate model of Reich, Fuentes, and Dunson (2011) to all-practice corn yields from Illinois. Data from all Illinois counties were obtained from the National Agricultural Statistics Service for the years 1926 to 2015. In terms of acres of corn harvested in 2016, Illinois ranks second only to Iowa. The yield in each county is mapped to the centroid of its corresponding areal unit. Only those counties with a complete yield history were retained; this results in 76 spatial locations (out of 102 counties in Illinois) across a period of 91 years. Figure 3a plots the temporal evolution of yields by year while figure 3b plots the unconditional density of crop yields. Both figures contain complementary information on the development of crop yields over time. From 1925 through 1950, gains in yields were modest and observed yields tended to cluster about the mean. Beginning in the late 1960s, the mean yield began to increase rapidly along with variation in yields. Over the period of analysis, Illinois has been subject to several catastrophic events, most notably in 1983, 1988, and 2012.

The choropleth map in figure 3c shows the mean yield in each county over the entire sample, while figure 3d shows the minimum yield over time. Neither map is adjusted for time trends, revealing several noticeable spatial patterns. Counties with the highest mean yields cluster in the northern and central parts of the state, concentrated around Logan County. These areas of Illinois are classified as having a hot summer - humid continental climate according to the Koppen climate classification. Lower yielding areas are typically categorized as humid subtropical climates.

If counties across Illinois were roughly subject to the same weather, the same rate of technolog-

ical diffusion, and the effects of weather and technology were constant across quantiles of the yield distribution, the maps in figures 3c and 3d would be similar. Ranking counties in terms of minimum and mean yield, little variation between the two rankings would be apparent. Dissimilarity in the patterns of the maps suggests that weather, technology adoption, and effects across quantiles vary by county. Accounting for spatial correlation by smoothing estimates across space incorporates variation across locations in a sensible way.

We consider explanatory weather variables similar to those used by Tolhurst and Ker (2014) and Rejesus et al. (2015), although in this application, the weather variables directly affect the yield distribution. Weather data are taken from the National Climatic Data Center's Time Bias Corrected Divisional Temperature-Precipitation-Drought Index Data. One complication from using long series of historical yields is that weather covariates are typically not available at the county level for all years. Both cooling degree days and precipitation are only available at the climate division level before 1955. There are nine climate divisions within Illinois and each Illinois county is assigned to a single climate division. It would be preferable to have weather variation across all counties, but the use of division data allows for some spatial variation and the use of more historical yields. Plots of yields by cooling degree days across the growing season and precipitation across the growing season are shown in figure 4. The relationships in the plots are consistent with previous findings of the effects of CDD and PCPN on corn yields.

Due to constraints necessary to produce a valid quantile function, the variables must be mapped to the unit interval. For time, the variable used in the estimation is given by year minus 1925 divided by 91. Both CDD and PCPN were normalized using the normal distribution cumulative density function. The estimated model includes a linear time trend, cooling degree days, precipitation, precipitation squared, cooling degree days interacted with precipitation, and cooling degree days interacted with the square of precipitation. The model is estimated using the approximate method of Reich, Fuentes, and Dunson (2011). The prior for the spatial range is Gamma(0.5, 0.5) and the prior for the variances are Inverse Gamma(0.1, 0.1). A total of 20,000 draws were taken from the Markov chain Monte Carlo algorithm with 10,000 discarded as burn-in. The model was evaluated at all quantiles from 0.05 to 0.95 by 0.05 as well as the octiles 0.125, 0.376, 0.625, and 0.875.

Results of the estimation are summarized in the plots of figure 5. Plotting the effect of time for all counties in the sample reveals that the effect of technological change, as embodied in the coefficient

associated with time, is greater at higher quantiles for almost all counties in the sample. At least for Illinois corn, results support the hypothesis of Lobell et al. (2014). Nonetheless, substantial variation across counties remains. Some counties have seen more rapid increases in yields at all quantiles and greater dispersion in yields. We are also interested in statistically testing the validity of these results.

Figure 5b shows actual yields in Knox County, Illinois as well as posterior mean quantile curves with all covariates except year fixed at their means over the sample. The quantile process provides a close fit to the bulk of the unconditional distribution and weather covariates have the effect of shifting the quantiles up or down. As in the plot of the year effect for each county, the slopes of the quantile curves at higher quantiles tend to be larger than the slope at lower quantiles. The conditional distribution appears to be becoming more negatively skewed over time. Slightly more than 5% of the observations lie above the 95 quantile curve and 5 quantile curve, but the fit is not egregious given that the other predictors have been fixed. Were extremely low precipitation included, for instance, the fitted quantile curves would shift down.

The effects of weather also differ across the quantiles of the distribution. In Knox County, IL the number of cooling degree days decreases yields at all quantiles of the conditional distribution while yields increase in precipitation. The interaction between CDD and PCPN is positive indicating that precipitation mitigates the effects of extreme temperatures. This effect is strongest at the upper quantiles of the yield distribution. Estimated effects of weather are largely consistent with those found in previous studies.

Conditional densities across years are shown for four representative counties where, again, all covariates except time are fixed at their sample means. The conditional densities are constructed using estimates of the quantile process and the adaptive kernel density estimator of Silverman (1986). A similar application can be found in Portnoy and Koenker (1989). This approach also allows for forecasted yield distributions so the quantile method provides a hypothetical method for pricing crop insurance policies. A recent study by Gaglianone and Lima (2012) demonstrates conditional density forecasting in a financial context.

The densities are evaluated in 1925, 1955, 1970, and 2015 and can be compared to the densities obtained in Tolhurst and Ker (2014) and Zhu, Goodwin, and Ghosh (2011). Two important qualifications must be given, the length of time considered in those studies is substantially different from

ours and they only considered distributions conditional on time. An advantage here is that the distribution is conditional on time and weather and we can then make statements about changes in the distribution with the effects of weather having been at least partially controlled. The importance of controlling for weather is the focus of Lusk, Tack, and Hendricks (2017) who found that failure to control leads to inaccurate measurement of the effect of technological change.

Because we model the full quantile process in a Bayesian framework, we can obtain posterior distributions of functions of model parameters. This facilitates simple tests of the impact of technological change across the conditional distribution. As an example, consider a single county where we are interested in whether the coefficient on time is greater at the 75th quantile compared to the 25th quantile. Within each Monte Carlo markov chain draw, the difference between $\beta_1(.75, s)$ and $\beta_1(.25, s)$ is obtained and a posterior distribution of the difference is constructed. This posterior distribution is shown in figure X for Adams County, IL. If the posterior mean is greater than zero, and zero is not within the given credible set, then we view the effect of technological change as being different across the two quantiles. Based on the posterior distribution of the coefficients, we find that 75 out of 76 counties had greater technological change at the 75th quantile.

Our results correspond closely with those of Tolhurst and Ker (2014) who utilized a frequentist framework. At least for corn, changes in production practices, seed technologies, and other technologies, have had important effects on the yield distribution. More importantly, these effects have not been confined to the mean of the distribution. More rapid increases in upper quantiles imply that the distribution has become left-skewed over time; dispersion of the condition distribution of yields increases with time. The use of quantile regression does not exclude more formal measures of the skewness and kurtosis of the conditional distributions. While conventional skewness and kurtosis are estimated by sample averages, there exist several quantile based alternatives. These alternatives replace, more or less, the mean with the median for location and the variance with an interquartile range for dispersion. Based on the estimated quantiles, we calculated the skewness coefficient of Hinkley (1975) and the kurtosis coefficient of Crow and Siddiqui (1967). Kim and White (2004) found that both measures compare favorably to classical skewness and kurtosis coefficients and are robust to outliers.

Table 1 shows summary statistics for the estimated coefficient of variation, skewness, and kurtosis, using quantile-based measures for the last year in the sample. In all counties, the median yield

increased over time and the dispersion of the data has increased. At the beginning of the sample, nine of the counties had positive skew while only 4 had positive skew at the end of the sample. Mean measurements of skewness and kurtosis across the counties indicate negative skew and excess kurtosis. Growth at lower quantiles of the distribution has lagged growth at middle and upper quantiles, resulting in significant changes to the yield distribution. The effects of technological change have not been confined to the location of the yield distribution but involve higher moments.

## 4 Conclusion

The econometric study of crop yields is characterized by several salient features. Yields are subject to technological change, the effects of weather and other covariates, spatial correlation, and a scarcity of data in many locations. We implement a Bayesian spatial quantile regression model for the conditional distribution of yields that allows technological change to affect different quantiles in different ways. Other covariates (namely weather) can have nonlinear effects on the conditional distribution. The model borrows strength across locations, resulting in more accurate inference, and partially ameliorating the statistical problems inherent in the estimation of yield densities.

The Bayesian framework avoids potential complications that can arise from the use of frequentist quantile regression, namely quantile crossing. Because the entire quantile process is modeled, inference concerning technological change is greatly simplified. We are able to test whether the rate of technological change differs across quantiles. Direct statements concerning the probability of observing an extreme yield can be obtained. For the Illinois corn yields that we consider, technological change at the 75th quantile was greater than at the 25th quantile in almost all counties. The magnitude of the difference was typically large and meaningful, although smaller than what would be obtained if the effect of weather was not controlled.

Fitting nonparametric densities to the quantile estimates, it is possible to generate density forecasts. Quantile-based measures of dispersion, skewness, and kurtosis, can also be obtained. Consistent with previous findings, the median Illinois yield and interquartile range have increased over time. Throughout the period, yields have tended to be negatively skewed and skewness has decreased over time. A finding of increasing kurtosis does not align with previous studies, but this study has several important differences from earlier works. Even though excess kurtosis is evident

in this application, it tends to be mild.

After controlling for the number of growing degree days and precipitation, we find that the effects of technological change are not limited to the mean of the yield distribution. Weather variables have different effects across quantiles of the distribution and locations. Increased precipitation mitigates the effects of increased growing degree days. Variation across locations highlights the importance of location-specific considerations when looking toward strategies intended to mitigate the effects of extreme weather. One aspect of yield modeling that have not considered is that spatial correlation could change with the weather regime. Capturing state dependent correlation would require a more sophisticated dynamic model.

Several potentially beneficial adjustments to the model might be made. We have characterized the model as almost ideal, and although it has many desirable properties, the model still involves tradeoffs in terms of flexibility. Several consequences for other areas of study are evident. Concerning the diffusion of technology, previous studies have largely confined the profitability of technology in terms of the increase in the mean yield that may occur. This ignores considerations of risk and motivations for technology adoption that extend beyond the first moment of the yield distribution. Likewise, research predicting the economic impacts of climate variation should consider the effects that weather variables have on the entire conditional distribution of yields. Research is already being undertaken in these directions and may benefit from the methods implemented here.

# Bibliography

Alston, J.M., M.A. Andersen, J.S. James, and P.G. Pardey. 2010. *Persistence Pays: U.S. Agricultural Productivity Growth and the Benefits from Public R& D Spending*. New York: Springer.

Antle, J.M. 2010. "Asymmetry, Partial Moments, and Production Risk." *American Journal of Agricultural Economics* 92:1294–1309.

Barnwhal, P., and K. Kotani. 2013. "Climatic impacts across agricultural crop yield distributions: An application of quantile regression on rice crops in Andhra Pradesh, India." *Ecological Economics* 87:95–109.

Benoit, D.F., and D.V.D. Poel. 2012. "Binary Quantile Regression: A Bayesian Approach Based on the Asymmetric Laplace Distribution." *Journal of Applied Econometrics* 27:1175–1188.

Bondell, H., B. Reich, and H. Wang. 2011. "Non-crossing quantile regression curve estimation." *Biometrika* 97:825–838.

Botts, R.R., and J.N. Boles. 1958. "Use of Normal-Curve Theory in Crop Insurance Ratemaking." *Journal of Farm Economics* 40:733–740.

Buchinsky, M. 1994. "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression." *Econometrica* 62:405–458.

Chak, P., N. Madras, and B. Smith. 2005. "Semi-nonparametric estimation with Bernstein polynomials." *Economics Letters* 89:153–156.

Crow, E., and M. Siddiqui. 1967. "Robust Estimation of Location." *Journal of the American Statistical Association* 62:353–389.

Deaton, A., and J. Muellbauer. 1980. "An Almost Ideal Demand System." *American Economic Review* 70:312–326.

Dixon, R. 1980. "Hybrid Corn Revisited." *Econometrica* 48:1451–1461.

Gaglianone, W., and L. Lima. 2012. "Constructing Density Forecasts from Quantile Regressions." *Journal of Money, Credit, and Banking* 44:1589–1607.

Gallagher, P. 1987. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *American Journal of Agricultural Economics* 69:796–803.

Goodwin, B.K., and A.P. Ker. 1998. "Nonparametric Estimation of Crop Yield Dsitributions: Implications for Rating Group–Risk Crop Insurance Contracts." *American Journal of Agricultural Economics* 80:139–153.

Griliches, Z. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25:501–522.

Hallin, M., Z. Lu, and K. Yu. 2009. "Local Linear Spatial Quantile Regression." *Bernoulli* 15:659–686.

Hinkley, D. 1975. "On Power Transformations to Symmetry." *Biometrika* 62:101–111.

Hurley, T.M., X. Rao, and P.G. Pardey. 2014. "Re-examining the Reported Rates of Return to Food and Agricultural Research and Development." *American Journal of Agricultural Economics* 96:1492–1504.

Jalan, J., and M. Ravallion. 2001. "Behavioral responses to risk in rural China." *Journal of Development Economics* 66:23–49.

Ker, A.P., and K. Coble. 2003. "Modeling Conditional Yield Densities." *American Journal of Agricultural Economics* 85:291–304.

Ker, A.P., and B.K. Goodwin. 2000. "Nonparametric Estimation of Crop Insurance Rates Revisited." *American Journal of Agricultural Economics* 82:463–478.

Ker, A.P., T.N. Tolhurst, and Y. Liu. 2015. "Bayesian Estimation of Possibly Similar Yield Densities: Implications for Rating Crop Insurance Contracts." *American Journal of Agricultural Economics* 98:360–382.

Kim, T.H., and H. White. 2004. "On more robust estimation of skewness and kurtosis." *Finance Research Letters* 1:56–73.

Koenker, R., and G. Bassett. 1978. "Regression Quantiles." *Econometrica* 46:33–50.

Lancaster, T., and S.J. Jun. 2010. "Bayesian quantile regression methods." *Journal of Applied Econometrics* 25:287–307.

Li, Q., R. Xi, and N. Lin. 2010. "Bayesian Regularized Quantile Regression." *Bayesian Analysis* 5:533–556.

Lobell, D.B., M.J. Roberts, W. Schlenker, N. Braun, B.B. Little, R.M. Rejesus, and G.L. Hammer. 2014. "Greater Sensitivity to Drought Accompanies Maize Yield Increase in the U.S. Midwest." *Science* 344:516–519.

Lusk, J., J. Tack, and N.P. Hendricks. 2017. "Heterogeneous Yield Impacts from Adoption of Genetically Engineered Corn and the Importance of Controlling for Weather." In W. Schlenker, ed. *Understanding Productivity Growth in Agriculture*. University of Chicago Press.

Nelson, C.H., and P.V. Preckel. 1989. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics* 71:370–378.

Ozaki, V.A., S.K. Ghosh, B.K. Goodwin, and R. Shirota. 2008. "Spatio-Temporal Modeling of Agricultural Yield Data with an Application to Pricing Crop Insurance Contracts." *American Journal of Agricultural Economics* 90:951–961.

Portnoy, S., and R. Koenker. 1989. "Adaptive L Estimation of Linear Models." *Annals of Statistics* 17:362–381.

Reich, B.J. 2012. "Spatiotemporal quantile regression for detecting distributional changes in environmental processes." *Journal of the Royal Statistical Society: Series C* 61:535–553.

Reich, B.J., M. Fuentes, and D.B. Dunson. 2011. "Bayesian Spatial Quantile Regression." *Journal of the American Statistical Association* 106:6–20.

Reich, B.J., and L.B. Smith. 2013. "Bayesian Quantile Regression for Censored Data." *Biometrics* 69:651–660.

Reich, L.S..B. 2013. *BSquare: Bayesian Simultaneous Quantile Regression*. R package version 1.1.

Rejesus, R.M., K.H. Coble, M.F. Miller, R. Boyles, B.K. Goodwin, and T.O. Knight. 2015. "Accounting for Weather Probabilities in Crop Insurance Rating." *Journal of Agricultural and Resource Economics* 40:306–324.

Roberts, M.J., W. Schlenker, and J. Eyer. 2013. "Agronomic Weather Measures in Econometric Models of Crop Yield with Implications for Climate Change." *American Journal of Agricultural Economics* 95:236–243.

Ryu, H.K., and D.J. Slottje. 1996. "Two flexible functional form approaches for approximating the Lorenz Curve." *Journal of Econometrics* 72:251–274.

Sanglestsawai, S., R.M. Rejesus, and J.M. Yorobe. 2014. "Do lower yielding farmers benefit from Bt corn? Evidence from instrumental variables quantile regressions." *Food Policy* 44:285–296.

Schlenker, W., W.M. Hanemann, and A.C. Fisher. 2006. "The Impact of Global Warming on U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions." *Review of Economics and Statistics* 88:113–125.

Schlenker, W., and M. Roberts. 2006. "Nonlinear effects of weather on corn yields." *Review of Agricultural Economics* 28:391–398.

Sherrick, B.J., C.A. Lanoue, J.D. Woodard, G.D. Schnitkey, and N.D. Paulson. 2014. "Crop yield distributions: fit, efficiency, and performance." *Agricultural Finance Review* 74:348–363.

Silverman, B. 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Smith, L.B., B.J. Reich, A.H. Herring, P.H. Langlois, and M. Fuentes. 2015. "Multilevel Quantile Function Modeling with Application to Birth Outcomes." *Biometrics* 71:508–519.

Sparger, J.A., G.W. Norton, P.W. Heisey, and J. Alwang. 2013. "Is the Share of Maintenance Research Rising in the United States?" *Food Policy* 38:126–135.

Tack, J., A. Harri, and K. Coble. 2012. "More than Mean Effects: Modeling the Effect of Climate on the Higher Order Moments of Crop Yields." *American Journal of Agricultural Economics* 94:1037–1054.

Tack, J., and M. Holt. 2016. "The Influence of Weather Extremes on the Spatial Correlation of Corn Yields." *Climatic Change* 134:299–309.

Tack, J., and D. Ubilava. 2015. "Climate, Risk, and Insurance: Measuring the Effects of ENSO on U.S. Crop Insurance." *Agricultural Economics* 46:245–257.

—. 2013. "The Effect of El Nino Southern Oscillation on U.S. Corn Production and Downside Risk." *Climatic Change* 121:689–700.

Tokdar, S., and J. Kadane. 2012. "Simultaneous linear quantile regression: a semiparametric Bayesian approach." *Bayesian Analysis* 7:51–72.

Tolhurst, T.N., and A.P. Ker. 2014. "On Technological Change in Crop Yields." *American Journal of Agricultural Economics* 97:137–158.

Wang, J., and S. Ghosh. 2014. "Shape restricted nonparametric regression with Bernstein polynomials." *Computational Statistics & Data Analysis* 56:2729–2741.

Zhu, Y., B.K. Goodwin, and S.K. Ghosh. 2011. "Modeling Yield Risk Under Technological Change: Dynamic Yield Distributions and the U.S. Crop Insurance Program." *Journal of Agricultural and Resource Economics* 36:192–210.
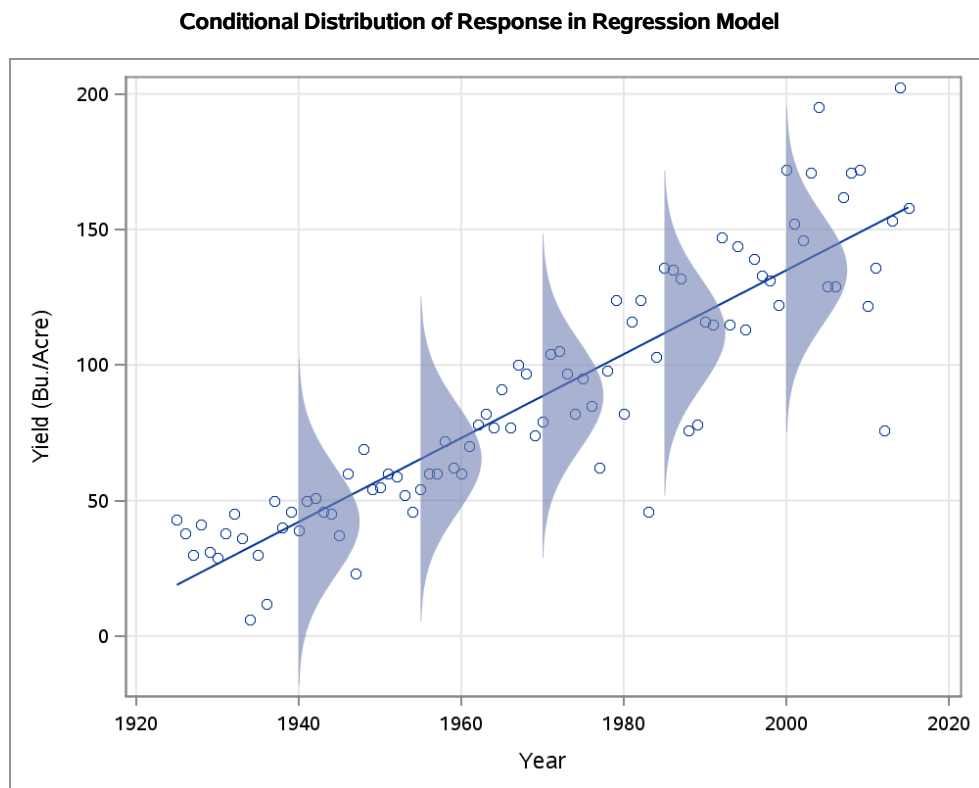
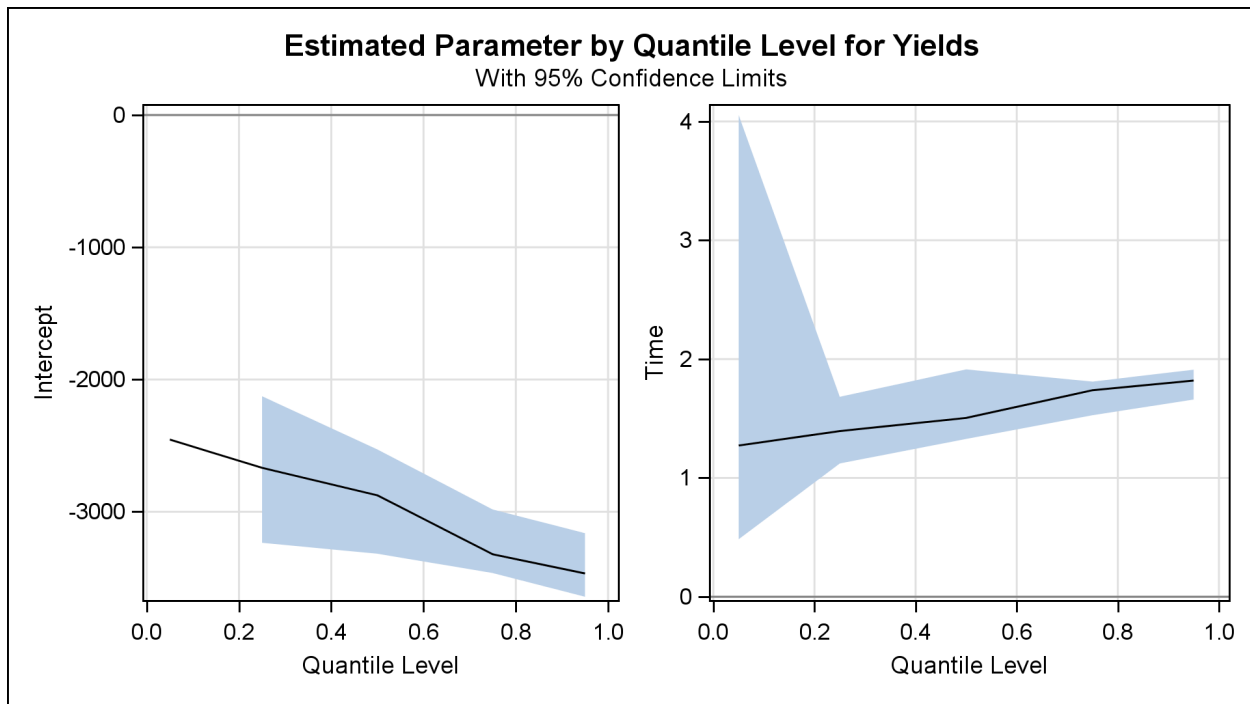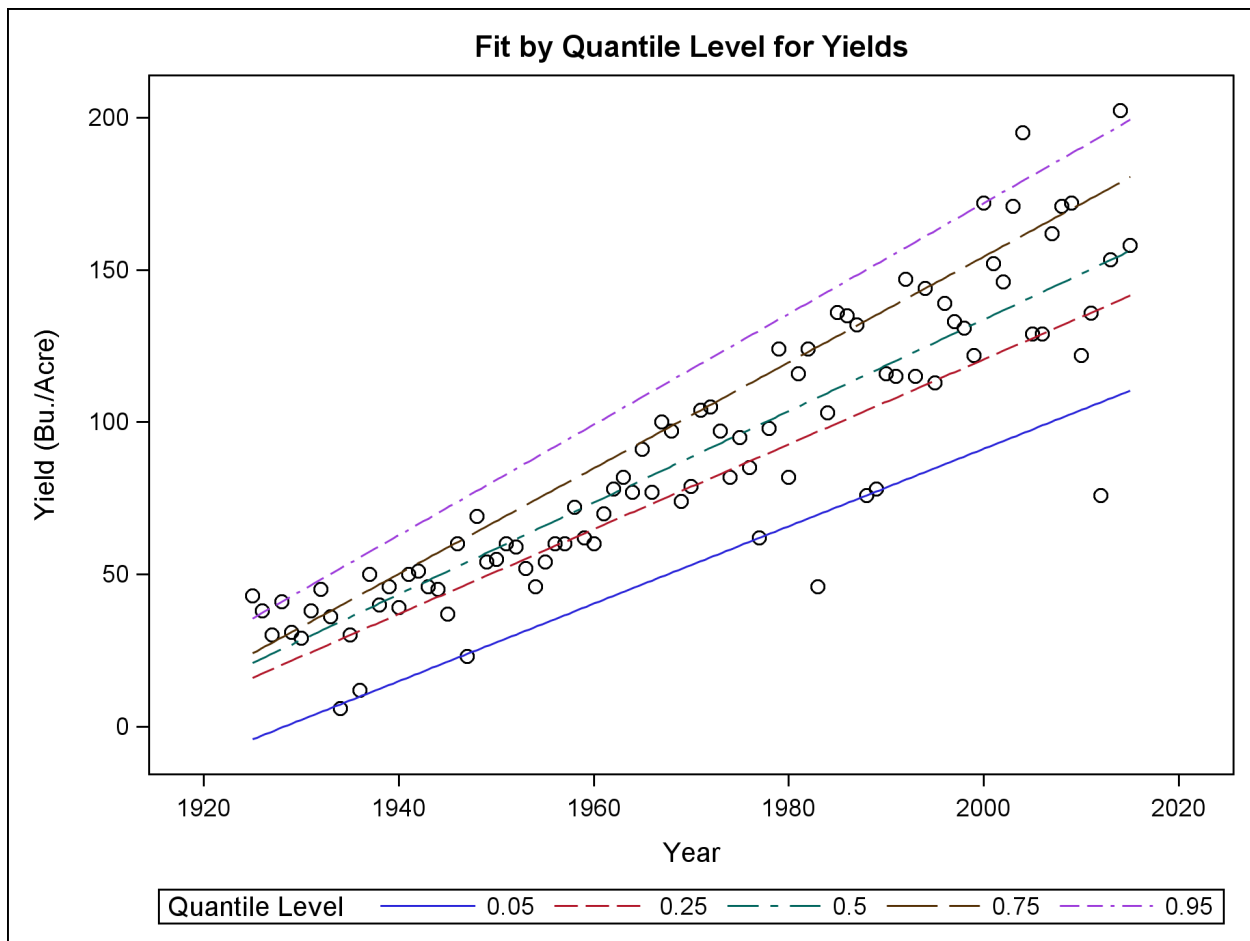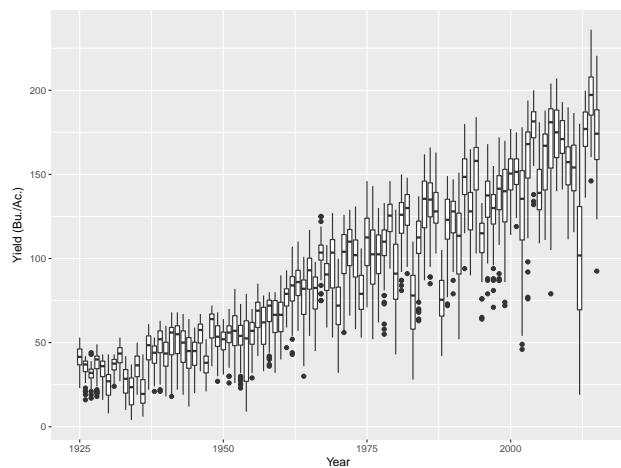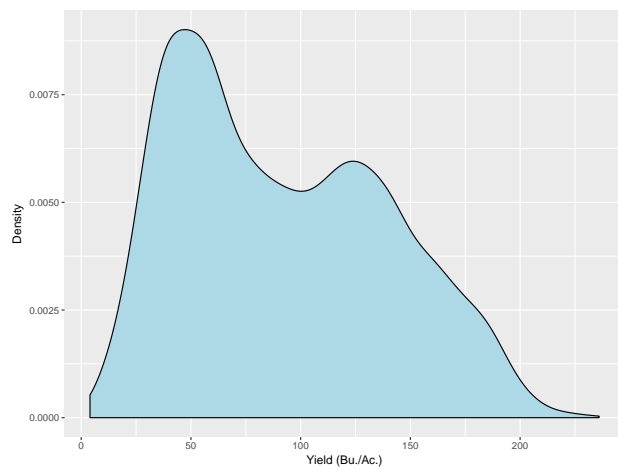Figure 1: Ordinary Linear Regression for Adams County, IL Corn Yields

Figure 2: Quantile Regression for Adams County, IL Corn Yields

(a) Illinois Yields Over Time

(b) Unconditional Distribution of Illinois Yields
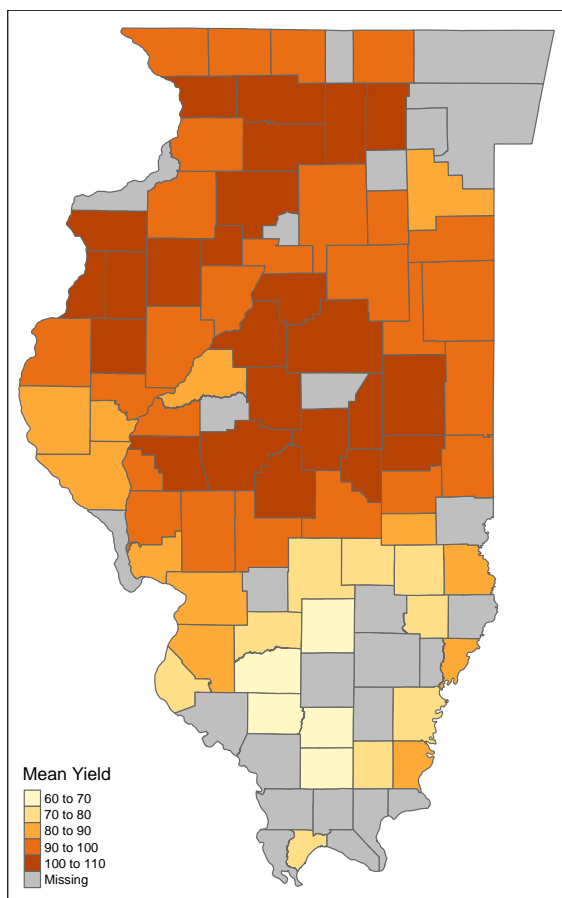
(c) Mean Yield Over Sample
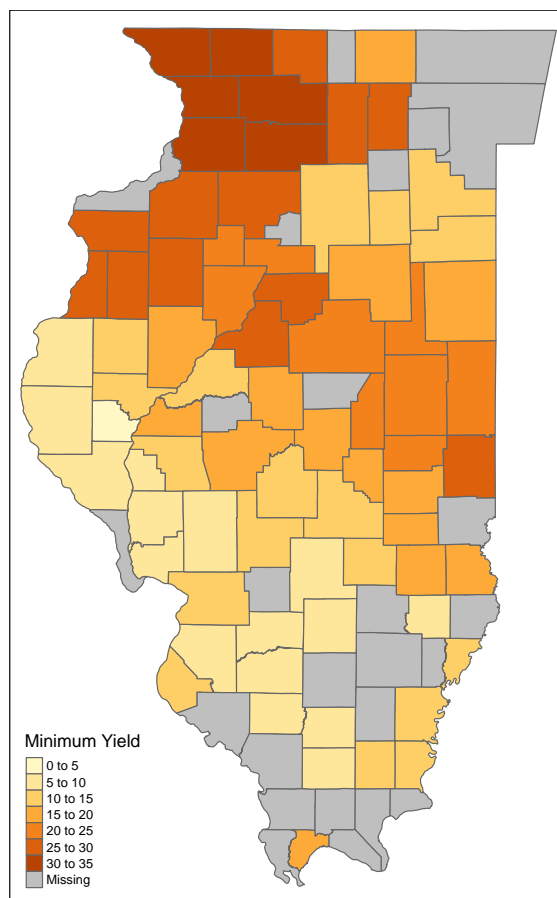
(d) Minimum Yield Over Sample

Figure 3: Illinois All-Practice Corn Yields

(a) Cooling Degree Days

(b) Precipitation
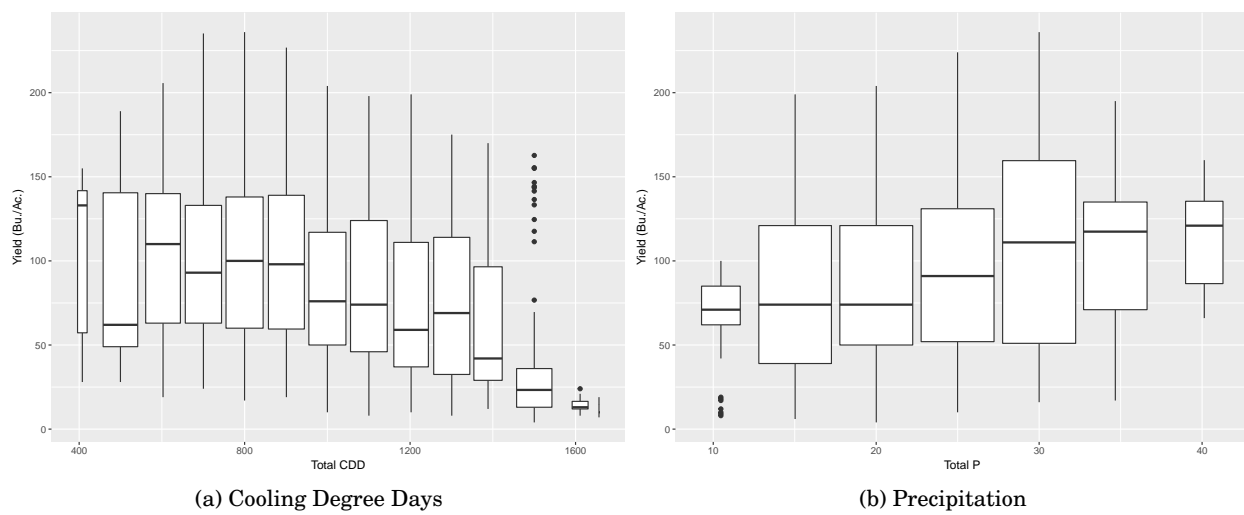
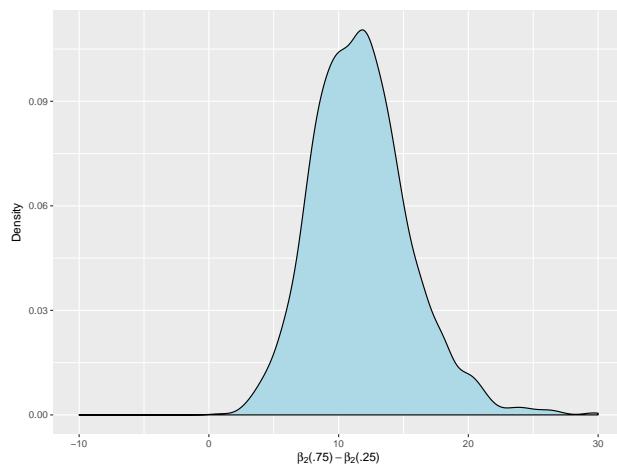Figure 4: Weather and Illinois Corn Yields
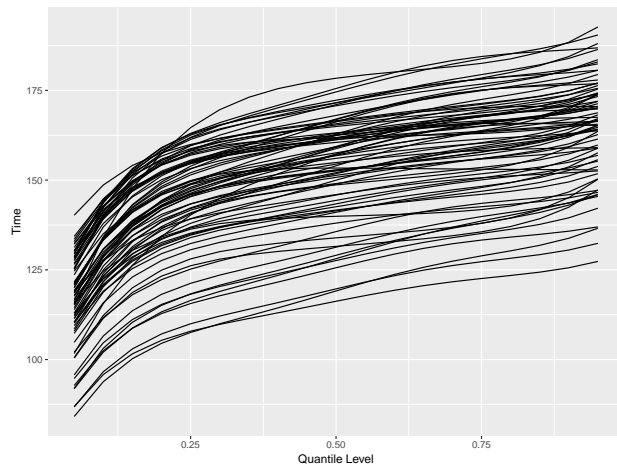
(a) Year Effect by County
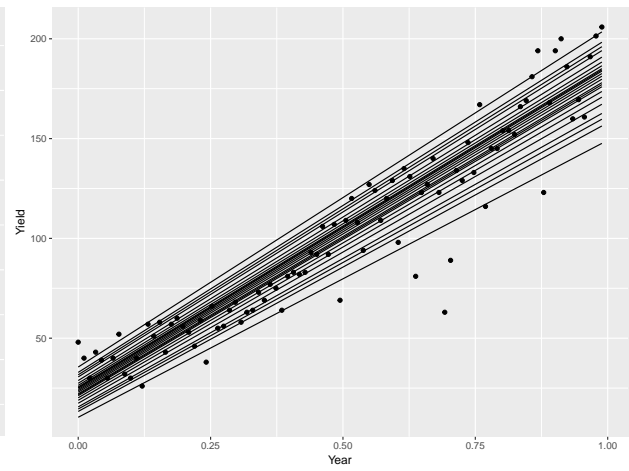
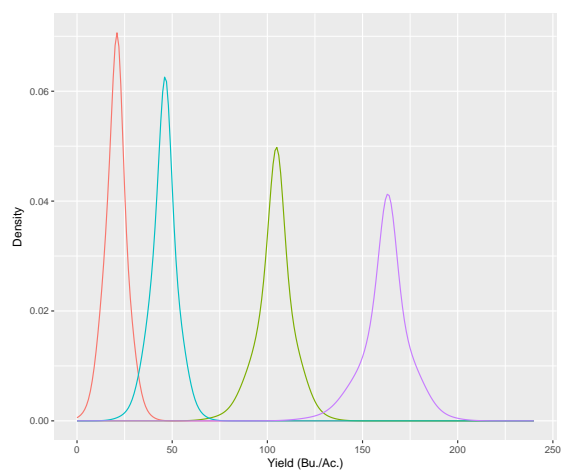(b) Fit by Year for Knox, IL
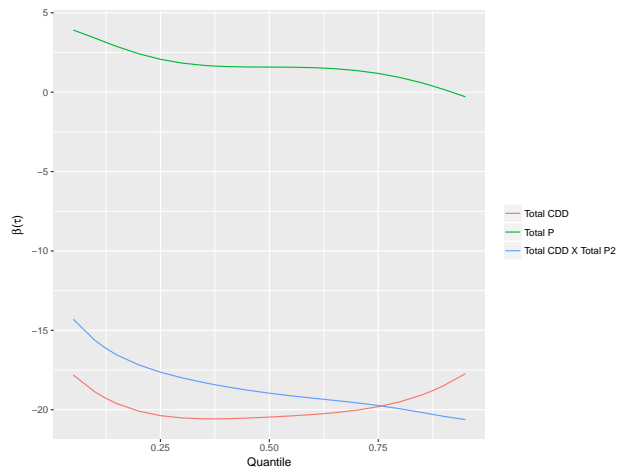
(c) Coefficients by Quantile for Knox, IL
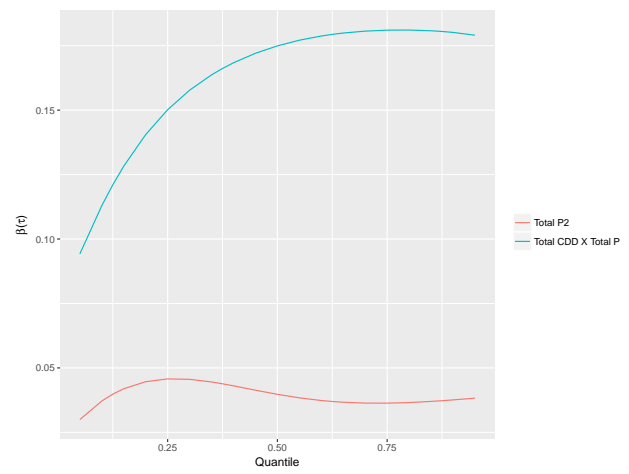
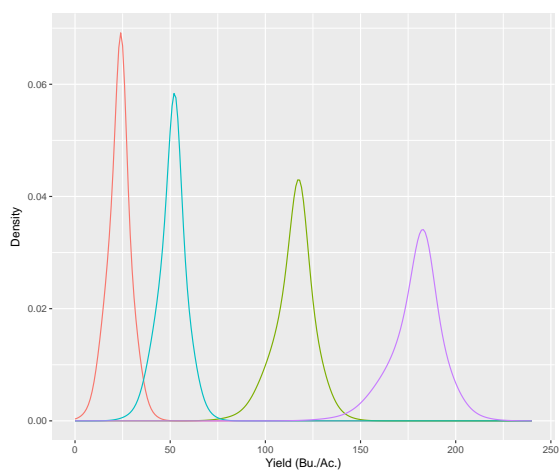(d) Coefficients by Quantile for Knox, IL
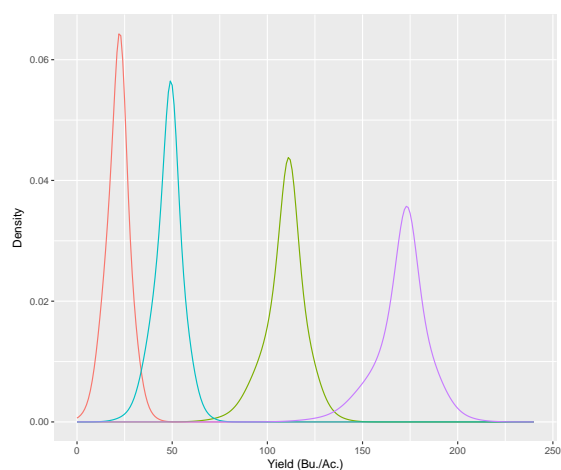
(e) Posterior for Difference between quantiles
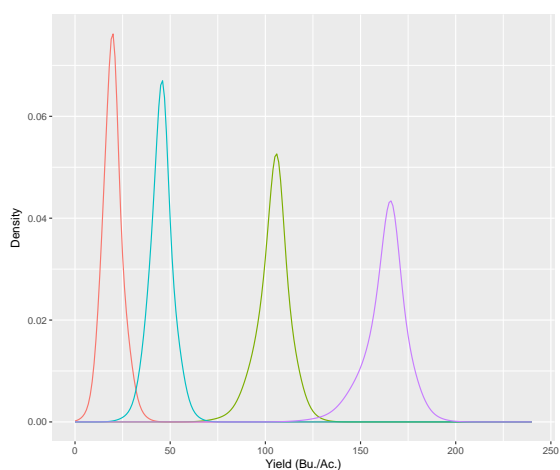
Figure 5: Estimation Results

(a) Adams, IL

(b) Knox, IL

(c) Greene, IL

(d) Wabash, IL

Figure 6: Conditional Yield Distributions Over Time

| Measure | Definition | Crop | Mean | Std. Dev. |
|---------|-----------|------|------|-----------|
| Dispersion | $\frac{Q_{75}-Q_{25}}{Q_{75}+Q_{25}}$ | Corn | 16.59 | 3.44 |
| | | Soybeans | | |
| | | Wheat | | |
| Skewness | $\frac{(Q_{75}-Q_{50})-(Q_{50}-Q_{25})}{Q_{75}-Q_{25}}$ | Corn | -0.15 | 0.07 |
| | | Soybeans | | |
| | | Wheat | | |
| Kurtosis | $\frac{(Q_{87.5}-Q_{62.5})+(Q_{37.5}-Q_{12.5})}{Q_{75}-Q_{25}}-1.23$ | Corn | 0.41 | 0.21 |
| | | Soybeans | | |
| | | Wheat | | |

Table 1: Quantile-based Moments Across County Distributions in 2015