A Learning Rule for Updating the Distribution of Crop Fields over Space and Time

Stephen M. Stohs and Jeffrey T. LaFrance

September 25, 2003

Department of Agricultural and Resource Economics

University of California, Berkeley

Abstract

The U.S. government operates the multi-peril crop insurance (MPCI) program to provide farmers with comprehensive protection against yield risk due to weather-related causes of loss and certain other unavoidable perils. Coverage is available on over 75 crops in primary production areas of the U.S. Producers may elect a coverage level on the range from 50 to 75 percent of the actual production history (APH) mean yield, where APH mean yield is defined as the farm-level average of a minimum of four, and a maximum of up to ten, consecutive years of yield[1].

A widely recognized feature of crop yield data are the high levels of spatial and temporal dependence. The MPCI ratemaking procedure currently in force makes little use of this dependency in the data, thereby failing to utilize information which could be used to more accurately price the insurance. We develop a methodology for estimating farm-level yield distributions that incorporates the spatial and temporal dependencies in yield data. The approach also permits actual learning across space and time, and from aggregate (county and regional) to disaggregate (individual farm) levels of crop production.

The U.S. federal crop insurance program has long been plagued with low participation rates, and taxpayer-funded subsidies are routinely employed to encourage higher participation levels. Current provisions of the insurance contract allow for premium subsidies on the range from 64 percent at the 75 percent coverage level up to 100 percent at the 50 percent coverage level. Besides covering over 50 percent of the premiums, additional subsidies are employed to provide incentives for private insurers to deliver the coverage, in the form of delivery expenses (marketing and service cost reimbursements), and reinsurance protection. Unlike private insurance, where the insurance company must charge a premium to cover the cost of claims payments and administrative expenses, government-provided MPCI relies on the taxpayer to cover well over half the cost of the program.

The current ratemaking procedure employed by the Federal Crop Insurance Corporation applies a pooled rate to the ten year farm-level average yield to determine farm-level premiums. Pooling results in adverse selection, as low-risk producers will to pay too much, and high-risk producers will pay too little. Further, the ten year average yield is subject to a high level of stochastic volatility; setting premiums proportional to the ten-year average thus results in a high level of intertemporal variance in premiums.

According to insurance theory, with actuarially fair insurance a risk-averse producer would prefer participating to foregoing coverage, risk-neutral producers would be indifferent between participating and not participating, and risk-loving individuals would prefer exposure to the full range of possible yield outcomes to the smoothing effect of insurance on income. Hence with unsubsidized but actuarially fair premiums, theory predicts full participation among risk-averse producers whose expected utility would increase with insurance, while risk-loving producers could forego coverage. Low participation rates with subsidized premiums suggest either that many farmers are risk-lovers or the premium rates are not perceived to be calculated correctly.

This chapter focuses on the latter of these two issues. We combine data sets on Kansas winter wheat yields—annual county-level yields over the period from 1947 through 2000, and farm-level sample moments, based on ten years of reported APH yield. We develop an information theoretic learning rule to combine statewide, county, and farm data to estimate individualized farm-level distributions for crop yields. Maximum entropy is used to estimate farm-level yield densities

[1]T–yields, Plugs.

from these moments. Actuarially fair premiums are computed by using numerical quadrature to integrate the claim function over the farm-level yield distribution.

The spatial and temporal dependence in crop yield data represent information which should be reflected in premium calculations. Our approach is designed to set rates which explicitly reflect the dependency structure of the data. We anticipate premiums that are more temporally stable and which better reflect farm-level risk.

A Learning Rule for Updating the Distribution of Crop Fields over Space and Time

Introduction

The U.S. government operates the multiple-peril crop insurance (MPCI) program to provide farmers comprehensive protection against the risk of weather-related causes of income loss and certain other unavoidable perils. Insurance payments under MPCI are a function of individual farm-level realized crop yield. As a result, the cost of MPCI depends on the distributions of farm-level yield for insured farms. A widely recognized feature of crop yields is high levels of both spatial and temporal dependence. The current ratemaking procedure makes little use of this.

Until recently, the federal crop insurance program had been plagued with low participation rates. Taxpayer-funded subsidies are routinely employed to encourage higher participation levels.[2] Current provisions allow premium subsidies of 64 percent at the 75 percent coverage level to 100 percent at the 50 percent coverage level. Additional subsidies are employed to provide incentives for private insurers to market MPCI. These subsidies include payments for delivery expenses (marketing and service cost reimbursements), and reinsurance protection provided by the FCIC. Unlike private insurance, where the private insurance company charges premiums to finance the costs of indemnity payments and administrative expenses, federally-subsidized MPCI relies on the taxpayer to cover more than half the costs of this form of insurance.

The existence of large premium subsidies to foster participation in MPCI is an economic puzzle. Theory implies that, with actuarially fair insurance[3] , a risk-averse producer would prefer

[2]For example, in 1996, only 75 percent of eligible acres were insured, despite an effective subsidy rate of 60 percent. Government outlays for 1996 were a record $1.76 billion, more than triple their level in the mid 1980s.

[3]By actuarially fair, we mean $P = E[I]$, where P is the premium paid by an individual, and I is a stochastic indemnity which depends on a producer's yield by a formula stated in the insurance contract. A producer's expected gain from purchasing the insurance is zero if the premium is actuarially fair.

participating to foregoing coverage, risk-neutral producers would be indifferent between participating and not participating, and risk-loving individuals would prefer gambling over the full range of possible yield outcomes to the smoothing effect of insurance on income. With unsubsidized actuarially fair premiums, we expect full participation among risk-averse producers who capture increases in expected utility by insuring. Low participation rates without subsidies suggest either that most farmers are risk-lovers or that the insurance is perceived to be too costly for producers who would otherwise insure at actuarially fair rates.

Many agricultural economists argue that a key factor causing low participation is a failure of premium rates to accurately reflect individual producer risk. There are several reasons that may be so, aside from the *prima facie* evidence of low participation levels:

1. The formula used to set rates is designed and computed by a single actuarial firm under government contract. All providers of insurance charge the same centrally-determined rates, eliminating any competitive incentive to improve ratemaking efficiency through more accurate rate calculations.
2. The rates are computed by first pooling producers into discrete risk classes delineated by county, crop, practice, and elected coverage levels. Aggregate loss cost ratios[4] (LCRs) are then computed within each risk class using the years 1975 to the present. The result is an estimate of the average claim as a percentage of the maximum potential premium within each risk class. The pooled average LCR is multiplied by an adjustment factor which reflects production practice, coverage level, and yield differential to arrive at a premium rate for each particular risk class. There is no *a priori* reason to assume this *ad hoc*

[4]The loss cost ratio is defined as the total of indemnities paid divided by total of liabilities for a pool of insured producers. Liability for an individual producer may be simply described as the indemnity which would be paid in the event of a total loss, taking into account all factors which determine coverage level.

classification scheme produces risk pools with the average LCR as a good estimate of expected claims for any given producer. Within any particular risk class, some farms will expect to profit by insuring and others will expect to lose money.

3. Although historical county-level yield data shows strong spatial correlation in contemporaneous yields, the current ratemaking methodology makes limited use of this property.
4. The premium charged an individual producer is proportional to a ten-year average of farm-level yields. Due to the high variance of farm-level yields[5], a ten-year average is an imprecise measure of expected farm-level yield and does not reliably measure the prospective field risk. For example, assuming a trend-reverting stochastic process for farm-level yield, producers with relatively poor recent production experience will tend to be charged too high a premium, as their true expected yield will be higher than the ten-year average. Conversely, producers with favorable recent production experience will have a lower expected yield than predicted from the ten-year average, resulting in underpriced coverage.

We develop methods to obtain more reliable estimates of farm-level crop yield distributions by incorporating information from regional and county-level yield data into the farm-level estimates. These methods are based on information theory and optimal learning rules (Zellner 2000) and minimum expected loss combinations of two estimators (Judge and Mittlehammer 2003). This new approach offers a number of potential advantages over the existing method:

- It makes efficient use of the spatial and time dependencies in the data;

[5]For illustration, the sample coefficient of variation of farm-level yield for 20,720 Kansas winter wheat farms over the period from 1991-2000 generally fell in a range from 20 to 80 percent, with an average of 39.45 percent.

- The method of combining information from the regional and county levels and from farm levels of data corrects for aggregation bias estimates based solely on county yield data;
- The density estimates are robust to departures from normality;
- Sensitivity of premiums to volatility in the ten-year average farm-level mean yield is reduced or eliminated;
- The learning rule offers an efficient mechanism for updating our estimates as new information becomes available.

Related Work

The provisions and operation of the U.S. federal multiple-peril crop insurance program are discussed in a number of sources. The survey by Knight and Coble (Knight and Coble 1997) provides a broad overview of the relevant literature. Wright and Hewitt (Hueth and Furtan, eds. 1994) compare the operation of federal crop insurance in various countries which have implemented programs, and suggest reasons why insurance schemes fail to thrive without the support of large taxpayer-funded subsidies. Harwood *et al*. (Harwood, Heifner, Coble, Perry, and Somwaru 1999) discuss agricultural risk management in general, including a section which explains the provisions of the U.S. federal multiperil crop insurance program. Details of the procedures for determining premium rates are provided in Josephson, Lord and Mitchell (Josephson, Lord, and Mitchell 2000).

An efficient method for estimating seemingly unrelated regression equations (SUR) was originally introduced by Zellner (1962) and is described at length in Davidson and MacKinnon (Davidson and MacKinnon 1993) and Greene (Greene 2000). SUR is applicable to estimating a linear regression model such as ours with panel data that is subject to contemporaneous correlation across cross-sectional observation units. In the present case, the data are Kansas county winter wheat yields over the years 1947–2000. The data exhibit a high degree of contemporaneous spatial correlation. This spatial correlation is estimated and used to implement feasible generalized least squares within the SUR framework.

The approach we use to combine information from the county and farm-level yield data farm-level moment data is based on an application of Bayes' rule to the normal distribution. This approach is explained in Gelman *et al.* (Gelman, Carlin, Stern, and Rubin 1995). Gelman *et al*. make an analogy between classical analysis of variance (ANOVA) and empirical Bayes estimation, characterizing the estimates produced by the latter as a compromise between the null hypothesis that all units share a common mean and the alternative hypothesis that the mean differs across units. The Bayesian approach produces a convex combination of the two estimates that balances the relative precisions of the estimates made under each of the two competing hypotheses.

The information theoretic method that we employ to update our yield distributions from one stage in the analysis to the next is based on the principle of maximum entropy. This theory is described in Jaynes (Jaynes 1982), Zellner and Highfield (Zellner and Highfield 1988), Ormoneit and White (Ormoneit and White 1999), and Tagliani (Tagliani 1993). Jaynes dispels the common misconception that the maximum entropy method is equivalent to maximum likelihood estimation; in the latter case, the choice of statistical model is *ad hoc*, while the maximum entropy criterion provides a rationale for both selecting and estimating the exponential polynomial form[6]. Zellner and Highfield apply the maximum entropy method to the problem of estimating a probability density function on the real line subject to a set of moment constraints—referred to elsewhere in the literature as the Hamburger moment problem (Tagliani 1984). Ormoneit and White provide a

[6]A number of authors including Zellner have demonstrated that the solution to the maximum entropy problem assumes the form of an exponential polynomial function, $f\left(x\right)=\exp\left(-\sum_{i=0}^{n}\lambda_i x^i\right)$.

numerical algorithm for computing maximum entropy densities on the real line and provide a table of numerical results. Tagliani wrote a series of papers in which he considered both the case investigated by Zellner and Highfield as well as the Stieltjes moment problem, where the support of the density is restricted to the non-negative half line. Because crop yields are by nature nonnegative, the Stieltjes case applies to the problem of estimating crop yield densities.

A number of researchers have focused on the role of ratemaking inaccuracies in fostering low participation in the crop insurance program (e.g., Just, Calvin, and Quiggin 1999). The general theme is that a key factor in determining whether a producer will participate is whether the expected return to insuring is positive. If insurance rates are inaccurate, some farms will face an expected gain to insuring, while others will face an expected loss. The adverse selection problem arises when only farms which expect a positive return to insuring choose to participate. The results are a pattern of persistent aggregate losses to the program, and the need to employ subsidies in order to foster participation among producers for whom the expected return to insuring is negative.

One factor leading to ratemaking inaccuracy is the pooling of producers used in the current ratemaking methodology. This pooling procedure is *ad hoc* and creates groups of individuals with heterogeneous risk profiles. Skees and Reed (Skees and Reed 1986) suggest farm-level ratemaking as a means of avoiding the problems due to risk pooling. They point out that the coefficient of variation is a key statistic in measuring farm-level risk, and that the pooling procedure currently employed likely results in grouping producers with differing coefficients of variation. They also conjecture that farms with a higher level of expected yield tend to have a lower coefficient of variation in yields, suggesting that they should be charged lower rates. The data and methodology of this chapter are conducive to addressing all of these kinds of questions, including whether crop yields exhibit negative skewness, whether farms with a higher mean yield tend to have a lower coefficient of variation, and whether a higher coefficient of variation is tantamount to a higher expected insurance claim.

The Modeling Approach

One farm-level ratemaking approach is based on the first four moments of the individual farm-level yield distributions. We first decompose farm-level yield into the sum of the county-level yield and a farm-level residual. This decomposition facilitates separate estimation of the density parameters for the county-level yield and farm-level residual distributions, which may be combined to provide estimates of farm-level yield distribution. This approach offers several advantages:

1. Incorporating the information from the relatively long series of county-level yield data should result in a more reliable estimate than the current method based on at most only ten years worth of farm-level data.
2. Using the third and fourth moments in density estimation incorporates the widely accepted properties of negative skewness and excess kurtosis in the yield distributions, relaxing the assumption of normality frequently made in econometric estimation of crop yield distributions (Just and Weninger).

The first step models the stochastic dependence of farm-level yield on the county-level and farm-level data generating processes. Let Y_{jt}^{f} denote the period-*t* yield on farm *j* within county $i\left(j\right)$. The farm-level yield is modeled by

$$Y_{jt}^{f} = Y_{i(j),t} + \delta_{i(j),j}^{(1)} + \varsigma_{j}\eta_{jt}^{f} \tag{1}$$

County-level yield in county *I*, in turn, is modeled by

$$\mu_{jt}^{(1)} = EY_{jt}^{f} \tag{2}$$

where

$\omega_{it}^{(1)}$ = county-level yield trend at time *t*,

$\delta_{i(j),j}^{(1)}$ = the difference between farm-level and county-level trend (assumed constant),

η_{it} is a standardized county-level yield shock[7],

η_{jt}^{f} is a standardized, idiosyncratic farm-level yield shock,

$\sigma_i = \sqrt{\omega_i^{(2)}}$ is the standard deviation of the county-level yield shock, where $\omega_i^{(2)}$ is the county-level yield variance,

The county-level yield coefficient of skewness is $\omega_i^{(3)} = E\eta_{it}^3$,

The county-level yield coefficient of kurtosis is $\omega_i^{(4)} = E\eta_{it}^4$,

The farm-level residual standard deviation is $\varsigma_j = \sqrt{\delta_j^{(2)}}$, where $\delta_j^{(2)}$ is the farm-level residual variance,

The farm-level residual coefficient of skewness is $\delta_j^{(3)}$,

The farm-level residual coefficient of kurtosis is $\delta_j^{(4)}$,

$Cov\left(\eta_{it}, \eta_{jt}^{f}\right) = 0$, for all *i* and *j*, for all *t*,

The moments[8] of the farm-level yield distribution for farm *j* are given by

$$\mu_{jt}^{(1)} = EY_{jt}^{f} \tag{3}$$

$$\mu_j^{(2)} = Var\left(Y_{jt}^{f}\right) \tag{4}$$

$$\mu_j^{(3)} = Sk\left(Y_{jt}^{f}\right) \tag{5}$$

---

[7] $E\eta_{it} = 0$ and $Var(\eta_{it}) = 1$.

[8] We apply the sobriquet *moment* loosely here, in the sense that the strict definitions of the (noncentral) moments are $E\left(Y_{jt}^{f}\right)^k, k = 1,2,3,4$. The characterizations by noncentral moments or by the mean, variance, skewness and kurtosis are equivalent, due to the general existence of a bijective mapping between the two sets of moments.

$$\mu_j^{(4)} = Ku\left(Y_{jt}^f\right) \tag{6}$$

All higher-order moments $\mu_j^{(k)}$, for $k = 2, 3, 4$, are assumed to be time-invariant.

Given these assumptions about the county- and farm-level yield, it can be shown that the moments of farm-level yield are given by

$$\mu_{jt}^{(1)} = \omega_{i(j),t}^{(1)} + \delta_{i(j),j}^{(1)} \tag{7}$$

$$\mu_{jt}^{(2)} = \omega_{i(j)}^{(2)} + \delta_j^{(2)} \tag{8}$$

$$\mu_{jt}^{(3)} = \frac{\sigma_i^3 \omega_{i(j)}^{(3)} + \varsigma_j^3 \delta_j^{(3)}}{\left(\sigma_{i(j)}^2 + \varsigma_j^2\right)^{3/2}} \tag{9}$$

$$\mu_{jt}^{(4)} = \frac{\sigma_{i(j)}^4 \omega_{i(j)}^{(4)} + 6\sigma_{i(j)}^2 \varsigma_j^4 \delta_j^{(4)}}{\left(\sigma_{i(j)}^2 + \varsigma_j^2\right)^2}. \tag{10}$$

These formulas provide the basis for combining estimates of the moments of the county-level yield distribution with the moments of the farm-level residual distribution. Estimates of the moments of the individual farm-level yield distribution are then based on these combined moment estimates.

Seemingly unrelated regression (SUR) was used to estimate county-level yield trends using National Agricultural Statistics Service (NASS) county-level yield data for Kansas winter wheat over the period from 1947-2000 (T = 54 time periods and M = 105 counties). The SUR method used an iterative two-stage feasible generalized least squares procedure, where the first stage used SUR to estimate a simple linear trend model[9] of the county-level yield trend:

$$Y_{it} = \alpha_i + \beta_i t + \varepsilon_{it}. \tag{11}$$

[9] Higher-order polynomial trends were also considered, but the improvement in fit was negligible, as measured by the Bayes Information Criterion (BIC). On grounds of parsimony, the linear trend model was selected.

The linear time term captures the predictable pattern of technological growth, while the error term includes weather-related productivity shocks as well as the stochastic component of technological change. The trend specification implicitly assumes the absence of a stochastic unit root in yields; the effects of past stochastic shocks are assumed to be transient. The first estimation step is to use ordinary least squares to estimate each county field trend:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_M \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix},$$

where $X_i \equiv \begin{bmatrix} 1_T & t \end{bmatrix}$,

$$1_T \equiv \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, t \equiv \begin{bmatrix} 1 \\ 2 \\ \vdots \\ T \end{bmatrix}, B_i \equiv \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}, and\ \varepsilon_i \equiv \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}.$$

The second stage estimates the covariance structure across counties using an exponential quadratic function of the distance between centroids of individual counties for the specification of the covariance between counties. Let $\hat{\varepsilon}_{it}$ be the estimated residual for county *i* in period *t*, let

$$\hat{\sigma}_{ij} = \frac{1}{T}\sum_{y=1}^{T} \hat{\varepsilon}_{it}\hat{\varepsilon}_{jt} \tag{12}$$

be the estimated sample covariance and let

$$\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}} \tag{13}$$

be the estimated sample correlation coefficient between counties *i* and *j* in period *t* for *i*, *j* = 1, 2, . . . , *M* and *t* = 1, 2, . . ., *T*. The Singh-Nagar procedure is used to estimate the exponential quadratic function:

$$\hat{\rho}_{ij} = \exp\left\{\delta_1 d_{ij} + d_{ij}^2\right\} \tag{14}$$

where d_{ij} represents the distance between the centroids of counties i and j. A weighted second-stage regression is used to correct for heteroskedasticity in the residuals in the correlation equation. Due to the large sample size (5670 observations) and the consistency of the first-stage procedure in the presence of heteroskedasticity, this second-stage correction had little effect.

The predicted correlation function is assumed to provide a reasonable approximation to the contemporaneous covariance structure for the linear trend regression residuals. We assume that the correlation structure across counties is stationary over time: $E\left[\varepsilon_i \varepsilon_j'\right] = \sigma_{ij} I$, so that the covariance matrix took the following form:

$$V = \Sigma \otimes I$$

where I is a $T \times T$ identity matrix and σ_{ij} is a typical element in the contemporaneous correlation matrix Σ which describes the correlation structure across counties.

Letting $\hat{\hat{\rho}}_{ij}$ denote the exponential quadratic fit of the correlation between counties i and j,

$$\hat{R} = \begin{bmatrix} 1 & \hat{\hat{\rho}}_{12} & \cdots & \hat{\hat{\rho}}_{1M} \\ \hat{\hat{\rho}}_{21} & 1 & \cdots & \hat{\hat{\rho}}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\hat{\rho}}_{M1} & \hat{\hat{\rho}}_{M2} & \cdots & 1 \end{bmatrix},$$

and

$$\hat{\hat{S}} = \begin{bmatrix} \hat{\sigma}_{11} & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\sigma}_{MM} \end{bmatrix}$$

then

$$\hat{\Sigma} = \hat{S}\hat{R}\hat{S} \tag{15}$$

and

$$\hat{V} = \hat{\Sigma} \otimes I \tag{16}$$

is the estimated variance-covariance matrix in our FGLS estimates of the country-level field equations.

The estimated residuals are defined by

$$\hat{\varepsilon} = \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_M \end{bmatrix},$$

where $\hat{\varepsilon}_i$ is the vector whose components are the estimated residuals

$$\hat{\varepsilon}_{it} = y_{it} - \left(\hat{\alpha}_i + \hat{\beta}_i t\right) \tag{17}$$

where $\hat{a}_i$ and $\hat{\beta}_i$ are the coefficient estimates from the previous iteration. The regressor matrix is defined by $X \otimes I_M$ with $X \equiv \begin{bmatrix} 1_T & t \end{bmatrix}$.

After updating $\hat{\beta}$ to reflect the estimated covariance matrix, the algorithm re-estimates the correlation structure, using the residuals from the revised trend estimates, iterating[10] until the relative percentage change in the norm of the coefficient estimates for the correlation function estimation iss less than 10^{-6}.

The predicated correlation function, $\hat{\hat{\rho}} = \exp\left\{\hat{\gamma}_1 d_{ij} + \hat{\gamma}_2 d_{ij}^2\right\}$, produced an $R^2 = 0.835$ with the following parameter estimates, standard errors, and t-statistics(insert footnote):

[10]We iterate over the nonlinear least squares estimates of the restricted correlation terms. Although the regressor matrix is the same for all equations, the correlation matrix is restricted to be an exponential function of a quadratic function of the distance between each pair of counties. There are more equations (*M*=105) than time-series observations (*T*=54), so that the unrestricted covariance matrix would be singular and therefore numerically impractical.

Parameter	Estimate	Standard Error	t-statistic
γ_1	-3.370×10^{-3}	6.088×10^{-5}	-55.4
γ_2	1.067×10^{-5}	3.718×10^{-7}	-28.7

A scatter plot of the final estimated and predicted county-level yield correlations are presented in Figure 1. This illustrates the high level of spatial correlation between contemporaneous county-level yields and that the exponential quadratic form is a reasonable approximation.

We applied the Jarque-Bera (J-B) test to the transformed residuals $L\hat{\varepsilon}$, where we have defined $L = chol\left(\left(SRS\right)^{-1}\right)$ as the Cholesky factor for the inverse covariance matrix, *SRS*, with *S* defined as the diagonal matrix of estimated standard errors and *R* defined as the estimated correlation matrix. Under the null hypothesis of normality, the J-B test statistic is asymptotically distributed as a Chi-square random variable with two degrees of freedom. The calculated value of the J-B test statistic is 767.7, while the 1 percent critical value for a $\chi^2\left(2\right)$ random variable is 9.21. Hence, the test rejects the normal distribution at all reasonable levels of confidence. We conclude that the county-level yield data is not normally distributed. It follows that farm-level densities also are not normally distributed (Feller 1971).

In principle, given our estimated county-yield trends and estimates of the farm-level residual moments, we can calculate estimates of the individual farm-level yield distributions. However, the farm-level residual moments are based on at most ten years of data, and the high variation in crop yields over time makes these estimates unreliable. To illustrate this issue, we simulated twenty-thousand simulated samples of size 10 from a standard normal distribution. The population mean, variance, skewness and kurtosis are well known to be 0, 1, 0 and 3, respectively.

Figure 2 is a scatter diagram for the simulated sample variance on the sample mean. Figure 3 is a scatter diagram for the sample kurtosis on the sample skewness.

A similar pattern of variation can be seen in the corresponding graphs of the raw farm-level residual moment estimates. Figure 4 is a scatter of the sample farm-level residual variances on the sample means, based on ten-year actual production history (APH) yields. These yield data are not normalized, so that the farm-level statistics are positive-valued and have large ranges for both the mean and variance. Figure 5 is a scatter of farm-level residual kurtosis on the skewness, again based on ten-year APH yields. The simulation results suggest that a substantial share of the variation in these moments is due to sampling error.

The remainder of this chapter focuses on our development and implementation of an empirical learning rule to reduce the uncertainty surrounding the individual farm-level crop yield distributions. The objective of this learning rule is to efficiently exploit the large data set across farms and the longer time series data set for all counties to obtain better estimates of the shape parameters for the individual farm-level yield distributions. These pooled estimates are used to sequentially update the moments for the yield distribution, first from the state level to the county level, and then from the county level to the farm level. A similar procedure is used to obtain mean and variance estimates of the distributions of the unknown yield distribution shape parameters for the average farm in each county and to update these estimates to the individual farms within each county. Figure 6 presents a diagram of the learning rule. An outline of the method is the following:

1. We begin with a diffuse prior and a normal likelihood for each of the mean, variance, skewness and kurtosis parameters of farm-level residual distributions[11]. We interpret farm-level residual distribution moments as independent random draws from the likelihood, and

[11]The normal prior may be justified by an appeal to the asymptotic distribution of our moment estimators, which is normal, and to the large data samples on which our prior estimates are based.

apply Bayes' rule sequentially in two updating steps, first to the average farm-level yield within a county, then from the county-level average to the individual farm-level residual moment estimates. The formula for updating the mean of a normally distributed random variable with new information is

$$\mu_1 = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}, \tag{18}$$

where μ_0 and τ_1^2 are the parameters of the prior distribution, y and σ^2 are the observation and its variance, and μ_1 is the posterior mode. The corresponding formula for updating the variance is

$$\tau_1^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}, \tag{19}$$

where τ_1^2 is the posterior variance. In each step, these formulas are used to update the prior information using the likelihood function to reflect the new information for the case at hand.

2. A similar procedure is used to obtain moment estimates for county-level yield distributions.
3. The county-level yield moments are combined with the farm-level residual moments to obtain estimates of the moments of farm-level yield distributions. These moment estimates provide the input for the maximum entropy estimation of the individual farm-level yield densities.

In the more detailed discussion below, w and d denote the sample estimates of county-level yield moments and farm-level residual moments, $\hat{\omega}$ and $\hat{\delta}$ denote the respective posterior updates, and $\hat{\mu}$ denotes posterior updates of the farm-level yield distribution moments. The corresponding population parameters are ω, δ, and μ, a subscript j indexes farm-number, $i(j)$ indexes the county in which farm j is located, and a superscript (k) denotes the order of the moment under consideration.

Constructing the Farm-Level Moments

The farm-level residuals are defined by the difference between the farm-level yield deviation from the ten-year average farm-level yield, and the county-level yield deviation from the ten-year average county-level yield,

$$\varepsilon_{jt} = \left(Y_{jt}^{f} - \overline{Y}_{j}^{f}\right) - \left(Y_{i(j),t} - \overline{Y}_{i(j)}\right), \tag{20}$$

where Y_{jt}^{f} is the farm-level yield on farm j in period t, $\overline{Y}_{j}^{f}$ is the ten-year farm-level APH mean yield on farm j, $Y_{i(j),t}$ is the county-level yield in county i in period t, and $\overline{Y}_{i(j)}$ is the ten-year average county-level yield for county $i(j)$.

The maintained assumption about the relationship between county-level and farm-level yield is

$$Y_{jt}^{f} = Y_{i(j),t} + \delta_{i(j),j}^{(1)} + \varsigma_{j}\eta_{jt}^{f}. \tag{21}$$

Rearranging and taking the expectation of both sides shows that

$$\delta_{i(j),j}^{(1)} = E\left(Y_{jt}^{f} - Y_{i(j),t}\right). \tag{22}$$

Under the assumption that the expected difference between an individual farm-level yield and its corresponding county-level yield trends is time-invariant, an unbiased estimate of the difference is given by

$$\delta_{i(j),j}^{1} = \overline{Y}_{j}^{f} - \overline{Y}_{i(j)} = \frac{\sum_{s=t-9}^{t} Y_{js}^{f} - Y_{i(j),s}}{10}. \tag{23}$$

The farm-level residuals are then defined to be the remaing farm-level error terms,

$$\varepsilon_{jt} = Y_{jt}^{f} - Y_{i(j),t} - \hat{\delta}_{i(j),j}^{(1)} = \left(Y_{jt}^{f} - \overline{Y}_{j}^{f}\right) - \left(Y_{i(j),t} - \overline{Y}_{i(j)}\right). \tag{24}$$

The farm-level residual moments were computed from actual farm-level data. The average difference between a farm-level yield and the county-level yield trend is

$$d_j^{(1)} = \overline{Y}_j^{f} - \overline{Y}_{i(j)}, \tag{25}$$

which is the excess of the 10-year farm-level mean yield over the corresponding county-level mean yield. Higher individual farm-level residual sample moments estimates are calculated by

$$d_j^{(k)} = \frac{\sum_{t=1}^{10} \varepsilon_{jt}^{k}}{10}, \tag{26}$$

for k = 2, 3, 4. Under the hypothesis of common residual moments across all 20,720 farms in the sample, pooled momentz for the farm-level residual moments are

$$d_p^{(k)} = \frac{\sum_{j=1}^{N} d_j^{(k)}}{N} \tag{27}$$

with the associated variance estimates

$$V_{d_p}^{(k)} = \frac{\sum_{j=1}^{N} \left(d_j^{(k)} - d_p^{(k)}\right)^2}{N-1}, \tag{28}$$

for k = 1, 2, 3, 4. These sample moments and their associated sample variances were used to construct the pooled prior distribution for each residual moment

$$\delta^{(k)} \sim N\left(d_P^{(k)}, V_{d_p}^{(k)}\right), \tag{29}$$

as the first step in our learning rule.

An alternative to the common-moment hypothesis is that the moments vary geographically by county, but farms within each county share common residual moments. This alternative would be supported by the data if the cross-county variation in moment estimates were relatively large compared to within-county variation. On the other hand, if within-county variation were relatively large compared to cross-county variation, the pooled moment hypothesis would be supported. The classical resolution of this dichotomy applies analysis of variance. If the null hypothesis is accepted, the pooled sample moments are is used, while if it is rejected, the county-specific moment estimators are used instead.

A Bayesian approach resolves the dichotomy through a compromise rather than an absolute choice between the two hypotheses. Under the alternative hypothesis, the county-specific moments and their variances may be estimated using

$$d_i^{(k)} = \frac{\sum_{j \in J_i} d_j^{(k)}}{N_i}, \tag{30}$$

and

$$V_{dj}^{(k)} = \frac{V_i^{(k)}}{N_i}, \tag{31}$$

where J_i is the set of farms in county *i*, N_i is the number of farms in county *i*, and

$$V_j^{(k)} = \frac{\sum_{j \in J_i} \left(d_j^{(k)} - d_i^{(k)} \right)^2}{N_i - 1} \tag{32}$$

is the sample variance of the corresponding farm-level sample moment for county *i*. This method computes the posterior mode as the precision-weighted average between the sample pooled moment and the county-level sample moment,

$$\hat{\delta}_i^{(k)} = \frac{\frac{1}{V_{d_j}^{(k)}} d_i^{(k)} + \frac{1}{V_{d_P}^{(k)}} d_P^{(k)}}{\frac{1}{V_{d_t}^{(k)}} + \frac{1}{V_{dp}^{(k)}}}, \tag{33}$$

with posterior variance given by

$$V_{\delta_i}^{(k)} = \frac{1}{\frac{1}{V_{d_j}^{(k)}} + \frac{1}{V_{d_p}^{(k)}}}. \tag{34}$$

It is straightforward to show[12] that the posterior variance is less than the smaller of the prior variance, $V_{d_p}^{(k)}$, and the likelihood variance, $V_{d_j}^{(k)}$. Hence the posterior distribution has a mean which is the precision-weighted average of the prior mean and the likelihood mean, and a variance which is smaller than the variances of either the prior or the likelihood.

After updating the farm-level residual moments from the statewide pooled level to the county level, the question remains whether there are significant differences in these moments at the individual farm level within each county. In principle, if the variation between farm-level moments is large relative to the sampling variation, then farm-level moment estimates should differ across the farms in each county. Conversely, if the sampling variance in the farm-level moment estimates is large relative to the inter-farm variation in moments for farms within a given county, then there is little basis for separate estimates across farms. An attractive aspect of the updating procedure we have developed and applied to this problem is that, at any stage, new information can be taken into account by treating the posterior from the previous update as the new prior and entering the new

[12]Suppose $V_i > 0$ for i = 1, 2, 3, and, without loss of generality, that $V_1 \geq V_2$. If

$$V_3 = \frac{1}{\frac{1}{V_1} + \frac{1}{V_2}},$$

then since $\frac{1}{V_1} > 0$, it follows that

$$V_3 < \frac{1}{0 + \frac{1}{V_2}} = V_2 = \min\{V_1, V_2\}$$

information through the likelihood function. We therefore update from county-level to farm-specific residual moments with the formulas:

$$\hat{\delta}_j^{(k)} = \frac{\frac{1}{V_{\delta_{i(j)}}^{(k)}}\delta_{i(j)}^{(k)} + \frac{1}{V_{d_j}^{(k)}}d_j^{(k)}}{\frac{1}{V_{\delta_{i(j)}}^{(k)}} + \frac{1}{V_{d_j}^{(k)}}} \tag{35}$$

and

$$V_{\delta_j}^{(k)} = \frac{1}{\frac{1}{V_{\delta_{i(j)}}^{(k)}} + \frac{1}{V_{d_j}^{(k)}}}. \tag{36}$$

County-level Sample Moments

Initial estimates of the moments of the county-level yield distributions are based on the trend regressions. The estimated county-specific trends provide estimates of the county-level mean yield, and exhibit a small standard error. These estimated county-specific means are not updated in our procedure. The estimated county-level yield trend is

$$\omega_i^{(1)} = \hat{\alpha}_i + \hat{\beta}_i t_0 \tag{37}$$

with corresponding variance equal to

$$V_{\omega_i}^{(1)} = Var\left(\hat{\alpha}_i + \hat{\beta}_i t_0\right) = Var\left(\hat{\alpha}\right) + 2t_0 Cov\left(\hat{\alpha}_i, \hat{\beta}_i\right) + t_0^2 Var\left(\hat{\beta}_i\right) \tag{38}$$

The parameter estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the SUR trend regression coefficient estimates, $t_0 T + 1$ is the next period after the last observation in the data, and the estimated variances and covariance in the variance formula are obtained from the county-specific elements of the SUR covariance matrix.

Higher moments for the county-level yield distribution are initially estimated using the sample moments of county-specific residuals. These consistently estimate the corresponding population moments when the trend regression is correctly specified. The hyperparameters (mode and variance) of the estimated variance are computed as

$$w_i^{(2)} = \frac{\sum_{t=1}^{T} \hat{\varepsilon}_{it}^{2}}{T-K} \tag{39}$$

and

$$V_{wi}^{(k)} = \frac{\sum_{t=1}^{T} \hat{u}_{it}^{2k} \Big/ (T-K) - \left[w_i^{(k)} \right]^2}{T} \tag{40}$$

The county-level sample skewness and kurtosis are calcuted with the normalized residuals, $\hat{u}_{it} = \hat{\varepsilon}_{it} \Big/ \sqrt{\omega_i^{(2)}}$. The formulas for the residual skewness ($k = 3$) and kurtosis ($k = 4$) are

$$w_i^{(k)} = \sum_{t=1}^{T} \hat{u}_{it}^{k} \Big/ (T-K) \tag{41}$$

and

$$V_{w_i}^{(k)} = \frac{\sum \hat{u}_{it}^{uk} \Big/ (T-K) - \left[w_i^{(k)} \right]^2}{T} . \tag{42}$$

The updating procedure for the variance, skewness and kurtosis of the county-level trend residuals is the same for k = 2, 3, 4, and is described generically for all three cases. Under the hypothesis of a common k^{th} moment across all counties, a pooled prior mean and variance are computed from the county-specific moment estimates:

$$w_p^{(k)} = \frac{\sum_{i=1}^{M} w_l^{(k)}}{M} , \tag{43}$$

and

$$V_{w_p}^{(k)} = \frac{\sum_{i=1}^{M} \left(w_i^{(k)} - w_p^{(k)} \right)^2}{M(M-1)} \tag{44}$$

These pooled priors are combined with the county-specific sample moments, again using Bayes' rule, generating posterior distributions for the higher order moments of the county-level yield distributions,

$$\hat{\omega}_i^{(k)} = \frac{\frac{1}{V_{w_i}^{(k)}} w_i^{(k)} + \frac{1}{V_{wp}^{(k)}} w_i^{(k)}}{\frac{1}{V_{w_i}^{(k)}} + \frac{1}{V_{wp}^{(k)}}}, \tag{45}$$

and

$$V_{\hat{\omega}_i}^{(k)} = \frac{1}{\frac{1}{V_{w_i}^{(k)}} + \frac{1}{V_{w_p}^{(k)}}}. \tag{46}$$

Combining the County-level and the Farm-level Moments

The means and variances of the posterior county-level yield moments and the posterior farm-level residual moments may be combined by the moment decomposition formulas described previously to obtain posterior modes and variances of the farm-level density parameters. For the mean and variance of farm-level yield, the formulas are additive

$$\hat{\mu}_j^{(k)} = \hat{\omega}_{i(j)}^{(k)} + \hat{\delta}_j^{(k)} \tag{47}$$

and

$$V_{\hat{\mu}j}^{(k)} = V_{\hat{\omega}_{i(j)}}^{(k)} + V_{\hat{\delta}_j}^{(k)} \tag{48}$$

define the updates of the mean and variance hyperparameters for the respective mean (k=1) and variance (k=2) parameters of the farm-level yield distribution.

However, for the subsequent calculation of the farm-level maximum entropy yield distributions, it is useful to characterize the farm-level yield distributions in terms of dimensionless statistics, namely, the coefficients of variation, skewness, and kurtosis. The calculation of posteriors for these three cases is somewhat more complicated, as the formulas depend on ratios of random variables. We use the delta method as it applies to a univariate function of a multivariate distribution.

The coefficient of variation is defined as the ratio of the standard deviation to the mean. The posterior mode may be found by computing

$$\hat{\gamma}_j = \left(\hat{\mu}_j^{(2)}\right)^{1/2} \Big/ \hat{\mu}_j^{(1)} . \tag{49}$$

and the posterior variance is

$$V_{\hat{\gamma}j} = \frac{\hat{\mu}_j^{(2)}}{\left[\hat{\mu}_j^{(1)}\right]^4} V_{\hat{\mu}j}^{(1)} + \frac{1}{4\left(\hat{\mu}_j^{(1)}\right)^2 \hat{\mu}_j^{(2)}} V_{\hat{\mu}}^{(2)} . \tag{50}$$

The posterior mode for the coefficient of skewness is

$$\mu_j^{(3)} = Sk\left(Y_{jt}^f\right) = \frac{\delta_{i(j)}^3 \omega_{i(j)}^{(3)} + \varsigma_j^3 \delta_j^{(3)}}{\left(\sigma_{i(j)}^2 + \varsigma_j^2\right)^{3/2}} , \tag{51}$$

and the posterior mode for the coefficient of kurtosis is

$$\hat{\mu}_j^{(4)} = Ku\left(Y_{ft}\right) = \frac{\sigma_{i(j)}^4 \omega_{i(j)}^{(4)} + 6\sigma_{i(j)}^2 \varsigma_j^4 \delta_j^{(4)}}{\left(\sigma_{i(j)}^2 + \varsigma_j^2\right)^2} . \tag{52}$$

Updating Results

Figure 6 shows that the updating procedure begins with two separate sequences of calculations. The updating sequence at the county level results in posterior moments (the mean, variance, skewness, and kurtosis) of the yield distribution in each county. A separate sequence of steps results in farm-level posterior moments of the farm-level residual deviation from county-level yield. The final stage in the process uses equations (49) – (52) to combine the posterior moments for the county-level yield distribution with the posterior moments for the farm-level residual distribution to obtain the updated farm-level yield distribution.

To provide insight to the operation of this updating procedure, we present two collections of four graphs each. The first collection of four graphs illustrates estimates of the moments of the farm-level residual distribution at each stage of the updating process. The second collection of four graphs compares estimates of the moments of the county-level yield distributions on the horizontal scale to estimates of moments of the farm-level yield distributions on the vertical scale.

In the first group of four graphs, the cross-hairs in each case represent the mean of the initial, pooled prior distribution. Each point in the plot represents an estimate of the applicable moment from an individual farm at two different stages in the process. The horizontal position of a point corresponds to the estimate after updating to the county-level; hence the vertical position of each point represents the estimate after updating to the farm-level.

Figure 7 illustrates the updating process for the farm-level residual mean. The solid horizontal and vertical line segments which intersect in the interior of the graph indicate the position of the pooled mean of -0.82 on the horizontal and vertical scales[13]. The horizontal position of each point in the scatter represents the county-level update of the farm-level residual mean; in other words, each vertical cluster of points represents a group of farms from within one particular county. The vertical position of each point represents the farm-level update of the farm-level residual mean.

The initial stage of updating from the pooled-level to the county-level results in a range of values on the horizontal scale from slightly below -4 to slightly below 5. There apparently is a

[13]The farm-level residual mean indicates the extent to which mean farm-level yield for an insured farm falls below mean county-level yield; the negative value for the pooled mean indicates the degree to which the average yield for insured farms falls short of the overall average yield.

significant difference across counties in respect to the deviation of farm-level mean for insured farms from the mean level of county yield.

The second stage of updating is indicated by the vertical spread of points about the dashed "45°" line segment that is superimposed. With the exception of one large outlier, the updated farm-level residual means lie within a very similar range to that of the county-level updates. Each vertical cluster of points about the dashed line represents the variation in farm-level estimates for insured farms in a specific county. Points which lie above the dashed line segment have a higher residual mean than average for the county in which they are located, while points below the line have a below-average residual mean for their respective county.

To summarize the results of farm-level updating of the residual mean displayed in Figure 7, we comment on the key qualitative features of the graph:

1. For the majority of farms, the county-level update was the most significant determinant of the residual mean, as indicated by the fairly tight vertical clusters of points about the dashed line segment. The indication is that geographical variation across counties is an important explanatory variable for the difference between the farm-level mean of insured farms and the mean of county-level yield.
2. For most farms, the additional variation captured in the update from county-level to farm-level estimates is negligible, resulting in a moderate degree of vertical spread within county groups. But for a handful of farms, the departure of farm-level experience from the county average is significant, producing the smattering of outliers in the vicinity of the origin which lie as far as three units above or below the county-level estimate. These outliers illustrate the ability of Bayesian updating to distinguish atypical cases.

Figure 8 shows the updating process for the farm-level residual variance, which measures the residual variance of farm-level yield about the county-level yield. Under the independence assumption, this quantity represents the portion of farm-level variance which is smoothed out, and hence missing, from the county-level yield variance. The construction of this graph is similar to that of Figure 7. The pooled variance, represented by the horizontal and vertical solid line segments which cross in the interior of the graph, assumes a value near 90, indicating that the farm-level yield variance significantly augments the measurable variation in county-level yields.

The qualitative features of Figure 8 are quite similar to those of Figure 7. In both cases, the majority of the variation in farm-level variance is accounted for by the update to the county-level mean, with a negligible additional amount of variation estimated at the farm-level. Again, for a smattering of cases, the farm-level update produced an estimate which differed significantly from the county-level update, as evidenced by the points which lie more than ten units above or below the 45° line.

Figures 9 and 10 show the respective updating processes for farm-level residual skewness and kurtosis. The qualitative features of these graphs are similar to those for the mean and variance, and hence the same descriptive comments of the updating process apply. It is interesting to note that the average farm-level skewness is below -0.2, while the average farm-level kurtosis is 3.5. Given that these averages represent consistent estimates of the pooled mean residual skewness and kurtosis, and that they were computed over 20,720 farms, significant departure from the normal distribution is suggested[14].

Figures 11 – 14 are scatter diagrams of the final updates of the county yield moments on the horizontal scale and the final updates of the farm-level moments on the vertical scale. The dashed line segment in each graph is a 45° line; points on the dashed line have identical county-level and

[14]The normal distribution features a skewness of 0 and a kurtosis of 3.

farm-level moments. The county-level moments are the result of applying the updating process to the county-yield regression results, without taking into consideration the individual farm-level data. The farm-level moments utilize the formulas for combining county-level moment estimates with farm-level residual moments to obtain the final four moments of each farm-level yield distribution.

Figure 11 compares the updated county-level yield means to the final updates of the farm-level yield means. Although both sets of yield means have similar ranges, the farm-level mean yields are on average somewhat below the county-level mean yields. This implies that insured farms tend to have mean yields which are below average compared to the county-level yield means, and supports the argument for adverse selection in the federal crop insurance program.

Figure 12 compares the updated county-level variances to the final updates for the farm-level variances. The farm-level variances exhibit a striking departure from the county-level yield variances, demonstrating a relatively large contribution of farm-level variability that is masked by aggregating yields into county-level averages. The county-level variances fall within a quite narrow range, suggesting little ability to measure differences in yield variability across counties. In contrast, the farm-level variances exhibit a much wider range of variation, reflecting the large differences in the farm-level residual variance even after our updating procedure. This provides support for the argument that there are substantial differences in yield risk across individual farms and that the farm-level yield distribution is much more uncertain than the distribution of county-level average yields.

Figure 13 compares the updated county-level skewness to the final updates for the farm-level skewness. The county-level skewness is on average below and has a narrower range than the farm-level skewness. This again reflects the greater variation in the individual farm-level yields, as well as the interesting fact that the sum of two independent random variables which both exhibit negative skewness generally tends to be less skewed[15]. Similarly, Figure 14 compares the updated county-level kurtosis to the final updates of the farm-level kurtosis. In the majority of cases, the county-level kurtosis falls in the range 3.0 – 3.2, while the farm-level kurtosis typically is between 3.0 and 3.5, although in a few cases the farm-level kurtosis exceeds 3.5 and in several where it is considerably below 3.0.

Overall, there are appreciable differences across farms in the moments that we have chosen to characterize their individual yield distributions. In addition, for most farms the hypothesis that crop yields follow a normal distribution is not supported by these data. In light of these two observations, we calculate individual farm-level yield distributions using the information theoretic method of maximum entropy. The resulting family of estimates nests the normal distribution as a special case, while allowing for departures from normality such as those that appear to characterize these crop yield data.

Maximum Entropy Estimation of Farm-level Yield Densities

Taking the final updates for the farm-level moments, the next step is to calculate dimensionless counterparts (coefficients of variation, skewness, and kurtosis) for each farm. We take these dimensionless shape parameters as the information set and apply the principle of maximum entropy to construct individualized farm-level crop yield distributions. As we explain in detail below, the dimensionless coefficients represent a set of sufficient statistics for the crop insurance premium rate when it is expressed as a percentage of the expected yield, when the distribution for crop yields is a

[15]The limiting case of this would be to add a large number of independent draws on a highly skewed random variable. By the central limit theorem, the root-n normalized average of a large number of independent draws is normally distributed with a skewness of zero, regardless of the level of skewness in the underlying distribution.

member of the exponential quartic family.[16] Because the final updates for the coefficients of skewness and kurtosis both fall on a narrow range, while the final updates for the coefficient of variation have a wide range, we consider three cases here:

1. the farm with the minimum coefficient of variation;
2. the farm with the median coefficient of variation; and
3. the farm with maximum coefficient of variation.

In each case, we use the corresponding final updates of the skewness and kurtosis for the chosen farm. The maximum entropy density parameters in each are computed by identifying the exponential quartic density that minimizesthe dual objective function, given these four moments[17].

Calculation of Farm-level Premiums for the Representative Cases

We next computed the farm-level crop insurance premiums for each of the three representative cases using the maximum entropy density. For illustration, the premiums were computed as a percentage of the coverage level, c, at 65, 75, and 85 percent. Let X denote the random variable for yield. Then an actuarially fair insurance premium may be computed using the formula

---

[16]When the information set includes the first four sample moments, the solution to the maximum entropy problem is a member of the exponential class of distributions, $f(x) = \exp\left(-\lambda_0 - \lambda_1 x - \lambda_2 x^2 - -\lambda_3 x^3 - -\lambda_4 x^4\right)$. This family includes the normal distribution, which is obtained when $\lambda_3 = \lambda_4 = 0$, and the exponential distribution, which is obtained when $\lambda_2 = \lambda_3 = \lambda_4 = 0$.

[17] Given arbitrary values for the first four sample moments as the information set, a solution to the maximum entropy problem may well not even exist. However, whenever a solution does exists, it is unique, and minimizing the dual objective function identifies this unique exponential quartic solution. One characterizing property of this distribution is that its first four moments equal those of the moments conditions taken as the information set.

$$
\begin{aligned}
P &= \frac{E\left[c\mu - X \mid X < c\mu\right]\Pr\left[X < c\mu\right]}{c\mu} \\
&= \frac{c\mu \Pr\left\{X < c\mu\right\} - \int_0^{c\mu} xf\left(x\right)dx}{c\mu} \times 100\% \\
&= \left(\int_0^{c\mu} f\left(x\right)dx - \frac{\int_0^{c\mu} xf\left(x\right)dx}{c\mu} \right) \times 100\%
\end{aligned} \tag{53}
$$

where $f\left(x\right)$ is the probability density function. Equivalently, changing variables to rescale X to a unit mean, $Y = X/\mu$ where $\mu = EX$, it is possible to write[18]

$$
P = \left(\int_0^{c} g\left(y\right)dy - \frac{\int_0^{c} yg\left(y\right)dy}{c} \right) \times 100\% \tag{54}
$$

The fact that the premium can be computed from this rescaled density implies that the coefficients of variation, skewness and kurtosis are a set of complete sufficient statistics for the actuarially fair insurance premium for a member of the exponential quartic class of distributions.

The table below displays the premium calculation results. As would be expected, the calculations are increasing in coverage level and in coefficient of variation.

[18] It is easy to see that $\partial x/\partial y = \mu$, and $y = c$ when $x = c\mu$. Making the appropriate substitutions leads to the premium formula in terms of the density of Y, $g\left(y\right) = f\left(\mu y\right)\mu dy$.

Coverage Level	Minimum CV	Median CV	Maximum CV
65%	0.8%	1.7%	2.2%
75%	1.8%	3.1%	3.7%
85%	3.6%	5.2%	5.9%

Exponential Quartic Premiums

Conclusion

In this paper, we have presented and provided examples of a new approach to computing farm-level crop yield distributions and the associated insurance premiums. This method offers a number of innovations over current practice:

1. The farm-level yield distributions systematically incorporate information from a long panel of county-levels yields. Incorporating county-level yield data into the farm-level yield distributions is facilitated by a model that decomposes the farm-level yield into the sum of the county-level yield and a farm-level residual. By definition, the county-level yield in any given year is computed as the total production of the crop in the county divided by total harvested acres in the county. The individual farm-level residual is modeled as an independent deviation from this average. This lead directly to formulas that express the moments of the farm-level yield distribution in terms of the corresponding moments of the county-level yield distribution and the distribution of the farm-level residual. County-level yields comprise a large share of farm-level yields. There also are several more available observations at the county-level than at the individual farm-level. Including the county-level

information therefore produces substantially more reliable premium calculations than those that are based on (at most) ten years of farm-level APH yields.

2. County-level yields exhibit a high degree of contemporaneous spatial correlation. The SUR estimates of the county-level yield trends and the corresponding calculations of county-level yield distributions both exploit this spatial correlation in an efficient manner. When the weather in one county is bad in a given year, there is a better than even chance that the surrounding counties also will experience similarly adverse conditions. To the extent that weather affects harvests, these local weather effects will be manifested in spatial correlations between crop yields in nearby counties. The decomposition of county-level yields into a mean trend and a residual exploits the spatial correlation at the county level. In principle, the resulting GLS estimates will provide more efficient estimates of the yield trends and residual decomposition than those obtained when spatial correlation is ignored[19].
3. The Bayesian updating approach used to obtain the county-level and farm-level moments systematically updates the farm-level yield distributions with information from county-level and the other farms in the available data set. Insurance premiums calculated using these methods will reflect the extent to which individual farms depart from the characteristics of surrounding farms. Current methods obscure these local differences, thereby increasing the risk of farm-level moral hazard.
4. The maximum entropy approach to density estimation is robust to significant departures from the normal distribution. Our results suggest the presence of significant excess kurtosis in contrast to the normal distribution. The maximum entropy principle offers one way to capture the impact of such a departure on the underlying probability distribution.

[19]Current APH-based ratemaking practice makes limited use of spatial correlation in computing premiums.

5. The premium calculations advocated here are derived from first principles. We first estimate a farm-level yield distribution. This theoretically determines the probability distribution of crop insurance claims. We then compute actuarially fair premiums as the expected indemnity payment that corresponds to this yield distribution. Current practice is an *ad hoc* procedure that pools and averages loss-cost ratios, and then multiplies this figure by a ten-year APH yield to determine an insurance premium. As we have seen, the ten-year APH yield averages are extremely volatile measures of farm-level crop yields.

Graphical Comparison of Maxent and APH Premiums

The advantage of our approach is illustrated graphically in Figure 15, which displays kernel density estimates for the premium distributions for all wheat in Morton County, Kansas. Three different premium distribution are superimposed in the figure:

1. the distribution of actuarially fair premiums, based on the maximum entropy distributions;
2. the distribution of APH premiums, based on the reported APH mean yields; and
3. the distribution of APH premiums, based on the trend-adjusted APH mean yields.

The actuarially fair premiums were computed by applying quadrature to the maximum entropy distributions, treating the maximum entropy distributions for farm-level yields as the true yield distributions. The distribution of the (unadjusted) APH insurance premium was estimated by first computing an estimate of the average loss cost ratio which enters the APH premium calculation. For each farm in the Morton County sample, a loss ratio equal to the expected loss as a percent of APH mean was computed by

$$L\left(\overline{y}_i\right) = \frac{E\left[c\overline{y}_i - X \mid X < c\overline{y}_i\right] \Pr\left\{X < c\overline{y}_i\right\}}{\overline{y}_i}; \qquad (55)$$

that is, by computing the expected premium as a percentage of mean yield, based on the ten-year APH mean $\bar{y}_i$. These farm-level expected loss ratios were averaged to obtain a proxy for the average LCR. This average LCR was multiplied by the farm-level APH means to estimate the distribution of APH-based premiums.

A similar procedure was used to compute adjusted APH-based premiums, except that the APH means were adjusted for trend growth. Because the APH means are computed in the current FCIC ratemaking process without a trend adjustment, a ten-year average of APH yields actually estimates the expected yield net of 4.5 years of trend growth. To see this, note that

$$Y^f_{j,t-k} = \alpha_{i(j)} + \beta_{i(j)}\left(t-k\right) + \delta^{(1)}_{i(j),j} + \varepsilon_{j,t-k} + \varepsilon^f_{j,t-k} \tag{56}$$

where $\varepsilon_{i(j),t-k}$ is the county-level yield shock, $\varepsilon^f_{j,t-k}$ is the farm-level yield shock,

and $E\varepsilon_{i,t-k} = E\varepsilon^f_{j,t-k} = 0$. The APH mean is given by

$$\bar{Y}^f_j = \sum_{k=0}^{9} Y^f_{j,t-k}\ , \tag{57}$$

which has expected value

$$\begin{aligned} E\bar{Y}^f_j &= \alpha_{i(j)} + \beta_{i(j)}\left(t-4.5\right) + \delta^{(1)}_{i(j),j} \\ &= EY^f_{j,t} - 4.5\beta_{i(j)} \end{aligned}. \tag{58}$$

In light of this downward bias, an adjustment for 4.5 years of trend growth was added to the APH mean for each farm before computing premiums in the adjusted case.

Kernel density estimates of the distributions of these three premium calculations illustrate the differences between the three approaches. The variation in the distribution of the maximum entropy based premiums reflects differences in the farm-level moments. The range of variation in the premiums is relatively narrow, from about 3.25 percent up to 4.2 percent of mean yield. In contrast, the two APH premium distributions show a considerably larger range of variation than the corresponding premium distribution obtained with information theoretic methods. The unadjusted APH kernel estimate also has a downward bias due to the failure to reflect trend growth in the premium estimates.

We can measure the differences between the adjusted APH and information theoretic premium calculations cases with a decomposition of the mean square error that treats the maximum entropy premium as the true premium and the APH premium as an estimator. Let P represent the actual premium and $\hat{P}$ the adjusted APH premium. The mean square error decomposition is

$$MSE = E\left(\hat{P} - P\right)^2 = E\left(\hat{P} - E\hat{P}\right)^2 + \left(E\hat{P} - P\right)^2 = \mathrm{var}\left(\hat{P}\right) + \left(E\hat{P} - P\right)^2 \qquad (59)$$

The sample values of these terms for the adjusted APH premium calculations are a variance of 1.14 percentage points, and a bias of 0.17 percentage points. Though it is impossible to know the extent that these results represent actual experience, two conclusions are suggested. First, the variance of APH premiums is large compared to the magnitude of the premiums. Because yields exhibit a large measure of intertemporal variation, a large share of the variance in APH premiums represents a fluctuation in premiums purely due to sampling variation, rather than actual variations in yield risk. Second, the current approach, which uses the unadjusted 10-year average APH yield to estimate the expected current yield results in significant negative bias in premiums relative to the actuarially fair values. Finally, the bias in APH premiums can be significantly reduced by adjusting the APH mean for trend, although the variance problem is not mitigated by this adjustment.

The current practice of computing premiums based on multiplying pooled average loss cost ratios by 10-year APH average yields is subject to bias and high intertemporal variance. Both help explain the historically low rate of participation in the MPCI crop insurance program, and the use of large subsidies to induce farmer participation. The bias is due to farms with heterogeneous risk profiles being pooled together for the purpose setting the MPCI insurance rates. The result is a premium subsidy from low-risk farms to high-risk farms within the pooled group. Low-risk farms thus optimally forego participation, unless they are enticed to participate with premium subsidies. The volatility in premiums is due to the fact that the APH premium is effectively proportional to the 10-year average APH yield, and the high intertemporal variance of yields translates into a high variance in this 10-year average. The result is premiums that exhibit a high intertemporal variance. If farmers are risk averse, highly variable premiums create another disincentive for participation. The method developed in this paper is a step in the direction of crop insurance premiums that more accurately reflect individual farm-level risk, and that are more intertemporally stable than those under current approach used by the FCIC. Again, if farmers are risk averse, then actuarially fair crop insurance rates that more accurately reflect individual farmer's crop production rist and that are more intertemporally stable should result in lower in premium subsidies necessary to induce the same level of participation rates under the federal crop insurance program.

References

Berck, Peter, Jacqueline Geoghagen, and Stephen M. Stohs, "A Strong Test of the von Liebig Hypothesis," *American Journal of Agricultural Economics*, 2000, *82* (4), 948–955.

Cobb, Loren, Peter Koppstein, and Neng Hsin Chen, "Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family," *Journal of the American Statistical Association*, March 1983, *78* (381), 124–130.

Conte, Samuel Daniel, and Carl de Boor, *Elementary Numerical Analysis—An Algorithmic Approach*, third ed., McGraw-Hill Book Company, 1980.

Cover, Thomas M., and Joy A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

Davidson, Russell, and James G. MacKinnon, Estimation and Inference in Econometrics, Oxford University Press, 1993.

Feller, William, *An Introduction to Probability Theory and its Applications*, second ed., Vol. Two, John Wiley and Sons, 1971.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Daniel B. Rubin, *Bayesian Data Analysis*, CRC Press, 1995.

Greene, William H., *Econometric Analysis*, fourth ed., Prentice Hall, 2000.

Harwood, Joy, Richard Heifner, Keith H. Coble, Janet Perry, and Agapi Somwaru, "Managing Risk in Farming: Concepts, Research and Analysis," Technical Report, Economic Research Service, U.S. Department of Agriculture, March 1999.

Hueth, Darrell L. and William H. Furtan, eds., *Economics of Agricultural Crop Insurance: Theory and Evidence*, Kluwer Academic Publishers, 1994.

Jaynes, Edwin T., “On the Rationale of Maximum-Entropy Methods,” *Proceedings of the IEEE*, 1982, 70(9), 939-952.

Josephson, Gary R., Richard B.Lord, and Charles W. Mitchell, “Actuarial Documentation of Multiple Peril Crop Insurance Ratemaking Procedures,” Technical Report, Risk Management Agency, August 2000.

Judge, Geroge G., and Ron C. Mittlehammer, “A Semi-parametric Basis for Combining Estimation Problems Under Quadratic Loss,” Working Paper, Dept. of Agricultural and Resource Economics, University of California at Berkeley, 200.

Just, Richard E., and Quinn Weninger, “Are Crop Yields Normally Distributed?,” *American Journal of Agricultural Economics*, 1999, *81* (2).

Just, Richard E., Linda Calvin, and John Quiggin, “Adverse Selection in Crop Insurance: Actuarial and Asymmetric Information Incentives,” *American Journal of Agricultural Economics*, 1999, *81* (4), 834–849.

Knight, Thomas O., and Keith H. Coble, “Survey of U.S. Multiple Peril Crop Insurance Literature Since 1980,” *Review of Agricultural Economics*, 1997, *19* (1), 128–156.

Ormoneit, Dirk, and Halbert White, “An Efficient Algorithm to Compute Maximum Entropy Densities,” *Econometric Reviews*, 1999, *18* (2) 127–140.

Skees, Jerry. R., and Michael R. Reed, “Rate Making for Farm-Level Crop Insurance: Implications for Adverse Selection,” *American Journal of Agricultural Economics*, 1986, *68*, 653–659.

Stohs, Stephen M., “A Bayesian Updating Approach to Crop Insurance Ratemaking,” forthcoming, Fall 2003.

Taglaini, Aldo, “On the Application of Maximum Entropy to the Moments Problem,” *Journal of Mathematical Physics*, 1993, *34* (1), 326–337.

Zellner, Arnold, "An Efficient Method of Estimating Seemingly Unrelated Regressions, and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 1962, *57*, 500–509.

_______, "Information Processing and Bayesian Analysis," Working Paper, University of Chicago Graduate School of Business, 2000.

Zellner, Arnold, and Richard A. Highfield, "Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions," *Journal of Econometrics*, 1988, *37*, 195–209.

Figures

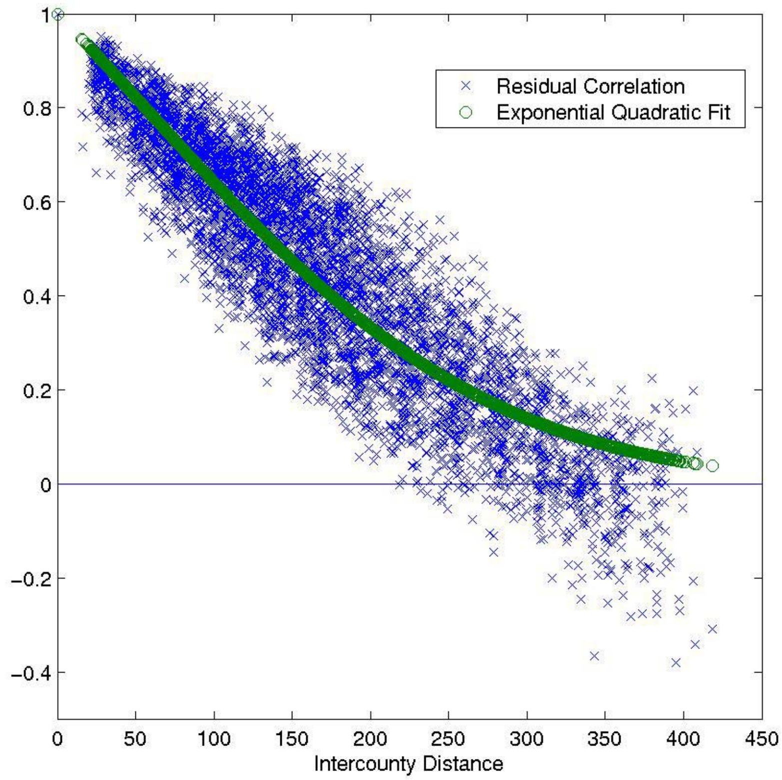

Figure 1. Estimated and Predicted County Yield Correlations

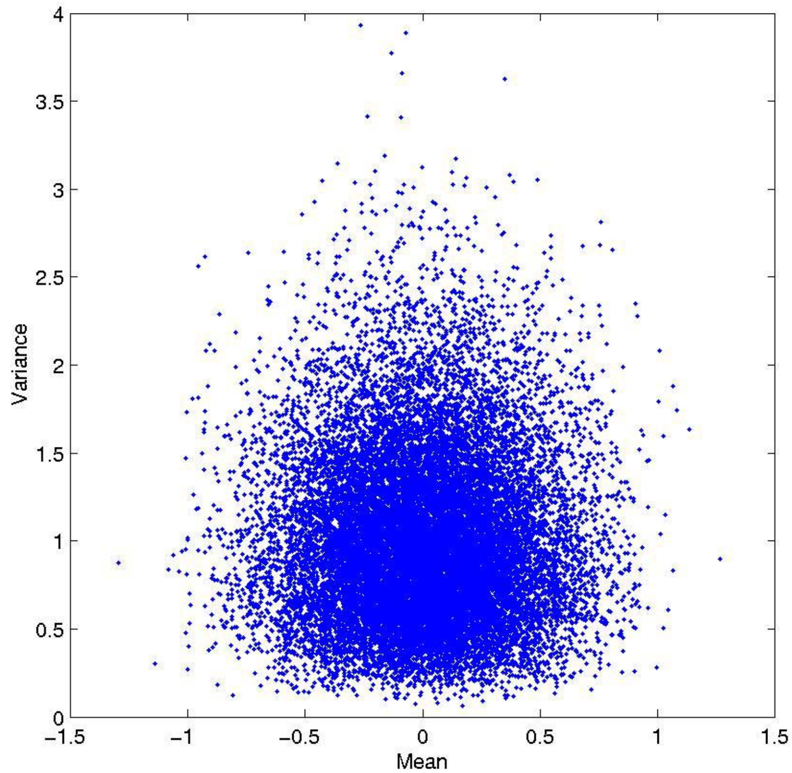

Figure 2. Mean and variance for 20,000 simulated standard normal samples

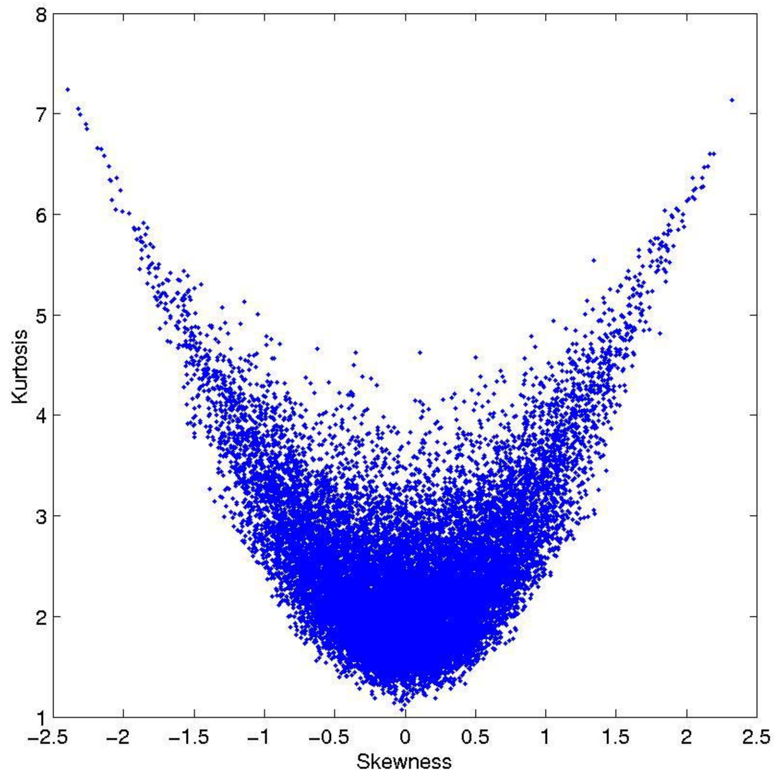

Figure 3. Skewness and kurtosis for 20,000 simulated standard normal samples

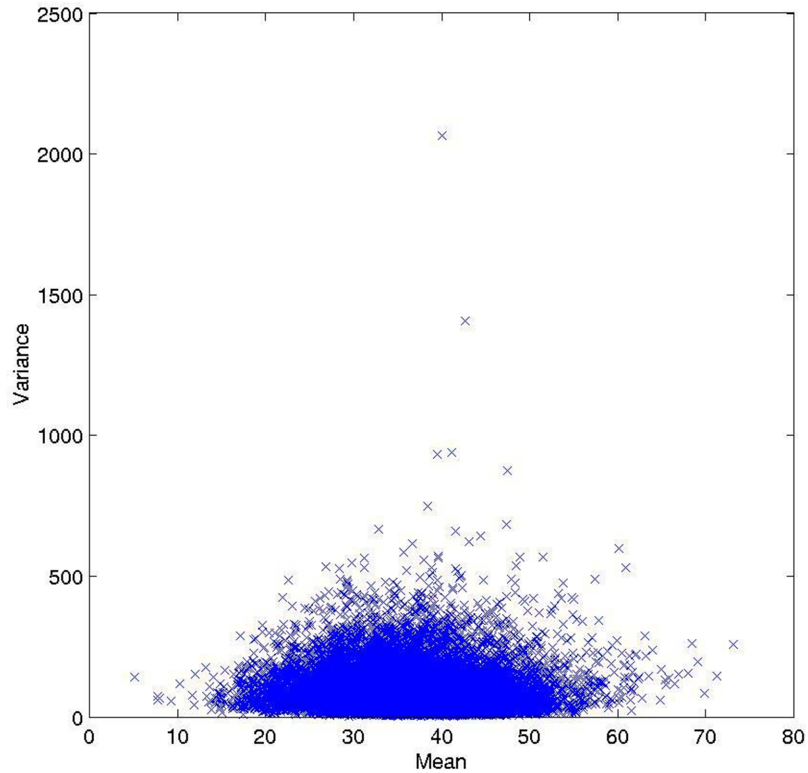

Figure 4. 10-year APH residual variance on APH mean yield

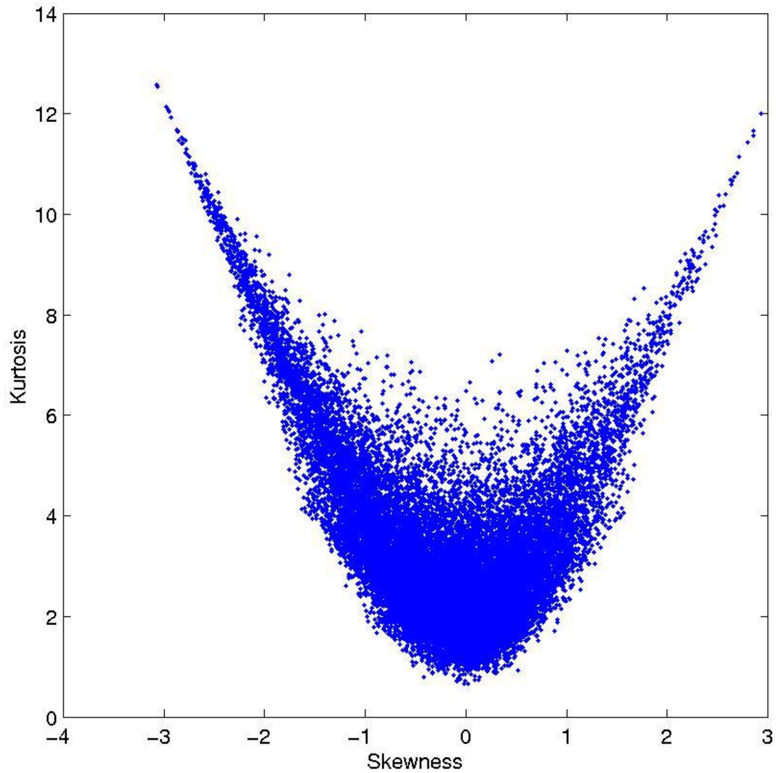

Figure 5. 10-year APH residual kurtosis on residual skewness

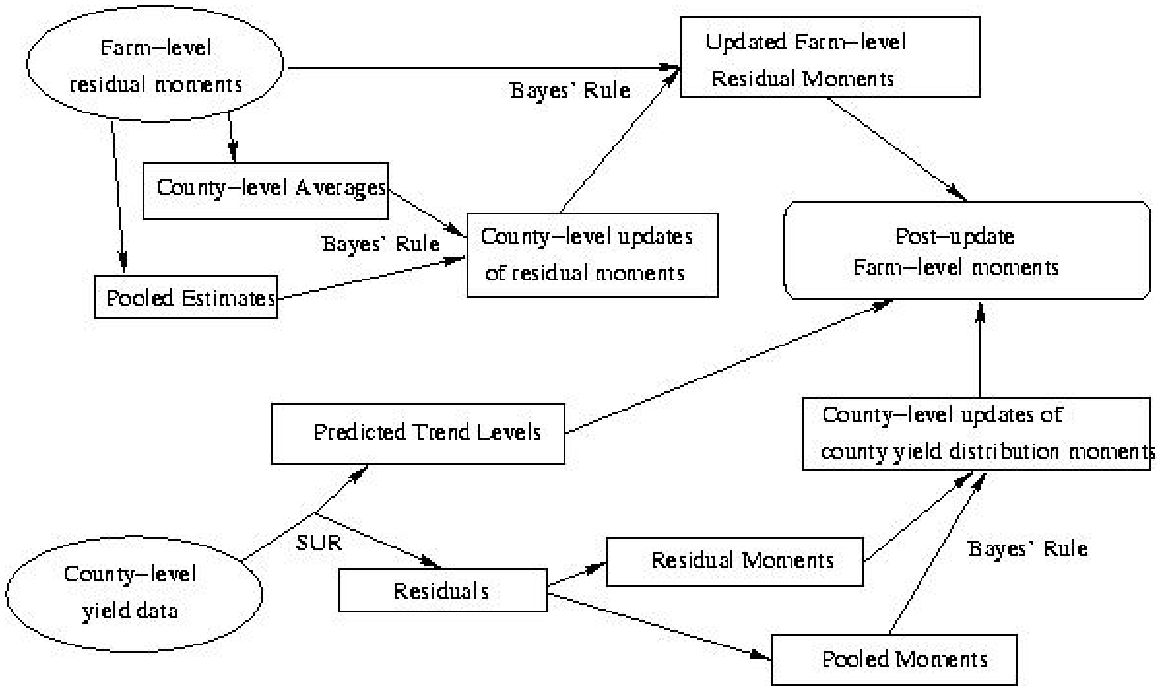

Figure 6. Schematic Diagram of Bayesian Updating Procedure

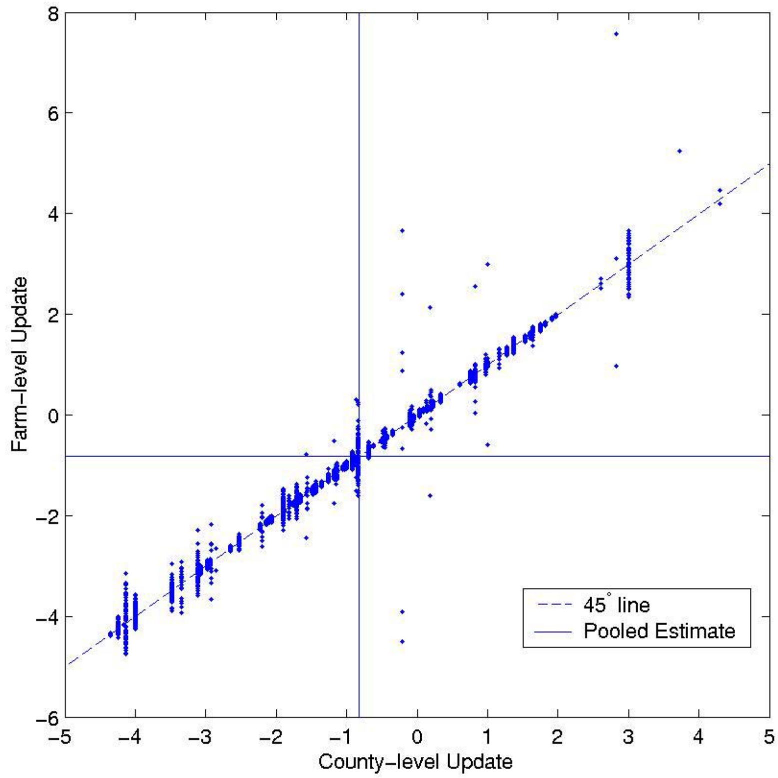

Figure 7. Update of Farm-level Residual Mean

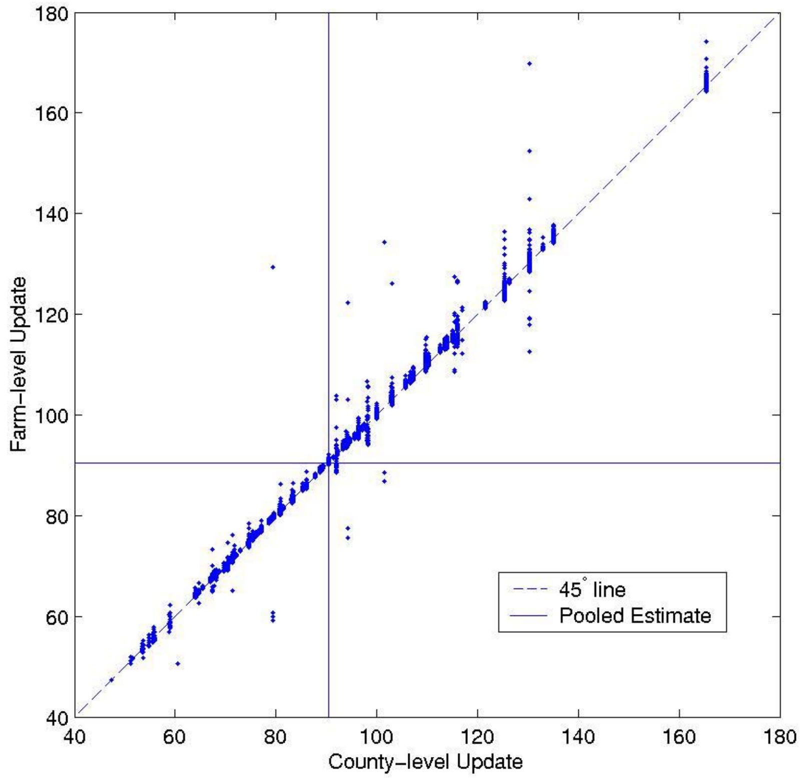

Figure 8. Update of Farm-level Residual Variance

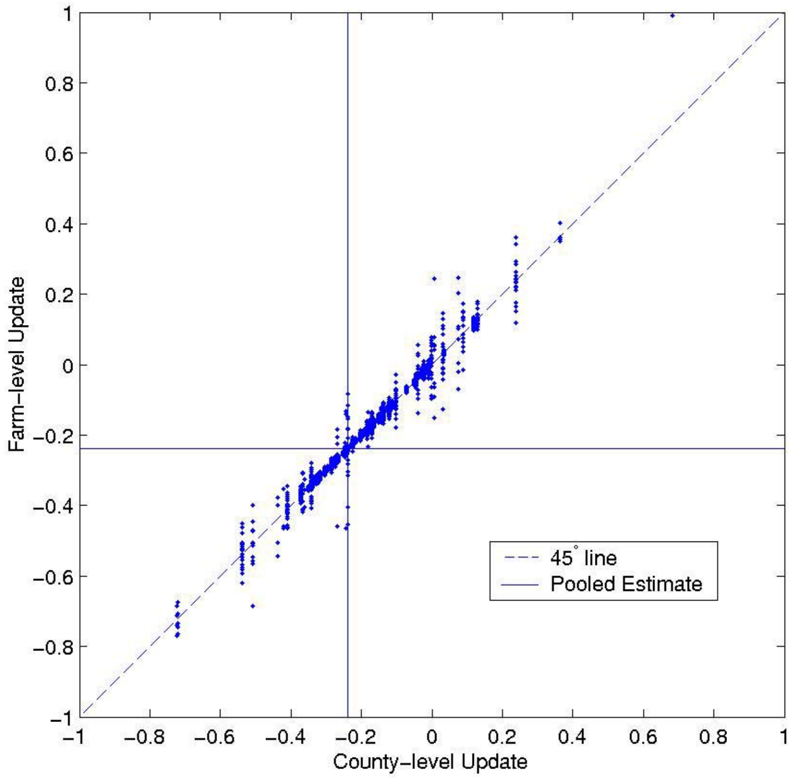

Figure 9. Update of Farm-level Residual Skewness

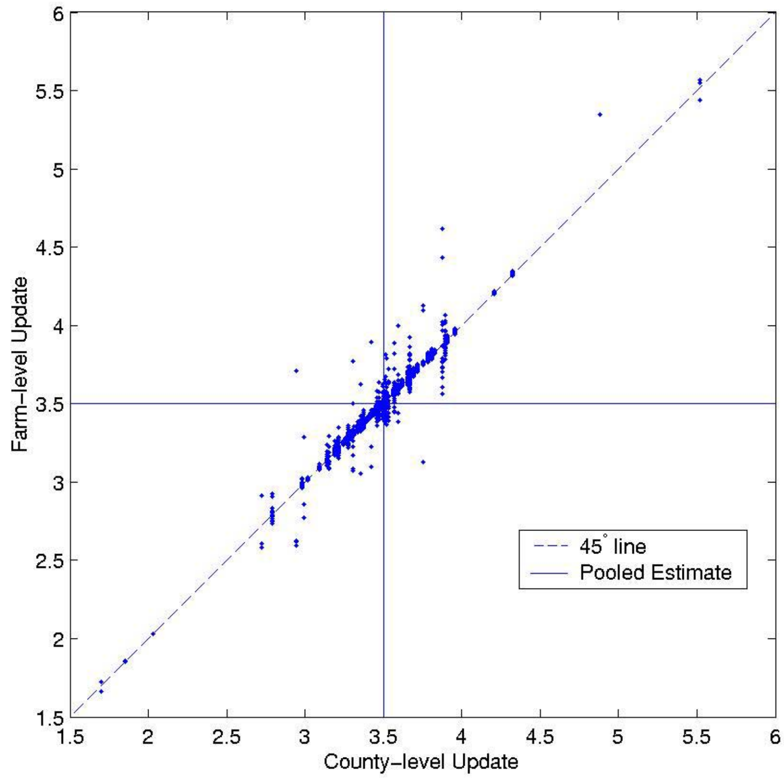

Figure 10. Update of Farm-level Residual Kurtosis

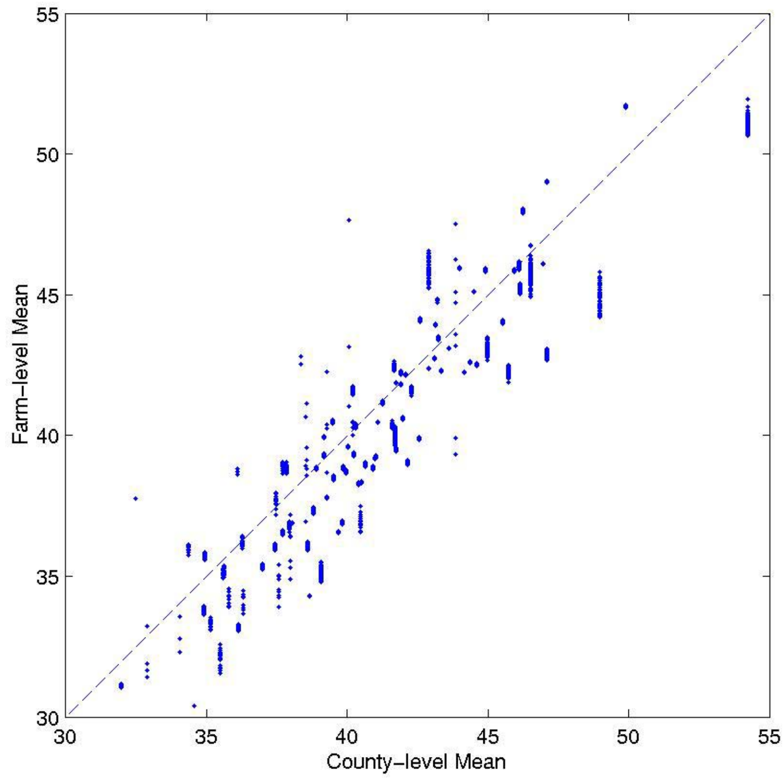

Figure 11. Complete Updates of Farm-Level Mean

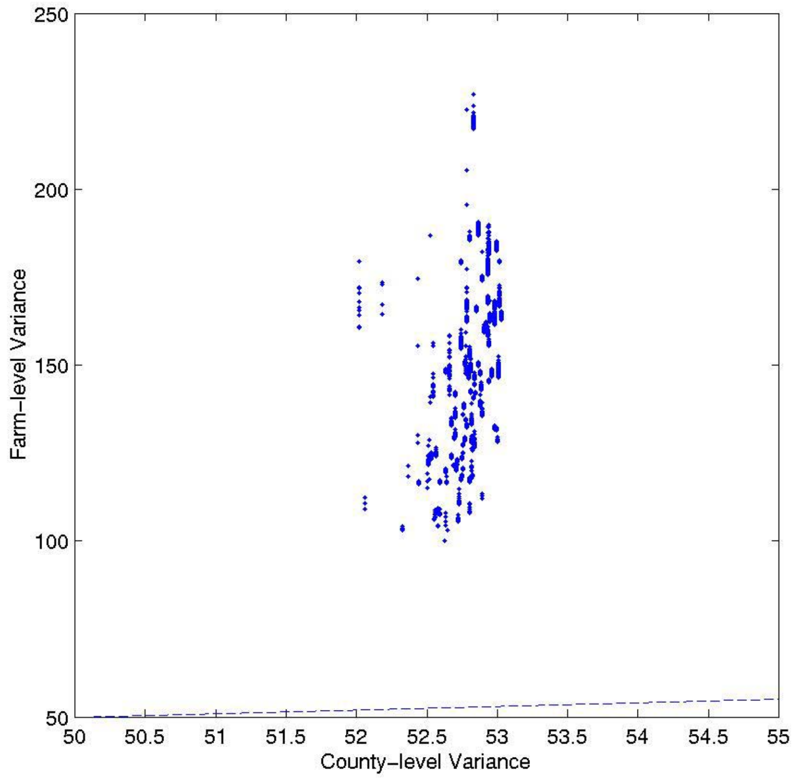

Figure 12. Complete Updates of Farm-Level Variance

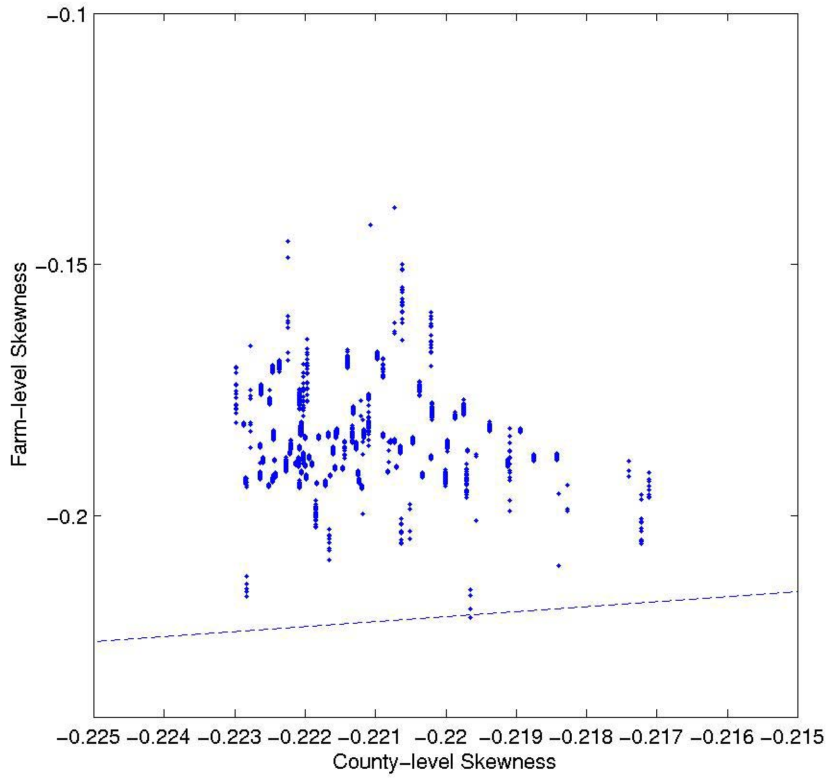

Figure 13. Complete Updates of Farm-Level Skewness

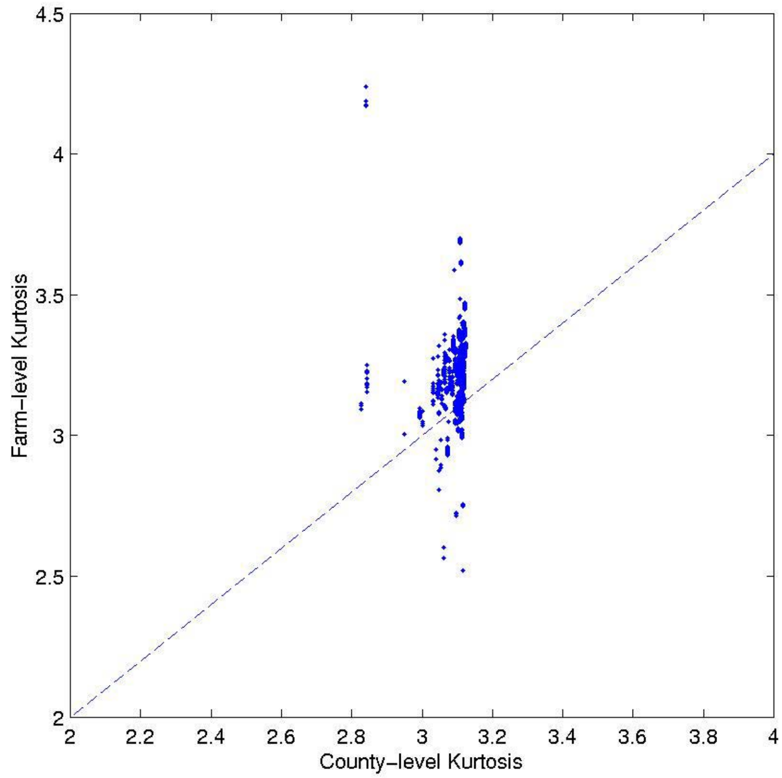

Figure 14. Complete Updates of Farm-Level Kurtosis

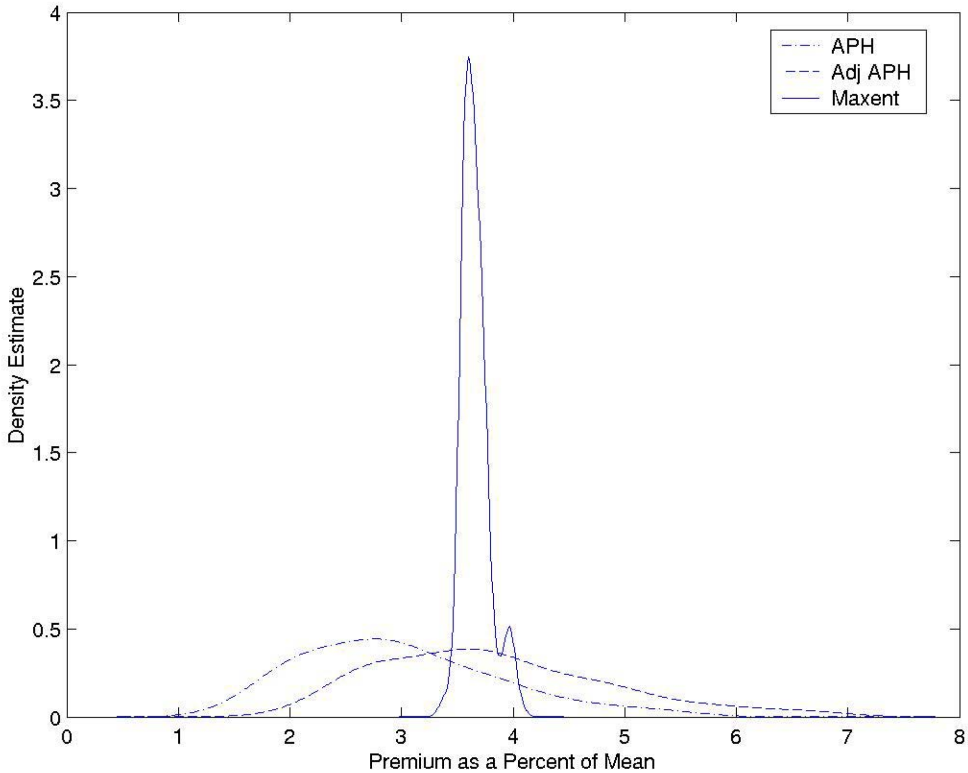

Figure 15. Kernel Density Estimates of Premium Distributions for Morton County