



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Using a Farmer's Beta for Improved Estimation of Expected Yields

Miguel A. Carriquiry, Bruce A. Babcock,
and Chad E. Hart

Effects of sampling error in estimation of farmers' mean yields for crop insurance purposes and their implications for actuarial soundness are explored using farm-level corn yield data in Iowa. Results indicate that sampling error, combined with nonlinearities in the indemnity function, leads to empirically estimated insurance rates that exceed actuarially fair values. The difference depends on the coverage level, the number of observations used, and the participation strategy followed by farmers. A new estimator for mean yields based on the decomposition of farm yields into systemic and idiosyncratic components is proposed, which could lead to improved rate-making and reduce adverse selection.

Key words: actual production history (APH), adverse selection, crop insurance, mean yields estimation, sampling error

Introduction

The U.S. crop insurance program offers farmers subsidized multiple-peril insurance against declines in farm-level yields or farm-level revenues.¹ Yield insurance is available through Actual Production History (APH) insurance. The two most utilized revenue insurance products are Revenue Assurance (RA) and Crop Revenue Coverage (CRC). The insurance guarantees and premium rates for all three products depend directly on a producer's APH yield.

The exact current rules for calculating APH yield are complicated, but in their simplest form, APH yields equal the simple average of past yields.² However, use of a simple average introduces two potential problems. First, when technology improvements increase expected yields, a simple average of past yields will tend to be lower than a farmer's expected yield. The existence of such a lag has been well recognized (Sherrick et al., 2004) and its impact on the crop insurance program has been analyzed previously (Skees and Reed, 1986; Just, Calvin, and Quiggin, 1999).

The second problem with the use of a simple average is that it can introduce excessive sampling error, especially when based on few observations. Sampling error can provide

Miguel A. Carriquiry is an assistant scientist with the Center for Agricultural and Rural Development (CARD); Bruce A. Babcock is Director of CARD and professor of economics; and Chad E. Hart is a scientist with CARD, all at Iowa State University.

Review coordinated by David K. Lambert.

¹ Single-peril insurance for farmers is available from the private sector. Crop hail insurance, fire insurance, and insurance against particular weather events, such as early freezes or inadequate rainfall, are the most important forms of single-peril crop insurance.

² For a complete description of APH yield rules, see http://www.rma.usda.gov/FTP/Publications/directives/18000/pdf/05_18010.pdf.

opportunities for adverse selection and increase premiums. Adverse selection will occur when farmers have better information about the average value of their sampling error and their yields than do insurers. Farmers who know they have positive average sampling error (past yields which, on average, were above expected yields) will know that the insurance guarantees they are offered will represent a greater percentage of their true expected yield than assumed by the insurer. Such farmers will have an increased propensity to buy crop insurance because they will tend to receive insurance payments greater than the level assumed by the insurer. Conversely, farmers who know they have negative average sampling error likely will not buy insurance because they will tend to receive insurance payments that are, on average, lower than the level assumed by the insurer. The net effect is an adversely selected insurance pool. This impact of sampling error has been noted previously by Barnett et al. (2005) and by Wang, Hanson, and Black (2003). But the magnitude of sampling error has not been previously estimated nor have its implications as a source of adverse selection and its impacts on insurance rates been examined.

Even in the absence of adverse selection, sampling error will increase premium rates when rates are based on historical insurance payouts and insurance payouts are convex in insurance coverage. A standard procedure for estimating premium rates is the loss-cost approach whereby historical losses (expressed as a fraction of the insurance guarantee) at a particular contracted coverage level are averaged over time and over policy holders. Sampling error implies actual coverage levels will differ from contracted coverage levels, although the average actual coverage level across policy holders is likely to be close to the contracted coverage level. By Jensen's inequality, because insurance payments are convex in coverage level, the average insurance payout across policy holders will be greater than the insurance payout to those policy holders who have contracted coverage levels equal to their actual coverage levels—i.e., convexity in payouts means that an increase in sampling error will increase premium rates at every coverage level.

This study makes two important contributions. The first is an improved understanding of the magnitude of potential problems caused by the amount of sampling error involved in basing crop insurance guarantees and premium rates on a simple average of past yields. The analysis abstracts somewhat from reality and is conducted on detrended yield data to allow the effects of sampling error to be isolated from the effects of yield trends, a topic that has been previously studied. The second contribution is the introduction of a new estimator that could replace a simple average of past yields in the crop insurance program. The new estimator is based on a commonly used decomposition of farm yield into systemic and idiosyncratic components. Systemic shocks affect farm yields through a "beta," which measures the sensitivity of farm yields to changes in area yields. The conditions under which the proposed estimator will reduce sampling error are identified, and a combination of bootstrapping and Monte Carlo simulations is used to estimate the magnitude of the reduction in sampling error for corn producers in Iowa. After results have been presented, the potential usefulness of the new estimator for crop insurance is discussed.

Conceptual Overview

Consider an insurance product whereby an indemnity payment is made if yields are below a yield guarantee that is the product of a producer-selected coverage level and the

producer's expected yield (μ).³ Since the expected yield of a farmer is unknown, it must be estimated through an estimator μ_a . Different estimators will in general result in different levels of sampling variability. Producers choose a coverage level (α) and a price guarantee level (p_g). It is assumed that farmers' beliefs about their yield distribution are captured by a yield distribution $F(y)$, which may have an expected yield that differs from the expected yield (μ_a) estimated by the insurer. For any expected yield assigned by the insurer, the farmer expects to receive an indemnity given by:

$$E_y(I(\mu_a)) = E_y(p_g * \max(\alpha\mu_a - y, 0)) = \int_0^{\alpha\mu_a} p_g(\alpha\mu_a - y)f(y) dy,$$

where y denotes yield, E_y is the expectations operator, and $f(y)$ is the density of yields.

To examine the role that sampling variability in estimation of expected yields plays in facilitating adverse selection, a better understanding of how expected net returns from crop insurance vary with the realization of a farmer's mean yield estimator is needed. Assume for now that farmers know their true mean yield and insurers act as if their estimates of farmers' mean yields used to determine insurance coverage equal their actual mean yields. For any given coverage level, a farmer's expected indemnities will increase when the estimated mean yield increases relative to a farmer's actual mean yield. If premiums decrease or do not rise as fast as expected indemnities, then sampling error that increases a producer's estimated mean yield will increase the producer's expected net returns from the program. As implied by this result, when the realization of the expected yield estimator is below (above) the farmer's true mean yield, farmers will tend to find insurance relatively overpriced (underpriced) and will tend to expect relatively lower (higher) returns from participation.^{4,5}

The misalignment between a farmer's estimated mean yield and his/her expected yield captures the asymmetric information effect quantified by Just, Calvin, and Quiggin (1999), where farmers with estimated expected yields above their true mean yield will be offered "fair premiums" (from the insurer's standpoint) that are lower than their expected indemnities. The result should be that farmers with estimated mean yields exceeding their true mean will have higher participation rates in the program compared to farmers whose estimated mean yields are lower than their true mean. In this situation, the insurer ends up with an adversely selected risk pool and with indemnities likely exceeding collected premiums. Increases in rates to mitigate losses will only result in a more adversely selected risk pool, because only producers with estimated mean yields that exceed their mean yield by a larger amount will stay in the program.

The above discussion assumes farmers know their true mean yield, which is not generally likely. However, the argument's validity only needs farmers to be better informed about their yield distribution than the insurer, which is certainly likely.

As noted in the introduction, even in the absence of adverse selection, the combination of sampling variability in estimation of expected yields and indemnity functions that are

³ For simplicity, we present our analysis in terms of yield insurance only. However, our results apply equally to the revenue insurance products that base their insurance guarantees and premium rates on a simple average of past yields.

⁴ A farmer's expected returns to crop insurance depend on more than the relationship of the farm's true mean yield and the farm's insurable yield. Other factors that may influence a farmer's assessment include premium subsidies, premium loads, and an incorrect risk assessment of a farmer's land.

⁵ The previous assertion is true for unsubsidized insurance. Subsidies of the premium rates will shift (lower) the realization of the expected yield estimator that will lead farmers to find the insurance to be relatively underpriced.

convex in the yield guarantee will cause insurance payouts to exceed the indemnities expected if the actual mean yield was known. Equation (1) demonstrates the well-known convexity of expected indemnities on yield guarantees:

$$(1) \quad \frac{\partial^2 I(\mu_a)}{\partial \mu_a^2} = p_g \alpha^2 f(\alpha \mu_a) \geq 0.$$

In the empirical section of the paper, the analysis is conducted for both the case where farmers know their true expected yield, and the case where potential informational asymmetries are not exploited by producers. The latter case includes the situation where informational asymmetries do not exist.

As suggested in the previous discussion, a decrease in sampling error could decrease adverse selection and bring expected indemnities closer to expected insurance payouts were the actual mean yields known. Current crop insurance rules allow insurance guarantees and premium rates to be determined by a simple average of as few as four years of yield data. Clearly, in this case, large sampling errors exist. Before presenting an alternative estimator that can reduce sampling error, the economic magnitude of the problems caused by sampling errors associated with use of a simple average is explored using Iowa corn farm-level yield data.

Impact of Uncertainty in the Estimation of Expected Yields

A Model of Yields

Farm yields at a point in time can be decomposed into systemic and idiosyncratic components (Miranda, 1991; Mahul, 1999; Vercauteren, 2000) as:

$$(2) \quad y_{it} = \mu_i + \beta_i(y_{ct} - \mu_c) + \varepsilon_{it} = \mu_c + \delta_i + \beta_i(y_{ct} - \mu_c) + \varepsilon_{it},$$

where μ_c is the area mean yield, δ_i is the difference between the area mean yield and farm i mean yield, y_{ct} is the area yield in year t , and ε_{it} is the farm-yield deviation in year t . It is also assumed that $E(\varepsilon_{it}) = 0$, $E(y_{ct}) = \mu_c$, $E(y_{it}) = \mu_i = \mu_c + \delta_i$, $\text{Var}(y_{ct}) = \sigma_c^2$, $\text{Var}(\varepsilon_{it}) = \sigma_\varepsilon^2$, $\text{Cov}(\varepsilon_{it}, y_{ct}) = 0$, and $\text{Var}(y_{it}) = \beta_i^2 \sigma_c^2 + \sigma_\varepsilon^2$.

This decomposition is appealing because of its simplicity and its similarity to the capital asset pricing model (CAPM) of finance, which provides clear interpretations for its main parameter β_i .⁶ In equation (2), β_i measures the sensitivity of farm yields to the systemic factors affecting area yield. Recently, Ramaswami and Roe (2004) derived this linear form from the aggregation of microproduction functions.

A variant of this model was used by Atwood, Baquet, and Watts (1996) to rate Income Protection. Their rating method uses equation (2) with a slight modification in order to obtain estimates of the residuals. These residuals are then used to construct simulated farm yields. While recognizing that premiums are affected by the precision involved in the estimation of a farmer's expected yields, Atwood, Baquet, and Watts do not elaborate

⁶ The CAPM model as used in finance relates the expected rate of return of an asset (r_i) to a risk-free rate of return (r_f) and the "market risk" given by the difference between r_f and the expected return on the market (r_m) as $r_i = r_f + \beta(r_m - r_f)$, where β captures the asset's expected returns sensitivity to systemic risks.

further on this point and do not investigate the implications of using this widely employed model for obtaining a better estimator of mean yield than the simple average of farm-level yield history.

Other authors (Miller, Kahl, and Rathwell, 2000a, b), rating insurance products for peaches in Georgia and South Carolina, have restricted beta to be one. Miranda (1991) showed that the area-weighted average of the betas of all farmers in the relevant region must equal one, but there is no reason why beta should equal one for all farmers.

Quantifying the Impacts of Sampling Error on Expected Indemnities

Monte Carlo simulation is used to obtain an estimate of the amount of sampling error arising from use of a simple average yield in a crop insurance program. Farm-level yields are modeled as in equation (2) for corn in the nine crop reporting districts (CRDs) in Iowa. Thus, the average CRD yield is the relevant area yield. For the analysis, 49 years (1956–2004) of CRD-level corn yield data were obtained from the National Agricultural Statistics Service (NASS). This longer series of data was used to: (a) fit trends for corn yields at the CRD level (which are used to detrend both CRD-level and farm-level yields as described below), (b) obtain the expected yield for each CRD (μ_c), and (c) estimate the model presented in equation (2) using the years in which both farm-level and CRD-level yields overlap.

Iowa corn is chosen for this analysis because Iowa corn farmers have a long history of high participation rates in the crop insurance program, and actual insurance unit-level yield data for participating farmers in each CRD of Iowa from 1991 to 2000 were made available by USDA's Risk Management Agency (RMA). The original data (at the optional-unit level) were aggregated to the enterprise-unit (or farm) level because of concerns about the quality of the data at the most disaggregate level, and to obtain data series across producers. Because insurance participation data are used in the study, concerns may arise about the presence of selection bias. However, as shown by table 1, the average yield from the farmers in the RMA sample for each CRD is similar to the average CRD yield as estimated by NASS.

To focus solely on the issue of sampling error, both the farm-level data and the CRD data were detrended. Plots of the CRD yields over time (not presented) showed significant trend terms, which were estimated as linear functions of time (separate trends for each CRD) using all 49 years of data. Farm yields were also detrended using the appropriate CRD trends. All yields were multiplicatively detrended to a base year (2004).⁷ The impact of the time trend specification on our results was checked by repeating the analysis for one CRD under three alternative time trends—locally weighted regression (linear and quadratic), and a two-knot linear spline regression. The results presented here were largely unaffected by the time trend selected.

Following Atwood, Baquet, and Watts (1996), equation (2) is estimated using OLS for each enterprise unit for which at least eight years of farm-level yield observations exist. Following Miranda (1991), and Atwood, Baquet, and Watts (1996), detrended data are used in estimation because the estimation equation can then be used to estimate

⁷ The detrending procedure used involves multiplying the actual yield at year t by the ratio of the CRD base year trend (2004 in this study, $\hat{y}_{c,04}$) to the trend at that year ($\hat{y}_{c,t}$). Specifically, all year t yields in CRD c are multiplied by $\hat{y}_{c,04}/\hat{y}_{c,t}$. In this way, all yields within the same year and CRD are multiplied by the same yield ratio, and relative yields are preserved.

Table 1. Summary Statistics of the Sample

CRD	Average Yield (bushels/acre)		Average Number of Farmers by Year	Std. Dev. of CRD Yields (bushels/acre)	Std. Dev. of Idiosyncratic Shocks ^c (bushels/acre)
	NASS ^a	RMA ^b			
10	150.8	153.2	159.8	22.3	20.2
20	148.8	149.0	149.9	19.3	20.3
30	144.7	141.8	80.2	20.4	22.9
40	145.1	145.2	127.3	23.3	20.5
50	152.5	152.7	313.0	22.2	20.2
60	142.9	142.3	153.3	22.3	23.1
70	138.2	137.7	102.2	23.8	21.3
80	123.5	131.4	48.9	27.9	27.7
90	134.3	134.4	105.0	26.8	22.7

^a Average of the detrended 1991–2000 CRD-level corn yields reported by NASS.

^b Average of the detrended 1991–2000 farm-level corn yields from the participation data used.

^c Idiosyncratic shocks are the farm-level yield deviations [ϵ_{it} from equation (2)] that are uncorrelated with area yields.

expected yield directly. The variance of the idiosyncratic shock is assumed to be equal for all enterprise units within a CRD. The regression residuals are used to estimate this variance as a weighted average of the error variance estimated for each farmer. The weights are determined by the number of yield observations of each farmer. Specifically, the variance of the idiosyncratic shock in district c is estimated by:

$$\hat{\sigma}_c^2 = \frac{\sum_{i=1}^I \sum_{t=1}^{T_i} (y_{it} - (1 - \hat{\beta}_i)\mu_c - \hat{\delta}_i - \hat{\beta}_i y_{ct})^2}{\sum_{i=1}^I (T_i - 2)},$$

where T_i denotes the number of yield observations available for farmer i , and I is the total number of farmers in the CRD. The estimated standard deviations of idiosyncratic shocks and CRD yields are presented in table 1. Note that the variance of idiosyncratic shocks would increase if the model were estimated using optional-unit data rather than enterprise-unit data.

The estimated variance $\hat{\sigma}_c^2$ captures enterprise-unit poolable risk. Systemic risk is captured by drawing from each CRD's detrended yield history. Together, the two sources of risk can be used to simulate the impact of sampling errors on crop insurance using equation (2) for any given β_i and δ_i . For simplicity, we begin by simulating expected indemnities for an average farmer in a CRD with $\beta_i = 1$ and $\delta_i = 0$. For each CRD, 10,000 yield samples, each of size 10, are constructed by bootstrapping from the set of detrended yields for that CRD with residuals independently drawn from a normal distribution with zero mean and variance $\hat{\sigma}_c^2$. For each of the 10,000 samples, the simple average of a subset of the observations (from 4 to 10 observations) is calculated. This step results in seven sets (each of size 10,000) of simple average yields for each CRD. Each set can be used to measure the expected amount of sampling error from use of a simple average yield in a crop insurance program that insures Iowa corn during this time period. By construction, simple average distributions based on fewer observations are mean-

preserving spreads of the distributions that average more years of data, and the mean of each of these distributions is the expected value of the farm yield which equals the expected CRD yield. Hence, the benefits of obtaining more precise estimators of a farmer's expected yield can be estimated.

From the insurer's perspective, insurance guarantees and expected indemnities will be based on the estimated μ_a . For any given contracted coverage level, insurers will expect to pay out more (less) indemnities than they actually pay out when $\mu_a < (>) \mu_c$ because, in this simulation experiment, all farmers have an actual expected yield equal to μ_c . The magnitude of the difference between what insurers expect to pay out and what they actually pay out depends on the contracted coverage level and μ_a . Actual average indemnities that will be paid out for each combination of μ_a and contracted coverage level can be obtained by simply calculating the corresponding actual coverage level and then using Monte Carlo integration on a sample of yield draws from the true yield distribution. For example, if μ_a is 20% greater than μ_c , then the contracted coverage levels of 65%, 75%, and 85% correspond to actual coverage levels of 78%, 90%, and 102% of μ_c . If μ_a is 20% less than μ_c , then the contracted coverage levels of 65%, 75%, and 85% correspond to actual coverage levels of 52%, 60%, and 68% of μ_c . A systematic sample, based on the quantiles of the distributions in 1% increments, is obtained for each of the seven expected yield estimator distributions (based on 4 to 10 yield observations). Using equation (2), 200,000 farm-level yields for each CRD are obtained by setting $\delta_i = 0$ and $\beta_i = 1$. Those simulations are used to compute the expected indemnities for each realization of expected yield estimated with a simple average. Averaging over realizations of the simple average provides the actual expected indemnities under sampling variability.

One measure of the impact of sampling error is to compare indemnities that would actually be paid out under sampling error to indemnities that would be paid out if the contracted coverage level were equal to the actual coverage level. Table 2 reports two sets of ratios. The top half of the table reports the ratio of the average actual indemnities paid to the average indemnities that would be paid under no sampling error for three different coverage levels and two different sample sizes under the assumption of no adverse selection. The bottom half reports the ratios of average indemnities under perfect adverse selection—i.e., only those farmers who have actual coverage levels greater than contracted coverage levels ($\mu_a > \mu_c$) buy crop insurance.

All entries in table 2 are greater than one because the expected indemnity schedule is convex. For example, at the 75% coverage level in CRD 10 with four observations, sampling error increases average indemnities paid by 21% under no adverse selection, and by 73.7% with adverse selection. Increasing the number of observations used to estimate the simple average yield to 10 still results in an 8.7% increase in indemnities with no adverse selection and a 40.6% increase with adverse selection in this CRD.

As observed from the table 2 results, the large amount of sampling error associated with the use of simple average yields in crop insurance programs can meaningfully affect actuarial soundness. Premium rates determined by historical indemnities will be significantly greater than they would be in the absence of sampling error, particularly when a large proportion of losses are paid to farmers who use relatively few observations to calculate their simple average, and when the program is subject to adverse selection. Of course, all estimators involve sampling error. But, as shown in the next section, an estimator based on equation (2) can significantly reduce sampling error, which can lead to less adverse selection and more accurate premium rates.

Table 2. Ratio of Actual Indemnities Based on Simple Average Yields to Expected Indemnities at the Farmer's Mean Yield

CRD	Simple Average Yield Observations					
	4 Observations			10 Observations		
	65% Coverage	75% Coverage	85% Coverage	65% Coverage	75% Coverage	85% Coverage
— Participate Independently of the Realized Simple Average —						
10	1.198	1.210	1.205	1.079	1.087	1.088
20	1.111	1.203	1.221	1.036	1.073	1.083
30	1.135	1.177	1.189	1.052	1.074	1.077
40	1.175	1.173	1.158	1.074	1.080	1.070
50	1.119	1.140	1.171	1.051	1.063	1.075
60	1.134	1.165	1.193	1.052	1.057	1.073
70	1.163	1.169	1.166	1.067	1.068	1.071
80	1.126	1.151	1.158	1.050	1.064	1.067
90	1.114	1.145	1.163	1.046	1.059	1.070
— Participate Only for Simple Average Realizations that Exceed Expected Yields —						
10	1.724	1.737	1.729	1.399	1.406	1.409
20	1.491	1.707	1.793	1.265	1.371	1.432
30	1.568	1.665	1.709	1.315	1.366	1.394
40	1.733	1.713	1.680	1.415	1.411	1.391
50	1.559	1.576	1.647	1.329	1.333	1.368
60	1.512	1.603	1.680	1.279	1.320	1.368
70	1.670	1.684	1.680	1.381	1.386	1.390
80	1.559	1.620	1.644	1.310	1.346	1.363
90	1.499	1.573	1.626	1.289	1.327	1.363

An Alternative Estimator

In the absence of yield trends, equation (2) shows that a simple average of past yields will provide an unbiased estimate of expected yield on a farm. In any given year, the estimator (\bar{y}_{it}) is simply the average of yields in the preceding T years:

$$(3) \quad \bar{y}_{it} = T^{-1} \sum_{t=1}^T (\mu_c + \delta_i + \beta_i(y_{ct} - \mu_c) + \varepsilon_{it}) = \mu_c + \delta_i + \beta_i(\bar{y}_{ct} - \mu_c) + \bar{\varepsilon}_{it}.$$

The expectation of average farm ($\bar{\varepsilon}_{it}$) and area ($\bar{y}_{ct} - \mu_c$) deviations are both zero. Sampling error arises when these average deviations are not zero.

Equations (2) and (3) allow for identification of the factors (other than the time trend) that make a farmer's simple average estimate of expected yields different from his/her expected yield. Based on T observations of farm- and area-level yields, the difference between a farmer's simple average yield and expected yield is $\bar{y}_{it} - E(y_{it}) = \beta_i(\bar{y}_{ct} - \mu_c) + \bar{\varepsilon}_{it}$. Thus, if an area has had a positive (negative) average shock during the T years considered and β_i is positive, then the farmer's simple average yield will tend to be greater (less) than his/her expected yield. If the farmer has a β_i of zero, then area shocks will have no influence on his/her simple average yield. The last term shows the effect of "on-farm shocks" in the difference between the simple average and expected yields.

The decomposition presented in equation (2) suggests an alternative to the simple average rule for calculating a farmer's expected yield, taking advantage of the information embedded in the area yield data. Equation (2) can be rewritten as:

$$(4) \quad y_{it} = \mu_c(1 - \beta_i) + \delta_i + \beta_i y_{ct} + \varepsilon_{it} = \alpha_i + \beta_i y_{ct} + \varepsilon_{it},$$

where $\alpha_i = \mu_c(1 - \beta_i) + \delta_i$, and hence, $E(y_i) = \alpha_i + \beta_i \mu_c$. Assuming the aggregation in the area yield is large enough, and after detrending farm- and area-level yields to a common base year, the parameters (α_i, β_i) can be estimated by applying OLS to equation (4). By the law of iterated expectations, and assuming $\hat{\mu}_c$ is unbiased, an unbiased estimator of farmer i 's expected yield is given by $\hat{y}_i = \hat{\alpha}_i + \hat{\beta}_i \hat{\mu}_c$.⁸ Thus, the right-hand side of equation (4) provides the basis for the estimator proposed in this study, \hat{y}_i .

To compare yield estimators, we use the mean squared error (MSE) criterion. The MSE of the simple average estimator is given by $MSE_{SA} = E(\bar{y}_{it} - \mu_i)^2 = \text{Var}(\bar{y}_{it}) + (E(\bar{y}_{it}) - \mu_i)^2$. Again, since yield trends are not considered in this study, $MSE_{SA} = \text{Var}(\bar{y}_{it})$. Because the proposed estimator based on equation (4) is also unbiased, its MSE equals its variance, given by $MSE_{NEW} = \text{Var}(\hat{y}_{it})$, and the proposed estimator is an improvement over the simple average estimator whenever $MSE_{NEW} < MSE_{SA}$, or, equivalently, whenever $\text{Var}(\hat{y}_{it}) < \text{Var}(\bar{y}_{it})$.

The variance of the simple average estimator can be derived from equation (3) as $\text{Var}(\bar{y}_{it}) = T^{-1} \text{Var}(y_{it}) = T^{-1} \text{Var}(\alpha_i + \beta_i y_{ct} + \varepsilon_{it}) = T^{-1}(\beta_i^2 \sigma_c^2 + \sigma_\varepsilon^2)$, where T is the number of observations used in estimation of \bar{y}_{it} . Under the assumption that the parameters α_i and β_i can be estimated through OLS, the variance of the proposed yield estimator can be derived. The regression parameters are estimated using the T_i observations for which paired data (y_{it}, y_{ct}) for farmer i are available. Assuming there is sufficient information whereby the expected yields for the area can be assumed to be known (or estimated with negligible error), the estimator for farmer i 's expected yield becomes $\hat{y}_i = \hat{\alpha}_i + \hat{\beta}_i \mu_c$ (i.e., the regression equation evaluated at $\check{y}_c = \mu_c$).

Considering that the regressor (area yields) is stochastic, the estimator just presented has an associated variance of:

$$\text{Var}(\hat{y}_i | \check{y}_c = \mu_c) = \sigma_\varepsilon^2 \left(\frac{1}{T_i} + E_{y_c} \left(\frac{(\mu_c - \bar{y}_{cT_i})^2}{T_i \sum_{j=1}^{T_i} (y_{cj} - \bar{y}_{cT_i})^2} \right) \right).$$

The variance of the proposed estimator for cases where the variance of the estimator of expected area yields cannot be ignored is easily derived.⁹ As is well known, the OLS estimates are the minimum variance linear unbiased estimators for α_i and β_i even if the regressors are stochastic (Greene, 1999).

⁸ $E(\hat{y}_i) = E_{\mu_c} E(\hat{\alpha}_i + \hat{\beta}_i \hat{\mu}_c | \hat{\mu}_c) = E_{\mu_c}(\alpha_i + \beta_i \hat{\mu}_c) = \alpha_i + \beta_i \mu_c$.

⁹ The variance of the proposed estimator when sampling variance of area yield is accounted for equals

$$\text{Var}(\hat{y}_i) = \sigma_\varepsilon^2 \left(\frac{1}{T_i} + E_{\mu_c} E_{y_c} \left(\frac{(\hat{\mu}_c - \bar{y}_{cT_i})^2}{T_i \sum_{j=1}^{T_i} (y_{cj} - \bar{y}_{cT_i})^2} \right) \right) + \beta_i^2 \text{Var}(\hat{\mu}_c).$$

The MSE of both estimators for farmer i 's expected yield can now be compared. In this case, the proposed estimator results in a lower MSE whenever

$$\sigma_e^2 E_{y_c} \left(\frac{(\mu_c - \bar{y}_{cT_i})^2}{T_i} \right) < T_i^{-1} (\beta_i^2 \sigma_c^2),$$

$$\left(\sum_{j=1}^{T_i} (y_{cj} - \bar{y}_{cT_i})^2 \right)$$

which after some rearrangement yields:

$$(5) \quad \sigma_e^2 E_{y_c} \left(\frac{\mu_c - \bar{y}_{cT_i}}{s_c / \sqrt{T_i}} \right)^2 < \beta_i^2 \sigma_c^2 (T_i - 1),$$

where s_c is the sample standard deviation of area yields. Tables 1 and 3 provide estimates of the standard deviations of idiosyncratic shocks and CRD yields, as well as information about the distribution of betas.

Equation (5) can be used to identify conditions under which the proposed estimator will perform better than the simple average estimator in the sense of MSE. Increasing the number of periods used will likely make MSE_{NEW} lower than MSE_{SA} .¹⁰ Consequently, the proposed rule will tend to perform better than the simple average rule when the number of observations increases and when the idiosyncratic risk (σ_e^2) is small. Equation (5) also indicates that the simple average rule has a greater chance of outperforming the proposed estimator when β_i is close to zero and/or it cannot be estimated precisely. The previous observations make sense because substantial noise will be introduced estimating β_i that controls for a systemic component having little effect on the farm-level yield. The effect of controlling and removing the systemic variability may be outweighed by introduced sampling variability.

Figures 1 and 2 help to illustrate when the proposed estimator outperforms the simple average rule. Since both estimators are unbiased, only their variances are compared. Only figures for Iowa CRD 50 (central Iowa) are presented because the other eight districts display the same general patterns. Farm-level yields used to compute the variance of the estimators are constructed employing the same numerical procedures detailed in the previous section. For every case, δ_i is fixed at zero and pairs of farm- and area-level yields are simulated and used to compute the bootstrapped variances for both estimators.

As illustrated in figure 1, increasing the number of farm yield observations increases the precision for both estimators. When $\beta_i = 0.5$ (panel A), the proposed estimator performs worse than the simple average rule when few farm yield observations are available. The relatively poor performance of the proposed estimator in this situation is expected as shown by equation (5). If a short series of area yields used in the estimation of β_i has low variability, then precise estimation of the regression parameters

¹⁰ As the sample size increases, the central limit and Slutsky's theorems indicate equation (5) reduces to $\sigma_e^2 < \beta_i^2 \sigma_c^2 (T_i - 1)$, since

$$\frac{\sqrt{T_i}(\mu_c - \bar{y}_{cT_i})}{\sigma_c} \frac{\sigma_c}{s_c} \rightarrow N(0, 1).$$

Also, if it can be argued that the average of farm-level yields (based on few observations) is approximately normally distributed, equation (5) reduces to $\sigma_e^2 \leq \beta_i^2 \sigma_c^2 (T - 3)$. These formulas point to the same conclusions presented in the body of the paper.

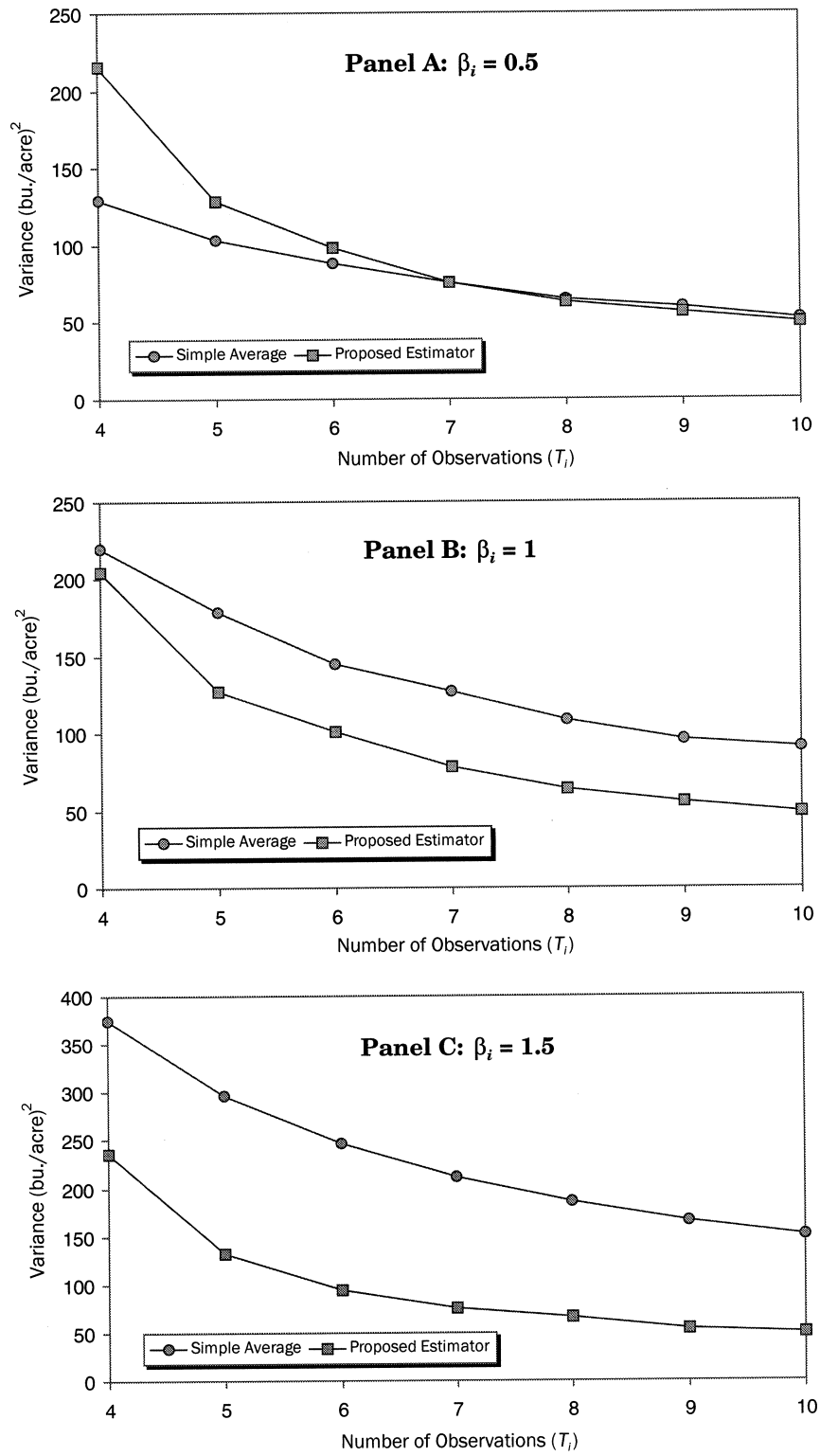


Figure 1. Relationship between the variance of a simple average rule and the proposed estimator

is impossible. This problem is mitigated when a longer time series of area-level yield data is used. Panels B and C of figure 1 show that the proposed estimator has a uniformly lower variance than the simple average rule when β_i increases, except perhaps when $T_i = 4$, where (as just mentioned) the proposed estimator may exhibit erratic behavior. Note that an increase in β_i increases the systemic component of variance; thus, the total farm-yield variance also increases because the idiosyncratic component of the variance is held constant.

Figure 2 depicts the relationship between a farmer's beta and variance of the estimators. As beta increases, farm-level variance also increases. As shown, the variance of the proposed estimator is largely invariant to farm-level variance, making it more robust than the simple average rule for farms with large betas. It is worth noting that for low values of beta, the simple average yield rule performs marginally better than the proposed estimator, whereas the latter greatly outperforms the former for larger betas. This finding suggests the distribution of farm betas in a region will determine the extent to which aggregate variance across farmers will be reduced by moving to the proposed estimator.

Table 3 reports sample statistics of the betas estimated through OLS for farmers with at least eight yield observations. As observed in table 3, the average beta for each CRD is close to one, as expected, and the standard deviations are approximately 0.4 for all districts. Consistent with Miranda (1991), the distribution of betas possesses no discernable skewness. Tests for normality fail to reject the null hypothesis (at a 5% confidence level) for all CRDs, and hence the normal distribution is used (with the average and standard deviations shown in table 3) to model the distribution of betas by region.¹¹

Table 4 presents the expected variance change for all Iowa CRDs when the yield estimators are based on 4 to 10 yield observations. Based on the results, the average variance of the proposed estimator is substantially lower than the average variance of the simple average estimator, except when there are just four observations available. The table reveals that gains in variance reduction beyond those resulting from increasing the number of observations can be achieved by switching estimators.

One measurement of the impact this change in variance could have on a crop insurance program is a reduction in "excessive" indemnities paid out because of sampling errors. As reported in table 2, the presence of sampling error increases indemnities above the level that would be paid out if there were no sampling error. The amount by which the reported ratios exceed 1.0 is taken as the measure of excessive indemnities. The table 2 ratios were calculated for a farmer with a beta of 1.0. A better measure would account for the actual distribution of betas.

Figure 3 charts the percentage reduction in excessive premiums that would accrue from moving to the proposed estimator taking into account the distribution of betas. These estimates are the average reduction across the nine Iowa CRDs for 65%, 75%, and 85% coverage levels for the situation in which there is no adverse selection. Consistent with table 4, an increase in the sample size increases the percentage reduction. At the 75% coverage level with 10 observations, the average reduction in excess indemnities is about 35%. At six observations, the reduction is 21%. Also consistent with table 4, with only four observations, the proposed estimator performs worse than a simple average.

¹¹ The number of farms used by CRD ranged from 58 to 368 with an average of 162, and the results from the test were the same for all CRDs. Additionally, and to check the impact of the parametric form imposed on the betas, the simulations were repeated using the empirical distribution of betas per CRD and yielded very similar results.

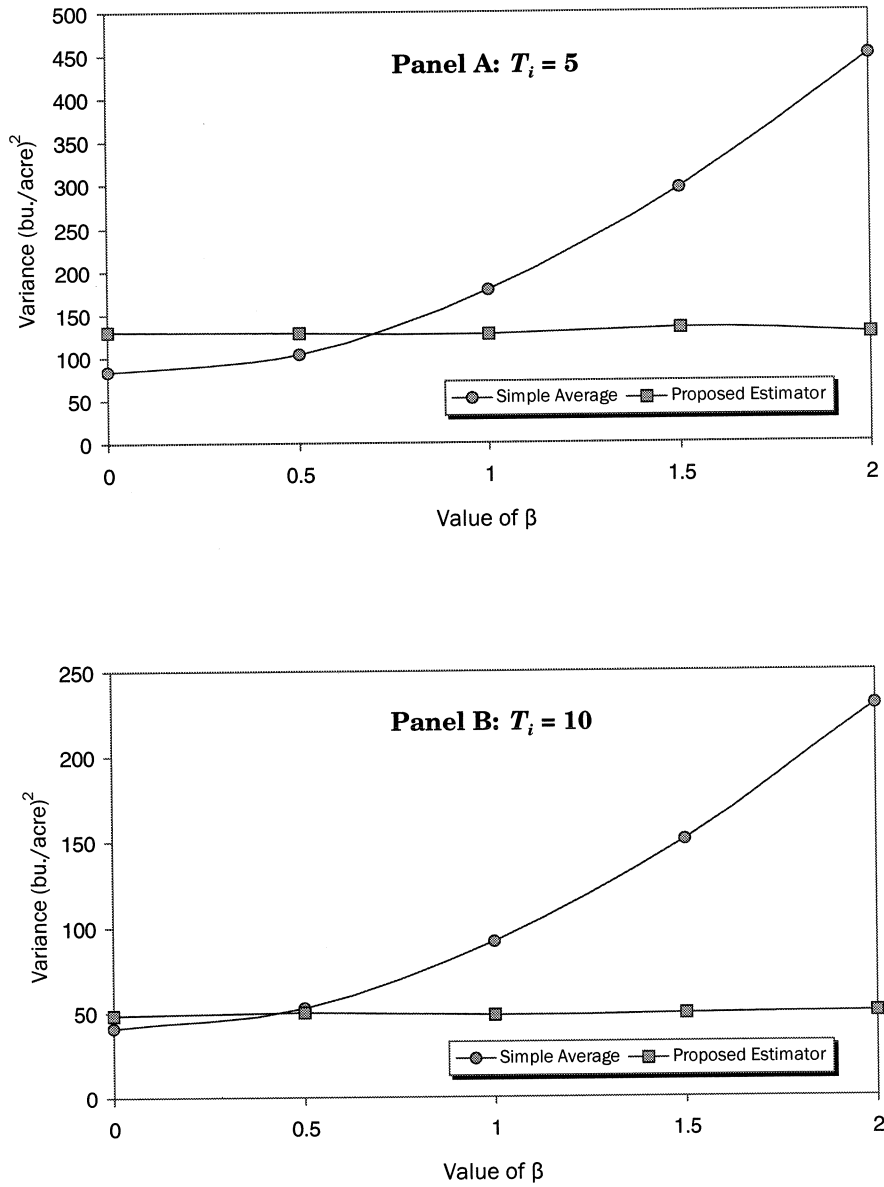


Figure 2. Variance of the simple average rule and the proposed estimator for different values of beta and number of observations

Table 3. Summary of Estimated Betas and the Distribution of the Betas for the Nine Iowa CRDs

CRD	No. of Farmers ^a	Average of Betas	Std. Dev. of Betas	Sample Quantiles of Betas		
				0.25	0.50	0.75
10	189	0.96	0.38	0.687	0.937	1.218
20	177	1.06	0.32	0.833	1.040	1.298
30	93	1.14	0.41	0.878	1.162	1.448
40	153	1.02	0.35	0.757	1.022	1.256
50	368	1.02	0.33	0.769	1.014	1.223
60	179	1.04	0.38	0.760	1.016	1.288
70	119	0.95	0.34	0.713	0.909	1.126
80	58	0.87	0.37	0.583	0.926	1.083
90	120	1.03	0.38	0.775	1.038	1.328
Average	162	1.01	0.36	0.750	1.010	1.250

^a Note, this column does not coincide with the number of farmers reported in table 1. This table presents all the farmers in the sample (with at least eight years of yields), whereas table 1 reports the average number of farmers in every year.

Table 4. Percentage Change in the Expected Variance by Moving from a Simple Average to Proposed Yield Estimator

CRD	Number of Observations in the Yield Estimator						
	4	5	6	7	8	9	10
	— % Change —						
10	289.0	-23.1	-34.9	-40.5	-43.5	-45.4	-47.0
20	125.8	-15.1	-32.3	-37.3	-41.4	-42.2	-44.6
30	20.5	-17.3	-30.4	-35.3	-40.2	-42.6	-44.0
40	2.2	-26.9	-39.6	-43.9	-47.7	-50.0	-51.7
50	4.7	-31.0	-40.3	-42.0	-45.6	-48.0	-49.2
60	17.7	-18.2	-28.5	-33.1	-38.0	-39.6	-41.9
70	143.2	-27.0	-36.9	-40.9	-44.2	-46.2	-47.3
80	149.0	2.6	-21.0	-28.7	-33.1	-35.3	-37.1
90	4.9	-39.9	-47.4	-51.0	-54.0	-55.2	-56.8

Potential Impacts on APH Rating

The analysis shows the proposed yield estimator has a smaller variance for expected yield estimates than can be achieved from a simple average of past yields if more than four observations are used. This smaller variance would lead to less adverse selection and improved APH premium ratemaking if the proposed yield estimator were used to determine insurance guarantees.

Currently, APH premium rates are based on a moving average of historical loss-costs. The ratemaking process blends historical loss-costs for an area and crop with similar data for surrounding areas to create a premium rate for APH. Given the current mechanisms for setting APH premium rates, the proposed yield estimator could impact insurance rates in two ways.

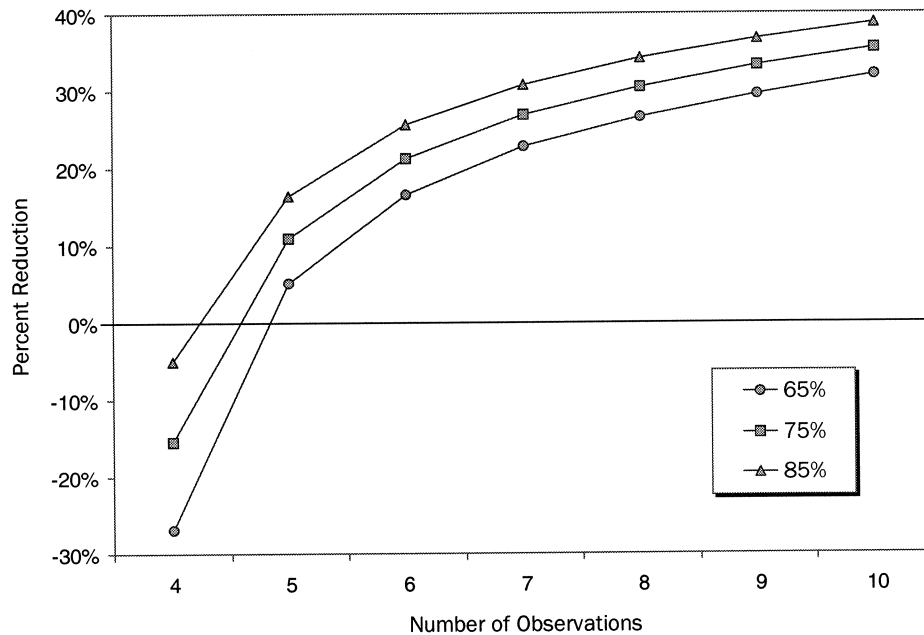


Figure 3. Average percentage reduction in excess premiums under the proposed estimator across the nine Iowa CRDs

First, if the proposed estimator were adopted as a replacement for APH yields, then future insurance indemnities with the new yield estimator would begin to systematically enter the loss-cost data used to determine APH premium rates. Because the proposed yield estimator would reduce the deviations between insurable yields and true mean yields, its usage should reduce adverse selection. Also, in the presence of insurance payouts that are convex in the yield guarantee, a lower variance estimator will bring expected payouts from the program closer to expected payouts under no sampling error. Specifically, excessive indemnities would be reduced. Loss-costs for the policies with the proposed yield estimator would reflect the decreased effects of adverse selection and the convexity of the expected indemnities on coverage levels. And as those loss-costs are eventually worked into the APH rate calculations, premium rates would be adjusted to better match with expected yields, which should work to reduce overall premium rates.

Second, given the data at RMA's disposal, the proposed yield estimation procedure could be applied to the historical insurance data. Historical loss-costs could be computed under the proposed yield estimator and these new historical loss-costs could be utilized in the current APH rate calculations to more fully reflect the impact of the proposed yield estimator on rate structure. In the first case, the proposed yield estimator would have a gradual impact on premium rates as each year of insurance experience is added to the historical base for APH premium rate formation. In the second case, the impact would be much quicker as all of the loss-cost data used to construct the premium rates would capture the effects of the proposed yield estimator on indemnities and liabilities.

The results from the comparison of expected yield estimators reveal that a reduction in the variance of the estimates translates into a reduction in indemnities that are paid out. Over the last decade, crops in the U.S. Corn Belt have generated insurance losses

less than are consistent with current crop insurance rates, leading some to suggest crop insurance is overrated in Corn Belt states (e.g., Schnitkey and Sherrick, 2007). Because adoption of the proposed yield estimator would lead to reductions in premium rates, it is a partial “remedy” to the overrating issue. While the issue of increasing yield trends is likely the larger component to the perceived overrating, variability in insurable yield estimates also contributes to overrating. The work in this analysis is targeted specifically at the variability issue. However, it should be noted that the yield estimator proposed here can easily be adapted to handle yield trending at the county or farm levels.

Conclusions

Barnett et al. (2005) demonstrated that errors in estimating mean farm yields potentially have a large impact on the ability of the U.S. crop insurance program to deliver efficient risk management tools to farmers. In our study, the magnitude of the problem for one crop and one state (Iowa corn) and the reduction in sampling variance that would come about from adoption of an alternative estimator are estimated. The new estimator will reduce sampling variance by an average of up to 57% when 10 observations are available to estimate the expected yield. Basing insurance guarantees on a simple average of past yields outperforms the new estimator only when four observations are used.

A reduction in sampling error should reduce the potential for adverse selection in crop insurance programs that would be possible if asymmetric information exists between insurance companies and farmers. Basing insurance guarantees on a lower-variance estimator would reduce the difference between contracted and actual coverage levels, thereby reducing the magnitude of the incentive for adverse selection. The reduction in the difference between contracted and actual coverage levels would also decrease the average level of indemnities paid out for any given contracted coverage level. This reduction would improve the actuarial performance of insurance programs. In addition to the benefit of more accurate premium rates, adoption of the proposed estimator would provide RMA with a method to reduce the impact of multiple-year yield losses on producers' insurance guarantees and premium rates because the proposed estimator ties a producer's expected yield directly to county yields.

Because of programming and learning costs, it is uncertain whether the benefits of the estimator would convince RMA and crop insurance companies to adopt the proposed regression-based estimator. However, both RMA and company personnel have successfully implemented much more complicated rating mechanisms with current products such as Revenue Assurance, Livestock Gross Margin, and the new Combined Product. The current system also involves complicated programming and education. A detailed understanding of yield cups, yield caps, and *t*-yields is required before one can calculate an APH yield. In fact, two yield calculations must be done before both an insurance guarantee and a premium rate can be calculated. The insurance guarantee is based on an APH yield, whereas the premium rate is based on a “Rate Yield.”¹²

¹² APH yields reflect yield adjustments (cups, caps, and transitional yields), whereas the “Rate Yield,” the one used for rating, is based on the actual production records without adjustment. The yield adjustments prevent the APH yield from increasing by more than 20% (cap) or decreasing by more than 10% (cup) in any one given year. APH yields are also prevented from falling below a floor (as a proportion of the county's transitional yield), which depends on the number of yield years in the farmer's record.

Finally, because this analysis was based on Iowa corn farmers' yields and crop insurance participation, caution must be taken in extrapolating the implications to other crops and states. Iowa corn yields are determined by systemic factors perhaps more than elsewhere. The degree of poolable risk affecting Iowa corn is relatively low. Both of these conditions favor the proposed estimator.

[Received December 2006; final revision received December 2007.]

References

- Atwood, J. A., A. E. Baquet, and M. J. Watts. "Income Protection." Tech. Rep., Dept. of Agr. Econ. and Econ., Montana State University. Prepared for USDA, Economic Research Service, Commercial Agricultural Division, February 1996. Online. Available at <http://www.rma.usda.gov/pilots/miscproj/techsum.pdf>. [Accessed November 2004.]
- Barnett, B. J., J. R. Black, Y. Hu, and J. R. Skees. "Is Area Yield Insurance Competitive with Farm Yield Insurance?" *J. Agr. and Resour. Econ.* 30,2(August 2005):285-301.
- Greene, W. H. *Econometric Analysis*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 1999.
- Just, R. E., L. Calvin, and J. Quiggin. "Adverse Selection in Crop Insurance." *Amer. J. Agr. Econ.* 81(1999):835-849.
- Mahul, O. "Optimum Area Yield Crop Insurance." *Amer. J. Agr. Econ.* 81(1999):75-82.
- Miller, S. E., K. H. Kahl, and P. J. Rathwell. "Evaluation of Crop Insurance Premium Rates for Georgia and South Carolina Peaches." *J. Agribus.* 18(Fall 2000a):303-317.
- . "Revenue Insurance for Georgia and South Carolina Peaches." *J. Agr. and Appl. Econ.* 32(2000b):123-132.
- Miranda, M. J. "Area-Yield Crop Insurance Reconsidered." *Amer. J. Agr. Econ.* 73(1991):233-242.
- Ramaswami, B., and T. L. Roe. "Aggregation in Area-Yield Crop Insurance: The Linear Additive Model." *Amer. J. Agr. Econ.* 86(2004):420-431.
- Schnitkey, G. D., and B. J. Sherrick. "GRIP in 2007." Pub. No. FEFO 07-04, Dept. of Agr. and Consumer Econ., University of Illinois, 28 February 2007.
- Sherrick, B. J., F. C. Zanini, G. D. Schnitkey, and S. H. Irwin. "Crop Insurance Valuation Under Alternative Yield Distributions." *Amer. J. Agr. Econ.* 86(2004):406-419.
- Skees, J., and M. Reed. "Rate Making for Farm-Level Crop Insurance: Implications for Adverse Selection." *Amer. J. Agr. Econ.* 68(1986):653-659.
- Vercammen, J. A. "Constrained Efficient Contracts for Area Yield Crop Insurance." *Amer. J. Agr. Econ.* 82(2000):856-864.
- Wang, H. H., S. D. Hanson, and J. R. Black. "Efficiency Costs of Subsidy Rules for Crop Insurance." *J. Agr. and Resour. Econ.* 28,1(April 2003):116-137.