ABSTRACT

Title of dissertation: ASYMMETRIC INFORMATION AND ALTERNATE PREMIUM RATING METHODS IN U.S. CROP INSURANCE: A COMPARISON OF HIGH AND LOW RISK REGIONS

Roger Claassen

Dissertation directed by: Professor Richard E. Just
Department of Agricultural and Resource Economics

Federally subsidized crop insurance has a long history of underwriting losses. These losses may be due to premium rating procedures that do not account, as fully as possible, for differences in yield loss risk across farms. If farmers understand their own yield loss risk in more detail than is reflected in crop insurance premium rates, an information asymmetry may be leading to adverse selection or moral hazard. Regional differences in underwriting losses suggest that the effect of asymmetric information is relatively large where inter-farm yield variability is also relatively large.

An econometric model is used to identify asymmetric information in crop insurance premium rating. A simulation model is used to compare existing crop insurance premium rates to alternative rates calculated using yield loss risk measures based on existing, farm-specific yield history data. Both models are applied to crop/region combinations where inter-farm yield variability is relatively low (non-irrigated corn in the Corn Belt region) and where inter-farm yield variability is relatively high (non-irrigated, continuously cropped wheat in the Northern Plains).

Region-wide asymmetric information effects are identified for both regions, but the asymmetric information effect is found to be larger in the high variability region. This difference explains at least part of the inter-regional difference in underwriting losses. The simulation analysis suggests that, on average, across an entire region, premium rates derived from a farm-specific measure of yield variability are closer to actuarially fair rates than RMA premium rates. At a county- and farm-level, however, it is much more difficult to say, with a high level of statistical confidence, whether these alternate premium rates are closer than RMA rates to the actuarially fair rates.

To provide a foundation for the crop insurance models, an econometric model of crop yields is estimated and used to separate total yield variation into systematic and random components. Random yield variation is tested against several common distributions, including normal, gamma, and beta. The effect of aggregation on the representation of both systematic and random yield variation is also investigated.

ASYMMETRIC INFORMATION AND ALTERNATE PREMIUM RATING METHODS IN U.S. CROP INSURANCE: A COMPARISON OF HIGH AND LOW RISK REGIONS

by

Roger Claassen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
doctor of philosophy
2008

Dedicated to Dr. Bruce L. Gardner

ACKNOWLEDGEMENTS

Many people helped me to complete this work. To my faculty advisor, Dr. Bruce Gardner, I owe a particularly large debt of gratitude. Over a number of years and through several setbacks, Bruce was always positive and encouraging. Given the demands of work and family, it would have been easy to drop out. Bruce's support and encouragement were crucial to the successful conclusion of my graduate program. I am deeply grateful to him for this.

I thank Dr. Richard Just and Dr. Erik Lichtenberg for many helpful comments and suggestions throughout my dissertation process. They shared freely of their insight and professional expertise and their guidance was critical in transforming rough ideas into a coherent analysis. Dr. Howard Leathers, Dr. Wesley Musser, and Dr. Daniel Vincent also agreed to serve as committee members for the dissertation defense and I thank them for their helpful comments and insights.

I wish to especially thank Dr. Just for stepping into the role of committee chair late in the process and Dr. Leathers for agreeing to join the committee only one week before the defense. Their willingness to step into these roles on short notice allowed the defense to go forward as scheduled.

I am also grateful to Marca Weinberg, who made it possible for me to successfully balance the demands of work and graduate study. Without her support and encouragement it would not have been possible to complete this dissertation.

To my wife, Cheryl, and our children Adam and Olivia: I could not have done this without you. You helped me in many ways, large and small, to reach this goal. Your faith in me is what has mattered most. Thank you.

TABLE OF CONTENTS

Chapter 1: Introduction

The Risk Management Agency (RMA) of the U.S. Department of Agriculture offers highly subsidized crop insurance to producers of a wide range of agricultural commodities. RMA rates crop insurance premiums using farm-specific data on yield mean but county-level data on yield variability. A number of analysts have argued that use of county-level rather than farm-level estimates of yield variability has resulted in asymmetric information, leading to adverse selection and/or moral hazard problems in the crop insurance program (e.g., Just *et al.*; Vandeveer and Lohman; Makki and Sumwaru). Regional differences in underwriting losses also suggest that asymmetric information problems could be more severe in some regions than in others (Glauber).

Can existing crop insurance yield history data, collected and maintained by RMA, be used to reduce asymmetric information in crop insurance premium rating? In this report, I investigate the information content of RMA crop insurance yield history data, particularly with respect to yield variability. I use these data, along with crop insurance contract data, to (1) identify asymmetric information in crop insurance premium rating, (2) show how consequences of asymmetric information vary by regions, and (3) assess whether the use of existing yield histories could reduce asymmetric information in crop insurance premium rating. I also use crop insurance yield data to investigate the distribution of farm-specific yields and the effect of spatial aggregation on the representation of yield variability.

In this context, use of the term “asymmetric information” is a bit of a misnomer. The real question is whether RMA can use existing data more effectively in rating crop insurance contracts. Nonetheless, previous authors have used the term in a similar context (particularly Makki and Somwaru) and I follow that precedent. Moreover, even if it is possible to reduce asymmetric information using farm-specific yield histories, some level of asymmetric information will likely remain. It is likely that farmers know more about their yield loss risk than can be deciphered from yield history data. Even when RMA possesses a full 10 year yield history, most farmers have site specific knowledge of soils, rainfall patterns, and other factors that may have contributed to past losses but may not be apparent in from the yield history. So, even though a reduction in asymmetric information could help to reduce underwriting losses and could allow policy makers to reduce premium subsidies, some level of asymmetric information is inevitable.

Crop Insurance: A (Very) Brief History

Crop insurance has been widely available to U.S. producers at subsidized premium rates since 1980. The Federal Crop Insurance Improvement Act of 1980 mandated broad availability of multi-peril crop insurance and premium subsidies of up to 30 percent. It was hoped that crop insurance would replace annual, ad hoc disaster assistance. During the 1980s and early 1990s, however, the crop insurance program was plagued by low participation and large underwriting losses. During these years, crop insurance purchase peaked at about 25 percent of eligible acreage, except in years when crop insurance was a condition of eligibility for disaster assistance (Glauber). Between 1981 and 1993, moreover, crop insurance indemnities exceeded premium payments by about 50 percent, a loss of about $2.3 billion (USGAO).

In 1994, Congress approved another major revision to the crop insurance program. The Crop Insurance Reform Act of 1994:

- increased premium subsidies for all levels of coverage (Table 1.1);
- provided for coverage levels in excess of 75 percent (80 and 85 percent);
- created catastrophic coverage (covers 50 percent of the yield at 55 percent of price) with a 100 percent premium subsidy (producers paid a flat $50 fee per crop); and
- required commodity program participants to obtain at least catastrophic (CAT) crop insurance coverage.

Table 1.1 Schedule of Premium Subsidies for Federally Subsidized Crop Insurance

Coverage Level		Premium Subsidy Rates		
Yield	Price	1981-94	1995-98[1]	2001-present[1]
Percent		*Percent of Premium*		
50	55	--[2]	100	100
55	100	30	46	64
65	100	30	42	59
75	100	17	24	55
85	100	--[2]	12	38

Source: Risk Management Agency, USDA

[1]Ad hoc premium subsidy increases were enacted for the 1999 and 2000 crop years, but were similar to premium subsidy in effect from 2001 to the present.

[2]CAT coverage and coverage over 75 percent were not offered before 1995

In 1995 crop insurance participation increased dramatically, reaching 221 million acres (about 80 percent of eligible acreage) including 115 million acres of CAT coverage, with much of the increase in response to the commodity program requirement (Dimukes and Glauber). After 1995, Congress dropped the insurance requirement for commodity program participants because producers complained that CAT coverage – despite low cost – provided little protection. Over the next several years CAT coverage declined while buy-up coverage (coverage in excess of CAT) slowly increased in terms of both acres and liability. Growth in newly created revenue insurance products (e.g., Crop Revenue Coverage) was particularly strong.

For the 1999 and 2000 crop years, Congress mandated additional premium subsidies and, beginning in the 2001 crop year, mandated permanent subsidy increases under the Agriculture Risk Protection Act of 2000. With the new, higher subsidies, crop insurance participation once again rose sharply as producers insured more acreage and bought higher levels of coverage (Dismukes and Vandeveer). By 2002, 80 percent of eligible acreage was covered with more than 50 percent of acreage covered at 70 percent or more. Between 1994 and 2003, moreover, indemnities paid were 98 percent of total premiums (not including the government subsidy). These aggregate figures led some to conclude that the crop insurance program was on actuarially sound footing, even though the cost of premium subsidies pushed the overall cost of the crop insurance program to more than $2 billion per year (Glauber).

The evidence, however, suggests that underlying actuarial problems continue. Despite large increases in crop insurance participation, large underwriting losses have persisted in parts of the Great Plains and the Southeast. These problems suggest that broader, subsidy-induced purchase of Federal crop insurance may have only masked underlying problems (Glauber). Large subsidies may have encouraged producers with low yield risk (who have a relatively small chance of receiving crop insurance indemnities) to purchase crop insurance. The underwriting gains from these producers may be offsetting the underwriting losses of others.

Did the crop insurance reforms of 1994 and 2000 simply mask actuarial problems with higher participation driven by premium subsidies? A number of researchers have argued that asymmetric information may be causing adverse selection or moral hazard problems in the Federal crop insurance program (e.g., Goodwin; Vandeveer and Lohman; Just et al., ; Makki and Sumwaru). Farmers who know more than RMA does about potential yield loss will adjust their crop insurance purchases and/or production practices accordingly. Producers who expect a positive net return to crop insurance are more likely to buy than producers who expected a negative return (adverse selection). Producers may also be less diligent in protecting crops against loss if losses are indemnified (moral hazard).

According to RMA actuarial documentation, premiums for yield-based insurance are rated using county-level data on potential yield loss, adjusted to farm–specific conditions using a farm-specific average yield obtained from yield histories maintained by RMA (Josephson *et al.*). Skees and Reed showed that (1) yield means are independent of standard deviations, (2) the coefficient of variation generally declines as average yield risees, and (3) producers with higher yields tended to be associated with lower theoretical premium rates. One way to make these three findings consistent with one another is to assume that yield variability (e.g., standard deviation) is roughly constant across producers. If that is true, lower premium rates are appropriate for high-yield producers as they are less likely to experience any given *percentage* loss in yield. RMA premium rating methods reflect these findings in the sense that they assume that expected loss is inversely related to average yield.

The RMA base premium rate is a measure of yield loss risk that underlies the most popular crop insurance products: Actual Production History (APH), Crop Revenue

Coverage (CRC), and Revenue Assurance (RA). In the simplest terms, the base premium rate for coverage level θ, on farm j, is calculated as:

$$BPR_{\theta j} = \varphi_{\theta}\left(\left(\frac{\tilde{y}_j}{y_R}\right)^{e}\rho_{REF} + F\right)$$

where $\tilde{y}_j$ is the RMA rate yield[1] on farm j, y_R is the county-level reference yield, e is the exponent, ρ_{REF} is the county-level reference rate (an adjusted version of the county-level loss-cost ratio or LCR), and F is the fixed rate load that accounts for prevented planting, catastrophic losses, etc. In its most basic form, the county-level LCR is total indemnity paid to producers in a given county divided by total crop insurance liability. Actuarial documentation also shows that a number of additional adjustments are used to spread risk across farms (see Josephson *et al.*). The exponent is always negative ($e < 0$) so that the base premium rate declines as the rate yield rises relative to the reference yield. A number of other adjustments related to high risk areas, special additional coverage, and other factors can also be made when calculating a farm-specific premium rate (see RMA 2000a and RMA 2000b, for more information).

Critics have argued that inter-farm variation in yield loss risk cannot be accurately assessed by looking only at differences in average yield. Skees and Reed argued that even if there is a general, statistically significant relationship between average yield and expected indemnity, heterogeneity in yield loss risk among farms with the same average yield can lead to asymmetric information problems and adverse selection.

Other researchers have reached similar conclusions. In a study of the relationship between the mean and variability of crop yields, Goodwin found little evidence to suggest

---

[1] The RMA rate yield is essential the mean yield for a producers production history. .

that yield losses are related to yield means and argued that premiums rated without farm-specific information on yield variability will not be correct. Vandeveer and Loehman found that yield loss risk (expected indemnity) varies significantly among farms with the same average yield and that crop insurance participants are likely to have a positive expected net return to crop insurance purchase. Just *et al.* used survey data on producers' subjective yield expectations to show that positive expected return is a key motivation for producers who purchase crop insurance and that negative return is a key barrier to producers who do not purchase insurance. They suggest that the lack of a suitable farm-specific measure of yield variability is at least part of the problem. Makki and Somwaru find that producer insurance purchase decisions are driven in part by the number of years in which yields (from crop insurance yield histories) have fallen below guarantee levels. Using a technique developed by Chiappori and Salanie, they find evidence of asymmetric information in corn and soybean insurance contracts in Iowa.

In theory, an actuarially fair premium rate is equal to the expected loss per dollar of crop insurance liability given a specific coverage level θ_j, which I denote as $E(L_j|\theta)$. The RMA premium rate is correct when $E(L_j|\theta) = \varphi_\theta \left(\frac{\tilde{y}_j}{y_R} \right)^e \rho_{REF} + F$. In practical terms, of course, expected loss can only be estimated as the loss-cost ratio over a period of years: $\overline{L}_j(\theta_j) = N_j^{-1} \sum_t L_{jt}(\theta_j)$, where $L_{jt}(\theta_j)$ is the ratio of indemnity to insurance liability on farm *j* in year *t*, The farm-specific ratio of indemnity to liability is defined as:

$$L_{jt}(\theta_j) = \frac{p_t(\theta_j \tilde{y}_j - y_{jt}) Z(\theta_j \tilde{y}_j > y_{jt})}{p_t \theta_j \tilde{y}_j} = \frac{(\theta_j \tilde{y}_j - y_{jt}) Z(\theta_j \tilde{y}_j > y_{jt})}{\theta_j \overline{y}_j}$$

where p_t is output price, θ_j is the coverage level expressed as a proportion (for 75% coverage θ=.75), $\tilde{y}_j$ is the RMA rate yield, y_{jt} is the actual yield, and Z(.) is an indictor function with value one when $\theta_j \tilde{y}_j > y_{jt}$, zero otherwise. Because farm-specific yield histories contain 10 or fewer yield observations, the real question is whether a farm-specific loss-cost ratio, which may be subject to substantial sampling error, comes closer to the theoretically correct premium rate than the RMA premium rate (once the farm-specific rate is adjusted for disaster reserves and prevented planting).

This Study

In this study, I develop (1) an econometric model of asymmetric information in crop insurance and (2) a simulation model to compare existing RMA premium rates to an alternative rating method that uses existing yield history data to incorporate farm-level information on yield loss risk. I apply both models to crop/region combinations where yield loss risk is low (non-irrigated corn in a portion of the Corn Belt region) and where yield loss risk is relatively high (continuously cropped non-irrigated wheat in a portion of the Northern Plains). With these models, I provide some insight on the role of RMA rating methods in regional differences in underwriting losses. I also assess the potential for using farm-specific data on yield loss risk in premium rating. Although there is little hope of accurately rating premiums for individual farms, it may be possible to reduce overall premium rating error on a broader scale (e.g., county or region), possibly reducing overall underwriting losses.

Chapter 2 provides a foundation for Chapters 3 and 4 by exploring crop insurance yield history data. I develop a model that identifies systematic variation in crop yields across producers and across time. I use the model to isolate farm-specific random yield variations and test the fit of several common distributions. Based on the results of this analysis, I develop a yield simulation model that underlies the simulation analysis in Chapter 4.

In Chapter 3, I build on the approach suggested by Chiappori and Salanie to (1) identify and quantify the effects of asymmetric information on the choice of crop insurance coverage level and returns to crop insurance purchase and (2) compare and contrast these asymmetric information effects on areas with relatively high yield loss risk (wheat in the Northern Plains) and relatively low yield risk (corn in the Corn Belt). While regional differences in the actuarial performance have been identified as an issue for the crop insurance program (Glauber), most existing research is limited in scope. For example, Just *et al*. use survey data on corn and soybean producers. Makki and Somwaru used data on corn and soybeans in Iowa. Vandeveer and Lohman studied corn producers in a single county in Indiana.

In Chapter 4, I explore the potential of alternate methods for rating yield loss risk. While it is important to identify and quantify the effects of asymmetric information, the real question is whether the information asymmetry can be reduced. In particular, can premium rating be improved using only the yield histories already maintained by RMA? Existing evidence appears to be mixed. On one hand, premium rating problems are pervasive enough to have been identified through empirical analysis which, in some cases, is based on crop insurance contract and yield history data (Makki and Somwaru). Persistent underwriting losses in some areas and the need for high premium subsidies to encourage broad participation also tend to support the notion that premium rating can be improved to reduce underwriting losses. On the other hand, crop insurance yield histories – which include, at most, 10 observations – are not sufficient to establish field-specific, actuarially fair premiums with a reasonable level of statistical confidence. Collecting additional data, moreover, is unlikely to solve the problem. Because 10 years already accounts for 20-25 percent of a farmer's working lifetime, collecting a longer time-series on individual fields would be impractical.

I develop a simulation model to test the hypothesis that using farm-specific yield histories to rate premiums – despite the uncertainties associated with small samples – could, on average, reduce overall deviation of premium rates from actuarial fair rates. "On average" means that premiums would be closer to actuarially fair for many or most farms while acknowledging that data limitations make accurate farm-specific premium rating impossible. To sort out the actuarial gains and losses from use of a farm-specific measure of yield variability, I focus on two sources of premium rating error. Premiums rated without the benefit of farm-specific information on yield variability may be in error because the rating model is only an approximation to the theoretically correct model. I refer to this type of error as "modeling error." On the other hand, premiums based on a theoretically correct model that includes a measure of farm-specific yield variability are subject to error because of sampling error in yield variability measures calculated from small samples. My simulation model, based on actual crop insurance yield history data, provides estimates of the relative size of modeling error in RMA premium rates and sampling error in alternative premium rates developed from a theoretically correct model.

Finally, I use the crop insurance yield data and the yield models to estimate the effect of aggregation on the representation of both systematic and random yield variation.

The effect of aggregation is important because a great deal of readily available agricultural data are aggregated to the county level. Disaggregated data, even when they do exist, can be difficult to obtain. For example, agricultural census data aggregated to the county level are readily available while use of underlying farm-specific data are restricted due to producer confidentiality concerns. Estimates of yield variation lost to aggregation of yield data from crop insurance units to the farm, county, and regional level may provide insight on the loss of variability when working with aggregated yield data from other sources.

Chapter 2: Crop Insurance Yields: Heterogeneity and Distribution

Representing heterogeneity in agricultural economics research and accounting for that heterogeneity in policy development are ongoing challenges. The importance of heterogeneity and problems associated with spatial and temporal aggregation have been addressed by numerous authors (e.g., Gardner and Kramer (1986), Just and Pope (1999)). In many agricultural settings, a key component of overall heterogeneity is the variability of crop yields. Both spatial and temporal dimensions of yield variability may be important in determining land use, cropping patterns, farm program participation, and crop insurance purchase decisions. The debate over yield distributions has spawned a large literature, including a number of articles that focus on methods of separating the systematic and random components of crop yields (e.g., see Just and Weninger (1999); Atwood et. al. (2002)).

In this Chapter, I develop crop yield models for non-irrigated corn in the Corn Belt, where yield risk is relatively low, and for continuously cropped, non-irrigated wheat in the Northern Plains, where yield risk is considerably higher. Using regression techniques, I develop fixed-effect models, possibly with time trends, designed to separate yield variation into systematic and random components. Model outputs are used to analyze the effect of aggregation on the representation of both systematic and random components of yield variation. I also test farm-specific random yield variation against three common distributions: normal, beta, and gamma (two versions of the gamma distribution are used); each of which has been advanced in a previous study as a candidate for describing random variation in crop yields. In Chapter 4, the yield models are used to simulate the potential for improving crop insurance premium rating using farm-specific information on yield variability.

Data

Yield data are obtained from Risk Management Agency (RMA) crop insurance yield histories for 2001, which include (at most) yield data for 1991-2000. Yield history data are for crop insurance “units.” A crop insurance unit can encompass a single field or an entire farm depending on the physical location of fields, their ownership, and the preferences of individual producers. Many farms include more than one crop insurance unit. The crop insurance yield histories reflect average annual yields for acreage planted to a given crop within the crop insurance unit. If fewer than 4 actual yield observations are available, local “transitional” yields are included to ensure that 4 yield “observations” are present in each yield history. For the purpose of this study, yield observations were considered valid only if yields and acreages are reported as "actual” (i.e., transitional yields are excluded).

For corn in the Corn Belt, data from 69 central and northern Illinois counties were selected. The study area includes counties that lie within Major Land Resource Areas (MLRA; USDA-SCS) 108, 111, and 115. For wheat in the Northern Plains, yield histories from 36 counties in North Dakota and South Dakota are included. These counties are part of MLRA 53A, 53B, 53C, 54, 55A, 55B, 55C, or 56.

Because of premium subsidy increases in 1995 and again in 1999-2000, the number of farms purchasing crop insurance has increased over time. Some producers – perhaps those with relatively low expected return to crop insurance or relatively low risk premiums – began purchasing crop insurance after one or both of the subsidy increases. Because the number of acres covered by crop insurance increased substantially between 1991 and 2000, many crop insurance units have yield histories of less than 10 years. As such, the data offers both challenges and opportunities. If farmers who insured only after 1995 differ from those with a longer crop insurance history, time trend parameters may reflect differences in "old" and "new" crop insurance participants rather than yield changes due to changes in production technology. On the other hand, if these more recent participants have lower risk of yield loss, the subsidy increase and subsequent increase in participation may hold evidence of asymmetric information. To inform the analyses planned for Chapters 3 and 4, I specify models for two datasets: One includes crop insurance units with 8 or more observations while the other includes units with 4 or more observations from 1995 or later. In the latter, pre-1995 data are excluded to ensure that a particularly good or bad year is not included in the data for some farms but not others (in the Corn Belt, for example, 1993 was a bad year because of extensive and prolonged flooding).

Before proceeding, I consider the possibility of outliers. There is no reason to believe that a low (or even zero) yield is an unreasonable outcome. During the period over which the data were collected, natural events (such as flooding and drought) resulted in widespread crop failure. Yields that are much higher than the large majority of yields within the same county in a given year, however, may be due to either particularly unusual conditions or erroneous reporting. In the Corn Belt, for example, some annual yield observations exceed 300 bushels per acre.

To identify observations that could be outliers, I used a standard "box plot" definition. An observation is identified as a potential outlier when $y_{ijkt} > q_{kt,75} + 1.5(q_{kt,75} - q_{kt,25})$, where y_{ijkt} is the yield for unit *i*, on farm *j*, in county *k*, at time *t* and $q_{kt,75}$ and $q_{kt,25}$ are the 75th and 25th percentiles, respectively, for county *k* at time *t*. In the time-series (with 8-10 observations per crop insurance unit), for example, this procedure identified 1,431 observations as potential outliers for corn in the Corn Belt and 1,258 for wheat in the Northern Plains. Because so many observations were identified, individual inspection was not possible. Because the number was small relative to the total size of the dataset, however, I elected to remove them all. When these observations were removed, some crop yield histories fell below the minimum length (4 or 8 years, depending on the dataset) and were entirely removed. For the long time-series of corn yields in the Corn Belt, I removed a total 3,215 observations, or about 0.6 percent (Table 2.1). In the Northern Plains, 4,233 wheat yields were removed from the long time-series data, about 1.7 percent. A smaller number of observations were removed from the short time-series datasets (with 4-6 observations per crop insurance unit). Descriptive statistics on the final datasets are given in Table 2.2.

Table 2.1. Number of Observations, Farms, and Crop Insurance Units by Region and Crop

	Observations	%Obs.	Farms	%Farms	Units	%Units	Outliers Removed
			Corn Belt Corn				
All Observations	729,699	100	45,758	100	101,474	100	1,431
Long Time-Series (8-10 obs.)	531,279	73	29,350	64	55,982	55	3,215
Short Time Series (4-6 obs.)	454,517	62	42,500	93	86,833	86	1,251
			Northern Plains Wheat				
All Observations	489,037	100	19,461	100	86,247	100	1,080
Long Time-Series (8-10 obs.)	244,481	50	10,154	52	26,984	31	4,233
Short Time-Series (4-6 obs.)	256,863	53	15,921	82	62,061	72	1,282

Table 2.2. Descriptive Statistics

	Observations	Mean	Standard Deviation	Minimum	Maximum
		Corn Belt Corn			
Long Time-Series (8-10 obs.)	528,064	144.6	33.1	0	245
Short Time-Series (4-6 obs.)	453,266	142.5	31.0	0	245
		Northern Plains Wheat			
Long Time-Series (8-10 obs.)	240,248	35.4	13.9	0	91
Short Time-Series (4-6 obs.)	255,581	34.9	13.8	0	89

Initial Yield Models

Yield models are intended to capture systematic variation in yields that could be reasonably anticipated by a government agency attempting to isolate the random variation in the error term. The appropriate method of identifying these systematic components, however, has been controversial. Just and Weninger (1999) emphasize econometric estimation of a flexible polynomial time trend that captures field-specific, possibly non-linear, trends. Atwood *et al.* present Monte Carlo evidence suggesting the procedure used by Just and Weninger tends to bias tests in a type II direction (failing to reject the null hypothesis when it is not true) because residuals from averaging or regression models tend to be "supernormal" or "normalized." They advocate using annual average yields within an area (e.g., region or county) to define the systematic component of crop yields, specifically rejecting the use of time trends, but acknowledge that the presence of a time trend in yields would invalidate their procedure.

In this study, I take a somewhat different approach. Like Just and Weninger, my initial model (the starting point for estimation) includes an econometrically estimated time trend albeit a far simpler one than employed by Just and Weninger. Linear and non-linear time terms are included but only at the region level to account for widespread changes in technology that tend to make higher yields more profitable. I do not use unit-, farm-, and county-specific time trends because they can be strongly influenced by particularly low or high yields at the beginning or end of a short time-series.

I do not use the Atwood *et al*. approach because the *ex post* use of annual county average yields assumes that government agencies or others can accurately predict county-average yields. The use of county averages to account for year-to-year variation implicitly assigns county-level random variation to the deterministic portion of the model.[2] The use of a time trend acknowledges that county yields cannot be accurately predicted and should be considered as part of random yield variation.

Initial models provide a starting point for model specification. The initial models are selected to be general enough to capture all potentially reasonable effects. The initial mean model is specified as:

(2.1) $y_{ijkt} = \beta_0^m + \beta_t^m t + \beta_f^m f(t) + \beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i + \varepsilon_{ijkt}$

where:

y_{ijkt} is the yield for unit *i*, on farm *j*, in county *k*, at time *t*;

$\beta_0^m, \beta_t^m, \beta_f^m, \beta_k^m, \beta_j^m$, and β_i^m are parameters (superscript *m* indicates mean model);

$f(t)$ is a non-linear function of time (defined below);

d_k =1 for county *k*, =0 otherwise;

d_j =1 for farm *j*, =0 otherwise;

d_i =1 for unit *i*, =0 otherwise; and,

ε_{ijkt} is an idiosyncratic error term.

Note that subscripts *i*, *j*, and *k* refer to unit-, farm-, and county-specific effects respectively. When only one subscript appears, the variable includes only variation specific to that level of the model. When the subscripts *ijk* appear together, the variable represents variation from all three levels.

[2] This approach is appropriate to the purposes of the Atwood *et al*. study as they are interested primarily in the distribution of purely random, farm-level errors. Later in this Chapter, before testing random yield

Assuming no serial correlation, squared residuals from equation (1) are the dependent variables in the variance model. The initial variance model is:

$$(2.2) \quad v_{ijkt} = \exp\left(\beta_0^v + \beta_t^v t + \beta_k^v d_k + \beta_j^v d_j + \beta_i^v d_i + \upsilon_{ijkt}\right)$$

Where:

$v_{ijkt} = \dfrac{\hat{\varepsilon}_{ijkt}^2}{\hat{y}_{ijkt}^2}$;

$\hat{\varepsilon}_{ijkt}$ is the residual from the mean model;

$\hat{y}_{ijkt}$ is the predicted yield from the mean equation;

$\beta_0^v, \beta_t^v, \beta_k^v, \beta_j^v$, and β_i^v are parameters (superscript *v* indicates variance model); and

υ_{ijkt} is an idiosyncratic error term.

Hypothesis Testing

I use statistical tests to determine whether parameters representing individual mean and variance model effects are significantly different from zero. Because model effects often involve multiple parameters, tests are based on F-statistics. The spatial effect of crop insurance units, for example, would include a dummy variable for each crop insurance unit in the sample. Formally, the null hypothesis in the unit-effect test is $H_0 : \beta_i^m = 0 \; \forall i \in I$ against the alternative $H_A : \beta_i^m \neq 0$ *for at least one* $i \in I$ where *I* is the set of all crop insurance units in the sample.

F-statistics are calculated from the model and residual sum of squares for unrestricted and restricted regressions. In the unit-effect example, the unrestricted model

---

errors against various parametric distributions, I use a similar adjustment to remove correlation among farms.

is estimated without restricting the unit-specific parameters while the restricted model is estimated with these parameters restricted to equal zero. Note that equations 2.1 and 2.2 are the most general models that I consider but are not necessarily the "unrestricted model" in a specific hypothesis test. As hypothesis testing proceeds, the unrestricted model in a specific test may contain parameter restrictions relative to equations 2.1 and 2.2.

Statistical tests are used to specify the final mean and variance models. For the time specification, I test the significance of several non-linear specifications (i.e., forms for the function, $f(t)$), as well as a linear time term. For spatial effects, nested tests of the significance of unit, farm, and county effects are used. The specific sequence of tests and interpretation of results are discussed in the next section. Details of specific mean model hypothesis tests are given in Appendix 2.1. Mean model effect tests are carried out on heteroskedasticity-corrected models. Thus, testing of variance model effects and variance model specification decisions are nested within the mean model tests. Details of specific variance model hypothesis tests are given in Appendix 2.2.

Calculation of the F-statistic for a specific parameter restriction in the mean model, therefore, involves several steps:

- The unrestricted mean model without heteroskedasticity correction is estimated to obtain predicted values and residuals that are used to specify the dependent variable in the variance model;
- The specification of independent variables in the variance model is determined by decision criteria based on a sequence of F-tests;

- Restricted and unrestricted mean models with heteroskedasticity corrections are estimated and the F-statistic is calculated from the model and residual sum of squares.

Because the dataset is very large, however, pre-set limitations in statistical software preclude the estimation of a single regression equation that includes all observations. The GLM procedure in SAS accommodates up to 32,000 fixed effects. The regional models easily exceed this preset limit when county, farm, unit, and time effects are included. In the Corn Belt, for example, there are more than 55,000 crop insurance units and more than 29,000 farms. Thus, a single F-statistic for a region-wide hypothesis (e.g., all unit-effect parameters are equal to zero) cannot be calculated. To accommodate the data, I divided them into 100 groups by placing each individual farm (with all associated crop insurance units) into one of the groups at random. Thus, the test procedure already described produces 100 F-statistics and associated *p*-values for each hypothesis test. To reach a well-defined conclusion, group-specific tests must be combined to form a single hypothesis test. Even if the null hypothesis is true, odds are that some of the 100 test statistics will lay in the critical region defined with a given significance level. Given that the null hypothesis is true, for example, one would expect roughly 5 percent of the data groups to produce test statistics with *p*-values between 0 and 0.05. If the null hypothesis is rejected when any 1 of the 100 *p*-values is in the range 0 to 0.05, then the size of the test, i.e., the probability of rejecting the null hypothesis when it is true, is $1-0.95^{100}$, which is much greater than 0.05.

A typical solution is to reduce the size of the critical region and reject the null hypothesis when any one test statistic falls in this reduced-size critical region (Hsu). A simple and well-known approach is to reduce the size of the critical region for any 1 test statistic to $\alpha^* = 1-(1-\alpha)^{1/N_{test}}$ where N_{test} is the number of tests. For $\alpha = 0.05$ and n=100, the size of the reduced critical region would be $\alpha^* = 0.00051$. A problem with this approach in the case of survey data is that the test is highly vulnerable to the presence of a few anomalous outliers. Correcting the size of the test in this way leads to rejecting the null hypothesis when only 1 of the 100 F-statistics falls in the critical region. As a result, if there are anomalies unique to a single group, then rejection may occur because of otherwise undetected errors in the data. Any test procedure in which results can be driven by only a small fraction of the data is potentially vulnerable to this testing bias.

An alternate solution is to develop a procedure that uses all 100 of the test statistics (and all of the data) rather than determining the test statistic based on the most extreme value (in this case the most extreme one percent of the data). One option is to use the average p-value (which I denote as $\overline{pval}$) as an overall test statistic. Each p-value is the result of a transformation that maps the F-statistic (or any test statistic) to a uniform distribution bounded by 0 and 1. An F-statistic that maps into a p-value of 0.05 indicates that the probability of observing an F-static at least as large is 0.05, given that the null hypothesis is true. Assuming that the individual p-values are independently and identically distributed, the distribution of $\overline{pval}$ can be approximated via the Central Limit Theorem (the Lindberg-Levy variant for a univariate distribution, see Greene p. 122).

Given that the underlying population of 100 *p*-values has a standard uniform distribution (e.g., a uniform distribution bounded by 0 and 1), the underlying population mean and variance are u_{pval} = 0.5 and σ^2_{pval} = 0.0833, respectively. Thus,

$$\frac{\overline{pval} - u_{pval}}{\sqrt{\sigma^2_{pval} / n}} = \frac{\overline{pval} - 0.5}{\sqrt{0.0833/100}} = \frac{\overline{pval} - 0.5}{0.02886}$$

is approximately normally distributed. The degree of approximation is highly accurate with 100 observations as can be verified by Monte Carlo methods. For the standard normal distribution, $\Phi^{-1}(0.05) = -1.645$, where Φ is the standard normal CDF. To check the approximation, 100,000 sets of 100 values where drawn from the standard uniform distribution and the average was calculated for each set. The 5th percentile of these 100,000 average values is -1.643.

Hypothesis testing proceeds as follows: A null hypothesis is rejected when $\overline{pval}$ falls into the critical region, defined as the lower 5 percent of the probability mass of $\overline{pval}$. A single-tailed test using the lower tail is appropriate because low *p*-values, by definition, indicate extreme values of the underlying test statistic (an F-statistic in this case), even for test statistics that could produce extreme values at either end of their distribution (e.g., a t-statistic can produce extreme values that are negative or positive depending on the sign of the parameter estimate). Formally, a null hypothesis is rejected when $\Phi\left(\frac{\overline{pval} - u_{pval}}{\sqrt{\sigma^2_{pval} / n}}\right) < \alpha$ where Φ is the standard normal CDF and α = 0.05.

Alternately, the critical value of $\overline{pval}$ associated with any given α is $\left(\sigma_{pval} \big/ \sqrt{n}\right)\Phi^{-1}(\alpha) + u_{pval}$ and the hypothesis is rejected when $\overline{pval} < \left(\sigma_{pval} \big/ \sqrt{n}\right)\Phi^{-1}(\alpha) + u_{pval}$. For $\alpha = 0.05$, the critical value of $\overline{pval}$ is equal to $0.02886\Phi^{-1}(0.05) + 0.5 = .4525$. This critical value is used throughout the next section to determine the statistical significance of specific model effects.

Estimation Results

Table 2.3 gives selected estimation results for the Corn Belt corn model using only crop insurance units with 8 or more yield observations. In the top portion of the table (*Time Effect Specification Test Results*), each numbered row represents a test related to the model time effect. Each of the four models reported in the top section of Table 2.3 includes spatial effects for county, farm, and crop insurance unit. The bottom portion (*Spatial Effect Specification Test Results*) reports test results for sequential tests of model spatial effects (unit, farm, and county), given the time specification selected base on statistical and other evidence. For all seven tests, the average marginal effect of time and the 1st, 10th, 50th, and 90th, and 99th percentiles of the marginal effect (in the unrestricted model for the test) are also shown.

Table 2.3. Specification Test Results for Corn Belt Corn--Long Time Series

Time Effect Specification Test Results

Test	Unrestricted Mean Model Time Terms*	Variance Model Test Results: Time Trend	Unit Effect	Farm Effect	County Effect	Mean Model Test Results: Parameter Restriction	Average p-value	Marginal Effect of Time (Unrestricted Model): Mean	Percentiles: 1	10	25	50	75	90	99
		Average p-values													
1	$\beta_t^m t + \beta_f^m t^2$	<0.0001	0.6173	<0.0001	<0.0001	$\beta_f^m = 0$	0.0024	-0.34	-2.60	-1.62	-0.91	-0.22	0.35	0.71	1.16
2	$\beta_t^m t + \beta_f^m \log(t)$	<0.0001	0.6267	<0.0001	<0.0001	$\beta_f^m = 0$	0.2227	1.47	0.27	0.75	0.98	1.25	1.62	2.53	4.99
3	$\beta_t^m t + \beta_f^m \log(\log(t))$	<0.0001	0.6351	<0.0001	<0.0001	$\beta_f^m = 0$	0.0005	2.10	-0.29	0.15	0.41	0.92	2.29	5.12	14.06
4	$\beta_t^m t$	<0.0001	0.6344	<0.0001	<0.0001	$\beta_t^m = 0$	0.0000	1.31	0.99	1.09	1.16	1.29	1.40	1.57	1.70

*All specifications include spatial effects for the unit, farm, and county.

Spatial Effect Specification Test Results

Test	Unrestricted Mean Model Spatial Terms*	Variance Model Test Results: Time Trend	Unit Effect	Farm Effect	County Effect	Mean Model Test Results: Parameter Restriction	Average p-value	Marginal Effect of Time (Unrestricted Model): Mean	Percentiles: 1	10	25	50	75	90	99
		Average p-values													
5	$\beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i$	<0.0001	0.6266	0.0001	<0.0001	$\beta_i^m = 0$	0.0119	1.31	0.99	1.09	1.16	1.29	1.40	1.57	1.70
6	$\beta_k^m d_k + \beta_j^m d_j$	<0.0001	0.9834	<0.0001	<0.0001	$\beta_j^m = 0$	<0.0001	1.31	0.99	1.10	1.20	1.30	1.39	1.56	1.71
7	$\beta_k^m d_k$	<0.0001	0.9294	0.0063	<0.0001	$\beta_k^m = 0$	<0.0001	1.32	0.98	1.09	1.20	1.31	1.42	1.58	1.76

*All specifications include a linear time term.

For corn in the Corn Belt, consider mean model test results, beginning with time effects. Each of the non-linear time terms tested is statistically significant at the 5 percent level. The average *p*-values associated with the relevant F-tests are less than 0.4525 for Tests 1-4 so the null hypothesis that associated parameters are equal to zero can be rejected. The sign and magnitude of the marginal effect of time in these models, however, is problematic. With the time-squared term in the model, for example, the marginal effect of time is negative for more than half of all observations. Because crop yields have generally been rising over time, the presence of a negative marginal effect likely indicates the presence of particularly large yields early in the time-series or particularly low yields late in the time-series. Because the time-series is short, these observations can exert a substantial influence.[3]

When the logarithm of time is included (in place of the time-squared term), the magnitude of the marginal effect varies from 0.27 to about 5 bushels per acre per year (bu/a/yr), although 90 percent of the marginal effects are 2.5 bu/a/yr or less. A marginal effect of 2.5 bu/a/yr implies that expected yield would increase 25 bushel per acre over a 10 year period. Given rapid advances in corn seed technology, such rapid yield increases seem possible but still dramatic. With the log-log term in the model, the range of time marginal effects is even larger, ranging from less than zero to more than 14 bu/a/yr. Ten percent of observations have marginal effect of 5 bu/a/yr, implying a ten-year increase in expected corn yields of 50 bushels per acre. The average marginal effect in the log-log case is more than 2 bu/a/yr.

[3] A change in economic conditions (e.g., lower input prices or higher input prices) could prompt producers to plan lower yields by using fewer inputs. Technical change, however, has generally caused increasing yields over the past 50 years, even as the real prices of agricultural commodities have declined.

The linear time term, absent non-linear time terms, is also significant at the 5 percent level (the null hypothesis that the time term parameter is equal to zero can be rejected as the average *p*-value of 0.0064 is less than 0.4525). With a linear time term only, the marginal effect of time appears to be more realistic, ranging between 1 and 1.7 bu/a/yr, with an average value of about 1.3. I adopt this specification for the time component and use it in testing spatial effects.

Results of spatial effect tests for the Corn Belt corn model are reported in the bottom section of Table 2.3. For the mean model, all three spatial effects (unit, farm, and county) are retained. Based on average *p*-values reported in the bottom section of Table 2.3, the null hypotheses that unit, farm, and county-effect parameters are equal to zero, respectively, are rejected. Average *p*-values are well below the critical value of 0.4525 in all three tests.

Finally, the variance model specification is the one implied by the variance model test results for mean model Test 5. Given the critical value of 0.4525, the null hypothesis that parameter estimates are equal to zero is rejected for the time trend, farm, and county effects while the null hypothesis that the unit effect is zero is not rejected in any of the models. Because I have no prior expectation about the sign or size of any of these effects, these test results were used to select the variance model. Thus, the variance model includes a time trend, farm, and county effect for each of the unrestricted models.

Table 2.3B. Specification Test Results for Northern Plains Wheat--Long Time Series

Time Effect Specification Test Results

		Variance Model Test Results				Mean Model Test Results		Marginal Effect of Time (Unrestricted Model)								
		Time Trend	Unit Effect	Farm Effect	County Effect				Percentiles							
Test	Unrestricted Mean Model Time Terms*	Average *p* -values				Parameter Restriction	Average *p* -value	Mean	1	10	25	50	75	90	99
1	$\beta_t^m t + \beta_f^m t^2$	0.24081	0.49658	0.00037	2E-112	$\beta_f^m = 0$	3.2598E-10	-2.6677	-5	-4.1	-3.5	-2.6	-1.8	-1.3	-0.8
2	$\beta_t^m t + \beta_f^m \log(t)$	0.25204	0.523	0.00019	4E-111	$\beta_f^m = 0$	1.4765E-08	-0.9268	-7.8	-4.2	-1.8	-0.1	0.69	1.14	1.78
3	$\beta_t^m t + \beta_f^m \log(\log(t))$	0.26714	0.58339	0.00013	1E-103	$\beta_f^m = 0$	1.8469E-06	-1.1329	-11	-4.1	-1.3	-0.1	0.36	0.65	1.12
4	$\beta_t^m t$	0.26393	0.60429	7.6E-05	6E-106	$\beta_t^m = 0$	0.00644367	-0.4223	-0.7	-0.6	-0.5	-0.4	-0.3	-0.3	-0.1

*All specifications include spatial effects for the unit, farm, and county.

Spatial Effect Specification Test Results

		Variance Model Specification Tests				Mean Model Test Results		Marginal Effect of Time (Unrestricted Model)								
		Time Trend	Unit Effect	Farm Effect	County Effect				Percentiles							
Test	Unrestricted Mean Model Spatial Terms*	Average *p* -values				Parameter Restriction	Average *p* -value	Mean	1	10	25	50	75	90	99
5	$\beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i$	0.24245	0.34429	0.02996	4.1E-28	$\beta_i^m = 0$	0.9668	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	$\beta_k^m d_k + \beta_j^m d_j$	0.21261	0.98442	9.3E-07	1.7E-55	$\beta_j^m = 0$	0.0000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	$\beta_k^m d_k$	0.33704	0.98663	0.00094	4E-108	$\beta_k^m = 0$	0.0000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

*Specifications do not include time terms.

For wheat in the Northern Plains, specification test results for the long time-series are given in Table 2.4. The set of tests is the same as in the Corn Belt corn model. The mean model tests for the non-linear time terms (Tests 1-3 in Table 2.4) show that each term is significant at (at least) the 5 percent level on the basis of average *p*-values. The marginal effect of time, however, would be negative for more than one-half of observation for each of the three unrestricted models (Tests 1-3). Test results for the linear time term (Test 4 in Table 2.4) are similar. While the average *p*-value associated with restricting the linear time-term parameter to zero is well under the 5 percent critical value, the marginal effect of time is negative for more than one-half of all observations. Given these results, time effects are excluded from the mean model.

Spatial effect specification test results, reported at the bottom of Table 2.4, are based on models that do not include time effects. The average *p*-values show that the unit effect parameters are not significantly different from zero and the unit effect is deleted from the model. The farm and county effects are significant and are retained in the model.

Finally the short time-series models are specified using the long time-series time effects as a starting point. For Corn Belt corn the initial time component is a simple, linear time trend. For Northern Plains wheat time effects are excluded from the initial model for the short time-series. Specification test results are shown in Tables 2.5 (Corn Belt corn) and 2.6 (Northern Plains wheat).

For the Corn Belt, the average *p*-value is low enough to reject the null hypothesis that the time trend variable parameter is equal to zero, but the marginal effect of time is very large for this specification (Test Model 1 in Table 2.5), implying an average increase in corn yields of 5.6 bushels per year. Over a 10-year period, that would imply corn yield increases of more than 56 bushels per acre – an ultimately unbelievable rise in corn yields. Because of the shortness of the time-series, the time parameter will be quite sensitive to high or low yields at the beginning or end of the time-series. I conclude that the data are not sufficient to estimate a time trend.

Spatially, the unit, farm, and county effects are all retained as the null hypothesis that all parameters associated with a given effect are all equal to zero, can be rejected on the basis of average p-values for both the Corn Belt corn model (Table 2.5, Test Models 2, 3, and 4) and the Northern Plains (Table 2.6, Test Models 2, 3, and 4). In the Northern Plains wheat model, the unit effect is retained in the short time-series model even though it was not retained in the long time-series model. That may be due to broader cross-sectional coverage of crop insurance units in the short time-series model.

Variance model specifications can be inferred from the variance model specification tests associated with Test Model 2 in Table 2.5 for Corn Belt corn and Test Model 1 in Table 2.6 for Northern Plains wheat. In both cases the final variance models include a linear time trend and spatial effects for unit, farm, and county.

Table 2.5. Specification Test Results for Corn Belt Corn--Short Time Series

Time Effect Specification Test Results

		Variance Model Test Results				Mean Model Test Results		Marginal Effect of Time (Unrestricted Model)							
		Time Trend	Unit Effect	Farm Effect	County Effect				Percentiles						
Test	Unrestricted Mean Model Time Terms*	Average p-values				Parameter Restriction	Average p-value	Mean	1	10	25	50	75	90	99
1	$\beta_t^m t$	0.0003	0.0453	0.0008	0.0000	$\beta_t^m = 0$	0.0000	5.668		5.360	5.511	5.667	5.779	6.000	

*All specifications include spatial effects for the unit, farm, and county.

Spatial Effect Specification Test Results

		Variance Model Specification Tests				Mean Model Test Results		Marginal Effect of Time (Unrestricted Model)							
		Time Trend	Unit Effect	Farm Effect	County Effect				Percentiles						
Test	Unrestricted Mean Model Spatial Terms*	Average p-values				Parameter Restriction	Average p-value	Mean	1	10	25	50	75	90	99
2	$\beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i$	0.0853	0.1343	0.0402	0.0178	$\beta_i^m = 0$	0.0029	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	$\beta_k^m d_k + \beta_j^m d_j$	0.1048	0.9009	0.0000	0.0014	$\beta_j^m = 0$	0.0000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	$\beta_k^m d_k$	0.1997	0.8548	0.0016	0.0000	$\beta_k^m = 0$	0.0000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

*Specifications do not include time terms.

Table 2.6. Specification Test Results for Northern Plains Wheat--Short Time Series

Spatial Effect Specification Test Results

Test	Unrestricted Mean Model Spatial Terms*	Variance Model Specification Tests: Time Trend	Unit Effect	Farm Effect	County Effect	Mean Model Test Results: Parameter Restriction	Average p-value	Marginal Effect of Time (Unrestricted Model): Mean	Percentiles: 1	10	25	50	75	90	99
		Average p-values													
2	$\beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i$	0.2437	0.0000	0.0000	0.0000	$\beta_i^m = 0$	0.1548	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	$\beta_k^m d_k + \beta_j^m d_j$	0.0128	0.8125	0.0000	0.0000	$\beta_j^m = 0$	0.0000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	$\beta_k^m d_k$	0.0011	0.9716	0.0000	0.0000	$\beta_k^m = 0$	0.0000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

*Specifications do not include time terms.

The Effect of Aggregation on the Representation of Crop Yield Variation

Much of the agricultural data available for economic research is aggregated to the county level. For example, a wide range of county aggregate data on production, land use, non-land farm input use, farm returns, and producer demographics is readily available to researchers through the agriculture census. The farm-specific data underlying county aggregates, however, can be used only under limited conditions to protect the confidentiality of producers who provide data. Likewise, annual estimates of acreage, crop production, livestock production, and other agricultural indicators are available by county but not at the farm or field level.

The use of county average yields will mask yield variation as idiosyncratic variation at the field- and farm-level is averaged out. To quantify the effect of aggregation on both systematic and random variation in crop yields within a single, consistent framework, I start with a basic yield model:

$$(2.3) \qquad y_{ijkt} = \hat{y}_{ijkt} + \varepsilon_{ijkt}$$

Where

y_{ijkt} is the yield for unit *i*, on farm *j*, in county *k*, at time *t*;

$\hat{y}_{ijkt}$ is the expected yield; and

ε_{ijkt} is the idiosyncratic yield error.

To develop a region-wide measure of yield variability that contains both systematic and random components, I subtract the region-average yield (averaged over both time and space) from both sides of (2.3) and square both sides:

$$(2.4) \qquad (y_{ijkt} - y)^2 = (\hat{y}_{ijkt} - y)^2 + (\varepsilon_{ijkt})^2 + 2(\hat{y}_{ijkt} - y)(\varepsilon_{ijkt})$$

where $y = a^{-1}\sum_i\sum_j\sum_k\sum_t a_{ijkt}\, y_{ijkt}$ and a is total acreage for the crop/region combination.

My measure of total, region-wide yield variation, calculated from the crop insurance unit-level data is:

$$\begin{aligned} a^{-1}&\sum_i\sum_j\sum_k\sum_t a_{ijkt}(y_{ijkt} - y)^2 \\ &= a^{-1}\sum_i\sum_j\sum_k\sum_t a_{ijkt}(\hat{y}_{ijkt} - y)^2 + a^{-1}\sum_i\sum_j\sum_k\sum_t a_{ijkt}(\varepsilon_{ijkt})^2 \\ &\quad + 2a^{-1}\sum_i\sum_j\sum_k\sum_t a_{ijkt}(\hat{y}_{ijkt} - y)(\varepsilon_{ijkt}). \end{aligned}$$

The LHS is an acre-weighted average of the total squared variation of unit-specific yields from a region-wide average yield. The first term on the RHS is a similar measure that represents systematic variation across farms. The second RHS term is a measure of total random variation – yield variance – represented by the crop insurance data. The last term is the interaction between systematic and random yield variation (which should be zero if systematic and random components have been successfully separated).

I consider the effect of three separate aggregations. The expected yield and residual are averaged across (1) crop insurance units within farms, (2) across all farms within a county, and (3) all counties within the region. Total yield variability, systematic variability, and random variability components are calculated at each level of aggregation. For example, averaging across crop insurance units (aggregating to the farm level) gives this estimate of farm-level yield variation:

$$\begin{aligned} a^{-1}&\sum_j\sum_k\sum_t a_{jkt}(y_{jkt} - y)^2 \\ &= a^{-1}\sum_j\sum_k\sum_t a_{jkt}(\hat{y}_{jkt} - y)^2 + a^{-1}\sum_j\sum_k\sum_t a_{jkt}(\varepsilon_{jkt})^2 \\ &\quad + 2a^{-1}\sum_j\sum_k\sum_t a_{jkt}(\hat{y}_{jkt} - y)(\varepsilon_{jkt}) \end{aligned}$$

Where $a_{jkt} = \sum_i a_{ijkt}$, $y_{jkt} = \sum_i a_{ijkt} y_{ijkt}$, and $\varepsilon_{jkt} = \sum_i a_{ijkt} \varepsilon_{ijkt}$. In this example, the yield, expected yield, and yield error are summed to the farm level before calculating the yield variability measure. The variability measures for other spatial aggregations are similarly derived.

To make the variability estimates comparable across regions and crops, I divide through by the squared region-average yield. To make the numbers easier to read, I multiply each by 100. I refer to these measures as "standardized variations." These measures for the Corn Belt corn data and Northern Plains wheat data can be found in Tables 2.7 and 2.8, respectively.

For corn in the Corn Belt, using crop insurance units with 8 or more years of data, 65 percent of total standardized yield variation ((3.13/4.83)*100) is from the random component. When data are aggregated to the farm level, total variation declines by about 9 percent (4.83 to 4.40). Systematic and random components each decline by about the same percentage. When yields are aggregated to the county level total variation declines by 55 percent (from 4.83 to 2.16). About 73 percent of the systematic component and 47 percent of the random component are lost.

Table 2.7. Effect of Aggregation on the Representation of Crop Yield Variation--Corn Belt Corn

Time-Series	Source of Variation	Data Not Aggregated	Data Aggregated to: Farms	Counties	Region
		Standardized Variations			
Long Time-Series (8-10 Obs.)	Total	4.83	4.40	2.16	1.19
	Systematic	1.68	1.52	0.45	0.06
	Random	3.13	2.82	1.66	1.12
	Interaction	0.03	0.05	0.06	0.00
		Number of Observations			
		528,064	306,468	690	10
		Standardized Variations			
Short Time-Series (4-6 Obs.)	Total	4.21	3.63	1.37	0.62
	Systematic	1.89	1.63	0.42	0.00
	Random	2.31	1.94	0.88	0.57
	Interaction	0.01	0.06	0.06	0.05
		Number of Observations			
		453,266	237,679	414	6

Using data for 1995-2001 (the short time-series), total variation declines by about 13 percent (from 4.83 with the longer time-series to 4.21 with the shorter). Systematic variation is larger than with the longer time-series while random variation is sharply decreased. These differences are likely explained by the broadening of the dataset to include a larger set of crop insurance units, which could account for larger systematic variation, and the shorter time-series, which could account for the reduction in random variation. Larger random variation in the long time-series model may reflect the effect of massive Midwest flooding in 1993, which is included in the longer time-series but not the shorter time-series.

When Corn Belt data are aggregated to the farm-level, loss of total yield variation for the short time-series is about 14 percent. Systematic and random variations are reduced by 14 and 16 percent, respectively. When aggregated to the county level, total variation, systematic variation, and random variation are reduced by 68, 78, and 62 percent, respectively. The loss of yield variation is also large for all components when data are aggregated from the county to the region level.

Table 2.8. Effect of Aggregation on the Representation of Crop Yield Variation--Northern Plains Wheat

Time-Series	Source of Variation	Data Not Aggregated	Data Aggregated to: Farms	Counties	Region
		Standardized Variations			
Long Time-Series (8-10 Obs.)	Total	14.56	12.91	6.66	3.34
	Systematic	3.71	3.71	1.55	0.01
	Random	11.22	9.57	5.33	3.42
	Interaction	-0.36	-0.36	-0.22	-0.10
		Number of Observations			
		240,248	95,736	474	10
		Standardized Variations			
Short Time-Series (4-6 Obs.)	Total	14.25	11.73	5.17	1.42
	Systematic	6.95	5.97	2.62	0.04
	Random	7.40	5.36	2.04	1.19
	Interaction	-0.10	0.41	0.51	0.20
		Number of Observations			
		255,581	81,025	306	6

The measures of standardized variation calculated for Northern Plains wheat are roughly three times those for Corn Belt corn (Table 2.7). For example, total variation in the long time-series models are 14.56 and 4.83 for Northern Plains wheat and Corn Belt corn, respectively. Random variation also makes up a larger share of total variation in the Northern Plains. For the long time-series models, random variation is 77 percent of total variation versus 64 percent for the Corn Belt.

Aside from overall larger overall and random variation, however, the effect of aggregation is similar in the Northern Plains (Table 2.8). Because the long time-series model of Northern Plains wheat does not contain a unit effect, aggregation from the unit to the farm level has no effect on the level of systematic variation represented by the model. The unit effect is significant in the short time-series model, however, indicating greater cross-sectional (and systematic) variation in the short time-series. Aggregation to the county- and region-level sharply reduces systematic and random variation in both the long and short time-series model, as in the Corn Belt models.

Normalizing Data and Testing Distributions

The appropriate distribution to use when modeling crop yields is controversial. Many researchers have argued that yields are not normally distributed due, in part, to negative skew that can be introduced by crop failure (Day; Gallagher; Buccola; Moss and Shonkwiler; and Ramirez; among others). Just and Weinegar, however, argue that studies rejecting normality suffer, in part, from methodological errors that could lead to type I error and provide empirical examples to make their point. Ultimately, Just and Weinegar argue that crop yields must be asymptotically normal due to the Central Limit Theorem.[4] As already noted, Atwood *et al.* suggest that the Just-Weninger approach may lead to type II error in tests for yield normality. More recently, in a study of out-of-sample predictions, Norwood et al, 2004, argue that use of an empirical distribution results in more accurate yield predictions when compared to a range of parametric distributions including normal, gamma, and beta. These results, however, are based on county-average yields, indicating that farm-specific idiosyncratic random variation has already been averaged out.

In this section, I test the distribution of farm-specific yield variation against the normal, gamma, and beta distributions. Because results depend on unit-specific

[4] They cite the work of White and Domowitz who argue that the Central Limit Theorem applies broadly to economic time-series, such as prices, as well as cross-sectional data.

parameters, only the longer time-series data are used. Before testing, moreover, variation that is unique to specific farms must be isolated, beginning with the normalized errors based on the mean and variance models already estimated:

(2.3) $$\hat{\varepsilon}^{*}_{ijkt} = (y_{ijkt} - \hat{y}_{ijkt}) / \hat{\sigma}_{ijkt}$$

where $\hat{\sigma}_{ijkt} = \left(\hat{y}_{ijkt}\hat{v}_{ijkt}\right)^{0.5}$ is the variance model estimate of yield standard deviation. These errors represent total random variation in crop yields, including the variation unique to crop insurance unit *i*, variation common to all units in farm *j*, and variation common to all farms in the county and region. Because these errors include these common variations, they are likely to be correlated across units within farms (because of common management) and across farms within counties (due to common weather, similar soil conditions, etc.).

To remove county- and region- level correlation, I calculate a county average for the normalized yield errors (implicitly the county average also includes errors common to the region) and subtract it from the normalized error:

(2.4) $$\hat{\varepsilon}^{*}_{ijt} = \hat{\varepsilon}^{*}_{ijkt} - \hat{\varepsilon}^{*}_{kt}$$

where $\hat{\varepsilon}^{*}_{kt} = I_k^{-1}\sum_{i\in k}\sum_{j\in k}\hat{\varepsilon}^{*}_{ijkt}$ and I_k is the number of crop insurance units in county *k*.

To eliminate the possibility of correlation among units within a farm, a single unit is selected at random from each farm. Analysis in the previous section shows that only a modest portion of variance is lost when aggregating from units to farms, so little is lost. In what follows, the unit-specific subscript is dropped; the selected unit is assumed to represent the farm.

After subtracting annual county-average errors, the resulting farm-specific error may no longer have a unit variance. To ensure unit variance across farm-specific errors, I divide each observation by its unit-specific standard deviation:

$$(2.5) \quad \hat{\varepsilon}^{**}_{jt} = \hat{\varepsilon}^{*}_{jt} / s_j$$

where $s_j^2 = N_j^{-1} \sum_t (\hat{\varepsilon}^{*}_{jt} - \hat{\varepsilon}^{*}_{j})^2$, $\hat{\varepsilon}^{*}_{j} = N_j^{-1} \sum_t \hat{\varepsilon}^{*}_{jt}$, and N_j is the number of yield observations for farm *j*.

To distinguish the normalized error from errors adjusted for other distributions, I denote $\hat{\varepsilon}^{n}_{jt} = \hat{\varepsilon}^{**}_{jt}$. For the non-normal distributions I am testing, normalization to constant mean and standard deviation across farms ensures that scale and shape parameters will also be constant. Using $\hat{\varepsilon}^{**}_{jt}$ as a starting point, normalized errors are relocating and/or rescaled to the range of values supported by each distribution:

- For the beta distribution, all observations must fit into the unit interval. I use the minimum and maximum $\hat{\varepsilon}^{**}_{jt}$ (across farms and time for the entire region) to move observations into the unit interval $\hat{\varepsilon}^{b}_{jt} = \dfrac{\hat{\varepsilon}^{**}_{jt} - \min(\hat{\varepsilon}^{**})}{\max(\hat{\varepsilon}^{**}) - \min(\hat{\varepsilon}^{**})}$.
- For the gamma distribution, observations need only to be translated to be positive: $\hat{\varepsilon}^{g}_{jt} = \hat{\varepsilon}^{**}_{jt} - \min(\hat{\varepsilon}^{**})$, i.e., the threshold parameter is the minimum value of $\hat{\varepsilon}^{**}_{jt}$ across farms and time for the entire region.
- Finally, because skew can only be positive in the gamma distribution, I also use an alternate, gamma-based normalization: $\hat{\varepsilon}^{g-}_{jt} = \max(\hat{\varepsilon}^{**}) - \hat{\varepsilon}^{g}_{jt}$ where $\max(\hat{\varepsilon}^{**})$ is the maximum value of $\hat{\varepsilon}^{**}_{jt}$. I refer to this distribution as gamma-minus.

Using the relocated and rescaled data, distribution parameters are estimated using maximum likelihood methods (estimated using PROC UNIVARIATE in SAS; see Table 2.9).

Table 2.9. Estimated Parameters for Common Parametric Distributions

Distribution	Number of Obs.	Location[1]	Scale[2]	Shape1	Shape2
		Corn Belt Corn			
Normal	303,177	0	0.96	--	--
Beta	303,177	0	1.00	14.9	12.8
Gamma	303,177	0	0.18	31.3	--
Gamma-Minus	303,177	0	0.20	23.9	--
		Northern Plains Wheat			
Normal	91,622	0.0063	0.98	--	--
Beta	91,622	0	1.00	17.8	24.3
Gamma	91,622	0	0.19	29.9	--
Gamma-Minus	91,622	0	0.13	59.0	--

[1]Location is the mean for the normal distribution, threshold parameter for gamma and beta distributions
[2]Standard deviation for normal distribution

Standard non-parametric methods are used to test the fit of the normalized, relocated, and rescaled data against normal, beta, gamma, and gamma-minus distributions. For a general test of the equality of distributions for samples, Hogg and Craig suggest converting data to a contingency table format and applying a chi-squared test. Multi-sample generalizations are straightforward (Hollander and Wolfe; Conover). For large samples, the resulting test statistic has a chi-squared distribution (approximately) with degrees of freedom equal to (r-1)(c-1) where r is the number of rows and c is the number of columns in the contingency table.

For this problem, a 2-row contingency table is set up to compare the observed, normalized, relocated, and rescaled data (placed in row 1 of the contingency table) against a simulated dataset drawn from one of the parametric distributions using parameters described in Table 2.9 (placed in row 2 of the contingency table). For example, the simulated normal distribution for Corn Belt corn is built through repeated Monte Carlo draws from a normal distribution with mean and standard deviation parameters of 0 and 0.96, respectively. The data are divided into 10 columns using criteria that are common to both rows. The range of normalized yield values included in each column is based on percentiles of the observed data. Column one, for example, includes all observations (for both rows) that are equal to or less than the 10th percentile of the observed data. Column two contains values that fall between the 10th and 20th percentiles, and so on. Chi-squared tests are carried out using PROC FREQ in SAS.

I also use a two-sample version of the Kolmogorov-Smirnov (KS) test to compare the distribution of the observed data against distribution-specific simulated datasets. In the KS test for goodness-of-fit, the test statistic is the maximum difference between the empirical distribution functions for the observed and simulated data. Because the KS test is formulated for continuous distributions, there is no need to place the data in arbitrary columns as in the chi-squared test. In large samples, the distribution of the test statistic can be found using an asymptotic approximation. KS tests are carried out using PROC NPAR1WAY in SAS. For details on the test statistic and the approximation of *p*-values see SAS Institute, 2004b.

Goodness-of-fit test statistics are reported in Table 2.10. None of the distributions investigated here are a good fit for the crop insurance yield data. All four distributions are strongly rejected for both chi-squared and KS tests. Based on the value of the test statistics, the gamma-minus distribution comes closest to fitting the data for both Corn Belt corn and Northern Plains wheat. That may be due to negative skew found in the normalized yield data for both Corn Belt corn (skew= -0.18) and Northern Plains wheat (skew = -0.20). Of the four distributions investigated, only the gamma-minus and beta distributions can accommodate negative skew.

Table 2.10. Results of Chi-Squared and Kolmogorov-Smirnov Tests

Distribution	Number of Observations	Chi-Squared Tests		KS Tests	
		Statistic	*p*-value	Statistic	*p*-value
		Corn Belt Corn			
Normal	303,177	888.5	<.00001	7.88	<.00001
Beta	303,177	704.8	<.00001	7.15	<.00001
Gamma	303,177	3,707.0	<.00001	17.53	<.00001
Gamma-Minus	303,177	251.6	<.00001	4.77	<.00001
		Northern Plains Wheat			
Normal	91,622	657.9	<.00001	7.30	<.00001
Beta	91,622	928.0	<.00001	8.47	<.00001
Gamma	91,622	1,825.7	<.00001	12.18	<.00001
Gamma-Minus	91,622	226.2	<.00001	3.17	<.00001

To provide additional insight, the normalized, relocated, and rescaled yield data are presented graphically. Histograms showing the Corn Belt data normalized for the four candidate distributions along with the theoretical curves for the distribution are shown in Figures 2.1-2.4. Figures 2.5-2.8 are for Northern Plains Wheat. The graphics confirm that the gamma-minus distribution appears to be the best fit for the data. Gallagher also used a gamma distribution, set up in a similar way, to model soybean yields. Gallagher argued that crop yields are likely to exhibit negative skew because maximum yields are limited by the genetic potential of the crop but could go as low as zero under extreme weather conditions. In the lower tail of each distribution (what appears to be the upper tail in gamma-minus) the parametric distributions appear to have less weight than the empirical distribution. The opposite is true of the upper tail, where the parametric distributions generally place greater weight. Closer to the mean, the parametric distributions appear to be slightly to the left of the empirical distribution. Both features are less pronounced in the gamma-minus distribution than in other the other three parametric options investigated.

Figure 2.1. Corn Belt Corn Yields and the Normal Density Curve

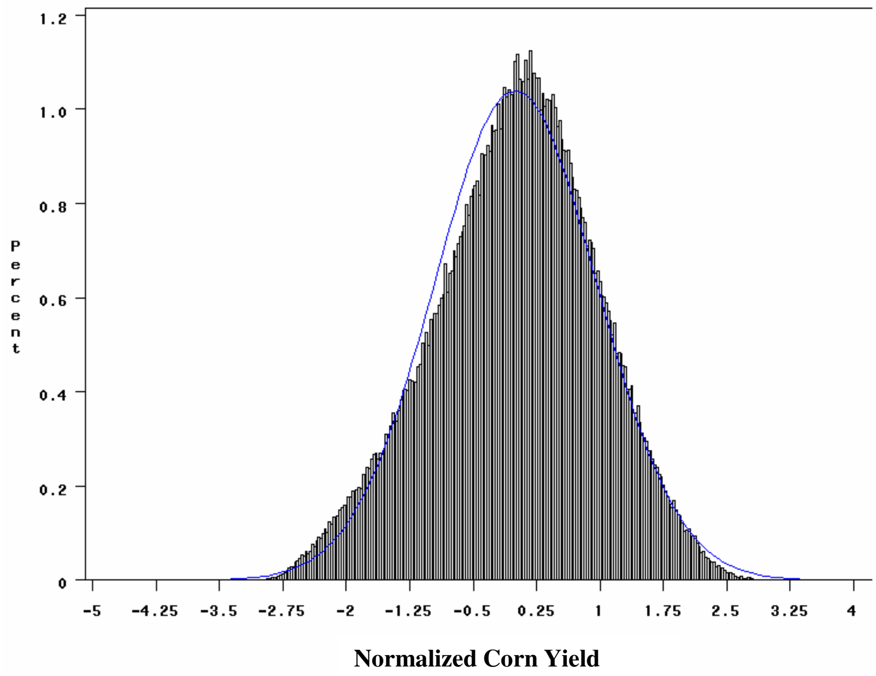
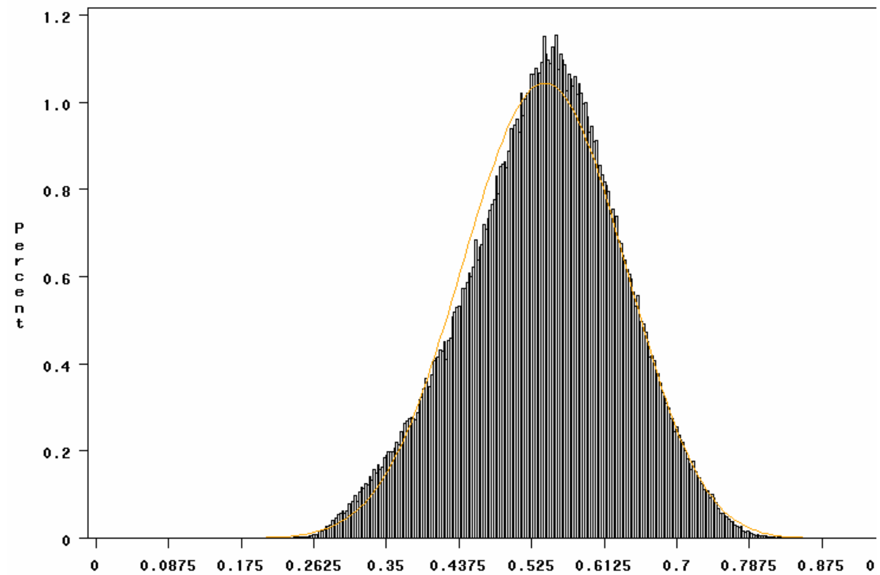
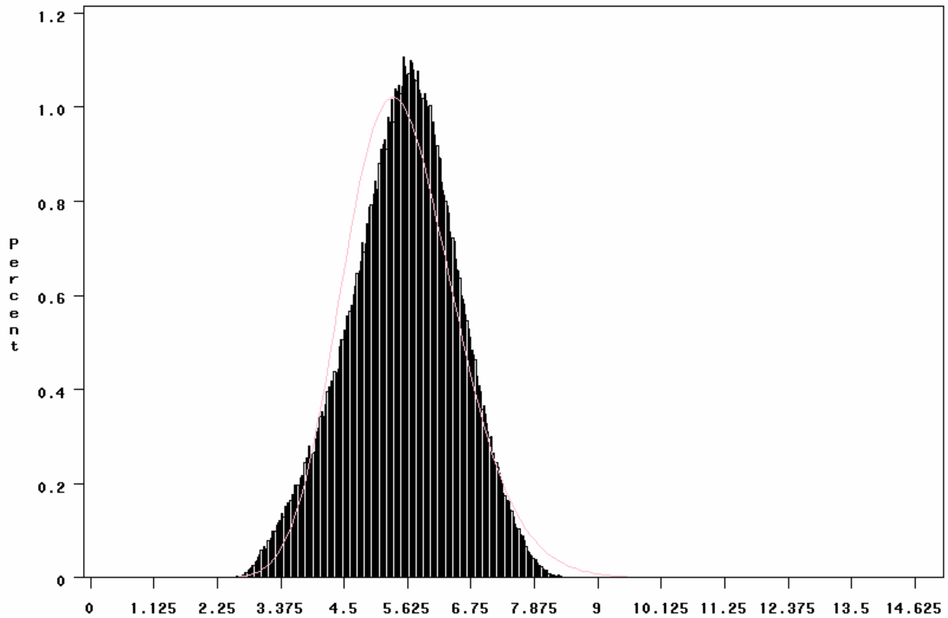


Figure 2.2. Corn Belt Corn Yields and the Beta Density Curve



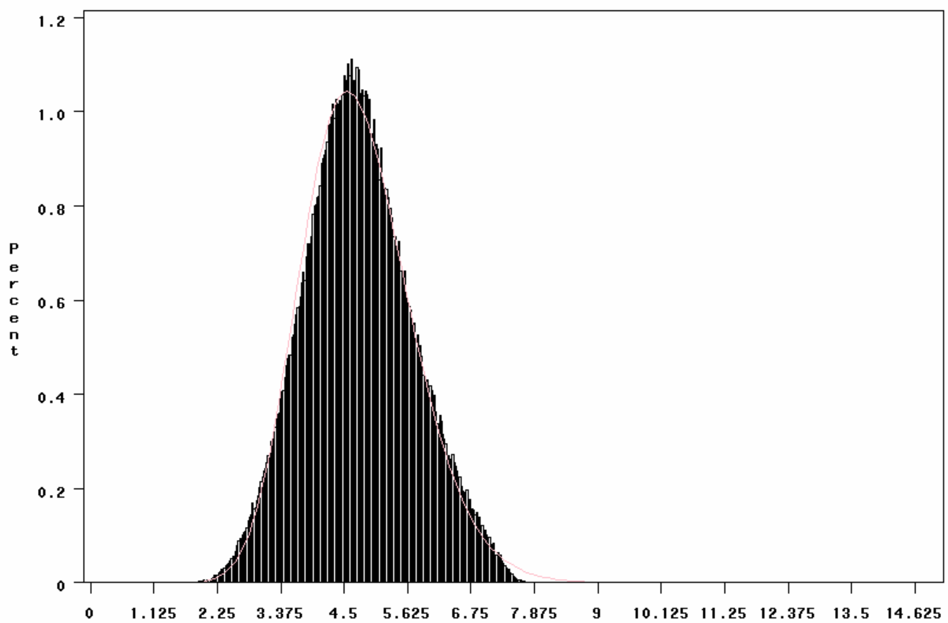
Normalized Corn Yield (Relocated and Rescaled for the Beta Distribution)

Figure 2.3. Corn Belt Corn Yields and the Gamma Density Curve



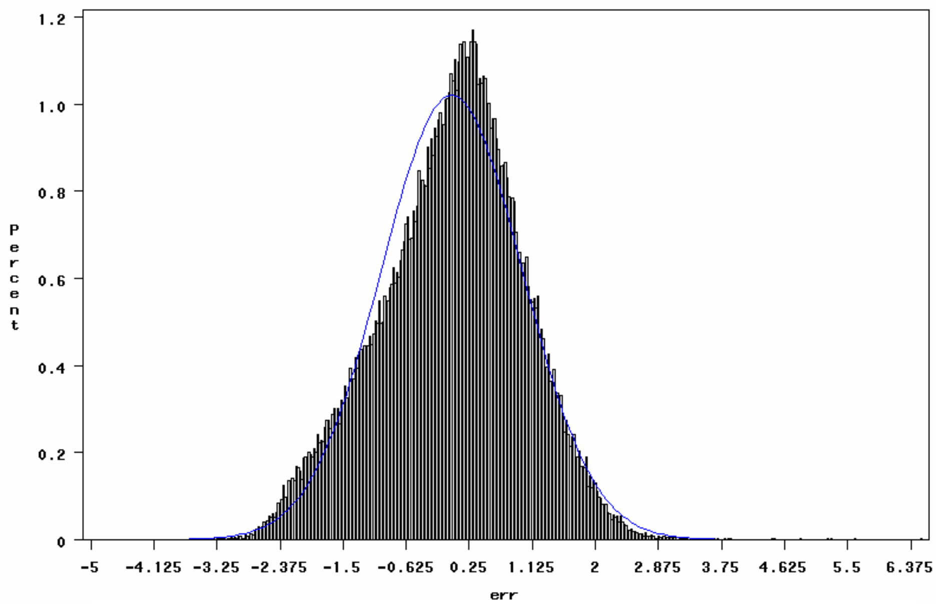
Normalized Corn Yield (Relocated and Rescaled for the Gamma Distribution

Figure 2.4. Corn Belt Corn Yields and the Gamma-Minus Density Curve



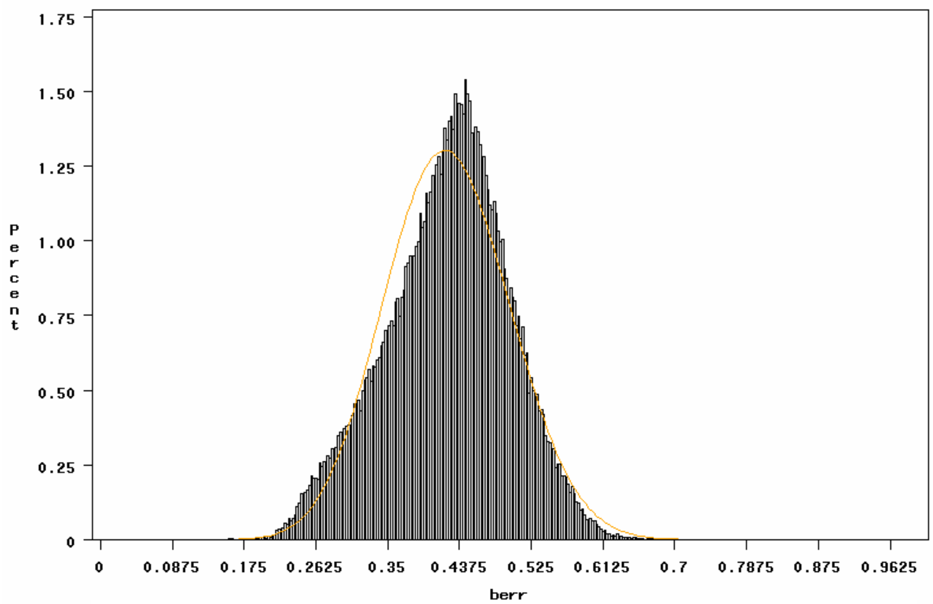
Normalized Corn Yield (Relocated and Rescaled for the Gamma-Minus Distribution)

Figure 2.5. Northern Plains Wheat Yields and the Normal Density Curve



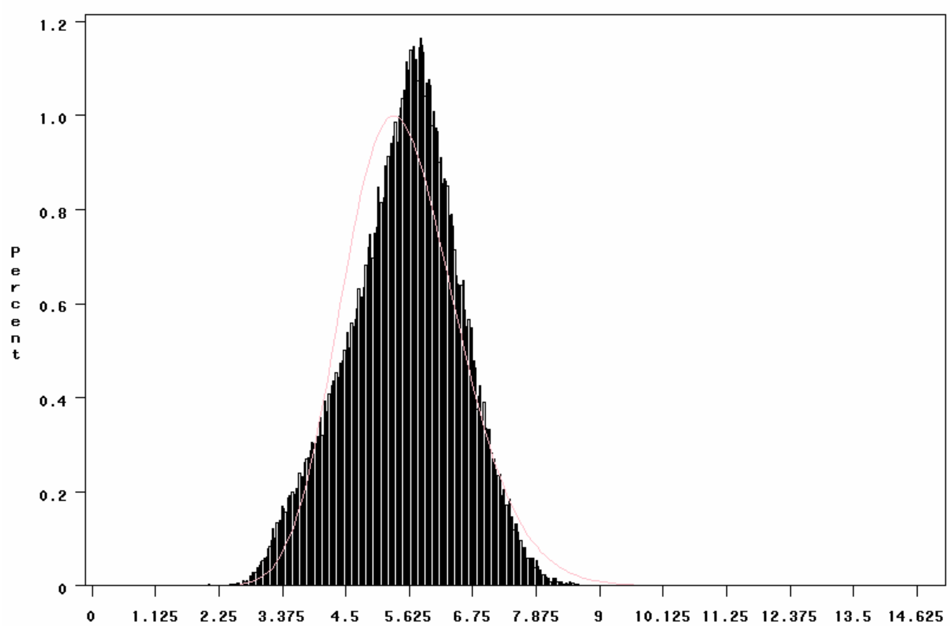
Normalized Wheat Yield

Figure 2.6. Northern Plains Wheat Yields and the Beta Density Curve



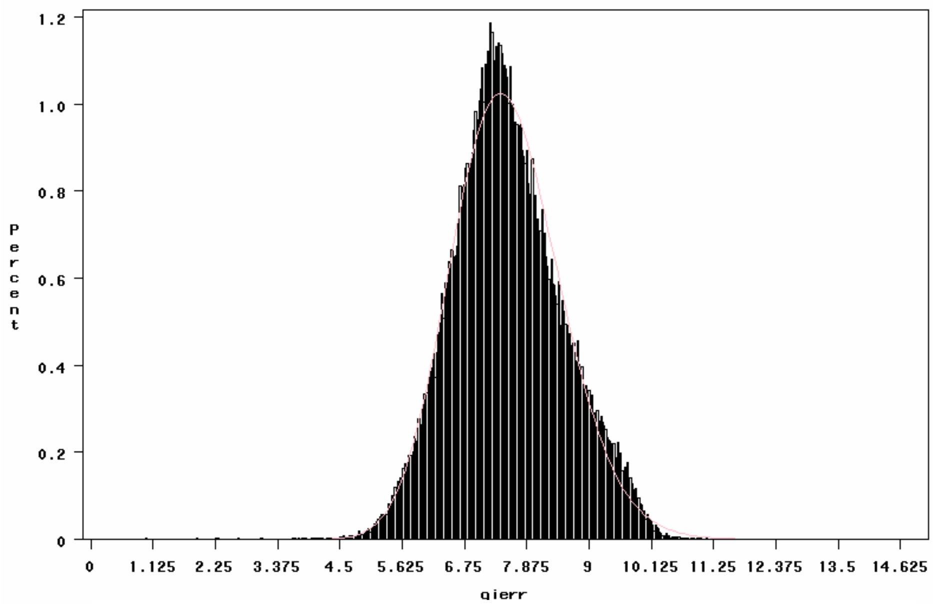
Normalized Wheat Yield (Relocated and Rescaled for the Beta Distribution)

Figure 2.7. Northern Plains Wheat Yields and the Gamma Density Curve



Normalized Wheat Yield (Relocated and Rescaled for the Gamma Distribution)

Figure 2.8. Northern Plains Wheat Yields and the Gamma-Minus Density Curve



Normalized Wheat Yield (Relocated and Rescaled for the Gamma-Minus Distribution)

Conclusion

The yield models and associated analysis are relevant to the ongoing debate on crop yields in two ways. First, estimation results indicate that time trends may be important, but that careful scrutiny is also important. The only time trend variable included in any of the four models estimated here is the linear term for the long time-series for corn in the Corn Belt. In most cases, scrutiny of time marginal effects showed evidence of over-fitting. While the parameter associated with every nonlinear and linear time term tested was found to be significantly different from zero, marginal effects were often negative or unreasonably large. Both results indicate that a time-series of 10 years is just long enough to estimate a linear time trend (if one does exist) but too short to reliably estimate a non-linear time trend, even if the long term trend is non-linear.

Second, in terms of the debate over yield distribution, the analysis presented here suggests that crop yields are not well represented by any of the common distributions tested. The best representation is based on the gamma-minus distribution, which can accommodate negative skew as noted by Gallagher. Normalized yield data for both crop/region combinations appear to exhibit a small negative skew. My overall results – that these parametric distributions are not a good fit for the yield data – is consistent with the recent results of Norwood *et al.*, who find that an empirical distribution is better suited to out-of-sample yield predictions than any of a range of parametric distributions, including those tested in this study.

Finally, the implications of aggregation for data use are extensive. While it has always been known that yield variation can be fully represented only through the use of farm-specific panel data, the loss of variability – particularly random variability – due to spatial aggregation appears to be quite large. In addition to losing a very large proportion of systematic yield variation, nearly one-half of random variation is lost when data are aggregated from farms to counties.

The loss of variation due to aggregation across farms within a county can have significant implications for the use of cross-sectional, farm-level data. While the loss of systematic variation would be expected when aggregating across farms, the loss in random variation appears to be just as large – a result that holds for both region/crop combinations and both the long and short time-series models. The result implies that, within a single year, there is a high level of random variation in yields across farms located in the same county. Farm level data that are collected only in cross section (e.g., USDA's Agricultural Resource Management Survey (ARMS)) generally provide only the current year's production. Given the large random variations across farms, yield from a single year may provide only a poor estimate of cross-sectional variation in intended production.

Likewise, characterizing yield variability using county average yield data could result in considerable error. While some researchers have used county data to calculate a yield variance (see Wu, for example), such a procedure would capture the effect of annual yield shocks common to all producers in the county (due, for example, to annual variation in weather), but not yield shocks that are specific to individual farms. Given that farm-specific shocks can account for one-half or more of total random variation, county-level estimates of yield variance could be quite different from actual farm-level variance.

Appendix 2.1: Hypotheses for Mean Model Specification Tests

In the mean model, calculating the average p-value for a single parameter restriction model involves five steps:

- Initial estimation of the unrestricted mean models (one for each of 100 data groups) to obtain residuals;
- Specification of a variance models for each group (see appendix 2.2);
- Estimation of heteroskedasticity corrected, unrestricted, and restricted models; and
- Calculation of the F-statistics and associated p-values from residuals of the unrestricted and restricted models;
- Calculation of average p-values.

A formal statement of the unrestricted mean model and parameter restriction for the test results given in Tables 2.3, 2.4, 2.5, and 2.6 is given in Appendix Table 2.1. The left-hand columns correspond to the table number in the text and the test number, respectively. For each test, the unrestricted model, null hypothesis (H_0), and alterative hypothesis (H_A) are shown.

Appendix Table 2.1. Formal Statement of Mean Model Hypothesis Tests

Table	Test	Unrestricted Model	Hypothesis Test
2.3 and 2.4	1	$y_{ijkt} = \beta_0^m + \beta_t^m t + \beta_f^m t^2 + \beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i + \varepsilon_{ijkt}$	$H_0 : \beta_f^m = 0 \quad H_A : \beta_f^m \neq 0$
	2	$y_{ijkt} = \beta_0^m + \beta_t^m t + \beta_f^m \log(t) + \beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i + \varepsilon_{ijkt}$	$H_0 : \beta_f^m = 0 \quad H_A : \beta_f^m \neq 0$
	3	$y_{ijkt} = \beta_0^m + \beta_t^m t + \beta_f^m \log(\log(t)) + \beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i + \varepsilon_{ijkt}$	$H_0 : \beta_f^m = 0 \quad H_A : \beta_f^m \neq 0$
	4	$y_{ijkt} = \beta_0^m + \beta_t^m t + \beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i + \varepsilon_{ijkt}$	$H_0 : \beta_t^m = 0 \quad H_A : \beta_t^m \neq 0$
	5	$y_{ijkt} = \beta_0^m + \beta_t^m t + \beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i + \varepsilon_{ijkt}$	$H_0 : \beta_i^m = 0 \; \forall i \in I$ $H_A : \beta_i^m \neq 0 \; for\, at\, least\, one\; i \in I$
	6	$y_{ijkt} = \beta_0^m + \beta_t^m t + \beta_k^m d_k + \beta_j^m d_j + \varepsilon_{ijkt}$	$H_0 : \beta_j^m = 0 \; \forall j \in J$ $H_A : \beta_j^m \neq 0 \; for\, at\, least\, one\; j \in J$
	7	$y_{ijkt} = \beta_0^m + \beta_t^m t + \beta_k^m d_k + \varepsilon_{ijkt}$	$H_0 : \beta_k^m = 0 \; \forall k \in K$ $H_A : \beta_k^m \neq 0 \; for\, at\, least\, one\; k \in K$
2.5	1	$y_{ijkt} = \beta_0^m + \beta_t^m t + \beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i + \varepsilon_{ijkt}$	$H_0 : \beta_t^m = 0 \quad H_A : \beta_t^m \neq 0$
2.5 and 2.6	2	$y_{ijkt} = \beta_0^m + \beta_k^m d_k + \beta_j^m d_j + \beta_i^m d_i + \varepsilon_{ijkt}$	$H_0 : \beta_i^m = 0 \; \forall i \in I$ $H_A : \beta_i^m \neq 0 \; for\, at\, least\, one\; i \in I$
	3	$y_{ijkt} = \beta_0^m + \beta_k^m d_k + \beta_j^m d_j + \varepsilon_{ijkt}$	$H_0 : \beta_j^m = 0 \; \forall j \in J$ $H_A : \beta_j^m \neq 0 \; for\, at\, least\, one\; j \in J$
	4	$y_{ijkt} = \beta_0^m + \beta_k^m d_k + \varepsilon_{ijkt}$	$H_0 : \beta_k^m = 0 \; \forall k \in K$ $H_A : \beta_k^m \neq 0 \; for\, at\, least\, one\; k \in K$

Appendix 2.2: Variance Model Specification and Heteroskedasticity Correction

For any given mean model specification, squared residuals and predicted yields are used to estimate the variance model, beginning with the initial variance model (equation (2.2)):

$$v_{ijkt} = \exp\left(\beta_0^v + \beta_t^v t + \beta_k^v d_k + \beta_j^v d_j + \beta_i^v d_i + \upsilon_{ijkt}\right)$$

where $v_{ijkt} = \dfrac{\hat{\varepsilon}_{ijkt}^2}{\hat{y}_{ijkt}^2}$, $\hat{\varepsilon}_{ijkt}$ is the residual and $\hat{y}_{ijkt}$ is the predicted yield from the mean model.

Specification testing proceeds as follows. First, I test whether the variance model time-effect parameter is equal to zero. Specifically, I test $H_0 : \beta_t^v = 0$ against the alternative $H_A : \beta_t^v \neq 0$.

- If $H_0 : \beta_t^v = 0$ is rejected, the time term is retained in the variance model
- If $H_0 : \beta_t^v = 0$ is not rejected, the time term is not retained in the variance model.

The next three tests involve nested hypotheses: that the unit-level effect parameters, the farm-level effect parameters, and county level effect parameters, respectively, are all equal to zero in the variance model.

Test 1 (crop insurance units) tests $H_0 : \beta_i^v = 0 \; \forall i \in I$ where I is the set of all crop insurance units (within a given group) against the alternative

$H_A : \beta_i^v \neq 0$ *for at least one* $i \in I$.

- If $H_0 : \beta_i^v = 0 \; \forall i \in I$ is rejected, then unit-, farm-, and county-level effects are retained in the variance model.

- If $H_0 : \beta_i^v = 0 \; \forall i \in I$ is not rejected, then unit-level effects are deleted from the model and Test 2 is performed.

Test 2 (farms) tests $H_0 : \beta_j^v = 0 \; \forall j \in J$ where J is the set of all farms (within a given group) against the alternative $H_A : \beta_j^v \neq 0$ *for at least one* $j \in J$.

- If $H_0 : \beta_j^v = 0 \; \forall j \in J$ is rejected then farm- and county-level effects are retained in the variance model.
- If $H_0 : \beta_j^v = 0 \; \forall j \in J$ is not rejected, then farm-level effects are deleted from the model and Test 3 is performed.

Test 3 (counties) tests $H_0 : \beta_k^v = 0 \; \forall k \in K$ where K is the set of all counties (within a given group) against the alternative $H_A : \beta_k^v \neq 0$ *for at least one* $k \in K$.

- If $H_0 : \beta_k^v = 0 \; \forall k \in K$ is rejected then county-level effects are retained in the variance model.
- If $H_0 : \beta_k^v = 0 \; \forall k \in K$ is not rejected, then county-level effects are deleted from the model.

If the variance model indicates the presence of heteroskedasticity, then the mean model is re-estimated correcting for it by multiplying both sides of the unrestricted mean model by $\hat{\sigma}_{ijkt}^{-1} = \left(\hat{y}_{ijkt}\hat{v}_{ijkt}\right)^{-0.5}$.

Chapter 3: Asymmetric Information in Federal Crop Insurance: Comparing High and Low Risk Areas

A number of researchers have argued that asymmetric information is a critical problem in the Federal crop insurance program (e.g., Skees and Reed; Vandeveer and Loehman; Just, Calvin, and Quiggen; Makki and Somwaru). Asymmetric information can take more than one form. In cases of hidden information, the insurer lacks good measures on the farm-specific likelihood of a yield or revenue loss needed to accurately rate crop insurance premiums. This type of asymmetric information can lead to adverse selection, where producers who are charged a lower than actuarially fair rate are more likely than other producers to purchase crop insurance or to purchase high levels of coverage. When asymmetric information takes the form of hidden action, the insurer cannot effectively monitor producer behavior, perhaps encouraging some producers to forgo some of the (costly) inputs and practices necessary to limit the likelihood of crop losses. The incentive to skimp on inputs is referred to as moral hazard.

In practice, these problems can lead to large underwriting losses, even as participation remains low. Between 1981 and 1993, crop insurance underwriting losses totaled $2.3 billion (GAO) while participation hovered at about 25 percent of eligible acreage (Glauber). Subsidy increases in 1995 and again in 1999/2000 pushed crop insurance participation to roughly 80 percent of eligible acreage (Dismukes and Vandeever) and appear to have brought crop insurance premium revenues into line with indemnity payments, at least when viewed at a national scale. This gain, however, came at a high cost: Government expenditures for premium subsidies – which are not considered underwriting losses – reached $2 billion in 2003. On a regional scale, moreover, underwriting losses persist in many areas of the Great Plains and the Southeast (Glauber).

Have premium subsidies simply masked chronic underwriting problems? High subsidies have successfully increased crop insurance participation, perhaps encouraging crop insurance purchase by low risk producers who rarely collect indemnities. Glauber's observations suggest the possibility of regional differences in the effectiveness of crop insurance premium rating. If regional differences in asymmetric information can be found, it may suggest that existing crop insurance premium rating methods are more effective in some regions than in others. It may also suggest that premiums collected in some regions are subsidizing underwriting losses in other areas of the county.

Previous studies of asymmetric information have been restricted in scope. For example, Just *et al.*, use survey data on corn and soybean producers. Makki and Somwaru study corn and soybeans in Iowa. Vandeveer and Lohman study corn producers in a single county in Indiana. To complement and extend the results of these and other previous studies, I estimate the effect of asymmetric information in a region where crop production is less risky (non-irrigated corn in the Corn Belt) and another where crop production is more risky (continuously cropped wheat in the Northern Plains).

In this Chapter, I modify an existing model of asymmetric information in insurance markets, based on the work of Chiappori and Salanie, for use with crop insurance contract, yield, and actuarial data. Because the crop insurance files are an extremely rich source of data on crop yield risk and crop insurance participation, I am able to extend their model in a number of ways. I estimate the model for Corn Belt corn and Northern Plains wheat, then compare and contrast the results in terms of overall production risk, coverage levels, and expected net return to crop insurance purchase, focusing on the effect of asymmetric information on crop insurance coverage.

Testing for Asymmetric Information: Theory

The theory presented here is based largely on Rothschild and Stiglitz (1976; referred to here as RS) except that I do not consider market supply of insurance contracts or market equilibrium. For federal crop insurance, the Risk Management Agency of the U.S. Department of Agriculture (RMA) sets the terms of sale to farmers and agrees to accept program losses, effectively voiding the zero-profit condition used by RS to characterize the supply of insurance contracts. I focus, instead, on crop insurance demand, given terms specified by RMA.

RS show that, among producers who are considered “observationally equivalent” by the insurer, those at higher risk of loss will demand more insurance than producers with lower loss risk. Individuals who are observationally equivalent are all offered the same menu of insurance contracts. For many RMA crop insurance products, including popular yield and revenue insurance products like Average Production History (APH) and Crop Revenue Coverage (CRC), producers who are located in the same county and have the same RMA yield average are generally presumed to be observationally equivalent—they are offered the same menu of insurance options and premium rates.[5] If these producers differ in terms of loss risk, however, this information asymmetry could lead to adverse selection or moral hazard.

[5] Some farms are also classified into “high risk areas” with additional premium charges. These differences are captured empirically by the RMA adjusted base rate used to characterize RMA yield loss risk rating in the econometric analysis.

In this section, I derive expressions for both yield- and revenue-based insurance indemnities and use them to show that yield loss risk is the key source of systematic variation in coverage choice for both yield and revenue insurance purchase decisions. While revenue insurance indemnities are triggered by low revenue rather than low yield, market prices and price risk are the same for all observationally equivalent farms, leaving yield loss risk as the key source of heterogeneity. Producers with relatively high yield loss risk – regardless of whether they choose yield or revenue insurance – will be more likely to purchase high levels of insurance coverage than observationally equivalent producers with low yield loss risk.

Consider a set of producers which share a common mean yield and a common yield distribution up to a scale factor (standard deviation, for example). Yields for producer *j* can be written as $y_{jt} = \bar{y} + \bar{\varepsilon} + \sigma_j \varepsilon_{jt}$, where $\bar{y}$ is mean yield, $\bar{\varepsilon}$ is a mean-zero error component common to all farms in a given area (e.g., county), σ_j is the farm-specific scale parameter (e.g., standard deviation), and ε_{jt} is a random variable drawn independently for each farm from a common distribution with mean zero ($E(\varepsilon) = 0$) and standard deviation equal to one. For practical purposes, producers who are located in the same county also face the same expected price and price risk: $\bar{p}$ is the expected price when insurance is purchased and μ is the (mean zero) change in price over the season, which is common to all producers in a county but unknown at the time of insurance purchase.

For yield-based insurance, the crop insurance indemnity (loss) per dollar of crop insurance liability (the loss-cost ratio or LCR) for coverage level θ on farm *j* in year *t* is:

(3.1) $L^Y_{jt}(\theta \mid \bar{y}, \sigma_j) = Z\left(\theta\bar{y} > \bar{y} + \bar{\varepsilon} + \sigma_j \varepsilon_{jt}\right)\left((\theta\bar{y})^{-1}(\theta\bar{y} - (\bar{y} + \bar{\varepsilon} + \sigma_j \varepsilon_{jt}))\right)$

where Z is an indicator function which is equal to one when actual yield falls below the guarantee level (i.e., when $\theta\bar{y} > \bar{y} + \bar{\varepsilon} + \sigma_j \varepsilon_t$); zero otherwise. The remaining part of the equation is the indemnity (dollars per dollar of liability) given Z=1. The expected value of the LCR can be written as:

(3.2) $E(L^Y(\theta \mid \bar{y}, \sigma_j)) = \Pr(\theta\bar{y} > \bar{y} + \bar{\varepsilon} + \sigma_j \varepsilon)\left((\theta\bar{y})^{-1}(\theta\bar{y} - E(y \mid \theta\bar{y} > \bar{y} + \bar{\varepsilon} + \sigma_j \varepsilon))\right)$

where $\Pr(\theta\bar{y} > \bar{y} + \bar{\varepsilon} + \sigma_j \varepsilon)$ is the probability of an indemnity and the remaining part of the equation is the expected indemnity per dollar of liability, given a yield low enough to qualify for an indemnity.

The RMA premium rate (dollars per dollar of liability) for coverage level θ, given mean yield $\bar{y}$, can be expressed as: $\rho^Y(\theta \mid \bar{y}, k) = E(L^Y(\theta \mid \bar{y}, \sigma_{\bar{y}k}))$ where $\sigma_{\bar{y}k}$ is the level of yield risk implicitly associated with $\bar{y}$ by RMA premium rating methods in local area k (e.g., county). Then, expected net return to yield insurance for producer j is

$$\begin{aligned} R^Y_{j\theta} &= E(L^Y(\theta \mid \bar{y}, \sigma_j)) - \rho^Y(\theta \mid \bar{y}, k) \\ &= E(L^Y(\theta \mid \bar{y}, \sigma_j)) - E(L^Y(\theta \mid \bar{y}, \sigma_{\bar{y}k})). \end{aligned}$$

Note that $R^Y_{j\theta} = 0$ when $\sigma_j = \sigma_{\bar{y}k}$. Because expected indemnity grows as yield variability rises[6], net return would be positive for producers with relatively high yield risk (i.e., $R^Y_{j\theta} > 0$ when $\sigma_j > \sigma_{\bar{y}k}$) and negative for producers with relatively low yield risk (i.e., $R^Y_{j\theta} < 0$ when $\sigma_j < \sigma_{\bar{y}k}$) regardless of coverage level.

[6] Indemnity grows because greater yield variability increases the probability of an indemnity (a specific percentage loss) and increases the average size of the indemnities that are paid.

A revenue-based indemnity would be triggered when actual revenue falls below some proportion of expected revenue. The first step to analyzing revenue-based insurance is to develop an expression for expected revenue:

$$E((\bar{p}+\mu)(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon)) = \overline{py} + E(\mu\bar{\varepsilon}) + \sigma_j E(\mu\varepsilon)$$

where $E(\bar{\varepsilon}) = 0$, $E(\varepsilon) = 0$, and $E(\mu) = 0$. The last two terms of the RHS represent the covariance between random movements in price and yield. Price-yield correlation varies spatially and can be large. Harwood *et al.* found correlation between corn prices and county average yields of less than -0.4 for most of Northern Illinois (the portion of the Corn Belt that is the focus of this study). These systematic correlations result from factors affecting crop production broadly, such as widespread drought or excessive rainfall that could affect market prices. Given these results, I assume that $E(\mu\bar{\varepsilon}) \leq 0$. Farm-specific yield variations, however, are not large enough to affect market prices that are determined on a national or even global scale, so I assume that $E(\mu\varepsilon) = 0$ and expected revenue reduces to $\overline{py} + E(\mu\bar{\varepsilon})$.

The crop revenue insurance indemnity (dollars per dollar of liability) on farm *j*, in year *t*, for a given level of coverage, θ, is:

(3.3)

$$\begin{aligned} L_{jt}^{R}(\theta \mid \bar{y},\sigma_j,\bar{p},\mu) = {} & Z\big(\theta(\overline{py}+E(\mu\bar{\varepsilon})) > (\bar{p}+\mu_t)(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon_{jt})\big) \\ & \times\big(\theta(\overline{py}+E(\mu,\bar{\varepsilon})\big)^{-1}\big(\theta(\overline{py}+E(\mu\bar{\varepsilon})) - (\bar{p}+\mu_t)(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon_{jt})\big) \end{aligned}$$

The expected value of the revenue-based indemnity can be written as:

(3.4)

$$E(L^R(\theta \mid \bar{y},\sigma_j,\bar{p},\mu) = \Pr\left(\theta(\overline{py}+E(\mu\bar{\varepsilon})) > (\bar{p}+\mu)(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon)\right)$$
$$\times\left((\theta(\overline{py}+E(\mu\bar{\varepsilon})))^{-1}\left((\theta(\overline{py}+E(\mu\bar{\varepsilon}))-E(py \mid \theta(\overline{py}+E(\mu\bar{\varepsilon}))>(\bar{p}+\mu)(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon))\right)\right).$$

To show the difference between yield and revenue-based indemnities, I modify the expression for the yield-based indemnity by multiplying each side of the inequality in the indicator function (*Z*) by $\bar{p}+\mu_t$ and multiplying the yield-based indemnity by $\frac{\bar{p}+\mu_t}{\bar{p}+\mu_t}$:

$$L^Y_{jt}(\theta \mid \bar{y},\sigma_j,\bar{p},\mu) = Z\left((\bar{p}+\mu_t)\theta\bar{y} > (\bar{p}+\mu_t)(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon_{jt})\right) \times\left((\bar{p}+\mu_t)\theta\bar{y}\right)^{-1}\left((\bar{p}+\mu_t)(\theta\bar{y}-(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon_{jt}\right). \tag{3.5}$$

Re-writing the yield and revenue indemnity expressions allows a direct comparison of terms:

$$L^Y_{jt}(\theta \mid \bar{y},\sigma_j,\bar{p},\mu_t) = Z\left(\theta\overline{py}+\mu_t\theta\bar{y} > (\bar{p}+\mu_t)(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon_{jt})\right) \times\left(\theta\overline{py}+\mu_t\theta\bar{y}-(\bar{p}+\mu_t)(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon_{jt})\right) \tag{3.6}$$

$$L^R_{jt}(\theta \mid \bar{y},\sigma_j,\bar{p},\mu_t) = Z\left(\theta\overline{py}+\mu_t\theta\bar{y} > (\bar{p}+\mu_t)(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon_{jt})-\theta E(\mu\bar{\varepsilon})+\mu_t\theta\bar{y}\right) \times\left(\theta(\overline{py}+\mu_t\theta\bar{y}-(\bar{p}+\mu_t)(\bar{y}+\bar{\varepsilon}+\sigma_j\varepsilon_{jt})-\theta E(\mu\bar{\varepsilon})+\mu_t\theta\bar{y}\right) \tag{3.7}$$

The difference between the yield (3.6) and revenue (3.7) insurance indemnities for a given producer and coverage level depends on the effect of two terms: $\theta E(\mu\bar{\varepsilon})$ and $\mu_t\theta\bar{y}$, which are included in revenue equation (3.7) but not yield equation (3.6). The first term accounts for the effect of price uncertainty on revenue given the coverage level and average yield. The second adjusts for the effect of price-yield covariance on expected revenue. Because these terms are constant within a small area (e.g., county) for farms with the same average yield, producers who are observationally equivalent for yield-based insurance are also observationally equivalent for revenue-based insurance. Given a specific coverage level, moreover, systematic variation in revenue-based expected indemnities across these producers is found only in the yield variability parameter (σ), as in the case of yield insurance.

The revenue insurance premium rate (dollars per dollar of liability) for coverage level θ, given mean price $\bar{p}$ and mean yield $\bar{y}$, can be expressed as $\rho^R(\theta \mid \bar{y}, k) = E(L^R(\theta \mid \bar{y}, \sigma_{\bar{y}k}, \bar{p}, \mu))$. Expected net return to yield insurance for producer j is

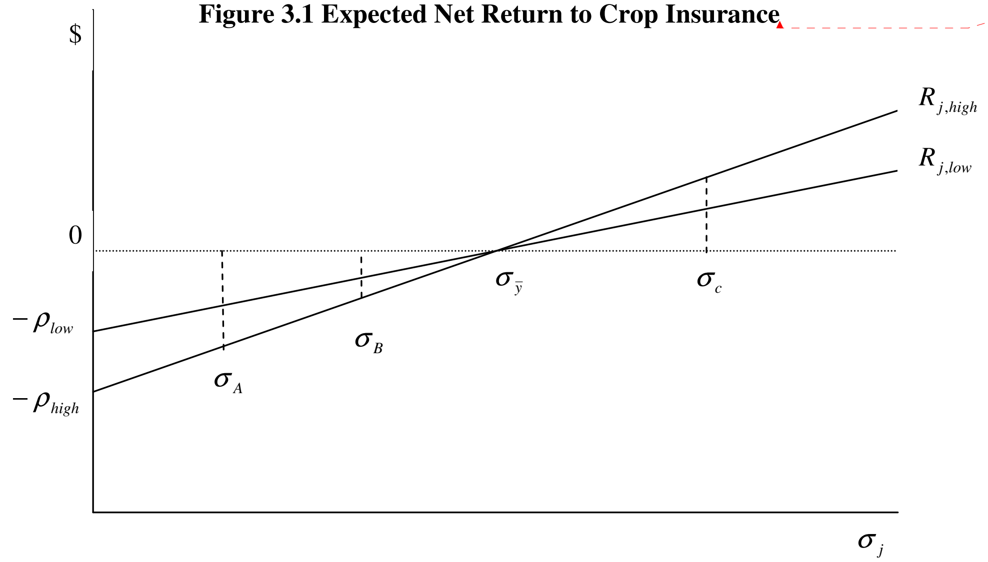
$$\begin{aligned} R_{j\theta}^R &= E(L^R(\theta \mid \bar{y}, \sigma_j, \bar{p}, \mu)) - \rho^R(\theta \mid \bar{y}, k) \\ &= E(L^R(\theta \mid \bar{y}, \sigma_j, \bar{p}, \mu)) - E(L_j^R(\theta \mid \bar{y}, \sigma_{\bar{y}k}, \bar{p}, \mu)). \end{aligned}$$

Because yield risk is the underlying source of variation in revenue insurance, $R_{j\theta}^R = 0$ when $\sigma_j = \sigma_{\bar{y}k}$, and expected indemnity grows as yield risk grows, net return is positive for producers with relatively high yield risk (i.e., $R_{j\theta}^R > 0$ when $\sigma_j > \sigma_{\bar{y}k}$) and negative for producers with relatively low yield risk (i.e., $R_{j\theta}^R < 0$ when $\sigma_j < \sigma_{\bar{y}k}$) regardless of coverage level.

Combining these observations with the theory stated by RS, Figure 3.1 traces out the implications of asymmetric information. The curves represent expected net return to crop insurance purchase for producers who are located in the same area (i.e., county) and have the same mean yield but differ in yield variability (i.e., observationally equivalent producers under current rating methods). As already noted, net returns are upward sloping and pass through zero when $\sigma_j = \sigma_{\bar{y}k}$ for all insurance types and coverage levels. At the vertical ($) axis yield risk is zero so the intercepts for high and low coverage are

$-\rho(\theta_{high})$ and $-\rho(\theta_{low})$, respectively. Finally, as RS show, for farmers with high yield loss risk ($\sigma_j > \sigma_{\bar{y}k}$) the expected net return to high coverage would be larger than expected net return to low coverage (i.e., $R_{j,high} > R_{j,low} > 0$), while for producers with lower yield loss risk ($\sigma_j < \sigma_{\bar{y}k}$) expected net return would be negative for all coverage levels, but would be more negative for high coverage than for low coverage (i.e., $R_{j,high} < R_{j,low} < 0$).

Figure 3.1 Expected Net Return to Crop Insurance



Formatted: Font: Bold

Producer *j* will purchase coverage of level θ only when: $R_{j\theta} + \eta_{j\theta} > 0$ and $R_{j\theta} + \eta_{j\theta} > R_{j\theta'} + \eta_{j\theta'}$ where $\eta_{j\theta}$ is the premium the producer *j* would be willing to pay for risk reduction due to insurance coverage *θ*. High yield loss risk producers ($\sigma_j > \sigma_{\bar{y}}$; σ_c in Figure 3.1) who seek to maximize returns from insurance purchase would select high coverage. Even risk neutral producers (i.e., for whom $\eta_{j\theta}$ = 0) would find it profitable, over time, to purchase high insurance coverage. These results hold for both yield and revenue insurance products.

For low yield loss risk producers ($\sigma_j < \sigma_{\bar{y}}$; σ_A, σ_B in figure 3.1), expected return to crop insurance purchase is negative for all insurance types and coverage levels. For these producers, overall return to crop insurance will be positive only if they are willing to pay a risk premium large enough to fully offset negative expected return. While high coverage offers a lower expected net return, it also offers greater risk reduction than low coverage would. So, highly risk averse producers may choose high coverage even when their yield risk is relatively low ($\sigma_j < \sigma_{\bar{y}k}$). Assuming that low yield risk producers are not systematically more risk averse than other producers, however, higher (less negative) net return for low coverage will mean that low yield risk producers are less likely than high yield risk producers to purchase high coverage.

Empirical Model

The theory suggests that asymmetric information is characterized by positive correlation between the level of (1) yield loss risk not accounted for by crop insurance premiums and (2) the level of crop insurance coverage purchased, given individuals who are "observationally equivalent" to the insurer. Chiappori and Salanie (referred to hereafter as CS) use these results to craft an empirical model that attempts to (1) identify loss risk not captured by insurance premiums (i.e., premium rating error) and (2) draw a systematic relationship between the premium rating error and the level of coverage purchased by the individual, accounting for observed differences across individuals. This systematic relationship, if it exists, is seen as evidence of asymmetric information.

In their empirical work on auto insurance in France, CS define two levels (high and low) for loss risk and coverage level. Loss risk is defined by whether or not the individual had at least one accident for which he or she was at fault—an accident was considered evidence of high loss risk. Motorists could purchase basic (low) coverage that is required by law and covers only liability or additional (high) coverage that also covers damage if the insured individual is found at fault. Loss risk not captured by insurance premiums (i.e., premium rating errors) is identified as the residual from a binomial probit model of loss risk regressed on the information used by the insurer to rate premiums (e.g., age, gender). In a second probit model, CS regress coverage level on the same insurer information. From this equation, the residual is a measure of variation in coverage accounting for insurer-observed differences in loss risk. Correlation between residuals from the loss risk and coverage equations is correlation between loss risk not captured by the insurer and the level of insurance coverage purchased, given observationally equivalent producers. Statistically significant correlation is seen as evidence of asymmetric information.

CS also estimated two other empirical models of asymmetric information. Using a bivariate probit model with the same dependent and independent variables used in the two-equation model, CS argue that asymmetric information is present if the correlation coefficient is significantly different from zero. Finally, given concern about the role of underlying functional forms used by insurers to translate motorist characteristics (e.g., age, gender) into insurance rates, they devised a test based on non-parametric methods.

First, the data was divided into a number of homogeneous groups using insurer data on motorist characteristics (the same data used as regressors in the probit models). Then a series of chi-squared tests (2X2 tables with loss risk in one dimension and coverage level in the other) were used to test for correlation between loss risk and coverage level. Row-column dependence was interpreted as evidence of correlation and, therefore, asymmetric information.

I use crop insurance contract, yield history, and actuarial data to develop a similar test of asymmetric information in Federal crop insurance. Because the crop insurance data are significantly more detailed than the automobile insurance data used by CS, I am able to extend their model in several ways. First, given data on multiple years of yield history, I develop a continuous measure of loss risk, rather than a binary measure as used by CS. The continuous measure should contain more information on yield risk than a binary measure. Second, I model multiple levels of crop insurance coverage using an ordered (rather than binomial) probit model, allowing estimation of model parameters from a richer, more detailed set of data. Finally, the crop insurance contract data contain a measure of yield loss risk which can be adjusted (using RMA actuarial data) to form an exogenous (or at least a predetermined) variable that represents RMA rating of yield loss risk (details below). Use of this data, rather than multiple variables representing individual characteristics, removes concern about functional form that led CS to develop a non-parametric test. Use of a bivariate regression model also offers a richer interpretation of the effect of asymmetric information.

My bivariate regression model of asymmetric information in the Federal crop insurance program is:

(3.8) $\bar{L}_j = \beta_0^L + \sum_k \beta_k^L d_k + \beta_\rho^L \rho_{ABR,j} + e_L$ (yield loss risk equation)

(3.9) $c_j^* = \beta_0^c + \sum_k \beta_k^c d_k + \beta_\rho^c \rho_{ABR,j} + e_c$ (coverage equation)

(3.10) $(e_L, e_c) \sim N(u_L, u_c, \sigma_L, \sigma_c, r)$

where $\bar{L}_j$ is a measure of yield loss risk for farm *j* (see data discussion for details); c_j^* is the producer's desired level of coverage for farm *j* (coverage is observed only at discrete values, so the underlying continuous variable, c_j^*, is latent); d_k is a dummy variable equal to 1 for county *k*, 0 otherwise, $\rho_{ABR,j}$ is the RMA adjusted base rate (an RMA measure of yield loss risk explained in the next section) for farm *j*; β_k^L and β_k^c are parameters representing county-effects in the yield loss risk and coverage equations, respectively; β_ρ^L and β_ρ^c are parameters representing the effect of the RMA measure of yield loss risk in the loss and coverage equations, respectively; e_L and e_c are error terms, with means u_L and u_c, respectively, and standard deviations σ_L and σ_c, respectively; and *r* is the correlation coefficient. Actual (discrete) coverage levels (e.g., .65, .75) are related to the latent variable *c** as follows:

(3.11)

$$c = 1 \quad if \quad c^* < c_1^{**}; \quad c = h \quad if \quad c_{h-1}^{**} \le c^* < c_h^{**} \quad (1 < h < H); \quad c = H \quad if \quad c_H^{**} \le c^* ,$$

Where *h* indexes coverage levels, c_h^{**} is the upper limit or threshold for coverage level *h*, and *H* equals the number of coverage levels, minus one. See Appendix (3.3) for detailed development of the likelihood function.

Although my equations are slightly different from those estimated by CS, their interpretations, particularly the interpretation of the residuals, is the same. In the yield loss risk equation (3.8) the residual represents the yield loss risk not accounted for by the RMA measure of yield loss risk and the county-specific dummy variables that account for any other county-specific differences. In other words, a large positive residual indicates a farm with loss risk above that implied by the RMA measure. The residual in equation (3.9) represents variation in the coverage level given observed differences across producers, which are accounted for by including the RMA measure of loss risk and county-level dummy variables. Positive correlation between the residuals from equations (3.8) and (3.9) would indicate that producers with high yield loss risk not accounted for by the RMA premiums are systematically more likely than other, observationally equivalent producers to choose high levels of coverage. This positive correlation implies asymmetric information.

A critical advantage of the bivariate regression model (over both single equation and non-parametric methods) is in using the model to estimate the effect of asymmetric information on crop insurance purchases. Once parameters are estimated, the model can be used to show how the level of asymmetric information affects coverage purchase decisions across observationally equivalent farmers who may, nonetheless, vary in terms of yield loss risk. The probability that a specific farm will select coverage level *h* is:

(3.12) Pr(coverage level *h* on farm *j*) =

$$\Pr(c_{h-1}^{**} \le c_j^{*} < c_h^{**}) = \Phi\left(\frac{c_h^{**} - E(c_j^{*} \mid \overline{L}_j)}{(1-r^2)^{0.5}}\right) - \Phi\left(\frac{c_{h-1}^{**} - E(c_j^{*} \mid \overline{L}_j)}{(1-r^2)^{0.5}}\right)$$

where $E(c_j^{*} \mid \overline{L}_j) = \beta_0^c + \sum_k \beta_k^c d_k + \beta_\rho^c \rho_{ABR.j} + \dfrac{re_L}{\sigma_L}$.

In this formulation, the parameter β_ρ^c represents the direct effect of premium cost on coverage levels, given that $\rho_{ABR,j}$ underlies premium rates. So, all else being equal, I expect $\beta_\rho^c < 0$; higher premium cost means less insurance. As already noted, the yield loss risk equation error term $e_L = (\overline{L}_j - \beta_o^L - \sum_k \beta_k^L d_k - \beta_\rho^L \rho_{ABR,j})$ represents the magnitude of asymmetric information. Variation in farm-specific yield loss risk ($\overline{L}_j$) that is not matched by variation in the adjusted base rate ($\rho_{ABR,j}$) is the asymmetric information effect. The term $\frac{re_L}{\sigma_L} = \frac{r}{\sigma_L}(\overline{L}_j - \beta_o^L - \sum_k \beta_k^L d_k - \beta_\rho^L \rho_{ABR,j})$ represents the effect of asymmetric information on coverage level. When $\beta_\rho^L > 0$ and $r > 0$, an increase in yield loss risk (an increase in $\overline{L}_j$) that is not matched by an increase in the adjusted base rate (ρ_{ABR}) results in an increase in coverage level (an increase in c^*).

Finally, the likelihood function can be developed using standard techniques (see Appendix 3.3). Parameters are estimated using PROC QLIM is SAS. In actual estimation the first threshold value is set equal to zero, i.e., $c_1 = 0$, and the other threshold values are estimated.

Data

Data on crop insurance contracts and yield histories for 2001 are used to estimate model parameters. For corn in the Corn Belt, data from 69 counties in central and northern Illinois are used (counties in Major Land Resource Areas (MLRA) 108, 111, or 115). For wheat in the Northern Plains, data from 33 counties in North and South Dakota is used (counties in MLRA 53A, 53B, 53C, 54, 55A, 55B, 55C, or 56). From this data, I distill 34,300 observations for Corn Belt corn and 11,600 observations for Northern Plains wheat (see Table 3.1 for more details).

Table 3.1 Number of Observations Used in Model Estimation

	Corn Belt Corn	Northern Plains Wheat
Total number of yield history observations (all crop insurance units)	729,699	489,037
Yield observations in units with at least 4 years history for 1995-2001	453,282	255,462
Yield observations in units with an APH, CRC, or RA contract in 2001	367,875	164,363
Yield risk observations (=number of units)	68,686	36,195
Yield risk observations retaining only 1 per farm	34,414	11,715

Crop insurance yield histories are used to define yield risk loss measures. Up to 10 years worth of crop insurance unit-specific yield history data are maintained by RMA. However, an increase in premium subsidies, beginning in 1995, prompted additional producers to purchase crop insurance. For 2001, roughly 10 percent of producers who had at least 4 years worth of actual yield history first purchased crop insurance in 1995 or later. If these additional producers have lower yield loss risk than producers who purchase insurance before the premium subsidy increase, these differences may help identify asymmetric information. To avoid including particularly good or particularly bad years for some farms but not others, however, I use data on crop insurance units with at least 4 years of actual yield data for the 6 years between 1995 and 2000.

I also restrict consideration to producers who purchased one of three popular RMA crop insurance products in 2001: Actual Production History (APH); Crop Revenue

Coverage (CRC); and Revenue Assurance (RA). Underlying RMA assessment of loss risk for these three insurance products is a common methodology for assessing yield risk (see USDA-RMA, 2000a; 2000b). Together, these insurance products account for 90 and 95 percent of crop insurance liability in the Corn Belt corn and Northern Plains wheat samples, respectively.

Underlying correlation among and within farms, if not accounted for, could undermine interference in the econometric estimation. The county-specific fixed effect already included in the model will account for any variation common to all farms within a county. Intra-farm correlation is eliminated by including only one randomly selected crop insurance unit per farm in the estimation. While some information is lost, the aggregation analysis in Chapter 2 suggests that information loss will be minimal because most intra-county spatial variation occurs between farms rather than within farms.

Yield loss risk can be summarized as the average or expected crop insurance indemnity for yield-based insurance. For farm *j* at time *t* the insured loss (per dollar of liability) is:

$$L_{jt} = \frac{p_t(\theta_j \tilde{y}_j - y_{jt})Z(\theta_j \tilde{y}_j > y_{jt})}{p_t \theta_j \tilde{y}_j} = \frac{(\theta_j \tilde{y}_j - y_{jt})Z(\theta_j \tilde{y}_j > y_{jt})}{\theta_j \bar{y}_j}$$

where p_t is output price, θ_j is the coverage level expressed as a proportion (for 75% coverage θ=.75), $\tilde{y}_j$ is the RMA rate yield[7], y_{jt} is the actual yield, and *Z*(.) is an indictor function with value one when $\theta_j \tilde{y}_j > y_{jt}$, zero otherwise. Note that the yield guarantee –

[7] The RMA rate yield is essential the mean yield for a producers production history. The rate yield is used to make farm-specific adjustments to county-level premium rates calculated from county-level data (see Josephson et al.; RMA 2000a; RMA 2000b).

the yield level below which losses are indemnified – is equal to $\theta_j \tilde{y}_j$. The average indemnity over a period of years is:

$$(3.13) \quad \overline{L}_j(\theta_j, \tilde{y}_j, N_j) = N_j^{-1} \sum_t L_{jt} = \frac{\sum_t Z(\theta_j \tilde{y}_j > y_{jt})(\theta_j \tilde{y}_j - y_{jt})}{N_j \theta_j \tilde{y}_j}$$

where N_j is the number of actual yield observations in yield history for farm *j*.

Depending on the value of θ_j, estimates of average indemnity may be zero on some farms, particularly in regions where crop production is less risky, even though the probability of receiving an indemnity is not actually equal to zero. To capture more information on yield variability, I define yield loss risk as the average loss when $\theta_j = 1$:

$$(3.14) \quad \overline{L}_j(1, \overline{y}_j, N_j) = \frac{\sum_t Z(\tilde{y}_j > y_{jt})(\tilde{y}_j - y_{jt})}{N_j \tilde{y}_j}.$$

In other words, considering all below-average yields provides more information on the lower tail of the yield distribution than could be obtained from looking only at yields that are a certain percentage (e.g., 25 percent) below average.

For all three RMA programs considered here, producers can choose coverage ranging from the catastrophic (CAT) coverage (50 percent yield guarantee indemnified at 55 percent of price) up to 85 percent of expected yield (indemnified at 100 percent of market price) or expected revenue. Depending on the region and crop, some coverage levels are infrequently purchased. For both Corn Belt corn and Northern Plains wheat, coverage levels of 50 (not including CAT), 55, and 60 percent were purchased by only a small percentage of producers and are combined in the analysis (Table 3.2). For the

Northern Plains, 80 and 85 percent coverage are also combined into a single level for the purpose of estimating the ordered probit model.

Table 3.2. Grouping of Coverage Levels for Estimation of Ordered Probit Model

Crop Insurance "Level" for Ordered Probit Model	Corn Belt Corn	Northern Plains Wheat
	Coverage, Percent of APH yield	
1	CAT[1]	CAT[1]
2	50[2], 55, 60	50[2], 55, 60
3	65	65
4	70	70
5	75	75
6	80	80, 85
7	85	--

[1]Catastrophic coverage; 50 percent of yield at 55 percent of price
[2]This is not catastrophic coverage because the producer can choose coverage at 100 percent of price.

The base premium rate (BPR) represents RMA assessment of yield loss risk and is available from crop insurance contract data (the BPR underlies APH, CRC, and RA premium rates). The BPR, however, is specific to the level of coverage purchased by the producer. Using it directly could bias parameter estimates. To obtain an exogenous (or at least predetermined) measure of RMA yield risk assessment, I use the adjusted base rate (ABR). The ABR is calculated from the producer's average yield (rate yield) and actuarial tables for the county, crop, and production practice (RMA, 2000a) and represents yield loss risk for a given level of insurance coverage. In 2001, for example, the ABR is calculated for 75 percent coverage. The BPR is then calculated from the ABR using multiplicative adjustment terms: $\rho_{j\theta} = \varphi_\theta \rho_{j,ABR}$, where $\rho_{ABR,j}$ is the ABR for farm *j*, $\rho_{j\theta}$ is the BPR for coverage level θ on farm *j*, and φ_θ is the coverage differential for the coverage selected by the producer. The ABR can be obtained from the BPR by reversing this process: $\rho_{ABR,j} = \varphi_{\theta}^{-1}\rho_{j\theta}$.

Results

Before proceeding to the econometric results, I calculate farm-specific expected indemnities for 75 percent coverage (using equation (3.13) with $\theta = 0.75$.) and expected net return to crop insurance (expected indemnity less insurance premium) with and without the premium subsidies. Key points in the empirical distribution of all three measures (the 5th, 25th, 50th, 75th, and 95th percentiles) for both crop/region combinations are reported in Table 3.3.

Crop production appears to be riskier in the Northern Plains, consistent with Chapter 2 results. In the Corn Belt, average loss is zero for more than half of crop insurance units, indicating that between 1995 and 2001 these producers did not report a yield of less than 75 percent of their current RMA rate yield. Seventy-five percent had average estimated losses of \$0.012/dollar of liability or less, while 95 percent had average estimated losses of \$0.078/dollar of liability or less. In the Northern Plains, average estimated losses are larger and vary more widely across producers. More than half of producers reported at least one yield below 75 percent of their rate yield. Average losses were estimated to be greater than \$0.10/dollar of liability for 25 percent of producers and greater than \$0.26/dollar of liability for 5 percent.

Table 3.3 Simulated Distribution of Expected Indemnity and Net Returns for Yield Insurance, 75 percent coverage, with and without premium subsidy

	Percentile				
	5	25	50	75	95
Corn Belt Corn					
	Dollars per dollar of liability				
Average loss[1]	0.000	0.000	0.000	0.012	0.078
Average net return without premium subsidy[2]	-0.088	-0.061	-0.046	-0.031	0.007
Average net return with premium subsidy	-0.045	-0.027	-0.018	-0.001	0.043
Northern Plains Wheat					
	Dollars per dollar of liability				
Average loss[1]	0.000	0.000	0.008	0.101	0.266
Average net return without premium subsidy[2]	-0.130	-0.090	-0.069	-0.011	0.136
Average net return with premium subsidy[2]	-0.053	-0.036	-0.023	0.053	0.209

[1]Calculated empirically using equation (3.13) from text

[2]Premium subsidy for 75 percent coverage was 55 percent in 2001

Average net return is also more variable for Northern Plains wheat producers than it is for Corn Belt corn producers. The 5th and 95th percentiles for average estimated net return to crop insurance (with the premium subsidy) for Corn Belt corn are -\$0.045/dollar of liability and \$0.043/dollar of liability, respectively, while the same percentiles are -\$0.053/dollar of liability and \$0.209/dollar of liability, respectively, for Northern Plains wheat. In the Corn Belt, at least 75 percent of producers had negative estimated net return, on average, for 1995-2001. In the Northern Plains at least 25 percent had estimated net returns of \$0.053/dollar of liability. These results suggest that loss risk is larger in the Northern Plains and that RMA premiums capture less of the variation in the Northern Plains than they do in the Corn Belt.

Econometric estimates of key parameters from equations (3.8) and (3.9) are reported in Tables 3.4 and 3.5. All parameter estimates, including county effects and coverage thresholds are given in Appendix Tables A3.1 and A3.2. Likelihood ratio tests reported in Tables 3.4 and 3.5 show that, for both region/crop combinations, model variables collectively improve the value of the full likelihood function ($\ell(\beta^L, \beta^c)$) over models with intercepts only ($\ell(\beta_0^L, \beta_0^c)$) and models with intercepts and county fixed effects only ($\ell(\beta_0^L, \beta_0^c, \beta_k^L, \beta_k^c)$). Moreover, all parameter estimates reported in Table 3.4 and 3.5 (and most reported in Appendix Tables 3.1 and 3.2) are significantly different from zero at the one percent level.

For both crop/region combinations, estimation results provide evidence of asymmetric information: (1) correlation is positive between the residuals for the risk loss and coverage equations and (2) the yield loss risk equation parameter associated with the adjusted base rate (ABR) is positive. Producers who have yield loss risk that is high relative to their ABR are likely to purchase higher coverage than producers who have relatively low yield loss risk. Differences in the probability of various coverage levels due to asymmetric information can be seen by comparing the asymmetric information effect across farms that are observationally equivalent or, at least, quite similar.

Table 3.4 Estimation Results for Bivariate Model of Corn Belt Corn[1]

Equation	Parameter	Estimate	Standard Error	t-ratio	p-value
Loss risk	Adjusted base rate (ρ_{ABR})	0.56	0.0005	1,077.5	<0.0001
Loss risk	Error standard dev. (σ_L)	0.04	0.0000	7,297.9	<0.0001
Coverage	Adjusted base rate (ρ_{ABR})	-10.75	0.0158	-678.3	<0.0001
Coverage	Correlation coefficient (r)	0.04	0.0002	194.7	<0.0001

Likelihood function values[2]:

Model with intercepts only	$\ell(\beta_0^L, \beta_0^c) =$	-1,681,712	
Model with intercepts and county effects only	$\ell(\beta_0^L, \beta_0^c, \beta_k^L, \beta_k^c) =$	645,234	
Full model	$\ell(\beta^L, \beta^c) =$	1,484,463	

Likelihood ratio tests:

$2[\ell(\beta^L, \beta^c) - \ell(\beta_0^L, \beta_0^c)] =$	3,363,424	(significant at 1% level)
$2[\ell(\beta^L, \beta^c) - \ell(\beta_0^L, \beta_0^c, \beta_k^L, \beta_k^c)] =$	-1,290,468	(significant at 1% level)

[1]Complete Results are given in Appendix Table 3.1
[2]For derivation of likelihood function, see Appendix 3.3

Table 3.5 Estimation Results for Bivariate Model of Northern Plains Wheat[1]

Equation	Parameter	Estimate	Standard Error	t-ratio	p-value
Loss risk	Adjusted base rate (ρ_{ABR})	1.68	0.0036	468.0	<0.0001
Loss risk	Error standard dev. (σ_L)	0.08	0.0000	1,931.5	<0.0001
Coverage	Adjusted base rate (ρ_{ABR})	-6.40	0.0462	-138.6	<0.0001
Coverage	Correlation coefficient (r)	0.14	0.0008	179.0	<0.0001
Likelihood function values[2]:					
Model with intercepts only	$\ell(\beta_0^L, \beta_0^c) =$	-725,029			
Model with intercepts and county effects only	$\ell(\beta_0^L, \beta_0^c, \beta_k^L, \beta_k^c) =$	-533,930			
Full model	$\ell(\beta^L, \beta^c) =$	-412,466			
Likelihood ratio tests:					
	$2[\ell(\beta^L, \beta^c) - \ell(\beta_0^L, \beta_0^c)] =$	625,126			
	$2[\ell(\beta^L, \beta^c) - \ell(\beta_0^L, \beta_0^c, \beta_k^L, \beta_k^c)] =$	242,928			

[1]Complete Results are given in Appendix Table 3.2
[2]For derivation of likelihood function, see Appendix 3.3

The asymmetric information effect estimated for farm *j* in county *k* can be written as:

$$(3.15) \qquad A_{kj} = \frac{r}{\sigma_L}(\bar{L}_j - \hat{\beta}_o^L - \hat{\beta}_k^L d_k - \hat{\beta}_\rho^L \rho_{ABR,j})$$

where the "hats" indicate parameter estimates. The asymmetric information effect is calculated for each farm in both datasets then classified as positive, near zero, or negative; "near zero" is defined as $-0.01 \leq A_{kj} \leq 0.01$. These bounds are constant across regions to facilitate inter-region comparison and were chosen with the mean and dispersion of the asymmetric information effects of each region in mind. (For the Corn Belt, the asymmetric information term has mean .0021 and standard deviation of .0461; in the Northern Plains, mean .0050 and standard deviation .1625). To ensure that comparisons are based on observationally similar producers, the datasets are divided (separately) into 5 equal-size groups based on the adjusted base rate (ABR).

For corn in the Corn Belt, Table 3.6 shows the predicted probabilities by (1) ABR quintile, (2) direction of the asymmetric information effect, and (3) coverage level. Only modest differences in the distribution of probability across coverage levels exist between producers with positive and negative asymmetric information effects, particularly for producers with adjusted base rates in the lower 3 quintiles. For producers in ABR quintile 1, the greatest difference in coverage probability between producers with positive and negative asymmetric information effects is for producers purchasing 85 percent coverage: 13.8 percent of producers with positive effects purchase compared with 12.1 percent for producers with negative effects. At 75 percent coverage – the most frequently purchased coverage level – the difference is only 0.3 (33.6 percent versus 33.9 percent). The asymmetric information effect is slightly larger for producers who have larger ABRs. Results from Table 3.6 for quintile 5 are shown in Figure 3.1. Producers with negative asymmetric information effects are much more likely to purchase CAT coverage (probability=0.32) than producers with positive asymmetric information (probability=0.24). Producers with "near zero" or positive asymmetric information are 2-3 percentage points more likely than other producers to buy coverage of 75 percent or greater (Figure 3.1).

Table 3.6. Effect of Asymmetric Information on the Level of Crop Insurance Coverage by RMA Adjusted Base Rate (quintiles)—Corn Belt Corn

Adjusted Base Rate[1] (quintiles)	Asymmetric Information Effect[2]	Number of Insurance Contracts	Coverage Level						
			CAT[3]	50/55/60	65	70	75	80	85
			Predicted probability of crop insurance coverage level[4]						
1	<0	3,086	0.094	0.015	0.139	0.129	0.336	0.166	0.121
1	=0	1,981	0.087	0.015	0.135	0.127	0.338	0.171	0.128
1	>0	1,804	0.080	0.014	0.129	0.124	0.339	0.177	0.138
2	<0	3,099	0.113	0.017	0.150	0.132	0.326	0.154	0.110
2	=0	1,920	0.105	0.016	0.145	0.130	0.329	0.159	0.115
2	>0	1,858	0.091	0.015	0.135	0.126	0.333	0.169	0.130
3	<0	2,959	0.123	0.018	0.156	0.134	0.321	0.146	0.101
3	=0	1,615	0.120	0.018	0.152	0.132	0.322	0.150	0.105
3	>0	2,291	0.106	0.016	0.144	0.129	0.327	0.159	0.118
4	<0	2,895	0.144	0.020	0.167	0.138	0.313	0.133	0.084
4	=0	1,467	0.130	0.019	0.160	0.136	0.319	0.142	0.094
4	>0	2,509	0.118	0.018	0.153	0.133	0.324	0.150	0.105
5	<0	2,618	0.322	0.027	0.189	0.127	0.226	0.073	0.036
5	=0	1,161	0.240	0.026	0.191	0.137	0.263	0.093	0.050
5	>0	3,091	0.235	0.025	0.184	0.135	0.267	0.099	0.056

[1]RMA premium rate for APH contracts before being adjusted for coverage level.
[2]Calculated using equation (3.15) from the text
[3]Catastrophic coverage; 50 percent of yield at 55 percent of price
[4]Calculated using equation (3.12) from the text

Figure 3.2. Effect of Asymmetric Information on the Probability of Crop Insurance Coverage -- High RMA Premium Rate Corn Producers in the Corn Belt

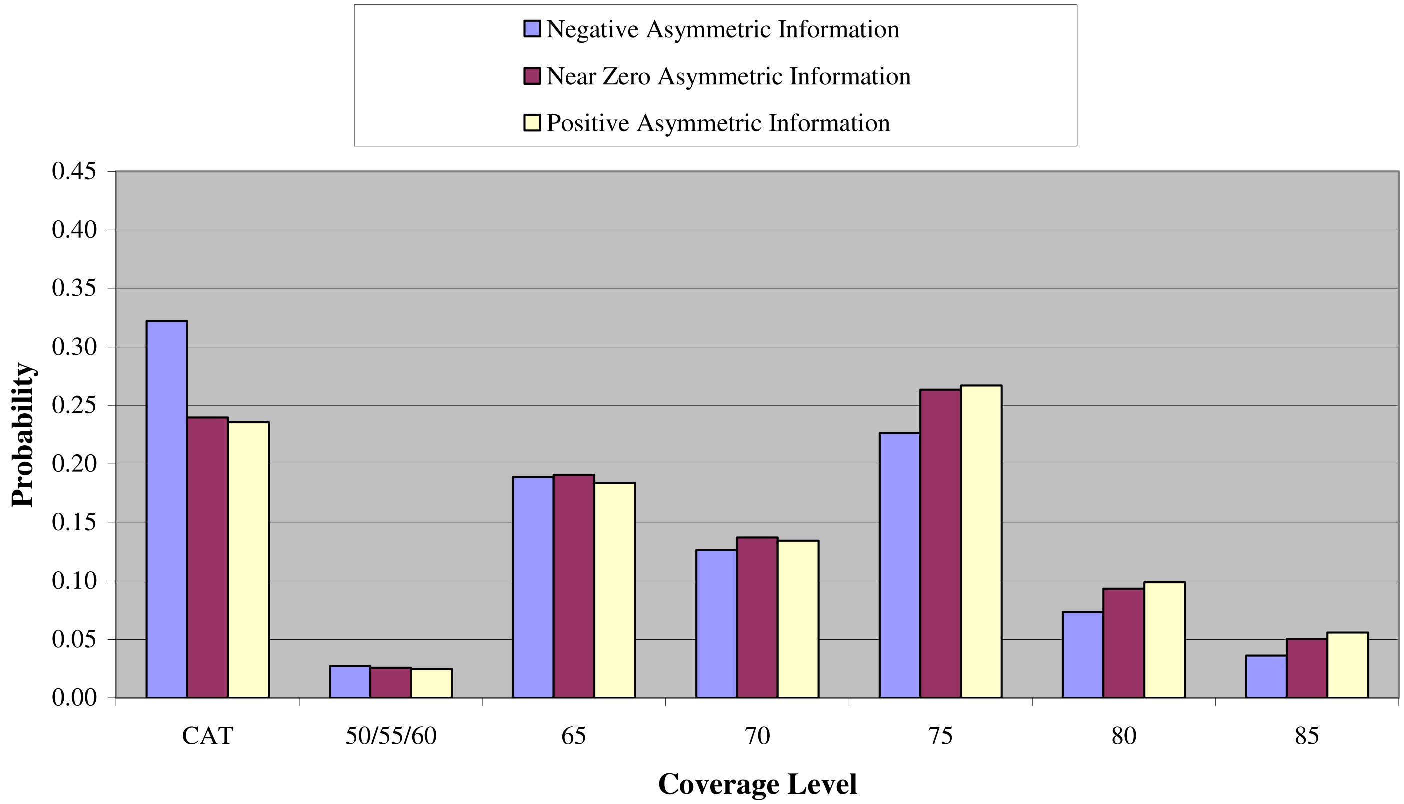


Table 3.7 shows the effect of asymmetric information on the coverage choices of wheat producers in the Northern Plains. Differences due to asymmetric information are much larger than they are for corn in the Corn Belt. Here, producers with positive asymmetric information effects are more likely to buy 70 percent or higher coverage than producers with negative asymmetric information effects. The opposite is true for 65 percent and lower levels of coverage. These results hold for all 5 quintiles, but become more pronounced as adjusted base rates rise. Results from Table 3.7 for quintile 5 are shown in Figure 3.3. Producers with negative asymmetric information effects are much more likely to purchase 65 percent coverage (probability=0.37) than producers with positive asymmetric information (probability=0.30). Producers with positive asymmetric information are more likely than those with negative asymmetric information effects to buy 70 percent coverage (0.43 vs. 0.38) and 75 percent coverage (0.19 vs. 0.11).

Table 3.7. Effect of Asymmetric Information on the Level of Crop Insurance Coverage—Northern Plains Wheat

Adjusted Base Rate[1] (quintiles)	Asymmetric Information Effect[2]	Number of Insurance Contracts	Coverage Level					
			CAT[3]	50/55/60	65	70	75	80/85
			Predicted probability of crop insurance coverage level[4]					
1	<0	1361	0.041	0.036	0.336	0.429	0.148	0.010
1	=0	162	0.039	0.034	0.329	0.432	0.155	0.011
1	>0	817	0.028	0.027	0.291	0.447	0.190	0.016
2	<0	1244	0.052	0.042	0.353	0.409	0.135	0.009
2	=0	195	0.055	0.043	0.358	0.406	0.130	0.008
2	>0	903	0.037	0.032	0.307	0.429	0.178	0.017
3	<0	1202	0.047	0.039	0.343	0.418	0.143	0.010
3	=0	138	0.040	0.034	0.324	0.429	0.161	0.012
3	>0	998	0.028	0.026	0.278	0.442	0.206	0.021
4	<0	1230	0.052	0.041	0.350	0.411	0.137	0.009
4	=0	123	0.042	0.035	0.325	0.426	0.159	0.012
4	>0	984	0.028	0.026	0.276	0.442	0.207	0.021
5	<0	1208	0.077	0.052	0.377	0.377	0.110	0.007
5	=0	92	0.048	0.039	0.341	0.416	0.145	0.010
5	>0	1039	0.036	0.031	0.298	0.431	0.187	0.018

[1]RMA premium rate for APH contracts before being adjusted for coverage level.
[2]Calculated using equation (3.15) from the text
[3]Catastrophic coverage; 50 percent of yield at 55 percent of price
[4]Calculated using equation (3.12) from the text

Figure 3.3. Effect of Asymmetric Information on Crop Insurance Coverage -- High RMA Premium Rate Wheat Producers in the Northern Plains

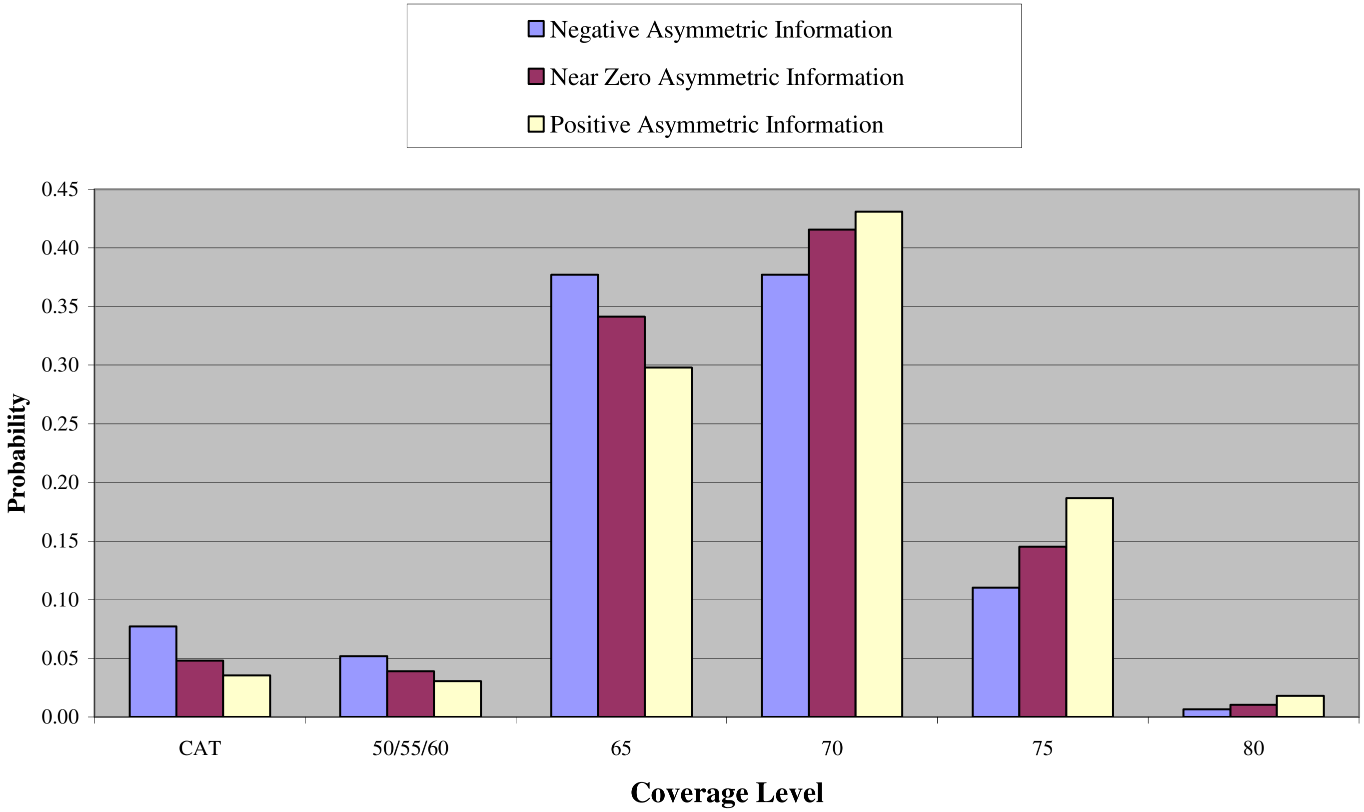


Table 3.8 gives expected net returns (expected indemnity less subsidized premium) for yield-based insurance for the three most popular coverage levels: 65, 70, and 75 percent. For the Corn Belt region estimated net returns for producers in the first 3 ABR quintiles range from $-0.030/$ liability to $-0.002/$ liability. For producers in ABR quintile 4, the effect of asymmetric information on net return is less than $0.0030 in all but 1 case (75 percent coverage). For producers in quintile 5, net returns range from -$0.07/$ liability to $0.010/$ liability, but are positive only for producers with positive asymmetric information effects.

Given that positive expected net return is the exception rather than the rule for Corn Belt producers, risk reduction appears to plays an important role in crop insurance purchase decisions for Corn Belt producers. Moreover, the difference in coverage levels between producers with positive and negative asymmetric information effects is small for the three levels of coverage reported in Table 3.8 (65, 70 and 75 percent). Even so, the asymmetric information effect appears to introduce inequity among producers. For 75 percent coverage, the difference in net return for producers with positive and negative asymmetric information effects varies from $0.02/$ liability for quintile 1 to $0.08 for producers in quintile 5.

For Corn Belt producers with negative asymmetric information, net return is generally more negative for higher levels of coverage. For producers with negative asymmetric information effects in ABR quintile 1, average net return for 65 percent coverage is -$0.017/$ liability, while average net return for 75 percent coverage is -0.023/$. For producers with negative asymmetric information in ABR quintile 5, average return is -\$0.050/\$ liability for 65 percent coverage and -0.070/\$ liability for 75 percent coverage. For producers with positive asymmetric information effects, on the other hand, returns generally rise with the coverage level. For positive asymmetric information producers in ABR quintile 1, average net return varies from -\$0.008/\$ liability (65 percent coverage) to -\$0.003/\$ liability (75 percent coverage). For producers in ABR quintile 5 average returns vary from \$0.009/\$ liability (65 percent coverage) to \$0.010/\$liability (75 percent coverage).

Table 3.8. Effect of Asymmetric Information on Average Net Return to Crop Insurance

Adjusted Base Rate[1] (quintiles)	Asymmetric Information Effect[2]	Corn Belt Corn Coverage Level 65	70	75	Northern Plains Wheat Coverage Level 65	70	75
		Average Net Return[3]			*Average Net Return*[3]		
1	<0	-0.017	-0.020	-0.023	-0.034	-0.041	-0.046
1	=0	-0.016	-0.018	-0.020	-0.031	-0.034	-0.035
1	>0	-0.008	-0.007	-0.003	0.021	0.023	0.027
2	<0	-0.019	-0.023	-0.027	-0.039	-0.047	-0.054
2	=0	-0.018	-0.021	-0.024	-0.032	-0.038	-0.041
2	>0	-0.007	-0.006	-0.004	0.026	0.027	0.029
3	<0	-0.021	-0.025	-0.030	-0.043	-0.051	-0.059
3	=0	-0.020	-0.024	-0.027	-0.026	-0.028	-0.029
3	>0	-0.006	-0.005	-0.002	0.059	0.061	0.064
4	<0	-0.024	-0.029	-0.034	-0.045	-0.054	-0.062
4	=0	-0.023	-0.026	-0.028	-0.004	-0.008	-0.011
4	>0	-0.002	0.000	0.003	0.071	0.073	0.076
5	<0	-0.050	-0.060	-0.070	-0.053	-0.064	-0.075
5	=0	-0.034	-0.039	-0.043	0.010	0.006	0.002
5	>0	0.009	0.009	0.010	0.110	0.112	0.113

[1]RMA premium rate for APH contracts before being adjusted for coverage level.

[2]Calculated using equation (3.15) and $-0.01 \leq A_{kj} \leq 0.01$ to define "near zero."

[3]Calculated as expected indemnity (equation 3.13) less subsidized premium

In the Northern Plains, expected returns depend more heavily on the asymmetric information effect. For all 5 quintiles of the adjusted base rate, net returns are negative for producers with negative asymmetric information effects and positive for producers with positive asymmetric information effects. For producers with negative asymmetric information effects, expected net return varies from -$0.075/$ liability (quintile 5, 75 percent coverage) to -$0.034/$ liability (quintile 1, 65 percent coverage) while those with positive asymmetric information effects enjoy expected returns varying from $0.021/$ liability (quintile 1, 65 percent coverage) to $0.113/$ liability (quintile 5, 75 percent coverage). Moreover, the differences in expected return between producers with positive and negative asymmetric information effects is large, ranging from $0.055/$ liability (65 percent coverage, ABR quintile 1) to as much as $0.188/$ liability (75 percent coverage, ABR quintile 5).

Incentives to buy low (high) coverage are also stronger for positive (negative) asymmetric information producers in the Northern Plains than for producers in the Corn Belt. For producers with negative asymmetric information effects in ABR quintile 1, the average net return to crop insurance purchase is -$0.034/$ liability for 65 percent coverage and -$0.046/$ liability for 75 percent coverage. For producers in ABR quintile 5, average net return varies from -$0.053/$ liability for 65 percent coverage to -$0.075/$ liability for 75 percent coverage. For producers with positive asymmetric information effects, returns generally rise with the coverage level. Producers in ABR quintile 1 enjoy average net returns of $0.021/$ liability (65 percent coverage) to $0.027/$ liability (75 percent coverage). For producers in ABR quintile 5, average returns vary from $0.110/$ liability (65 percent coverage) to $0.113/$liability (75 percent coverage).

Conclusion

Evidence of asymmetric information was found in both the Corn Belt and the Northern Plains, but the results of this study indicate that the consequences differ across regions. In the Corn Belt, estimates indicate that shifts in coverage due to the asymmetric information effect are modest, as are differences in expected net return between producers who have negative asymmetric information effects versus those with positive effects. In the Northern Plains, however, asymmetric information effects appear to be substantially larger, in terms of both coverage effects and differences in average net return.

A key difference between the regions is expected net return to crop insurance purchase. In the Corn Belt, very few producers are likely to realize a positive expected net return to crop insurance purchase. Even among producers who have positive asymmetric information effects, expected indemnities rarely exceed premium costs. These results suggest that premium rates are, on average, higher than actuarially fair premiums would be.

In the Northern Plains, many producers are likely to realize net expected gains to crop insurance purchase, and the gains are potentially large – much larger than could be realized by Corn Belt producers. On the other hand, those producers who would realize negative expected returns to crop insurance also face larger losses than their Corn Belt counter-parts. It is not clear whether premium rates are, on average, lower or higher than actuarially fair rates.

These results suggest that RMA premium rates reflect an average yield loss across producers who are charged the same or similar premium rate, but that actual indemnities vary within these groups. The econometric estimation suggests that there are behavioral consequences to "averaging" across producers: those who perceive high risk relative to their crop insurance premium rate will purchase higher coverage than those who do not. As was the case with other results, differences between producers with positive and negative asymmetric information effects tend to be larger in the Northern Plains than they are in the Corn Belt when compared on a dollar-per-dollar of liability basis.

These findings also suggest that asymmetric information may well play a role in creating inter-regional differences in underwriting loss. The aggregation analysis of Chapter 2 (see Tables 2.7 and 2.8) showed that the variability of yields across farms is larger in the Northern Plains than in the Corn Belt. Because the RMA rating mechanism groups producers by average yield, the producers in each group who face relatively high yield loss risk tend to realize a return based on asymmetric information. These producers tend to buy more coverage, increasing their private return but also increasing government underwriting loss. Other producers, who have lower yield loss risk, buy less coverage or opt out of insurance entirely. Because variation across producers is larger in the Northern Plains than in the Corn Belt, the behavioral consequences of asymmetric information (differences in coverage level) are also larger.

An important question is whether coverage and net return effects are due to adverse selection or moral hazard. In adverse selection the information asymmetry involves hidden information: Producers know more about production risk on their fields than RMA does. For example, farms with high risk of yield loss, compared to observationally equivalent farms, earn higher returns from crop insurance participation than do other producers, without altering production practices. In this case, a key underlying question is whether RMA could reasonably obtain additional information on producer yield loss risk from existing yield histories. An alternate explanation is moral hazard. Here, a producer who has purchased crop insurance uses production practices and input levels that are less effective in minimizing production risk than he or she would without insurance coverage. In this case, a key question is whether crop insurance terms can be altered to discourage cheating.

Current program design and economic research suggest that incentives for cheating, at least on a broad scale, are small. Given that 65, 70, and 75 percent are the most popular levels of coverage, however, a producer would have to absorb a 25-35 percent loss before any indemnity is paid. This large deductible would appear to make moral hazard a less likely explanation for the observed asymmetric information effect when compared with the possibility of adverse selection. The results of Roberts *et al.* tend to support this conclusion. They investigated the moral hazard effect of crop insurance purchase by estimating the effect of crop insurance purchase on crop yields, finding only limited evidence of moral hazard. In Texas, both wheat and soybean yields appear to have declined by about 10 percent due to crop insurance purchase. Elsewhere, however, even when yield changes were found to be statistically significant, yield changes were small.

The key question, of course, is what to do about asymmetric information. Farm-level expected loss data are based on only a handful of observations for each farm. While these data are useful in capturing broad patterns – particularly interregional differences – farm-specific data are subject to considerable sampling error. In other words, the small number of observations available for each farm may or may not be representative of the farm's actual yield distribution and estimates of error in rating premiums are noisy. Can these data be used to improve crop insurance premium rating? And, if so, with what level of confidence? I take up these questions in the next Chapter.

Appendix Table 3.1. Full Econometric Model Results for Corn Belt Corn

Equation	Parameter	County	Estimate	Standard error	T-statistic	*P*-value
Loss Risk	Intercept		0.0416	0.0001	743.0	<0.0001
Loss Risk	County effect	17001	0.0049	0.0001	62.8	<0.0001
Loss Risk	County effect	17003	0.0036	0.0004	9.9	<0.0001
Loss Risk	County effect	17009	0.0127	0.0001	87.6	<0.0001
Loss Risk	County effect	17011	0.0030	0.0001	41.7	<0.0001
Loss Risk	County effect	17013	-0.0017	0.0002	-8.8	<0.0001
Loss Risk	County effect	17015	-0.0193	0.0001	-168.3	<0.0001
Loss Risk	County effect	17017	0.0250	0.0001	218.6	<0.0001
Loss Risk	County effect	17019	0.0244	0.0001	374.6	<0.0001
Loss Risk	County effect	17021	0.0151	0.0001	203.2	<0.0001
Loss Risk	County effect	17029	0.0211	0.0001	239.9	<0.0001
Loss Risk	County effect	17031	0.0308	0.0004	84.1	<0.0001
Loss Risk	County effect	17039	0.0101	0.0001	118.1	<0.0001
Loss Risk	County effect	17041	0.0136	0.0001	157.5	<0.0001
Loss Risk	County effect	17045	0.0186	0.0001	221.0	<0.0001
Loss Risk	County effect	17053	0.0068	0.0001	91.6	<0.0001
Loss Risk	County effect	17057	-0.0016	0.0001	-18.7	<0.0001
Loss Risk	County effect	17059	0.0469	0.0002	289.1	<0.0001
Loss Risk	County effect	17061	0.0025	0.0001	24.4	<0.0001
Loss Risk	County effect	17063	0.0118	0.0001	146.7	<0.0001
Loss Risk	County effect	17067	0.0017	0.0001	22.0	<0.0001
Loss Risk	County effect	17071	-0.0043	0.0001	-46.5	<0.0001
Loss Risk	County effect	17073	0.0036	0.0001	49.8	<0.0001
Loss Risk	County effect	17075	0.0078	0.0001	118.4	<0.0001
Loss Risk	County effect	17077	0.0389	0.0003	131.3	<0.0001
Loss Risk	County effect	17083	0.0105	0.0001	82.3	<0.0001
Loss Risk	County effect	17087	0.0687	0.0004	168.2	<0.0001
Loss Risk	County effect	17091	0.0064	0.0001	87.1	<0.0001
Loss Risk	County effect	17093	0.0343	0.0001	329.8	<0.0001
Loss Risk	County effect	17095	0.0013	0.0001	15.1	<0.0001
Loss Risk	County effect	17099	0.0081	0.0001	121.1	<0.0001
Loss Risk	County effect	17103	-0.0005	0.0001	-6.4	<0.0001
Loss Risk	County effect	17105	0.0157	0.0001	248.7	<0.0001
Loss Risk	County effect	17107	0.0195	0.0001	269.1	<0.0001
Loss Risk	County effect	17109	0.0054	0.0001	64.7	<0.0001
Loss Risk	County effect	17113	-0.0017	0.0001	-26.0	<0.0001

Appendix Table 3.1 continued next page

Appendix Table 3.1 -- *continued from previous page*

Equation	Parameter	County	Estimate	Standard error	T-statistic	*P*-value
Loss Risk	County effect	17115	0.0079	0.0001	104.5	<0.0001
Loss Risk	County effect	17117	0.0234	0.0001	254.3	<0.0001
Loss Risk	County effect	17123	0.0031	0.0001	37.1	<0.0001
Loss Risk	County effect	17125	0.0010	0.0001	9.0	<0.0001
Loss Risk	County effect	17127	0.0140	0.0003	40.5	<0.0001
Loss Risk	County effect	17129	0.0240	0.0001	253.7	<0.0001
Loss Risk	County effect	17131	0.0036	0.0001	42.6	<0.0001
Loss Risk	County effect	17133	0.0244	0.0001	201.4	<0.0001
Loss Risk	County effect	17137	0.0029	0.0001	29.5	<0.0001
Loss Risk	County effect	17139	0.0257	0.0001	285.6	<0.0001
Loss Risk	County effect	17141	-0.0058	0.0001	-67.0	<0.0001
Loss Risk	County effect	17143	0.0025	0.0001	26.9	<0.0001
Loss Risk	County effect	17147	0.0206	0.0001	263.5	<0.0001
Loss Risk	County effect	17149	-0.0051	0.0001	-49.0	<0.0001
Loss Risk	County effect	17155	0.0000	0.0001	0.4	0.7193
Loss Risk	County effect	17157	0.0241	0.0002	145.4	<0.0001
Loss Risk	County effect	17161	-0.0018	0.0001	-15.8	<0.0001
Loss Risk	County effect	17163	0.0247	0.0001	212.1	<0.0001
Loss Risk	County effect	17167	0.0110	0.0001	129.9	<0.0001
Loss Risk	County effect	17169	0.0089	0.0001	84.4	<0.0001
Loss Risk	County effect	17171	0.0026	0.0001	20.1	<0.0001
Loss Risk	County effect	17175	-0.0011	0.0001	-12.9	<0.0001
Loss Risk	County effect	17179	0.0102	0.0001	126.6	<0.0001
Loss Risk	County effect	17181	-0.0235	0.0002	-95.4	<0.0001
Loss Risk	County effect	17183	0.0209	0.0001	284.3	<0.0001
Loss Risk	County effect	17187	0.0021	0.0001	27.3	<0.0001
Loss Risk	County effect	17193	0.0253	0.0001	215.8	<0.0001
Loss Risk	County effect	17195	-0.0055	0.0001	-61.7	<0.0001
Loss Risk	County effect	17197	0.0305	0.0001	361.3	<0.0001
Loss Risk	County effect	17199	0.0456	0.0003	153.5	<0.0001
Loss Risk	County effect	17203	0.0000			
Loss Risk	Adjusted Base Rate		0.5566	0.0005	1,077.5	<0.0001
Loss Risk	Error Standard Error		0.0411	0.0000	7,297.9	<0.0001
Coverage	Intercept		1.7559	0.0015	1,194.6	<0.0001
Coverage	County effect	17001	-0.6401	0.0020	-325.9	<0.0001
Coverage	County effect	17003	-0.9863	0.0116	-85.0	<0.0001
Coverage	County effect	17009	-0.5726	0.0037	-155.0	<0.0001
Coverage	County effect	17011	-0.2922	0.0018	-161.3	<0.0001
Coverage	County effect	17013	-0.8429	0.0052	-163.4	<0.0001
Coverage	County effect	17015	-1.0299	0.0030	-344.4	<0.0001

Appendix Table 3.1 continued next page

Appendix Table 3.1 -- c*ontinued from previous page*

Equation	Parameter	County	Estimate	Standard error	T-statistic	*P*-value
Coverage	County effect	17017	-0.7957	0.0029	-271.4	<0.0001
Coverage	County effect	17019	0.0898	0.0017	54.2	<0.0001
Coverage	County effect	17021	-0.3567	0.0019	-189.6	<0.0001
Coverage	County effect	17029	-0.1480	0.0022	-66.2	<0.0001
Coverage	County effect	17031	-0.3706	0.0091	-40.6	<0.0001
Coverage	County effect	17039	-0.2282	0.0022	-104.9	<0.0001
Coverage	County effect	17041	-0.1403	0.0022	-63.9	<0.0001
Coverage	County effect	17045	-0.1946	0.0021	-90.5	<0.0001
Coverage	County effect	17053	0.0128	0.0019	6.8	<0.0001
Coverage	County effect	17057	-0.6550	0.0022	-291.3	<0.0001
Coverage	County effect	17059	-0.4417	0.0041	-108.2	<0.0001
Coverage	County effect	17061	-0.6861	0.0026	-259.9	<0.0001
Coverage	County effect	17063	-0.1891	0.0020	-93.1	<0.0001
Coverage	County effect	17067	-0.6407	0.0020	-325.0	<0.0001
Coverage	County effect	17071	-0.5365	0.0023	-230.3	<0.0001
Coverage	County effect	17073	-0.5132	0.0018	-279.5	<0.0001
Coverage	County effect	17075	-0.1428	0.0017	-85.3	<0.0001
Coverage	County effect	17077	-0.4764	0.0078	-61.2	<0.0001
Coverage	County effect	17083	-0.6241	0.0032	-193.8	<0.0001
Coverage	County effect	17087	-0.5412	0.0111	-48.6	<0.0001
Coverage	County effect	17091	-0.3291	0.0018	-178.3	<0.0001
Coverage	County effect	17093	-0.2625	0.0026	-100.0	<0.0001
Coverage	County effect	17095	-0.4807	0.0021	-227.0	<0.0001
Coverage	County effect	17099	-0.4921	0.0017	-291.0	<0.0001
Coverage	County effect	17103	-0.6216	0.0020	-308.7	<0.0001
Coverage	County effect	17105	0.2117	0.0016	132.2	<0.0001
Coverage	County effect	17107	-0.2840	0.0018	-154.1	<0.0001
Coverage	County effect	17109	-0.5969	0.0021	-280.6	<0.0001
Coverage	County effect	17113	-0.0073	0.0017	-4.3	<0.0001
Coverage	County effect	17115	-0.1362	0.0019	-70.5	<0.0001
Coverage	County effect	17117	-0.3849	0.0023	-165.8	<0.0001
Coverage	County effect	17123	-0.1534	0.0021	-73.3	<0.0001
Coverage	County effect	17125	-0.9235	0.0029	-319.6	<0.0001
Coverage	County effect	17127	-0.8375	0.0094	-89.0	<0.0001
Coverage	County effect	17129	-0.3420	0.0024	-141.8	<0.0001
Coverage	County effect	17131	-0.0634	0.0021	-29.7	<0.0001
Coverage	County effect	17133	-0.8434	0.0031	-273.1	<0.0001
Coverage	County effect	17137	-0.5147	0.0024	-210.9	<0.0001
Coverage	County effect	17139	-0.0508	0.0023	-22.2	<0.0001
Coverage	County effect	17141	-0.5641	0.0022	-256.0	<0.0001

Appendix Table 3.1 continued next page

Appendix Table 3.1 -- *continued from previous page*

Equation	Parameter	County	Estimate	Standard error	T-statistic	*P*-value
Coverage	County effect	17143	-0.2406	0.0023	-103.7	<0.0001
Coverage	County effect	17147	-0.0016	0.0020	-0.8	0.4356
Coverage	County effect	17149	-0.6514	0.0026	-248.2	<0.0001
Coverage	County effect	17155	-0.4885	0.0029	-165.7	<0.0001
Coverage	County effect	17157	-0.7261	0.0042	-171.9	<0.0001
Coverage	County effect	17161	-0.6146	0.0029	-211.8	<0.0001
Coverage	County effect	17163	-0.6589	0.0029	-225.2	<0.0001
Coverage	County effect	17167	-0.2930	0.0021	-136.5	<0.0001
Coverage	County effect	17169	-0.7494	0.0026	-282.8	<0.0001
Coverage	County effect	17171	-0.6989	0.0034	-205.0	<0.0001
Coverage	County effect	17175	-0.0182	0.0023	-8.1	<0.0001
Coverage	County effect	17179	-0.5974	0.0020	-292.3	<0.0001
Coverage	County effect	17181	-1.5043	0.0078	-192.2	<0.0001
Coverage	County effect	17183	-0.0670	0.0019	-35.9	<0.0001
Coverage	County effect	17187	-0.1362	0.0019	-70.6	<0.0001
Coverage	County effect	17193	-0.9437	0.0032	-297.5	<0.0001
Coverage	County effect	17195	-0.6643	0.0023	-293.5	<0.0001
Coverage	County effect	17197	-0.4055	0.0021	-190.1	<0.0001
Coverage	County effect	17199	-0.4344	0.0078	-55.7	<0.0001
Coverage	County effect	17203	0.0000			
Coverage	Adjusted Base Rate		-10.7458	0.0158	-678.3	<0.0001
Coverage	Threshold2		0.0887	0.0001	724.0	<0.0001
Coverage	Threshold3		0.6507	0.0003	2,378.8	<0.0001
Coverage	Threshold4		1.0240	0.0003	3,276.6	<0.0001
Coverage	Threshold5		1.9141	0.0004	5,037.4	<0.0001
Coverage	Threshold6		2.5311	0.0004	5,687.3	<0.0001
	Correlation Coefficient		0.0397	0.0002	194.7	<0.0001

Appendix Table 3.2. Full Econometric Model Results for Northern Plains Wheat

Equation	Parameter	County	Estimate	Standard error	T-statistic	*P*-value
Loss Risk	Intercept		-0.0340	0.0004	-94.9	<0.0001
Loss Risk	County effect	38003	0.0730	0.0004	186.6	<0.0001
Loss Risk	County effect	38005	0.0652	0.0005	130.9	<0.0001
Loss Risk	County effect	38017	0.0700	0.0004	185.3	<0.0001
Loss Risk	County effect	38019	0.0873	0.0004	212.1	<0.0001
Loss Risk	County effect	38021	0.0527	0.0005	101.4	<0.0001
Loss Risk	County effect	38027	0.0723	0.0007	111.2	<0.0001
Loss Risk	County effect	38031	0.0927	0.0005	174.4	<0.0001
Loss Risk	County effect	38035	0.0815	0.0004	198.6	<0.0001
Loss Risk	County effect	38039	0.0979	0.0005	188.3	<0.0001
Loss Risk	County effect	38045	0.0690	0.0004	158.3	<0.0001
Loss Risk	County effect	38063	0.0790	0.0005	153.7	<0.0001
Loss Risk	County effect	38067	0.0368	0.0004	85.4	<0.0001
Loss Risk	County effect	38071	0.1067	0.0005	205.7	<0.0001
Loss Risk	County effect	38073	0.0501	0.0005	103.6	<0.0001
Loss Risk	County effect	38077	0.0481	0.0004	116.2	<0.0001
Loss Risk	County effect	38079	0.0870	0.0005	167.6	<0.0001
Loss Risk	County effect	38081	0.0373	0.0005	76.8	<0.0001
Loss Risk	County effect	38091	0.0919	0.0004	210.5	<0.0001
Loss Risk	County effect	38093	0.0896	0.0004	221.9	<0.0001
Loss Risk	County effect	38095	0.1132	0.0005	239.6	<0.0001
Loss Risk	County effect	38097	0.0644	0.0004	150.6	<0.0001
Loss Risk	County effect	38099	0.0677	0.0004	160.8	<0.0001
Loss Risk	County effect	38103	0.0648	0.0004	147.0	<0.0001
Loss Risk	County effect	46005	-0.0267	0.0008	-32.4	<0.0001
Loss Risk	County effect	46011	0.0322	0.0008	39.3	<0.0001
Loss Risk	County effect	46025	0.0036	0.0006	6.2	<0.0001
Loss Risk	County effect	46029	0.0085	0.0005	16.2	<0.0001
Loss Risk	County effect	46037	-0.0055	0.0005	-11.2	<0.0001
Loss Risk	County effect	46039	0.0233	0.0007	33.3	<0.0001
Loss Risk	County effect	46051	0.0054	0.0006	9.0	<0.0001
Loss Risk	County effect	46057	0.0405	0.0009	47.3	<0.0001
Loss Risk	County effect	46077	0.0114	0.0009	13.0	<0.0001
Loss Risk	County effect	46091	0.0048	0.0006	8.3	<0.0001
Loss Risk	County effect	46109	0.0000			
Loss Risk	Adjusted Base Rate		1.6799	0.0036	468.0	<0.0001
Loss Risk	Error Standard Error		0.0828	0.0000	1,931.5	<0.0001

Appendix Table 3.2 continued next page

Appendix Table 3.2 -- *continued from previous page*

Equation	Parameter	County	Estimate	Standard error	T-statistic	*P*-value
Coverage	Intercept		1.6217	0.0048	337.7	<0.0001
Coverage	County effect	38003	0.5356	0.0050	106.5	<0.0001
Coverage	County effect	38005	1.1282	0.0064	175.3	<0.0001
Coverage	County effect	38017	0.3646	0.0049	75.2	<0.0001
Coverage	County effect	38019	0.7727	0.0053	145.7	<0.0001
Coverage	County effect	38021	0.3054	0.0067	45.8	<0.0001
Coverage	County effect	38027	0.6338	0.0083	76.0	<0.0001
Coverage	County effect	38031	0.7416	0.0068	108.5	<0.0001
Coverage	County effect	38035	0.5312	0.0053	100.7	<0.0001
Coverage	County effect	38039	0.6829	0.0067	102.2	<0.0001
Coverage	County effect	38045	0.6454	0.0056	115.1	<0.0001
Coverage	County effect	38063	0.8545	0.0066	129.1	<0.0001
Coverage	County effect	38067	0.2866	0.0055	51.9	<0.0001
Coverage	County effect	38071	1.1692	0.0067	174.5	<0.0001
Coverage	County effect	38073	0.6582	0.0062	105.7	<0.0001
Coverage	County effect	38077	0.0758	0.0053	14.3	<0.0001
Coverage	County effect	38079	0.5377	0.0067	80.7	<0.0001
Coverage	County effect	38081	0.4018	0.0062	64.4	<0.0001
Coverage	County effect	38091	0.6516	0.0056	116.1	<0.0001
Coverage	County effect	38093	0.8806	0.0052	169.1	<0.0001
Coverage	County effect	38095	0.8030	0.0061	132.1	<0.0001
Coverage	County effect	38097	0.4946	0.0055	90.0	<0.0001
Coverage	County effect	38099	0.6349	0.0054	117.3	<0.0001
Coverage	County effect	38103	0.6118	0.0057	107.9	<0.0001
Coverage	County effect	46005	0.0807	0.0105	7.7	<0.0001
Coverage	County effect	46011	-0.2509	0.0105	-24.0	<0.0001
Coverage	County effect	46025	0.1195	0.0073	16.4	<0.0001
Coverage	County effect	46029	0.1074	0.0067	15.9	<0.0001
Coverage	County effect	46037	0.4732	0.0062	75.8	<0.0001
Coverage	County effect	46039	0.0451	0.0090	5.0	<0.0001
Coverage	County effect	46051	-0.0340	0.0077	-4.4	<0.0001
Coverage	County effect	46057	0.2016	0.0110	18.4	<0.0001
Coverage	County effect	46077	0.7625	0.0112	67.9	<0.0001
Coverage	County effect	46091	0.3642	0.0074	49.1	<0.0001
Coverage	County effect	46109	0.0000			
Coverage	Adjusted Base Rate		-6.3998	0.0462	-138.6	<0.0001
Coverage	Threshold2		0.3141	0.0012	271.9	<0.0001
Coverage	Threshold3		1.5279	0.0017	902.6	<0.0001
Coverage	Threshold4		2.7557	0.0019	1,430.1	<0.0001
Coverage	Threshold5		4.0912	0.0030	1,351.9	<0.0001
Coverage	Correlation Coefficient		0.1368	0.0008	179.0	<0.0001

Appendix 3.3: Development of Likelihood Function

To develop the likelihood function, I write the bivariate normal density as the product of the marginal density of ε_L and the conditional density of ε_c:

(A3.1) $\phi(e_L, e_c) = \phi(e_L)\phi(e_c | e_L)$

where $e_L \sim N(u_L, \sigma_L^2)$ and $(e_c \mid e_L) \sim N\left(u_c + r\dfrac{\sigma_c}{\sigma_L}(e_L), \sigma_c^2(1-r^2)\right)$. Because e_L and e_c are error terms in regression equations they will have zero means so long as the equations include intercept terms ($u_c = u_L = 0$). Moreover, because coverage is observed only discretely, however, it is necessary to assume that $\sigma_c = 1$ (Maddala). With these changes, the distributions become

(A3.2) $e_L \sim N(0, \sigma_L^2)$ and

(A3.3) $(e_c \mid e_L) \sim N\left(\dfrac{r}{\sigma_L}(e_L), (1-r^2)\right)$.

From the PDF and these distributions, a likelihood function can be developed.

The probability that c_j^* falls between limits c_{h-1} and c_h can be written as $\Pr(c_{h-1} \le c_j^* < c_h)$. Substituting for c_j^* (from equation (3.9)) yields $\Pr(c_{h-1} \le e_c + \beta_o^c + \sum_k \beta_k^c d_k + \beta_\rho^c \rho_{ABR,j} < c_h)$, which can be re-written as

(A3.4) $\Pr\left(c_{h-1} - \left(\beta_o^c + \sum_k \beta_k^c d_k + \beta_\rho^c \rho_{ABR,j}\right) \le e_c < c_h - \left(\beta_o^c + \sum_k \beta_k^c d_k + \beta_\rho^c \rho_{ABR,j}\right)\right)$.

Transforming e_c to zero mean and unit variance yields this expression:

(A3.5)

$$\Pr\left(\frac{c_{h-1} - \left(\beta_o^c + \sum_k \beta_k^c d_k + \beta_\rho^c \rho_{ABR,j}\right)}{(1-r^2)^{0.5}} - \frac{re_L}{\sigma_L (1-r^2)^{0.5}} \le \frac{e_c - \frac{re_L}{\sigma_L}}{(1-r^2)^{0.5}} < \frac{c_h - \left(\beta_o^c + \sum_k \beta_k^c d_k + \beta_\rho^c \rho_{ABR,j}\right)}{(1-r^2)^{0.5}} - \frac{re_L}{\sigma_L (1-r^2)^{0.5}} \right)$$

.

Rearranging terms and substituting gives

(A3.6) $\Pr\left(\tilde{c}_{h-1} \le e_c^n < \tilde{c}_h\right)$

where

$$\tilde{c}_h = \frac{c_h - \left(\beta_o^c + \sum_k \beta_k^c d_k + \beta_\rho^c \rho_{ABR,j}\right)}{(1-r^2)^{0.5}} - \frac{re_L}{\sigma_L (1-r^2)^{0.5}}$$

and

$$e_c^n = \frac{e_c - \frac{re_L}{\sigma_L}}{(1-r^2)^{0.5}}$$

and

$$e_L = \overline{L}_j - \beta_o^L - \sum_k \beta_k^L d_k - \beta_\rho^L \rho_{ABR,j} .$$

Then the contribution of farm *j* to the likelihood function can be written as:

(A3.7)
$$\ell_j = \frac{1}{\sigma_L} \phi\left(\frac{\overline{L}_j - \beta_o^L - \sum_k \beta_k^L d_k - \beta_\rho^L \rho_{ABR,j}}{\sigma_L} \right) \times \left(\int_{-\infty}^{\tilde{c}_1} \phi(e_{jc}^n) de \right)^{z_{j1}} \cdots \left(\int_{\tilde{c}_{h-1}}^{\tilde{c}_h} \phi(e_{jc}^n) de \right)^{z_{jh}} \cdots \left(\int_{\tilde{c}_H}^{\infty} \phi(e_{jc}^n) de \right)^{z_{jH}}$$

where ϕ is the standard normal density function, z_h =1 if coverage level h is selected, =0 otherwise, and other variables are as defined above. Evaluating the integrals yields:

$$\text{(A3.8)} \quad \ell_j = \frac{1}{\sigma_L}\phi\left(\frac{\overline{L}_j - \beta_o^L - \sum_k \beta_k^L d_k - \beta_\rho^L \rho_{ABR,j}}{\sigma_L}\right) \times \left(\Phi(\tilde{c}_1)\right)^{z_{j1}} \cdots \left(\Phi(\tilde{c}_h) - \Phi(\tilde{c}_{h-1})\right)^{z_{jh}} \cdots \left(1 - \Phi(\tilde{c}_H)\right)^{z_{jH}}$$

where Φ is the standard normal distribution function. Then the likelihood function can be written as

$$\text{(A3.9)} \quad \ell = \prod_j \ell_j \,.$$

Chapter 4: Using Farm-Specific Data in Crop Insurance Premium Rating: Can an Alternate Method of Premium Rating Reduce Asymmetric Information Effects?

Can premium rating methods be altered to reduce the asymmetric information effects identified in Chapter 3? Existing evidence appears to be mixed. On the one hand, premium rating problems are pervasive enough to be identified through empirical analysis which, in some cases, is based on crop insurance contract data and associated yield histories – information already available to the Risk Management Agency (RMA) (also see Makki and Somwaru). Persistent underwriting losses and the need for high premium subsidies rates to encourage broad participation (Glauber) tend to support the idea that premium rating can be improved. On the other hand, crop insurance yield histories – which include, at most, 10 years of data, are not sufficient to establish farm-specific, actuarially fair premiums with a reasonable level of statistical confidence. Collecting additional data, moreover, is unlikely to solve the problem. Because 10 years already accounts for 20-25 percent of a farmer's working lifetime, collecting a longer time-series on individual farms would be impractical.

Whether an alternate method of crop insurance premium rating can reduce asymmetric information effects depends, in part, on what constitutes a reduction. While the data constraints cited above imply that crop insurance premium rating errors are unavoidable at the farm level, previous research suggests that broad patterns of adverse selection do exist (e.g., Vandeveer and Loehman; Just, Calvin, and Quiggen; Makki and Somwaru). If alternative methods of premium rating are judged on the basis of their broad impact on actuarial fairness (at a county or even broader level), rather than farm-specific actuarial fairness, improvement may be possible. In this chapter, I test the hypothesis that using farm-specific yield histories to rate premiums – despite the uncertainties associated with small samples – could, on average, reduce overall deviation from actuarial fairness. "On average" means that premiums would be closer to actuarially fair for many or most farms while acknowledging that data limitations make accurate farm-specific premium rating impossible.

I focus on two types of error that could cause farm-specific premiums to deviate from actuarially fair levels. RMA premiums are rated using farm-specific information on yield mean but not yield variability. To the extent that farms which are uniform in yield mean are not uniform in terms of yield variability, RMA premiums are subject to error because the model used to rate premiums is not a theoretically correct model. I refer to this source of error as "modeling error." On the other hand, alternate premium rates are based on a theoretically correct model that incorporates a noisy measure of farm-specific variability. Because farm-specific estimates of yield variability are subject to sampling error, premium rates based on these estimates will also be subject to error. I refer to this source of error as "sampling error."

Is the modeling error inherent in the RMA procedure larger or smaller than the sampling error that would be inherent in an alternate method that incorporates farm-level estimated yield variability? I do not argue that this comparison should be made at the farm level (although I do report farm-level simulation results). I attempt only to discern whether there is an alternative set of premium rates that is likely to reduce rating error within a region or county (i.e., for which the average level of sampling error is less than the average level of modeling error at the region or county level).

In the next section, I develop more precise definitions of “modeling error” and “sampling error.” Following that, I develop a Monte Carlo simulation in which the modeling error in RMA premiums can be isolated and compared to the sampling error that would arise from use of farm-specific information on yield variability. To provide a realistic starting point for the simulations, I use a yield model developed from actual crop insurance yield histories. I show how the yield model can be used to define model and sampling error, then present and discuss results for the Corn Belt (non-irrigated corn) and Northern Plains (non-irrigated, continuous wheat).

Modeling Error and Sampling Error

Modeling error is the difference between the RMA premium rate and the theoretically correct premium rate, calculated with full knowledge of farm-specific yield distributions. Even if producers purchase revenue insurance rather than yield insurance (and many do), the driver of asymmetric information problems is limited information on the potential for yield loss[8]. I focus on yield-based premiums, but the results of the analysis would be useful in rating revenue insurance premiums, as well.

Because RMA premiums are intended to be actuarially fair, the theoretically correct premium is the expected indemnity, given a specific coverage level. The derivation of a theoretically correct premium rate begins with the definition of a yield

[8]Although crop insurance sales have shifted substantially to revenue-based products in recent years, yield risk is still critical in the rating of crop insurance premiums. The RMA base premium rate underlies producer premiums in Actual Production History (APH) (yield) insurance and also the most popular revenue insurance products, Crop Revenue Coverage (CRC) and Revenue Assurance (RA). Moreover, intra-county differences in producer revenue risk flow almost entirely from differences in yield risk. Thus, improving intra-county premium rating depends on improvement in the assessment of yield risk.

loss function. For any given farm in any given year the insurance indemnity per dollar of liability for yield-based coverage (*ex post*) can be written as:

$$(4.1) \qquad L_{jt} = \frac{p_t(\theta_j \tilde{y}_j - y_{jt})Z(\theta_j \tilde{y}_j > y_{jt})}{p_t \theta_j \tilde{y}_j} = \frac{(\theta_j \tilde{y}_j - y_{jt})Z(\theta_j \tilde{y}_j > y_{jt})}{\theta_j \tilde{y}_j}$$

where p_t is the guarantee price, θ_j is the coverage level (purchased by farm *j*) expressed as a proportion (e.g., for 75% coverage θ_j=.75), $\tilde{y}_j$ is the rate yield[9], y_{jt} is actual yield at time *t*, and *Z*(.) is an indictor function with value one when $\theta_j \tilde{y}_j > y_{jt}$, zero otherwise. Expected loss on farm *j* can be approximated by:

$$(4.2) \qquad E(L_j \mid \theta_j, \tilde{y}_j, N_j) = N_j^{-1} \sum_t L_{jt} = \frac{\sum_t Z(\theta_j \tilde{y}_j > y_{jt})(\theta_j \tilde{y}_j - y_{jt})}{N_j \theta_j \tilde{y}_j}$$

where N_j is number of yield observations. When N_j is large, equation (4.2) approximates the theoretically correct premium rate for yield-based insurance. Then modeling error can be defined as:

$$(4.3) \qquad ME_j^* = \rho(\theta_j \mid \tilde{y}_j) - E(L_j \mid \theta_j, \tilde{y}_j, N_j \to \infty)$$

---

[9] The RMA rate yield is essentially the mean yield for a producer's production history. The rate yield is used to make farm-specific adjustments to county-level premium rates calculated from county-level data (see Josephson et al.; RMA 2000a; RMA 2000b).

where $\rho(\theta_j \mid \tilde{y}_j)$ is the RMA premium rate for rate yield $\tilde{y}_j$ and coverage θ_j.

Unlike theoretical premiums, actual RMA premium rates include charges for loss reserves, prevented planting, or catastrophic events (Josephson *et al.*). To ensure a fair comparison between the RMA and theoretical rates, I use the minimum premium charged by RMA to any producer within a given county to represent these additional premium components. I add the county minimum to my theoretical premium rate and subtract the sum from the RMA premium rate to form ME_j:

$$(4.4) \qquad ME_j = \rho(\theta_j \big| \tilde{y}_j) - \Big(E(L_j \big| \theta_j, \tilde{y}_j, N \to \infty) + \min_k(\rho(\theta))\Big)$$

where $\min_k(\rho(\theta))$ is the lowest RMA premium for coverage θ in county *k*.

Sampling error is the difference between the theoretically correct premium rate, calculated from the correct yield distribution, and a theoretically correct premium rate calculated from a limited number of yield observations which are also drawn from the correct yield distribution. For any given sample, *s*, sampling error for farm *j* can be defined as:

$$(4.5) \qquad SE_{js} = E(L_{js} \big| \theta_j, \tilde{y}_j, N = N_j) - E(L_j \big| \theta_j, \tilde{y}_j, N \to \infty)$$

where N_j is equal to the number of yield observations in the yield history for farm *j*. Note that a comparison of absolute sampling error and absolute modeling error is, in fact, a comparison of the absolute distance of the RMA premium rate from the theoretically correct premium rate $E(L_j \big| \theta_j, \tilde{y}_j, N \to \infty) + \min_k(\rho(\theta))$ and the absolute distance of the alternate rate ($E(L_{js} \big| \theta_j, \tilde{y}_j, N = N_j)) + \min_k(\rho(\theta))$) from the theoretically correct premium rate. In other words, when absolute sampling error is less than absolute modeling error, the alternate premium is closer to the actuarially fair rate than is the RMA rate.

The distribution of SE_j depends on the underlying distribution of yields and the size of the yield samples drawn from the underlying distribution. (The crop insurance yield histories maintained by RMA include 10 or fewer yield observations. Only yield histories with 8 or more observations are used.) For any given yield distribution, the distribution of SE_j can be approximated using Monte Carlo methods:

$$\Pr\left(| SE_j | < X\right) = [S^{-1}\sum_s Z\left(| SE_{js} | < X\right)|S \to \infty)]$$

where S is the number of samples drawn from the yield distribution to make the approximation.

The probability that absolute sampling error is less than absolute modeling error at the farm level can be written as:

(4.6) $$\Pr\left(| SE_j | < | ME_j |\right) = [S^{-1}\sum_s Z\left(| SE_{js} | < | ME_j |\right)\big|S \to \infty)] \text{.}$$

As already noted, however, there is little hope of correctly rating premiums on a farm-by-farm basis. The probability that model error exceeds sampling error on a single farm gives little information on the potential for the broader success of a system of alternate premium rates. Taking liability-weighted averages of $| SE_j |$ and $| ME_j |$ for farms within a given county, however, allows calculation of county-level probabilities:

(4.7) $$\Pr\left(| SE_k | < | ME_k |\right) = [S^{-1}\sum_s Z\left(| SE_{ks} | < | ME_k |\right)\big|S \to \infty)]$$

where $|SE_{ks}| = \dfrac{\sum_{j\in k}\theta_j \tilde{y}_j |SE_{js}|}{\sum_{j\in k}\theta_j \tilde{y}_j}$ and $|ME_k| = \dfrac{\sum_{j\in k}\theta_j \tilde{y}_j |ME_j|}{\sum_{j\in k}\theta_j \tilde{y}_j}$.

Similar statistics can also be defined for a broader area (region):

(4.8) $\quad \Pr(|SE|<|ME|) = [S^{-1}\sum_s Z(|SE_s|<|ME|)\,|S\to\infty)]$

where $|SE_s| = \dfrac{\sum_k \sum_{j\in k}\theta_j \tilde{y}_j |SE_{js}|}{\sum_k \sum_{j\in k}\theta_j \tilde{y}_j}$ and $|ME| = \dfrac{\sum_k \sum_{j\in k}\theta_j \tilde{y}_j |ME_j|}{\sum_k \sum_{j\in k}\theta_j \tilde{y}_j}$.

Yield Model

Simulation of the theoretical premium rates (equations 4.4 and 4.5) and the probabilities in (4.6)-(4.8) requires a yield model which is representative of the premium rating situation faced by RMA. The yield models estimated in Chapter 2, based on actual crop insurance yield histories, are an obvious choice. While these models can represent the true yield distributions only to the extent that the yield histories accurately represent actual yield distributions, I assume that these are the true yield distributions for the purpose of the simulation model. While it is certain that these yield models are not accurate at the farm level, they are likely to be representative, overall, of the premium rating situation faced by RMA in the Corn Belt and Northern Plains.

Combining and rearranging equations (2.3)-(2.5) from Chapter 2 leads to the yield model:

(4.9) $\quad (\hat{\varepsilon}^{**}_{jt} s_j + \hat{\varepsilon}^{*}_{kt})\hat{\sigma}_{jkt} + \hat{y}_{jkt} = y_{jkt}$

where:

$\hat{\varepsilon}_{jt}^{**}$ is the normalized yield error for farm *j* at time *t*;

s_j is the standard deviation of the farm-specific error after removal of the county average error;

$\hat{\varepsilon}_{kt}^{*}$ is the county-specific error for county *k* at time *t*;

$\hat{\sigma}_{jkt}$ is the standard deviation, based on the variance model estimated in Chapter 2, for farm *j* in county *k* at time *t*;

$\hat{y}_{jkt}$ is the predicted yield, based on the yield model estimated in Chapter 2, and;

y_{jkt} is the observed yield.

Because the unit-specific errors, $\hat{\varepsilon}_{jt}^{**}$, are independent of one another and have zero mean (=0) and unit variance (=1) across farms (the result of normalization procedures described in Chapter 2), these observations can be used to specify an empirical distribution common to all farms. For datasets with relatively few observations, specific values are often under-represented in raw data; a problem that is typically resolved with smoothing techniques. The use of these techniques, however, requires additional assumptions about the distribution used for kernel smoothing, bandwidth, etc. Because the crop insurance yield datasets are so large (over 300,000 observations in the Corn Belt dataset), I elected not to smooth them. Smoothing was not necessary as a gap-filling measure and tended to add weight to outlying observations.

The other random term is the county-specific average error, $\hat{\varepsilon}^{*}_{kt}$. For the purpose of the simulation model I treat these errors as if they were fixed parameters. For a farm with a 10-year yield history, located in county *k*, 10 percent of draws from the empirical distribution are paired with each of the county errors calculated from the crop insurance yield data (there is one county-average yield per year over the 10-year yield history). This procedure ensures that broader weather patterns and other factors that may have affected yields on a broad scale are not over or under represented in the simulated yields for any specific farm. For example, many corn producers in Adams County, IL, experienced large yield losses due to flooding in 1993. Using a Monte-Carlo procedure to draw from the empirical distribution of county yield errors would risk including the 1993 county error more than once for some farms and not at all for others. An alternate procedure would be to introduce some county-level randomness by drawing "years" and assigning each farm its county-level error for that year. Over the course of 1,000 samples, however, it is unlikely that such a procedure would produce results that are significantly different from using the county average errors as if they were parameters, as proposed above.

Given the approach and assumptions outlined above, the yield model is

(4.10) $$(e^{**}s_j + \hat{\varepsilon}^{*}_{kt})\hat{\sigma}_{jkt} + \hat{y}_{jkt} = y^{*}_{jkt}$$

where e^{**} is a random draw from the empirical distribution and y^{*}_{jkt} is the simulated yield for farm *j*, county *k* and time *t*. Other terms are as previously defined.

Simulations

Simulations are based on yield insurance with 75 percent coverage, which is popular with producers in both the Corn Belt and the Northern Plains. Yield-based RMA premiums for 75 percent coverage are estimated from the base premium rate (BPR) found in the crop insurance contract data. The BPR underlies Actual Production History (APH), Crop Revenue Coverage (CRC), and Revenue Assurance (RA) premiums. Because these three products are very popular, the BPR is available for contracts covering about 90 percent of corn insurance liability in the Corn Belt corn data and 95 percent in the Northern Plains wheat data.

The BPR, however, is specific to the level of coverage purchased by the producer. To obtain an RMA premium rate for a common level of coverage, I use the adjusted base rate (ABR). The ABR is calculated from the producer's average yield (rate yield) and actuarial tables for the county, crop, and production practice (RMA, 2000a) and represents loss risk for a given level of insurance coverage. In 2001, for example, the ABR is calculated for 75 percent coverage. The BPR is then calculated from the ABR using multiplicative adjustment terms: $\rho_{j\theta} = \varphi_{\theta}\rho_{ABR,j}$, where $\rho_{ABR,j}$ is the ABR for farm *j*, $\rho_{j\theta}$ is the BPR for coverage level θ on farm *j*, and φ_{θ} is the coverage differential for the coverage selected by the producer. The ABR can be obtained from the BPR by reversing this process: $\rho_{ABR,j} = \varphi_{\theta}^{-1}\rho_{j\theta}$.

To simulate the theoretically correct premiums, I draw values of e^{**} from the empirical distribution, convert them to simulated yields using (4.10) and find the farm-specific expected indemnity (actuarially fair premium rate) using (4.2). In anticipation of the Monte Carlo analysis, I draw up to 10 simulated yield observations per sample for each farm (one for each year of a specific farm's yield history) and draw a total of 1000 samples. That is, for each farm I draw up to 10,000 simulated yields.

The actual theoretical premium rate (the second term on the RHS of (4.5); $E(L_j \mid \theta_j, \tilde{y}_j, N \to \infty))$ is estimated from (4.2) using all 10,000 draws (N=10,000). A total of 1,000 of the sample-specific theoretical premium rates (the first term on the RHS of (4.5); $E(L_{js} \mid \theta_j, \tilde{y}_j, N = N_j))$ are calculated for each farm using (4.2). Using these simulated values along with the RMA premium rate described above, farm-specific estimates of modeling error and sampling error can be calculated using (4.4) and (4.5), respectively. The probability that modeling error is greater than sampling error can be calculated for the farm, county, and region from (4.6), (4.7), and (4.8), respectively. The average alternate premium rate, used for purpose of discerning the general direction of premium rate changes under the alternate rating mechanism, is

$$S^{-1} \sum_s E(L_{js} \mid \theta_j, \tilde{y}_j, N = N_j) + \min_k (\rho(\theta)) .$$

Results

Regional results are given in Table 4.1. For both Corn Belt corn and Northern Plains wheat the probability that model error is larger than sampling error is reported to be 1.000 because the average absolute model error exceeds sampling error for each of the 1,000 samples drawn for the Monte Carlo analysis. In the Corn Belt the alternate premium rates are, on average, \$0.0024/\$ liability below RMA premium rates. In the Northern Plains alternate premium rates are, on average, \$0.0146/\$ liability higher than RMA rates.

At the county level, results are mixed. Average absolute modeling error is (statistically) significantly larger than average absolute sampling error in 59 of 69 Corn Belt counties (85 percent; see Appendix Table 4.1) and 18 of 36 Northern Plains counties (50 percent; see Appendix Table 4.2). The average difference between alternate and RMA premium rates also varies considerably among counties, ranging from \$0.001/\$ liability to more than \$0.010/\$ liability in the Corn Belt data (Appendix Table 4.1). A similar range is observed in Northern Plains counties (Appendix Table 4.2).

Table 4.1 Sampling Error, Model Error, and Difference Between Sampling and Model Error at the Region, County, and Farm Level

	Corn Belt Corn	Northern Plains Wheat
	Dollars per Dollar of Liability	
Average Absolute Sampling Error	0.0114	0.0331
Average Absolute Model Error	0.0210	0.0419
Avg. Difference, Sampling less Model Error	-0.0096	-0.0088
Pr (Sampling Error < Model Error)	1.000	1.000
Average Difference, Alternate less RMA Premium Rate	-0.0024	0.0146
Number of Counties	69	36
Counties with Pr (Sampling Error < Model Error) ≥ 0.95	59	18
Percent of Counties with Pr (Sampling Error < Model Error) ≥ 0.95	85.5	50.0
Number of Farms	25,491	7,244
Farms with Pr (Sampling Error < Model Error) ≥ 0.95	7,032	1,220
Percent of Farms with Pr (Sampling Error < Model Error) ≥ 0.95	27.6	16.8

Farm level results illustrate the difficulty in accurately rating premiums at the farm level. In the Corn Belt, modeling error is (statistically) significantly larger than sampling error for only 7,032 of more than 25,000 farms in the dataset (28 percent; Table 4.1). Of these 7,032 farms, the alternate premium rate is larger than the RMA premium rate for only 540 farms (Table 4.2). For these farms, the alternate premium rates are, on average, \$0.095/\$ liability higher than RMA premiums (see Table 4.2, top row in Corn Belt section). For other Corn Belt farms (6,492 farms), alternate premiums are, on average, lower than the RMA premium by \$0.031/\$ liability. For Corn Belt farms where modeling error was not found to be statistically significantly larger than sampling error, average premium differences are smaller. On farms where alternate premiums are higher than RMA premiums, the average difference is $0.026/$ liability. On farms where alternate premiums are lower, the average difference is $0.009/$ liability (see Table 4.2, second row). These premium rate changes, if implemented would be large relative to existing premiums. For Corn Belt corn the average ABR is $0.062/$ liability.

In the Northern Plains, the Monte Carlo analysis shows a statistically significant difference between modeling and sampling error on 1,220 of 7,244 (about 17 percent; Table 4.1, fourth row). Of those 1,220 farms, the alternate premium was higher than the RMA rate for 161 farms (Table 4.2, fourth row). For this group, alternate premiums rates are, on average, $0.26/$ liability higher than RMA premium rates. The large majority of 1,220 farms where modeling error is (statistically) significantly larger than sampling error – 1,059 farms – were assigned alternate premium rates which were, on average, $0.055/$ liability lower than the RMA rate (Table 4.2; fourth row).

For 6,024 farms where modeling error was not shown to be statistically significantly larger than sampling error, premium changes would be smaller, be still large relative to existing ABRs. For 3,719 farms where alternate premium rates are higher than RMA rates, the average difference in premium rates was $0.054/$ liability (Table 4.2, row 5). For the 2,305 farms with alternate rates lower than the RMA rate, the difference was about $0.022/$ liability. While these differences are less than on farms where there is a significant difference between modeling and sampling error, they are still large relative to the average ABR – $0.135/$ liability for Northern Plains wheat.

Table 4.2. Number of Farms, Insured Production, and Mean Difference Between RMA Premium Rates and Alternate Rates Based on a Farm-Specific Measure of Yield Variability

		Alternate Premium Rate > RMA Rate	Alternate Premium Rate < RMA Rate	All Farms
Corn Belt Corn				
Modeling Error > Sampling Error with 95% Confidence	Number of Farms	540	6,492	7,032
	Insured Production (mil. bu.)	4.5	89.58	94.08
	Mean Rate Difference[1]	0.0946	-0.0314	-0.0253
Modeling Error and Sampling Error NOT Significantly Different	Number of Farms	10,088	8,371	18,459
	Insured Production (mil. bu.)	96.833	93.183	190.016
	Mean Rate Difference	0.0264	-0.0091	0.0090
All Farms	Number of Farms	10,628	14,863	25,491
	Insured Production (mil. bu.)	101.333	182.763	284.096
	Mean Rate Difference	0.0294	-0.0200	-0.0024
Northern Plains Wheat				
Modeling Error > Sampling Error with 95% Confidence	Number of Farms	161	1,059	1,220
	Insured Production (mil. bu.)	0.68	5.09	5.77
	Mean Rate Difference	0.2603	-0.0552	-0.0180
Modeling Error and Sampling Error NOT Significantly Different	Number of Farms	3,719	2,305	6,024
	Insured Production (mil. bu.)	15.75	10.10	25.85
	Mean Rate Difference	0.0542	-0.0218	0.0245
All Farms	Number of Farms	3,880	3,364	7,244
	Insured Production (mil. bu.)	16.43	15.19	31.62
	Mean Rate Difference	0.0627	-0.0330	0.0146

[1]Difference is the average alternate premium rate less the RMA premium rate

The farm level results can be seen more clearly in histograms showing the density of crop insurance liability against the difference in alternate and RMA premium rates. In the Corn Belt (figure 4.1), premium rate differences ranged from -\$0.043/\$ liability to more than \$0.200/\$ liability, but nearly all the changes are confined to the range between -\$0.153/\$ liability (1st percentile) and \$0.125/\$ liability (99th percentile). The large majority of changes lie between -\$0.030/\$ liability (5th percentile) and \$0.055/\$ liability (95th percentile). The solid colored portion of the bars in Figure 4.1 represents farms for which modeling error is significantly larger than sampling error. As already noted, only 28 percent of crop insurance liability is on farms where modeling error is demonstrably larger than sampling error. Moreover, significance appears to be concentrated among farms that would receive a premium rate decrease if the alternate premiums were adopted. As already shown in Table 4.2, only a very small number of farms can be shown to have expected indemnities (alternate premium rates) that are statistically significantly higher than RMA rates. For about 20 percent of insurance liability, alternative premiums exceed RMA premiums by between \$0.004/\$ liability (75th percentile) and \$0.055/\$ liability (95th percentile), but almost none of these changes are statistically different from zero.

In the Northern Plains (figure 4.2), premium rate differences range from about \$-0.35/\$ liability to about \$0.50/\$ liability, although nearly all differences are contained between -\$0.095/\$ liability (1st percentile) and \$0.250/\$ liability (99th percentile). The large majority of changes lie between -\$0.050/\$ liability (5th percentile) and \$0.135/\$ liability (95th percentile). Although the premium changes for Northern Plains wheat cover a broader range than changes for Corn Belt corn, the story is quite similar: Only 16 percent of crop insurance liability is on farms where modeling error is demonstrably larger than sampling error and it appears to be concentrated among farms that would experience a net decrease in premium rate if the alternative premiums were adopted. Given that premium rates appear to be low, on average, improved rating may depend on increasing premiums on some farms. For about 20 percent of insurance liability, alternative premiums exceed RMA premiums by between \$0.035/\$ liability (75th percentile) and \$0.135/\$ liability (95th percentile), but almost none of these changes are statistically different from zero.

Figure 4.1. Distribution of Difference Between Alternate and RMA Premium Rates
Corn Belt Corn

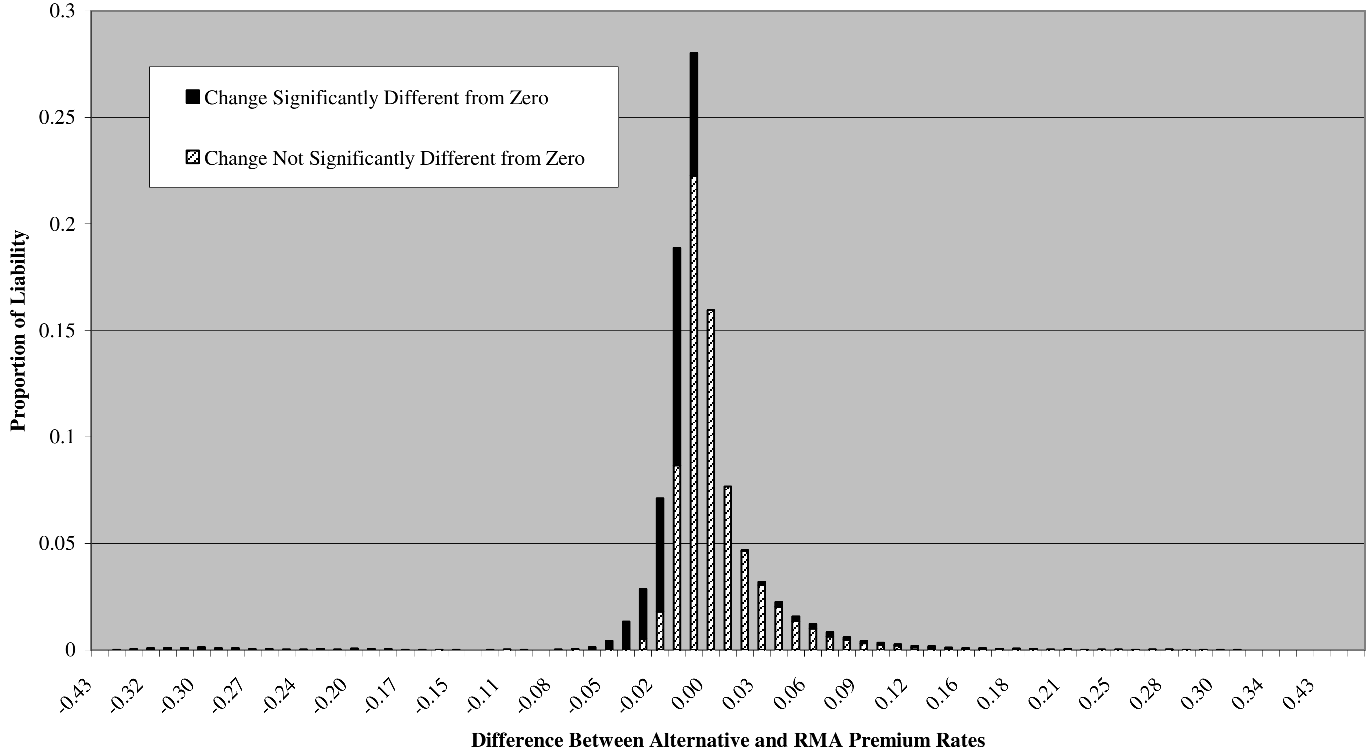
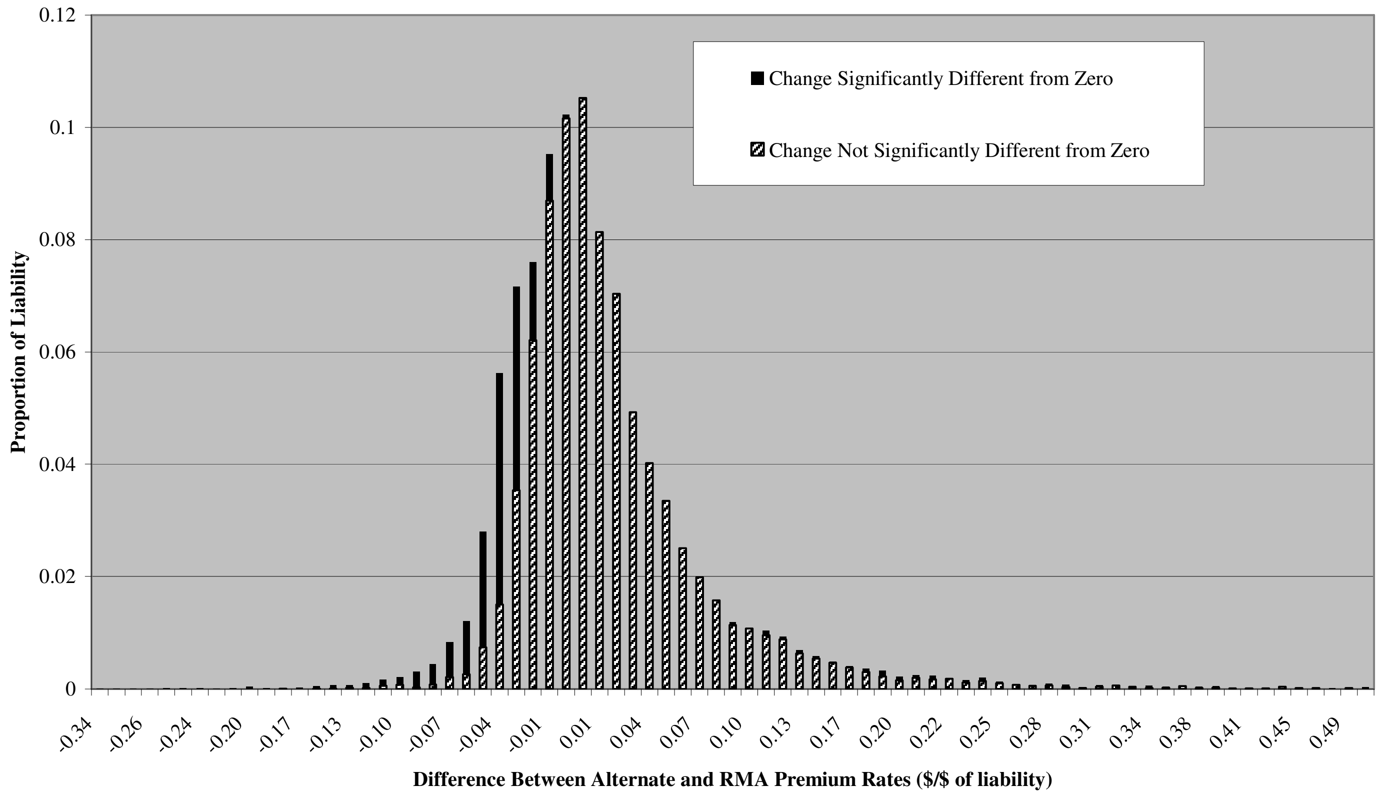


Figure 4.2. Distribution of Difference Between Alternate and RMA Premium Rates
Northern Plains Wheat



Conclusion

The analysis indicates that the use of farm-specific information on yield variability could improve the actuarial soundness of premium rating, if improvement is sought only at a sufficiently broad scale. At a regional scale, the result is a strong one: Modeling error is found to exceed sampling error in every one of the 1,000 Monte Carlo iterations for both the Corn Belt and Northern Plains simulations. At the county level, however, results are not as strong, particularly in the Northern Plains. In 85 percent of Corn Belt counties, modeling error exceeds sampling error with at least 95 percent confidence. Modeling error exceeds sampling error with at least 95 percent confidence in only 50 percent of Northern Plains counties. Accepting a lower level of confidence helps a little: At 90 percent and 80 percent confidence, modeling error exceeds sampling error in 56 and 61 percent of Northern Plains counties, respectively.

Nonetheless, Chapter 4 results suggest that premium rating does contribute to inter-regional differences in underwriting loss. In the Corn Belt, alternate premiums would, on average, be lower by about 0.24 cents per dollar of liability. On the 284 million bushels insured under the contracts in the data used for this study, valued at $2.50 per bushel, total premiums would be reduced by about $1.7 million per year. Because of the premium subsidy, of course, the total reduction in farm-paid premiums would be less. In the Northern Plains, on the 31.6 million bushels of insured wheat, if valued at $4 per bushel, adoption of alternate premiums would result in a $2 million per year increase in total premiums. Of course, these datasets account for a small portion of overall annual crop insurance liability in these crops and regions, let alone all crops and all regions. The analysis, moreover, does not account for possible changes in crop insurance purchase.

Farm level results highlight the difficulty of transitioning to and maintaining a set of alternate premiums. If the alternate premiums were adopted, large groups of producers in both regions would face substantial premium increases – increases for which RMA would have little statistical evidence. RMA would certainly have to grapple with the question of when to treat a bad year (or years) as simply a bad year – the kind of situation crop insurance is intended to protect against – and when to consider that bad year a signal of higher risk that should trigger a higher rate. Of course, current premium rating methodology implicitly answers these questions and the results of this analysis suggest that there is room to be more aggressive in adjusting premium rates.

Of course, roughly half of all producers would see premiums decline if the alternate premiums were implemented. While RMA is unlikely to be challenged by producers who would be offered lower rates, farm-level results show that RMA may also be on more solid statistical ground in offering discounts to customers with good loss histories. In fact, a system of experience-based premium discounts (but not surcharges) has been authorized by Congress and a possible schedule of discounts was recently proposed by Rejesus *et al*. Of course, the effectiveness of discounts without surcharges would vary considerably between the Corn Belt and the Northern Plains. In the Corn Belt, discounts could make sense in the absence of surcharges, so long as the discounts are not too large. In the Northern Plains, discounts without offsetting surcharges could exacerbate underwriting losses. This analysis, and the practical evidence of persistent underwriting losses in the Northern Plains (and elsewhere), indicate that premiums are, on average, too low.

Appendix Table 4.1. Simulation Model Results, by County, for Corn Belt Corn

County	Number of Farms	Insured Production (bushels)	Avg. Absolute Sampling Error	Avg. Absolute Model Error	Probability Modeling Error > Sampling Error	Avg. Diff., Alternate less RMA Rate
17001	495	3,830,342	0.0152	0.0328	1.000	-0.0222
17003	6	45,941	0.0267	0.1411	1.000	-0.0933
17009	93	1,023,569	0.0208	0.0483	0.996	-0.0277
17011	769	10,524,727	0.0072	0.0169	1.000	-0.0109
17013	20	147,754	0.0091	0.0288	1.000	-0.0236
17015	135	4,521,613	0.0066	0.0183	1.000	-0.0138
17017	140	1,731,345	0.0080	0.0249	1.000	-0.0204
17019	1304	12,381,163	0.0126	0.0150	1.000	-0.0014
17021	554	6,566,364	0.0052	0.0072	1.000	-0.0013
17029	320	3,216,151	0.0059	0.0116	1.000	-0.0059
17031	10	55,105	0.0394	0.1147	1.000	0.1040
17039	320	3,814,526	0.0070	0.0213	1.000	-0.0197
17041	350	3,895,828	0.0058	0.0130	1.000	-0.0102
17045	350	4,780,573	0.0046	0.0156	1.000	-0.0111
17053	743	6,395,798	0.0174	0.0282	1.000	0.0233
17057	308	3,626,575	0.0149	0.0311	1.000	-0.0151
17059	76	1,677,293	0.0206	0.0336	0.993	0.0252
17061	154	2,145,839	0.0166	0.0499	1.000	-0.0373
17063	523	5,387,674	0.0136	0.0202	1.000	0.0080
17067	565	4,926,040	0.0138	0.0318	1.000	-0.0260
17071	288	3,646,604	0.0107	0.0125	0.978	0.0010

Appendix Table 4.1 continued next page

Appendix Table 4.1 -- *continued from previous page*

County	Number of Farms	Insured Production (bushels)	Avg. Absolute Sampling Error	Avg. Absolute Model Error	Prob. Average Difference Less Than Zero	Avg. Difference, Alternate less RMA Rate
17073	745	9,566,408	0.0098	0.0149	1.000	-0.0002
17075	1346	14,415,353	0.0176	0.0238	1.000	0.0110
17077	21	150,004	0.0202	0.0930	1.000	-0.0890
17083	94	1,143,267	0.0110	0.0140	0.928	0.0046
17087	13	150,446	0.0303	0.1219	1.000	-0.0764
17091	793	8,163,252	0.0148	0.0207	1.000	-0.0001
17093	235	2,704,281	0.0180	0.0310	1.000	0.0286
17095	381	4,111,280	0.0087	0.0146	1.000	-0.0026
17097	4	173,329	0.0155	0.0201	0.768	0.0155
17099	1166	11,754,606	0.0118	0.0121	0.803	0.0054
17103	466	7,524,117	0.0082	0.0088	0.859	0.0037
17105	1975	16,729,360	0.0170	0.0267	1.000	0.0250
17107	664	7,903,640	0.0076	0.0144	1.000	-0.0013
17109	377	4,737,358	0.0073	0.0180	1.000	-0.0141
17113	1160	12,435,805	0.0084	0.0108	1.000	-0.0045
17115	510	6,085,786	0.0058	0.0091	1.000	-0.0034
17117	269	2,774,366	0.0094	0.0316	1.000	-0.0271
17123	423	4,529,180	0.0095	0.0119	1.000	0.0022
17125	176	1,978,256	0.0129	0.0928	1.000	-0.0876
17127	6	44,567	0.0359	0.0500	0.855	0.0120
17129	262	3,302,585	0.0071	0.0249	1.000	-0.0194

Appendix Table 4.1 continued next page

Appendix Table 4.1 -- *continued from previous page*

County	Number of Farms	Insured Production (bushels)	Avg. Absolute Sampling Error	Avg. Absolute Model Error	Prob. Average Difference Less Than Zero	Avg. Difference, Alternate less RMA Rate
17131	397	4,432,910	0.0129	0.0209	1.000	-0.0050
17133	145	941,713	0.0321	0.0605	1.000	0.0257
17137	181	1,914,212	0.0059	0.0137	1.000	-0.0077
17139	310	3,555,468	0.0088	0.0093	0.774	0.0085
17141	289	5,766,673	0.0099	0.0110	0.902	0.0087
17143	293	2,875,329	0.0118	0.0274	1.000	-0.0143
17147	487	6,496,606	0.0075	0.0087	0.998	0.0083
17149	145	2,276,467	0.0130	0.0294	1.000	-0.0164
17151	5	45,588	0.0120	0.1461	1.000	-0.1232
17153	4	19,114	0.0289	0.0046	0.000	0.0021
17155	159	1,791,669	0.0115	0.0116	0.606	0.0036
17157	47	444,936	0.0386	0.0324	0.189	-0.0018
17161	158	1,847,829	0.0150	0.0240	0.999	0.0006
17163	121	736,448	0.0149	0.0308	1.000	0.0061
17167	333	4,717,649	0.0073	0.0318	1.000	-0.0263
17169	184	1,586,157	0.0093	0.0315	1.000	-0.0290
17171	83	1,058,343	0.0063	0.0099	0.999	-0.0008
17175	327	3,922,749	0.0082	0.0091	0.922	0.0063
17179	430	4,612,378	0.0078	0.0342	1.000	-0.0279
17181	23	233,466	0.0100	0.0631	1.000	-0.0618

Appendix Table 4.1 continued next page

Appendix Table 4.1 -- *continued from previous page*

County	Number of Farms	Insured Production (bushels)	Avg. Absolute Sampling Error	Avg. Absolute Model Error	Prob. Average Difference Less Than Zero	Avg. Difference, Alternate less RMA Rate
17183	740	7,358,191	0.0165	0.0311	1.000	-0.0225
17187	538	6,556,423	0.0080	0.0098	1.000	-0.0012
17193	153	1,979,841	0.0199	0.1174	1.000	-0.1018
17195	298	5,184,436	0.0136	0.0326	1.000	-0.0158
17197	473	3,782,134	0.0215	0.0325	1.000	0.0247
17199	21	214,966	0.0444	0.1594	1.000	-0.1080
17203	538	5,025,099	0.0109	0.0161	1.000	0.0034

Appendix Table 4.2. Simulation Results, by County, for Northern Plains Wheat

County	Number of Farms	Insured Production (bushels)	Avg. Absolute Sampling Error	Avg. Absolute Model Error	Probability Modeling Error > Sampling Error	Avg. Difference, Alternate less RMA Rate
38003	523	2,122,674	0.0263	0.0423	1.000	-0.0298
38005	234	888,615	0.0322	0.0529	1.000	-0.0429
38017	483	2,800,233	0.0239	0.0318	1.000	-0.0151
38019	499	2,300,184	0.0671	0.1026	1.000	0.0993
38021	112	801,911	0.0189	0.0179	0.295	-0.0028
38027	103	382,862	0.0325	0.0298	0.258	0.0033
38031	163	810,144	0.0315	0.0323	0.633	0.0136
38035	329	1,726,486	0.0259	0.0276	0.890	0.0060
38039	186	751,668	0.0347	0.0404	0.986	0.0295
38045	284	1,269,369	0.0306	0.0272	0.047	0.0123
38063	193	738,695	0.0388	0.0421	0.881	0.0339
38067	272	1,217,504	0.0422	0.0601	1.000	0.0561
38071	191	1,035,185	0.0568	0.0736	0.950	0.0636
38073	189	717,172	0.0196	0.0218	0.940	-0.0035
38077	284	1,194,102	0.0190	0.0350	1.000	-0.0243
38079	171	617,210	0.0335	0.0376	0.898	0.0128
38081	183	715,865	0.0230	0.0331	1.000	-0.0226
38091	287	1,318,341	0.0286	0.0369	0.999	0.0358
38093	487	2,317,639	0.0299	0.0317	0.883	-0.0062
38095	275	1,358,020	0.0460	0.0542	0.991	0.0437
38097	262	1,149,627	0.0281	0.0271	0.273	0.0178

Appendix Table 4.2 continued next page

Appendix Table 4.2 -- *continued from previous page*

County	Number of Farms	Insured Production (bushels)	Avg. Absolute Sampling Error	Avg. Absolute Model Error	Prob. Average Difference Less Than Zero	Avg. Difference, Alternate less RMA Rate
38099	319	1,533,121	0.0361	0.0385	0.858	0.0287
38103	340	1,346,249	0.0304	0.0381	0.999	-0.0023
46005	28	129,690	0.0226	0.0250	0.750	0.0236
46011	24	36,743	0.0242	0.0213	0.244	0.0166
46025	92	225,884	0.0348	0.0348	0.528	0.0184
46029	108	228,713	0.0265	0.0278	0.729	0.0235
46037	143	484,563	0.0274	0.0242	0.084	0.0208
46039	41	58,943	0.0372	0.0591	0.997	0.0573
46051	56	115,090	0.0316	0.0434	0.996	0.0426
46057	20	35,870	0.0299	0.0496	0.998	0.0489
46077	17	34,095	0.0341	0.0337	0.512	0.0328
46079	2	2,182	0.0272	0.0555	0.959	0.0555
46091	86	302,475	0.0224	0.0273	0.976	-0.0151
46097	7	30,193	0.0420	0.0675	0.967	0.0675
46109	185	528,595	0.0195	0.0207	0.784	0.0201

Chapter 5: Conclusion

For much of its history, participation in the Federal crop insurance program has been low while underwriting losses have been large. Premium subsidy increases in 1995 and 2000 raised participation rates to roughly 80 percent of eligible acres and, on a national scale, brought premiums and indemnities roughly into balance. Nonetheless, problems remain. Public expenditure for premium subsidies is large ($2 billion in 2003 alone), even as some regions continue to experience underwriting losses.

The RMA premium rating model may be part of the problem. In theory, premiums should equal the expected crop insurance indemnity on a farm-specific basis. The RMA model uses a county-level estimate of yield loss risk with a farm-specific adjustment based on yield averages. If farms located in the same area have the same average yield but differ in terms of yield loss risk, the RMA premium rating model will not capture this difference, leading to asymmetric information problems. If premium rating error can be reduced, on average, through the use of farm-specific data, it may be possible to reduce underwriting losses and, perhaps, premium subsidy expenditures. The fact that underwriting losses persist in only some regions suggests that there are regional differences in asymmetric information.

The focus of Chapters 2, 3, and 4, was on understanding the relationship between RMA crop insurance yield history data and RMA rating of yield loss risk. Comparing yield loss history to yield based premiums shows wide variation across producers. In Table 3.3, for example, the difference between average yield loss (calculated from yield histories) and RMA yield-based premium rates (assuming 75 percent coverage for both) varies from $-0.045 to 0.043 in the Corn Belt and $-0.053 to $0.209 in the Northern

Plains. The statistical question underlying the analysis of Chapters 3 and 4 is this: Could these observed differences between farm-specific yield loss, implied by yield data, and RMA yield-based premium rates have arisen by chance, or are they likely to represent real differences? If these differences are real, could they explain differences in underwriting losses across regions? And, more to the point, can yield history data be used to improve crop insurance premium rating?

There is a lot of evidence to suggest that differences between farm-specific yield loss and RMA premium rates are real, particularly where crop production is relatively risky. Chapter 3 was constructed around a test of asymmetric information in crop insurance premium rating, although the term "asymmetric information" is a bit of misnomer given that all data used in the analysis was obtained from RMA. In this case, the term refers to information that may be embedded in the RMA data but is not used by RMA in premium rating.

Chapter 3 showed that producer behavior (selection of crop insurance coverage level) is linked to premium rating errors, which were isolated by regressing average yield loss (from yield histories) on an RMA measure of yield loss risk (the Adjusted Base Rate or ABR). Producers with high yield loss relative to their ABR tend to purchase more insurance (higher coverage) than other producers, as predicted by the theory developed by Rothsechild and Stiglitz. The affect of asymmetric information is larger in the Northern Plains, where crop production is riskier than it is in the Corn Belt (see Tables 2.7 and 2.8).

The analysis in Chapter 4 appears to confirm the basic results of Chapter 3; that average yield losses estimated from production history are systematically different from

RMA premium rates (the ABR). In the simulation model developed in Chapter 4, I assumed that existing yield histories accurately reflect farm-specific yield distributions. Given this assumption, theoretically correct premiums, adjusted for loss reserves, etc., were calculated and compared to (1) RMA premium rates and (2) theoretically correct "alternate" premiums calculated from small samples drawn from the farm-specific yield distributions.

These alternate premium rates came closer to the theoretically correct premium rates, on average when each region is viewed in whole. In every one of the 1,000 Monte Carlo simulations run on each region, the alternate premium rates calculated from small samples using theoretically correct rating methods were, on average, closer to theoretically correct rates than are the RMA rates (the ABR). Here, the results of Chapter 4 strongly support those of Chapter 3: In a broad sense, both models confirm that asymmetric information is an issue in both regions. The results suggest that it would be possible, on average, to improve the actuarial correctness of crop insurance premiums using farm-specific information on yield variability.

Broadly, the analysis suggests that premium rating does contribute to inter-regional differences in underwriting loss. In the Corn Belt, RMA premiums (ABRs) were estimated to be, on average across all farms in the region, \$0.0024/\$ liability too high. This result is consistent with the results drawn from Table 3.8, namely that net return to coverage is modestly negative, on average, for a wide range of producers. In the Northern Plains, RMA premiums (ABRs) were, on average, \$0.145/\$ liability too low. This result is also consistent with the Northern Plains results reported in Table 3.8, in the sense that positive net returns are the norm among producers with positive asymmetric information effects and were generally more common among Northern Plains producers than among Corn Belt producers.

At the county and farm level, however, simulation results paint a more mixed picture and point toward potential difficulties in implementation of alternate premium rates. Alternate premiums are, on average, closer to the theoretical (actuarially fair) rate with probability of at least 0.95 in 85 percent of Corn Belt counties but only 50 percent of Northern Plains counties. At the farm level, 33 percent of Corn Belt insurance liability (in this sample; 94 million bushels of 284 million bushels) is on farms where alternate premium rates can be shown (with probability of at least 0.95) to be closer to theoretical (actuarially fair) rates than are RMA premiums. In the Northern Plains, only 18 percent of insurance liability is on farms where alternate rates are demonstrably closer to actuarially fair rates when compared to RMA rates. In both regions, for a large majority of cases for which farm-level alternate premiums are demonstrably closer than RMA rates to actuarially fair rates, the alternate premium rate is, on average, lower than the RMA rate.

If RMA officials determine to go forward with a premium rating mechanism that incorporates farm-specific information on yield variability, transition rules would be a critically important component of the new policy. Assuming current rates (ABRs) are taken as starting point, what criteria would be used to judge whether alternate premium rates are "different enough" to warrant a rate adjustment? One could think of this question in terms of updating premium rates based on the additional information gathered through annual yield reports. When is a yield loss (the idiosyncratic portion only) large enough that it should be viewed as evidence of need for a premium increase, rather than a normal fluctuation that is to be expected? It is easy to imagine rules involving some type of Bayesian updating of yield distributions as additional yield information becomes available. Moreover, given that current premium rating rules limit producer premium increases to 20 percent per year, some additional mechanism may be needed to limit the rate of change.

The premium rate discounts recently proposed by Rejesus *et al*., would be at least a partial implementation of alternate premiums. Premium discounts would be based on 5 years worth of yield performance. The authors argue that 5 years is a compromise between sampling error associated with small samples and the need to provide discounts based on a reasonably short period of good performance. Evidence from Chapter 4 suggests the appropriateness of discounts, in a statistical sense, is difficult to judge even based on 8-10 years worth of data, let alone 5 years. The Rejesus *et al*., approach is an attempt to improve premium rating, on average, with the understanding that farm-specific premiums will still be erroneously rated. Reducing the overall level of deviation from actuarial fairness would also require premium surcharges, which would be somewhat more difficult to implement both statistically and politically, as highlighted by the potential large increases and lack of statistical evidence in support of farm level premium increases (see figures 4.1 and 4.2).

Updating rules could be explored using a dynamic version of the simulation model developed in Chapter 4. For each model farm, an initial draw of 4-10 yield observations could be used to calculate an initial premium rate. (RMA requires at least 4 actual yields before transitional or "T" yields are dropped from the rate yield calculation, but uses no more than 10 yields.) Subsequent yield draws (one yield at a time to simulate the inflow of annual yield data) could be used to update premium rates using various updating rules. Can updating rules be devised to encourage premium rates to converge toward the theoretically correct premium rate? Can premium rates come close to theoretical rates within a reasonable period of time (3-5 years) without being susceptible to following outliers? A Monte Carlo simulation that involved 1,000 or more of these simulation experiments could provide useful insight on the design of premium discount and surcharge mechanisms.

Two of the many challenges that would face policy makers in devising specific update rules would be the recognition of yield trends and identification of yield outliers. As shown in Chapter 2, it can be difficult to capture time trends. Moreover, parametric functional forms that could be useful in defining outliers are generally not an exact fit for crop yield data. Some parametric distributions (gamma-minus in particular) may come close enough to be useful in developing rules about the treatment of yield observation that are far from the bulk of previous observations.

Ironically, the transition to an alternate method of premium rating may be easier in the Corn Belt – where the effect of asymmetric information is relatively small – than in the Northern Plains – where asymmetric information effects are larger. Partly, that is true because Corn Belt yields are more consistent (less variable) over time and across farms. Premium rate adjustments, where they are made for Corn Belt farms, would be smaller than the adjustments that would be required in the Northern Plains. Moreover, farm-level differences between RMA and alternative rates are more likely to be statistically significant at the farm level for Corn Belt farm than for Northern Plains farms.

References

Atwood J., S. Shaik, and M. Watts. "Are Crop Yields Normally Distributed? A Reexamination." *Amer. J. Agr. Econ.* 85(November 2003): 888-901.

Babcock, Bruce A., Chad E. Hart, and Dermot J. Hayes. "Actuarial Fairness of Crop Insurance Rates with Constant Rate Relativities" *Amer. J. Agr. Econ.* 86(August 2004): 563-575.

Botts, R.R., and Boles, J.N. 1957. "Use of Normal-Curve Theory in Crop Insurance Rate Making." *J. Farm Econ*. 39:733-40.

Buccola, S.T. "Testing for Nonnormality in Farm Net Returns." *Amer. J. Agr. Econ.* 68(May 1986): 334-343.

Chiappori, P. and Salanie, B. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy* 108(1): 56-78).

Conover, W.J. 1999. *Practical Nonparametric Statistics*. New York: John Wiley & Sons, Inc.

Day, R.H. "Probability Distributions of Field Crop Yields." *J. Farm Econ.* 47(1965):713-741.

Dismukes, R. and J. Glauber. "Crop & Revenue Insurance: Premium Discounts Attractive to Producers." *Agricultural Outlook.* 269(March 2000): 4-6.

Dismukes, R. and M. Vandeveer. "U.S. Crop Insurance: Premiums, Subsidies, & Participation." *Agricultural Outlook.* 287(December 2001): 21-24.

Gallagher, P. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *Amer. J. Agr. Econ.* 69(November 1987):796-803.

Gardner, B.L. and R.A. Kramer. "Experience with Crop Insurance Programs in the United States." In *Crop Insurance for Agricultural Development: Issues and Experience*, P. Hazell, C. Pomareda, and A. Valdes, eds., Baltimore: The Johns Hopkins University Press, 1986.

Glauber, J.W. "Crop Insurance Reconsidered." *Amer. J. Agr. Econ.* 86(Number 5, 2003): 1179-1195.

Goodwin, B.K. "An Empirical Analysis of the Demand for Multiple Peril Crop Insurance." *Amer. J. Agr. Econ.* 75(May 1993):425-434.

Goodwin, B.K. 1994. “Premium Rate Determination in the Federal Crop Insurance Program: What Do Averages Have to Say About Risk?” *J. Agr. Res. Econ*. (19)2: 382-395.

Greene, W.H. *Econometric Analysis, Third Edition*. Upper Saddle River, NJ: Simon & Schuster, 1997.

Hogg, R.V. and A.T. Craig. 1978. *Introduction to Mathematical Statistics*. New York: Macmillan Publishing Co., Inc.

Hollander, M. and D. Wolfe. *Nonparametric Statistical Methods*. New York: John Wiley & Sons, Inc., 1973.

Hsu, J.C. 1996. *Multiple Comparisons: Theory and Methods*. New York: Chapman & Hall.

Josephson, G., R. Lord, and C. Mitchell. “Actuarial Documentation of Multiple Peril Crop Insurance Ratemaking Procedures.” Report prepared for USDA-RMA by Milliman and Robertson, 2001.

Just, R.E., L. Calvin, and J. Quiggen. "Adverse Selection in Crop Insurance: Actuarial and Asymmetric Information Incentives." *Amer. J. Agr. Econ.* 81(November 1999): 834-849.

Just R.E. and R.D. Pope. "Implications of Heterogeneity for Theory and Practice in Production Economics." Amer. J. Agr. Econ. 81(August 1999): 711-718.

Just, R. and Q. Weninger. 1999. "Are Crop Yields Normally Distributed?" *Amer. J. Agr. Econ.*. 81(May 1999): 287-304.

Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, 1983.

Makki, S. and A. Somwaru. “Asymmetric Information in the Market for Yield and Revenue Insurance Products. Technical Bulletin 1892, U.S. Department of Agriculture, Economic Research Service, April 2001.

Makki, S. and A. Somwaru. “Evidence of Adverse Selection in Crop Insurance Markets.” *The Journal of Risk and Insurance*. 68(4): 685-708.

Moss, C.B. and J.S. Shonkwiler. "Estimating Yield Distributions with a Stochastic Trend and Nonnormal Errors." *Amer. J. Agr. Econ.* 75(November 1993): 1052-62.

Norwood, B., M. Roberts, and J. Lusk. “Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions.” *Amer. J. Agr. Econ.* 86 (November 2004):1032-1043.

Ramirez, O.A. "Estimation and Use of a Multivariate Parametric Model for Simulating Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean, and Wheat Yields." *Amer. J. Agr. Econ.* 79(February 1997): 191-205.

Rejesus, R.M., K.H. Coble, T.O. Knight, and Jin, Y. "Developing Experience-Based Premium Rate Discounts in Crop Insurance." *Amer. J. Agr. Econ.* 88(May 2006): 409-419.

Roberts, Michael J., Nigel Key, and Erik O'Donoghue. "Estimating the Extent of Moral Hazard in Crop Insurance Using Administrative Data." *Review of Agricultural Economics*. 28(3):381-390.

Rothschild, M. and J. Stiglitz. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics*. 90(November 1976): 629-49.

SAS Institute Inc. 2004a. *SAS/ETS 9.1 User's Guide*. Cary, NC,: SAS Institute Inc.

SAS Institute Inc. 2004b. *SAS/STAT 9.1 User's Guide*. Cary, NC,: SAS Institute Inc.

Skees, J.R. and M.R. Reed. "Rate Making for Farm-Level Crop Insurance: Implications for Adverse Selection." *Amer. J. Agr. Econ.* 68(1986): 653-59.

U.S. Department of Agriculture, Risk Management Agency. *Premium Rate Calculations for the Continuous Rating Model*. April 10, 2000 (a).

U.S. Department of Agriculture, Risk Management Agency. *Crop Revenue Coverage Continuous Rate Calculation Guide*. August 28, 2000 (b).

U.S. Department of Agriculture, Soil Conservation Service. *Land Resource Regions and Major Land Resource Areas of the United States*. Agriculture Handbook 296, December 1981.

Vandeveer, M. and E. Lohman. "Farmer Response to Modified Crop Insurance: A Case Study of Corn in Indiana." *Amer. J. Agr. Econ.* 76 (February 1994): 128-140.

White, H. and I. Domowitz. "Nonlinear regression with Dependent Observations." Econometrica. 52(January 1984): 143-61.

Wu, JunJie. "Crop Insurance, Acreage Decisions, and Nonpoint-Source Pollution." *Amer. J. Agr. Econ.* 81 (May 1999): 305-320.