University of Hawai`i at Mānoa
Department of Economics
Working Paper Series

Saunders Hall 542, 2424 Maile Way,
Honolulu, HI 96822
Phone: (808) 956 -8496
www.economics.hawaii.edu

Working Paper No. 14-10

Separating Moral Hazard from Adverse Selection: Evidence from the U.S. Federal Crop Insurance Program

By

Michael J. Roberts,
Erik O'Donoghue,
And Nigel Key

March 2014

Separating Moral Hazard from Adverse Selection: Evidence from the U.S. Federal Crop Insurance Program

By Michael J. Roberts, Erik O'Donoghue and Nigel Key*

Draft: July 21, 2011

We use data from the administrative files of the U.S. Department of Agriculture's Risk Management Agency to examine how the distribution of crop yields changed as individual farmers shifted into and out of the federal crop insurance program. The large panel facilitates use of fixed effects that span each combination of farmer and production practice to account for unobserved differences in farmer abilities, risk preferences and soils, in addition to fixed effects for interactions between all years and all counties to account for geographically-specific technological change, local prices, and weather. We also account for farm-specific yield variances. Conditional on this large set of fixed effects, we estimate the mean shift in yield and non-parametrically estimate the shift in the distribution around the conditional mean associated with enrollment in crop insurance. Because differences between farmer and land types have been accounted for (i.e., controlling for adverse selection), the estimated shifts in yield distributions likely reflect moral hazard. For most crops in most states we find insurance is associated with statistically significant but small downward shifts in average yield. The largest shifts occur for cotton and rice, the highest-value of five crops considered. By integrating the estimated shift in yield distributions over actual indemnities paid, we provide estimates of the total indemnities paid due to moral hazard. Our results indicate moral hazard accounted for an estimated \$53.7 million in indemnities between 1992 and 2001, which amounts to 0.9% of indemnities paid to the insured crops and states considered.

JEL: D82, G22, Q18.
Keywords: Moral hazard, asymmetric information, insurance, agricultural policy.

* Roberts (correspondence): University of Hawai'i at Manoa, Saunders 542, 2424 Maile Way, Honolulu HI 96822, michael.roberts@hawaii.edu. O'Donoghue and Key: USDA Economic Research Service, 1800 M Street, Washington DC 20036. The authors gratefully acknowledge useful comments and suggestions from Rod Rejesus, Tom Vukina, and Xiaoyong Zheg and participants at seminars at Purdue University and North Carolina State University. Views expressed are the authors, not necessarily those of the USDA.

Design and assessment of public insurance programs hinge on whether moral hazard or adverse selection is the greater source of inefficiency. If the main problem is moral hazard—the insured, protected from failure, are enticed to shirk or take on excessive risk—then it seems unlikely that public contracts could mitigate perverse incentives any better than private contracts would. If, however, the larger problem is adverse selection—exogenous exposure to risk is observed by those seeking insurance but not by the insurer—then public policies can entice efficient pooling or sharing of risk that private markets may not otherwise achieve, due to the well-known lemons problem (Akerlof 1970). If both information problems are acute, then public policies that force pooling may exacerbate moral hazard. Thus, a welfare-improving public policy requires that moral hazard be sufficiently rectifiable using a schedule of deductibles, co-payments, and exclusions that is simple enough such that it can be implemented in a cost-effective manner.

These essential tensions between moral hazard and adverse selection underlie all public insurance programs. The key challenge of policy evaluation is measuring the incidence of unavoidable moral hazard in a manner that clearly separates its effect from adverse selection. Separating these effects empirically is difficult because in many ways the two problems are observationally similar: whichever the source, bad outcomes tend to be more prevalent among the insured as compared to the uninsured.

In this article we empirically investigate this issue for a lesser known but pervasive public insurance program: multi-peril crop insurance in the United States. Although federal crop insurance has been in place since the Great Depression, few farmers participated in the program before 1980 when the modern structure of the program was established. Lack of participation in the early federal program, and general nonexistence of private broad-coverage insurance, suggests adverse selection and moral hazard inhibited inception of this insurance market.[1]

The modern structure of the program involves USDA's Risk Management Agency (RMA) developing available insurance products, setting premiums and premium subsidy rates, and private insurance companies that market the plans to farmers. RMA then reinsures private insurance companies for most of their risk exposure. Perhaps the most significant development in recent decades was the 1994 Federal Crop Insurance and Reform Act (FCIRA). Beginning with this Act, Congress promoted significant expansion of coverage through large and steadily growing premium subsidies, and by increasing the number of farming activities covered. In most years since 1995, farmers have collectively enrolled over 200 million acres. Today, over 80 percent of US crop acreage is covered under the program.

The key contribution of this study is that we are able to exploit a large and unique panel data set to derive precise quantitative estimates of the incidence

[1]Private markets for insurance against narrow specific risks—notably hail—have long existed, but broad, multi-peril coverage of a crop has not existed in the United States or other countries without significant government subsidies.

and budgetary cost of moral hazard in a way that plausibly distinguishes these effects from adverse selection. Until now, we have generally observed that farmers with insurance have lower yields, but have not disentangled the degree to which this association follows because insured farmers manage land of lower quality or because farmers manage their land poorly because they are insured. The latter effect captures residual moral hazardmoral hazard not dissuaded by prevailing contract structure. We derive estimates for each of five crops and each of three to nine states, depending on the crop. The crops and states examined included the largest in the nation and comprise significant shares of US production and world exports.

To separate moral hazard from adverse selection, we consider how yields of farmers change as they cycle into and out of the program in comparison to farmers in the same county not cycling into or out of the insurance program. This difference-in-differences approach differs greatly from earlier cross-sectional research that compared farmers with insurance to those without it. By comparing how individual farmers' yields change with adoption of crop insurance we avoid making comparisons between farms of different types, so our estimates do not include the most natural source of adverse selection.

We develop these estimates using regressions with fixed effects for each combination of farmer and practice (irrigated or non-irrigated) to control for unobserved variations in land quality and producer skill and risk preferences. The model also includes fixed effects for each combination of county and year to control for technological change, local price effects, and a substantial share of weather variation, all of which may otherwise lead to dynamic adverse selections. Finally, we account for farm-specific yield variances and non-parametrically estimate the standardized residual conditional on fixed effects and insurance coverage. Results indicate how the non-parametrically estimated distribution of yields, conditional on fixed effects, shifts with insurance enrollment. This analysis provides an estimate of the shift in yield distributions due to residual moral hazard for each crop and state. To estimate the share of indemnities due to moral hazard, we then integrate the difference in estimated yield distributions over actual indemnities paid between 1992 and 2001.

The data are comprised of the administrative files of USDA's Risk Management Agency and include millions of observations—the entire population of insurance contracts—so statistical power is considerable, despite use of hundreds of thousands of fixed effects. The empirical approach is possible because most insurance records include yield histories that extend back to periods prior to first enrollment in the insurance program. Thus, for each farmer's crop and practice (irrigated or not), we observe yield outcomes in years with and without insurance coverage. Identification is further aided by focusing on the time period that experienced the largest increase in premium subsidies and coverage growth.

Results indicate small but statistically significant downward shifts in yield distributions of farmers when they have insurance as compared to when they do not have insurance in 28 of 32 crop-by-state regressions. Two of the remaining four crop-by-state combinations indicate negative shifts in yields that are not statistically significant. For most crops and states, the budgetary cost of moral hazard is estimated to lie between 0.5 percent and 2 percent of indemnities paid. The cost is largest for cotton in Arkansas and California, equal to 6.8 and 3.3 percent of indemnities paid, respectively. To the extent that subtle forms of adverse selections may remain, one may view our estimates as an upper bound on the incidence of moral hazard.

I. Related Literature

A. Separating Moral Hazard and Adverse Selection

B. Crop Insurance

Chiappori, Durand and Geoffard (1998) exploit a natural experiment in France to examine how a change in co-payments affected use of physician services in comparison to a control group that did not experience such a change. Abbring et al. (2003) and Abbring, Chiappori and Pinquet (2003) exploit variance in past insurance claims and induced changes in insurance contract terms to identify moral hazard separate from adverse selection. Finkelstein and Poterba (2006) do the converse of our investigation by exploiting *unused observables* to identify adverse selection separate from moral hazard.[2]

Do insured drivers drive more recklessly than uninsured drivers and thus increase the likelihood of an accident? (Cohen and Dehejia 2004) To what extent do those with health and life insurance take less care of themselves and thus increase the likelihood of sickness, injury or death?(Chiappori, Durand and Geoffard 1998) Does unemployment insurance cause workers to exert less job effort and thus increase the odds they will lose their jobs? (Chiu and Karni 1998) Does deposit insurance cause depositors to pay too little attention to the banks managing their investments, ultimately leading to bank mismanagement and failure?(Keeley 1990). In the literature on crop insurance, efforts to separate moral hazard from adverse selection mainly use cross-sectional identification strategies. We review the crop insurance literature in a separate section below.

Previous research on links between crop insurance and input use found mixed evidence of moral hazard. Horowitz and Lichtenberg found that crop insurance caused fertilizer and pesticide use to increase by 19% and 21% respectively. They explained this counterintuitive result by arguing that costly yield-enhancing inputs also increase yield risk. With insurance, farmers see the upside benefits of higher yields while sharing the downside risks of crop failure with the insurance company, which could cause them to intensify the use of risk-increasing inputs

[2]Finkelstein and Poterba define *unused observables* as "attributes of individual insurance buyers that are correlated both with subsequent claims experience and with insurance demand but that insurance companies do not use to set insurance prices."

despite their cost. If correct, their findings imply potentially large negative environmental implications stemming from crop insurance subsidies.

These findings, however, are countered by empirical work by Quiggin, Karagiannis and Stanton (1993), Smith and Goodwin (1996) and Babcock and Hennessy (1996), among others, who estimate modest declines in input use resulting from insurance adoption. Quiggin, Karagiannis and Stanton (1993) consider the joint-effects of moral hazard and adverse selection using a production-function based approach based on data from the 1988 Farm Cost and Returns Survey. They find input use and yields are negatively associated with insurance, a phenomenon indicative of both moral hazard and adverse selection. However, they made no attempt to separate these two phenomena. Smith and Goodwin (1996) simultaneously model the insurance decision and input use decisions of dryland farmers in Kansas and also find reduced input use. Babcock and Hennessy infer moral hazards from observed relationships between inputs, outputs and output risk.

These earlier studies generally use data from a time when crop insurance was less heavily subsidized and fewer farmers participated. Today, subsidized crop insurance is far more prevalent so adverse selection is a substantially smaller deterrent to program participation, making it less likely to explain differences in yields between insured and uninsured acres.

Mixed findings from earlier studies regarding the effects of insurance on input or output levels likely stem from the difficulty in identifying moral hazard separately from adverse selection or other unobserved farmer or land-related characteristics (Abbring et al. 2003). Typically, researchers have estimated insurance effects by regressing input use or yields on an insurance indicator and other controls. The key empirical challenge stems from the endogeneity of the insurance decision: the insurance decision is not randomly assigned. Indeed, because of adverse selection, one would expect that insurance adopters differ from non-adopters. Thus, the assumption of no correlation between unobserved factors driving input levels and the decision to insure is immediately called into question. The assumption is particularly strong in cross-sectional studies, where unobserved factors relating to land quality likely influence insurance decisions, input-use decisions, and yield outcomes. In these studies, confounding factors could be the main source of observed associations between insurance and behavior.

II. Model

Farmer i's decision to insure in year t, which we denote I_{it}, depends on time-varying state variables including subsides, prices and technology, which we denote by ϕ_t, as well as individual farmer characteristics, preferences and private information, which we denote by θ_i. To keep things simple, the variable I_{it} is assumed dichotomous, equal to 1 if insured and 0 if not insured. We thus have:

$$I_{it} = d(\theta_i, \phi_t). \tag{1}$$

As with insurance decisions, farmer yield outcomes depend on prices, technology and other time-varying state variables (ϕ_t), plus systemic and idiosyncratic weather outcomes (w_{it}). Yield outcomes also depend on farmer effort, which is unobserved and may depend on insurance decisions, denoted $e_i(I)$. This latter effect—the causal effect of insurance on yield via effort—is the moral hazard effect. We thus have:

$$Y_{it} = f(\theta_i, \phi_t, w_{it}, e_i(I)). \tag{2}$$

To simplify notation and clarify concepts of moral hazard and adverse selection, denote the set of farmers and years with insurance by $\mathcal{I}$ and the set of famers and years without insurance by $\mathcal{NI}$, and define

$$\begin{aligned}
Y_{1|1} &= \mathcal{E}\left[f(\theta_i, \phi_t, w_{it}, e_i(I=1))|\{i,t\} \in \mathcal{I}\right] \\
Y_{1|0} &= \mathcal{E}\left[f(\theta_i, \phi_t, w_{it}, e_i(I=1))|\{i,t\} \in \mathcal{NI}\right] \\
Y_{0|1} &= \mathcal{E}\left[f(\theta_i, \phi_t, w_{it}, e_i(I=0))|\{i,t\} \in \mathcal{I}\right] \\
Y_{0|0} &= \mathcal{E}\left[f(\theta_i, \phi_t, w_{it}, e_i(I=0))|\{i,t\} \in \mathcal{NI}\right]
\end{aligned}$$

Thus, the first binary number in the subscript indicates the insurance decision (1 is insured, 0 is uninsured) and the second binary number in the subscript indicates the subset of the population over which the expectation is taken (1 indicates $\mathcal{I}$, 0 indicates $\mathcal{NI}$). Using these definitions, we can express the observed incidence of *moral hazard* as

$$MH = Y_{1|1} - Y_{0|1} \tag{3}$$

and we can express the observed incidence of adverse selection as

$$AS = Y_{0|1} - Y_{0|0}. \tag{4}$$

Further note that the observed difference in yield between insured and uninsured farmers is

$$OD = Y_{1|1} - Y_{0|0} = MH + AS \tag{5}$$

which implies $AS = OD - MH$.

To complete the model and make identifying assumptions, we must assume a functional form for yield in relation to farmer characteristics and farmer, time and location effects, and insurance-induced effort effects. We consider a separate model for each crop and state. For each model, we develop an index i that spans all farmer and practice combinations.[3] The empirical model relates crop yield, Y_{it}, for farmer-practice i in year t to a variable I_{it} that indicates whether i has insurance in year t, plus a series of fixed effects controls, described below. The model also includes the farmer-practice specific standard deviation of yields, σ_i.

$$Y_{it} = \alpha_i + \gamma I_{it} + w_{ct} + u_{it} \tag{6}$$

$$\epsilon_{it} = u_{it}/\sigma_i \tag{7}$$

$$\epsilon_{it} \sim F(\epsilon|I_{it}) \tag{8}$$

In 6, α_i represents a farmer-specific intercept that accounts for land quality and farmer skill; γ is the average effect of moral hazard on yield; and w_{ct} denotes a county-by-year fixed effect that captures technological change, county-level weather, and local price effects. The error, u_{it}, captures other unobserved factors affecting yields, like within-county weather variations and pest infestations. Equations 7 and 8 give some limited structure for the error in 6. The standardized error, ϵ_{it}, is defined as the error divided by a farmer-specific standard deviation, σ_i, which accounts for the fact that some farmers may have intrinsically more variable yields than other farmers. Standardized errors have a general continuous distribution function $F(\epsilon|I_{it})$ that is conditional on whether or not yields are insured; we estimate this function nonparametrically.

The farmer-level fixed effects eliminate time-invariant heterogeneity of land and farmers. If year fixed effects were also included in the model, then insurance-related effects would be identified by yield changes over time on farms that cycle into or out of the insurance program in comparison to simultaneous yield changes on farms that either remain insured or remain uninsured. Instead of year fixed effects, however, we use county-by-year fixed effects. County-by-year fixed effects narrow these simultaneous difference-in-differences comparisons to farms within the same county.

The individual fixed effects (α_i) and individual variances (σ_i) account for adverse selection that gives rise to an observed relationship between yields and insurance that might otherwise reverse the direction of causality. That is, we expect farmers to be more likely to buy insurance if they operate marginal land that produces lower or more variable yields, all else the same. Adverse selection can thus give rise to a correlation in which causation goes from yields to insurance. With moral hazard the causal link goes in the opposite direction: insurance

[3]Because few farmers cultivate the same crop on both irrigated and non-irrigated land, we often refer to i as a farmer index rather than a farmer-practice index.

causes farmers to change their effort, inputs, or management practices, which leads to different yield outcomes. By including individual farmer fixed effects and individual farmer variances, we account for adverse selections, allowing us to isolate the effect of moral hazard from adverse selection.

The general and flexible model complements our rich administrative data set (described below), which includes several years of yield data for most farmers, including (for most farmers) observations both with and without insurance. Insured observations generally predominate after FICRA and uninsured observations generally predominate before FICRA, but there is wide individual heterogeneity. This allows us to identify the insurance shift parameter γ, even with farmer-specific intercepts and county-by-year fixed effects. The large number of observations also allows for a non-parametric estimation of the error distribution function conditional on insurance, $F(\epsilon|I_{it} = 0)$ and $F(\epsilon|I_{it} = 1)$. Non-parametric estimation of the error distribution is particularly useful given our focus on moral hazard, which may embody incentives to alter management practices and inputs in a way that increases yield variance more at the low-end of the distribution as compared to the high-end of the distribution.

In principle, there are many ways to estimate a model of the form given by equations 6 through 8. Random effects or other more structured models incorporating individual farm heterogeneity are typical when the number of individuals is relatively small. These models can account for individual heterogeneity using fewer degrees of freedom and can thereby increase statistical power. However, random effects models also require strong assumptions about the distribution of individual effects (α_i) and how they relate to other components of the model, particularly the error. In our application, which includes nearly the entire population of insured farmers, such an approach unnecessarily adds computational difficulty and modeling assumptions. In particular, because we expect adverse selection would lead to causation going from yields to insurance rather than from insurance to yields, the correlation between the random effects and the error would lead to bias. We therefore treat α_i and w_{ct} as parameters—fixed effects—rather than drawn from a prior distribution.

III. Identification

Although farm-specific fixed effects and county-by-year fixed effects remove the most obvious forms of confounding from omitted variables or adverse selections, the approach hinges on the assumption that, conditional on these controls, *timing* of insurance adoption is exogenous. Hence, this analysis does not constitute a natural experiment in which the timing of insurance adoption was randomly assigned across farmers. Indeed, all farmers had access to the same menu of insurance contracts in each year.

There are several reasons why assuming conditional exogeneity is reasonable in this context. First, because the data span a period of time when premium subsidies were rapidly increased, most of the cycling is into, not out of, the insurance program. It therefore seems likely that farmers' insurance decisions are driven mainly by increases in federal subsidies, which are exogenous to farmers' production decisions. Second, while farmers with different risk environments would plausibly find it optimal to enter the program at different times, farm-specific fixed effects would account for these differences. It is not clear how the timing of insurance decisions would systematically be correlated with farm-level changes in yields over time. Third, it seems plausible that idiosyncratic factors would cause different farmers to take different amounts of time to learn whether enrolling in the crop insurance programs was in their best interest. The timing of information-related factors are plausibly unrelated to other factors that might simultaneously influence changes in farmer effort, input use or yield outcomes, especially when these are conditioned on county-by-year fixed effects and individual farmer-by-practice fixed effects.

IV. Estimating Indemnities Due to Moral Hazard

Indemnity payments are a function of yield outcomes and, depending on the insurance contract, other factors like prices.[4] Expected indemnities integrate the indemnity payment function over the distribution of yield outcomes. To evaluate expected indemnities parametrically would be an extremely complicated procedure given (a) the wide variety of insurance policies and coverage levels, all with differing payment functions and (b) the widely varying distribution of yield outcomes. While it is technically feasible to construct payment functions for all individual contracts using our data, the process would be extremely labor-intensive and could be prone to error. The greater conceptual challenge is (b), since minor changes in the distribution of yields, particularly in the tails, might imply large differences in expected indemnities (Goodwin and Ker 1998). Moreover, yield distributions vary widely over time and space, which is why crop insurance is so susceptible to adverse selection (Just, Calvin and Quiggin 1999).

Rather than evaluate each indemnity payment function, we instead use the physical outcomes of those functions: indemnity payments actually received by farmers. We then consider only the estimated shift in the yield distribution function stemming from the insurance decision, which means we don't have to estimate separate conditional distributions for each farm and practice. This approach is simpler, requires few assumptions, and is robust to a wide range of unobserved heterogeneities. The approach is also non-parametric in the sense that we need not calculate the payment function for each insurance contract and yield history or make additional assumptions about the distribution of county-by-year effects, which include both random components like the weather, fixed factors like farmer ability, soil characteristics, and climate, and potentially complex interactions of these factors.

[4]Some insurance contracts insure revenue per acre (the product of price and yield) as opposed to yields.

To clarify the approach, define the indemnity payment function $P(Y_{it}|\theta_{it}, I_{it} = 1)$, where θ_{it} accounts for factors besides yield, including the farmer's insurance contract, yield history and, if applicable, prices or other factors; and define $g(Y_{it}|\theta_{it}, I_{it} = 1)$ as the probability density function of Y_{it} . Expected indemnities are then:

$$\int P(Y_{it}|\theta_{it}, I_{it} = 1)g(Y_{it}|\theta_{it}, I_{it} = 1)dY_{it} \tag{9}$$

This expression gives expected payments that occur both in the presence and in the absence of moral hazard. If there were no moral hazard, the yield distribution is instead given by $g(Y_{it}|\theta_{it}, I_{it} = 0)$. Thus, to calculate expected indemnities due to moral hazard (IMH) we need to subtract expected indemnities without moral hazard from expected indemnities with moral hazard

$$\begin{aligned}
(10) E[IMH] &= \int P(Y_{it}|\theta_{it}, I_{it} = 1)g(Y_{it}|\theta_{it}, I_{it} = 1)dY_{it} - \\
&\qquad \int P(Y_{it}|\theta_{it}, I_{it} = 1)g(Y_{it}|\theta_{it}, I_{it} = 0)dY_{it} \\
(11) \qquad &= \int P(Y_{it}|\theta_{it}, I_{it} = 1)\left[g(Y_{it}|\theta_{it}, I_{it} = 1) - g(Y_{it}|\theta_{it}, I_{it} = 0)\right] dY_{it}
\end{aligned}$$

The first term in equation 10 is expected indemnities; the second term is expected indemnities holding the payment function fixed and shifting the yield distribution from conditionally insured to conditionally uninsured, while holding the farmer type (embodied by θ_{it}) fixed. Since the payment function is identical in both terms we can simplify the expression.

Recall that our regression model provides an expression for Y_{it} conditional on insurance and individual farmer and time characteristics:

$$Y_{it}|\theta_{it}, I_{it} = \alpha_i + \gamma I_{it} + w_{ct} + u_{it} \tag{12}$$

where

$$u_{it}/\sigma_i \sim F(\epsilon|I_{it}) \qquad dF/d\epsilon = f(\epsilon|I_{it}) \tag{13}$$

This allows us to express the distribution function of Y_{it} in relation to the distribution function of the model error. Since we are interested in effects of moral hazard on indemnities paid, we need to consider instances where farmer i is observed to be insured at time t. In this case the density function $g(Y_{it}|\theta_{it}, I_{it} = 1)$ evaluated at the observed outcome (Y_{it}) equals the distribution of the error evaluated at its observed outcome $f(\epsilon_{it}|I_{it} = 1)$. These must be equal because the only way to achieve the observed outcome Y_{it} is if the error equals the observed error ϵ_{it}.

Next consider $g(Y_{it}|\theta_{it}, I_{it} = 0)$. Since observed Y_{it} is an insured yield, here we need to evaluate the density of a counterfactual, i.e., the relative odds of achieving the observed outcome if the farmer were not actually insured. Conditional on the same farmer not being insured ($I_{it} = 0$), a different error is needed to achieve the same observed level of Y_{it}. Denote this counterfactual error by ϵ'_{it}. We therefore have:

$$Y_{it} = a_i + \gamma + w_{ct} + \sigma_i \epsilon_{it}, \tag{14}$$

and

$$Y_{it} = a_i + w_{ct} + \sigma_i \epsilon'_{it}, \tag{15}$$

which implies

$$\epsilon'_{it} = \epsilon_{it} + \frac{\gamma}{\sigma_i} \tag{16}$$

Thus, to achieve the observed insured yield outcome we need to add to the observed error the lost effect of γ, scaled by the farm-specific variance σ_i. Finally, we need to account for the counterfactual error having a different distribution, as it is conditional on $I_{it} = 0$ rather than $I_{it} = 1$. Thus,

$$g(Y_{it}|\theta_{it}, I_{it} = 1) = f(\epsilon_{it}|I_{it} = 1) \tag{17}$$

$$g(Y_{it}|\theta_{it}, I_{it} = 0) = f(\epsilon_{it} + \frac{\gamma}{\sigma_i}|I_{it} = 0) \tag{18}$$

Substituting these expressions into 10 gives expected indemnities due to moral hazard:

$$\int P(Y_{it}|\theta_{it}, I_{it} = 1) \left[f(\epsilon_{it}|I_{it} = 1) - f(\epsilon_{it} + \frac{\gamma}{\sigma_i}|I_{it} = 0) \right] dY_{it} \tag{19}$$

Instead of evaluating expected indemnities, which would require specification of the payment function for each insurance contract, it is simpler and perhaps more useful to consider indemnities actually paid due to moral hazard. We do this by substituting observed indemnity payments for the payment function and replacing the integration with a sum. Specifically, if we denote the density function of the model's conditional error distribution by $f(\epsilon_{it}|I_{it} = 1)$ and $f(\epsilon_{it}|I_{it} = 0)$ and the total indemnities received by farmer i in year t by P_{it}, then the share of indemnities due to moral hazard is calculated as:

$$IMH = \sum_i \sum_t P_{it} \left[f(\epsilon_{it}|I_{it} = 1) - f(\epsilon_{it} + \frac{\gamma}{\sigma_i}|I_{it} = 0) \right] \tag{20}$$

The expression in 20 just multiplies observed indemnity payments by the relative likelihood of those outcomes with and without insurance, and then sums over all observed indemnities. It may be estimated using data on actual indemnity payments and estimates of $\gamma, f(\epsilon|I_{it} = 1)$ and $f(\epsilon|I_{it} = 0)$ from estimation of the regression model (6 - 8).

V. Data

Data obtained from the U.S. Department of Agriculture's Risk Management Agency (RMA) includes all U.S. crop insurance contracts from 1992 through 2002.[5] From these data we developed a history of yield outcomes in all years for each farm, crop, and practice (irrigated or non-irrigated). Each observation therefore contains an average yield for each crop over all fields an individual farmer operates, the level of insurance purchased, and all indemnity payments received (if any) in each year, county, and using a specific practice (irrigated or not).

We consider coverage in eleven states that grow a significant portion of the nation's five largest crops (in terms of production value): corn, soybeans, wheat, cotton, and rice. The states considered comprise at least half the total production of each commodity (to measure broad relevance) and also significant geographic variability (to see if farmers producing different crops in different regions exhibited different behavior). For corn and wheat, nine states were selected: Kansas, Illinois, Indiana, Iowa, Ohio, Nebraska, North Dakota, Montana, and Texas. For soybeans we selected the same set of states minus Montana, since that state produces very little. In 2000 these states produced almost 66% of total corn production in the United States, more than 60% of total soybean production, and over 55% of total wheat production. For cotton and rice, we selected three states: Arkansas, California, and Texas. In 2000 these states produced over 51% of total cotton production, and nearly 73% of total rice production.

Because the data set is derived from RMA administrative files, the observations include the population of insured yields but do not include the population of uninsured yields. Most uninsured yields come from yield histories on farms insuring their yields for the first time; some also come from farmers who have insured their fields in past and future years but not in the current year. All uninsured yields in each year were therefore insured at some later point in time.[6] The raw data files give insurance contract details for each insured unit and a yield history associated with that unit, where a unit might comprise one or more of a farmer's fields of a given crop and practice. Farmers must pay higher premiums if they choose to

[5] We were able to obtain insurance contract data back to 1989, in some cases with yield histories extending to the early 1980s. Unfortunately we could not use data prior to 1992 because these contracts did not include an identification number that would allow us to link files over time, which is critical for defining the insurance adoption variable, I_{it}. The 2002 data include information on insurance purchased and indemnities paid as well as yields in earlier years, but do not include yields in 2002. So while we use yield history data from the 2002 contract files, yield observations end in 2001.

[6] If farmers do not provide a verifiable yield history then they receive a "transitory yield" in place of actual yield history for purposes of premium determination until enough actual crop history has been accumulated. Transitory yields equal 60% of county-average yields. Transitory yields were dropped from the data analysis.

insure multiple fields having the same crop and practice separately. Over time, some farmers have combined or split units and, as a result, different insured units can have identical yield histories. This means we cannot observe yields over time on individual units and we therefore aggregate all units of each farmer and practice.[7] This allows us to merge data from different contract years and thus observe yields over time for the same farm in relation to changes in insurance coverage on that farm.

The model requires a dichotomous indicator of insurance (I_{it}). While each individual must have adopted insurance at some point to be in the RMA dataset, the timing of insurance adoption varies across farms. Operators could also adopt markedly varying levels of coverage. Many insured farmers, particularly in the mid 1990s, held only "catastrophic" coverage (CAT) for their crops. CAT coverage is fully subsidized by the government, requiring only a nominal administrative fee (typically $50 or $100, depending on the year) for each crop a farmer chooses to enroll in a given county, regardless of total acreage. Coverage provided by CAT is very low, most typically equal to 50 percent of expected yield at 55 percent of expected price, but this varies slightly depending on the year.

To focus on insurance more likely to influence behavior, we define "insured" yields as those having coverage in excess of CAT coverage, which is often informally referred to as "buy up" coverage. Most farmers with "buy up" have coverage that insures at least 65 percent of the expected yield at 75 percent or more of expected price. Many also have revenue insurance, which insures a price multiplied by an approved (expected) yield. Since we focus on a broad assessment of crop insurance, we do not differentiate between the alternative buy-up insurance products in this paper—all insurance coverage above the CAT level is considered "insured" ($I_{it} = 1$) and CAT or no insurance is considered "uninsured" ($I_{it} = 0$).[8]

Summary statistics by state, year, and disposition (insured or not) are shown in figures 1a-1e. Each figure summarizes a single crop for two representative states and shows the average yield and standard deviation of yields for both insured and uninsured farms. Note that the standard deviation is not a pure measure of risk because it measures variation in the cross-section of each year. Thus, this variance captures variation in land quality plus idiosyncratic risk that is orthogonal to state-level variations. The plots below the means and standard deviations show the number of observations for both insured and uninsured in each year. These plots show the large sample sizes that number in the many thousands for each crop, year, and disposition (insured or uninsured). They also show how insurance adoption increased over time: in earlier years, the number

[7]Aggregating insured units helps us to link yields and coverage over time and can obscure a form of insurance fraud called "crop switching." Crop switching involves furtively transferring output from one insured unit to another, giving the appearance of a higher-than-actual yields on some insured units and lower-than-actual yields on other units, giving rise to a fraudulent indemnity claim.

[8]In separate analysis (not reported) we found very small yield effects—much smaller than effects estimated here—stemming from adoption of CAT insurance.

of uninsured observations tends to be larger than the insured number and this reverses sharply over time, with much of the change occurring with the 1994 FCIRA.

For most crops, states, and years, the average yields of insured observations are lower than those of uninsured observations. A majority of insured observations also have a higher standard deviation across farms within state-years, and a larger share of insured observations have a higher coefficient of variation because their mean yields tend to be lower. These patterns, which are indicative of both adverse selection and moral hazard, are stronger and more prevalent in earlier years than in later years, when the government more heavily subsidized insurance and the share of insured cropland was considerably higher. In the most recent years, differences between insured and uninsured yields are small. The fact that yield differences nearly vanished as insurance adoption became more prevalent suggests that much of the observed yield differences are due to adverse selection rather than moral hazard.

VI. Estimation

A separate analysis is conducted for each crop and state. Estimation is undertaken in steps due to the large number of observations and fixed effects. First, we generate indicator variables for each county and year combination. Second, we remove individual farmer fixed effects (α_i) by subtracting each farmer's mean values from the dependent variable, the insurance indicator, and all county-by-year indicator variables. The sheer number of individual farmers in the data set makes it infeasible to estimate these fixed effects jointly with the other coefficients. Third, we regress de-meaned yields against the de-meaned county-year dummy variables and the de-meaned insurance indicator. Residuals from the OLS regressions in the third step are used to construct estimates of the errors, u_{it}. To adjust for individual farm-level yield variance, we divide each error by the sample standard deviation of residuals for each farmer and practice, as indicated in equation (2).[9] These adjusted residuals serve as estimates of ϵ_{it}.

In the final step we use a non-parametric kernel density to estimate $F(\epsilon|I_{it})$. Separate non-parametric densities are estimated for all insured ($I_{it} = 1$) and uninsured ($I_{it} = 0$) observations. Kernel density estimates were made using the software package "R" and the default bandwidth selection in the function "density."[10]

[9]Specifically, if we define n_i as the number of observations for a given farmer, crop and practice, and $\hat{u}_{it}$ as the corresponding residuals, then the standardized residuals for that crop, farmer and practice are $\frac{\hat{u}_{it}}{s_i}$ where $s_i = \sqrt{\sum_i \frac{u_{it}^2}{n_i - 1}}$.

[10]Each point on each density is calculated as a weighted average of frequencies "local" to each point. Locality is determined by bandwidth, which is calculated using Silverman's (1986) rule of thumb: 0.9 times the minimum of the standard deviation and the interquartile range divided by 1.34 times the sample size to the negative one-fifth power. Weights are determined using a Gaussian (normal) distribution centered on the point estimate.

One potential concern is the relatively few observations used in estimating individual farm variances. While most crop-states have tens or even hundreds of thousands of observations, there are only two to 10 observations per farm. Particularly for farms with less than five observations, this estimate is poor. In one respect the imprecision of these estimates is of little concern: we have little interest in individual farm variances themselves, we only wish to purge their influence on the level of indemnities paid so as not to confound the effects of moral hazard with adverse selection.

In another respect, however, imprecise estimation of σ_i does present a problem in estimation of the density function $f(\epsilon|I)$. When very few observations are used in estimation of σ_i, the standardized residual becomes a noisy (though unbiased) estimate of ϵ, which could cause the estimated density function to be biased too wide relative to the true distribution. To some, and perhaps a large, extent this problem is diminished by the fact that we calculated indemnities due to moral hazard using the difference between the estimated densities $f(\epsilon|I = 1)$ and $f(\epsilon|I = 0)$. The bias in each of these densities is similar because they are based on the same estimated σ_i. To the extent that the bias is similar in both distributions, differencing removes the bias. But since we do not evaluate the densities at the same location, a small amount of bias may remain. A Monte Carlo simulation exercise, described in the appendix, indicates this bias is small if we restrict data used for density estimation to farms with six or more observations. The appendix also shows how little results change if we instead restrict density estimation to farms with greater or fewer than six observations.

VII. Results

Tables 1-5 summarize results for corn, soybeans, wheat, cotton, and rice. Each row in each table gives results for the state given in column 1. Column 2 reports the average yield over all farms and years. Columns 3, 4 and 5 give the estimated mean shift in average yield due to moral hazard (the coefficient γ from equation 1), its corresponding standard error, and the associated t statistic. Therefore it is not surprising that we obtain statistical significance for most of the mean shift parameters. The reported R2 values in column 6 are for the regressions of de-meaned variables and therefore do not include variance explained by individual farmer-practice fixed effects. Despite this, the amount of variance explained remains fairly high due to the county-by-year fixed effects. Estimates for each crop-state employ a large dataset, ranging from 2,420 observations for corn in Montana up to 1.22 million observations for soybeans in Iowa (column 7). Columns 8 and 9 report the estimated share of indemnities due to moral hazard and associated total budgetary losses.

The estimated mean shift in yields with insurance is negative for 30 of the 32 state-crops examined, and all but 2 of the 30 negative estimates are statistically significant. The two positive coefficients are for corn in both Montana and North Dakota, which have comparatively few observations. Of these two estimates, only

Montana is statistically significant. While unexpected, note that it is theoretically possible for insurance to cause an increase in input use depending on variance effects of inputs (Cohen and Dehejia 2004) and because current yield outcomes affect future premium rates.

Despite statistical significance, the economic significance of the yield shifts are generally modest, typically equal to less than three percent of average yield. Larger shifts are observed in cotton and rice yields, especially in Arkansas (about 11 percent and 4 percent of average yield, respectively).

The shifts in yield distributions, which combine the mean shifts with the error distribution shifts, are illustrated in figures 6 and 7. These figures show distribution plots for 10 of the 32 crop-state combinations considered: two states for each of the five crops.[11] The error distributions generally appear similar for insured and uninsured yields. In some cases the error distribution of insured yields is slightly wider than it is for uninsured yields. In a few cases, and noticeably for rice, the insured distribution has a thicker lower tail. While the sizes of most estimated effects are modest in size, the pattern of effects is broadly consistent with the theory of moral hazard.

The estimated share of total indemnities paid due to moral hazard is typically between 0.5% and 2%, but ranges from -7.4% (an anomalous result for the small corn sample in Montana) to 6.4% (for Arkansas cotton). Summing over all states and crops, indemnities due to moral hazard were estimated to total 53.7 million dollars for the years 1992-2001, or about 0.9% of all indemnities paid out over those 10 years for the crops and states considered.

Looking across states and crops, the budgetary cost of moral hazard is naturally associated with the estimated shift in mean yield, but the relationship is far from perfect. For example, the share of indemnity costs for California cotton (3.3%) is about 50% higher than for Arkansas rice (2.3%), while the mean yield shift for California cotton is only slightly larger than Arkansas rice (4.7 versus 4.2%). The share of indemnities paid due to moral hazard is influenced by the distribution shift as well as the mean. It is also influenced by the insurance policies and coverage levels that farmers chose. Across all crops and states, the estimated share of indemnities due to moral hazard is positive in 29 of the 32 estimates. The estimate is negative in the two instances where the mean yield shift was positive plus one other case, rice in California. Note that the mean shift for rice in California was just 0.3 percent of average yield and not statistically different from zero, so a small estimated reduction in indemnities paid due to moral hazard is not surprising. Apart from these few exceptions, the results display remarkable consistency across crops and states, and especially across states with similar land and yield distributions.

[11] We present only 10 of the 32 crop-states in order to conserve space. The remaining 22 are available from the authors upon request.

VIII. Conclusions

In this paper we apply fixed-effects regression models and non-parametric density estimation to a rich administrative panel data set to show how yields change when farmers buy subsidized crop insurance. We identify the effect of insurance on yield outcomes by comparing how crop yields changed as farmers cycled into and out of the insurance program in comparison to yield changes on other farms in the same county growing the same crop that either remained in the program or had not yet enrolled. Though simple, this identification strategy controls for a tremendous amount of land heterogeneity and for time-related effects that might otherwise confound the effects of moral hazard. The large data set (all insurance contracts from 1992 to 2001) allows us to examine how insurance decisions influence the whole distribution of yield outcomes using separate analyses for different states and crops.

We find strong evidence of moral hazard in the great majority of crops and states; but in most cases, the effect appears quite modest, equal to less than 3 percent of average yield. Larger effects of 4.7 and 11 percent are found for cotton production in California and Arkansas. These estimated effects give an upper bound on the social cost of moral hazard because they do not count cost savings associated with the altered behavior. Since the changes are small, and decisions are made at the margin, observed output changes are likely offset by cost savings of a magnitude similar to observed output changes. We therefore expect the social cost of moral hazard to be only a small fraction of our estimated output effects.

The estimated effect on indemnities of $53.7 million amounts to less than one percent of the more than $6 billion in total insurance indemnities paid out for all five crops over the ten years examined. The estimated share of indemnities paid due to moral hazard may justify modest premium adjustments by state and crop commensurate with these effects.

Potential extensions of this research might more closely examine those crops and states where moral hazard appears more prevalent and test to see whether moral hazard has changed over time. It may also be useful to examine some of the more specialized crops that have been added to the program and should be amenable to similar analysis if appropriate records have been kept. However, because moral hazard effects appear relatively modest, perhaps a more useful direction for future work would be to examine the problem of adverse selection more thoroughly.

Earlier research has found large loss ratios (Coble and Knight 2002), large combined effects of moral hazard and adverse selection (Quiggin, Karagiannis and Stanton 1993), and incidence of adverse selection (Just, Calvin and Quiggin 1999). These earlier findings combined with ours suggest that adverse selection poses a far larger problem than moral hazard. It remains an open question whether and how much remaining adverse selection might be avoidable with better risk measurement and contract design. The rich administrative data now available might be particularly useful in developing insurance premiums more closely aligned with farmers' true expected indemnities and might thereby diminish the level of subsidies required to elicit broad participation in the program.

REFERENCES

Abbring, J.H., J.J. Heckman, P.A. Chiappori, and J. Pinquet. 2003. "Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish?" *Journal of the European Economic Association*, 1(2-3): 512–521.

Abbring, J.H., P.A. Chiappori, and J. Pinquet. 2003. "Moral hazard and dynamic insurance data." *Journal of the European Economic Association*, 1(4): 767–820.

Akerlof, G.A. 1970. "The market for" lemons": Quality uncertainty and the market mechanism." *The quarterly journal of economics*, 488–500.

Babcock, B.A., and D.A. Hennessy. 1996. "Input demand under yield and revenue insurance." *American Journal of Agricultural Economics*, 78(2): 416–427.

Chiappori, P.A., F. Durand, and P.Y. Geoffard. 1998. "Moral hazard and the demand for physician services: First lessons from a French natural experiment." *European Economic Review*, 42(3-5): 499–511.

Chiu, W.H., and E. Karni. 1998. "Endogenous adverse selection and unemployment insurance." *Journal of Political Economy*, 106(4): 806–827.

Coble, K.H., and T.O. Knight. 2002. "Crop insurance as a tool for price and yield risk management." 445–468. Boston: Kluwer Publishers.

Cohen, A., and R. Dehejia. 2004. "The effect of automobile insurance and accident liability laws on traffic fatalities." *The Journal of Law and Economics*, 47(2): 357–393.

Finkelstein, A., and J.M. Poterba. 2006. "Testing for Adverse Selection with "Unused Observables"." *NBER Working Paper No. 12112*.

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

Just, R.E., L. Calvin, and J. Quiggin. 1999. "Adverse selection in crop insurance: Actuarial and asymmetric information incentives." *American Journal of Agricultural Economics*, 81(4): 834–849.

Keeley, M.C. 1990. "Deposit insurance, risk, and market power in banking." *The American Economic Review*, 80(5): 1183–1200.

Quiggin, J., G. Karagiannis, and J. Stanton. 1993. "Crop insurance and crop production: an empirical study of moral hazard and adverse selection." *Australian Journal of Agricultural Economics*, 37(2): 95–113.

Smith, V.H., and B.K. Goodwin. 1996. "Crop insurance, moral hazard, and agricultural chemical use." *American Journal of Agricultural Economics*, 78(2): 428–438.

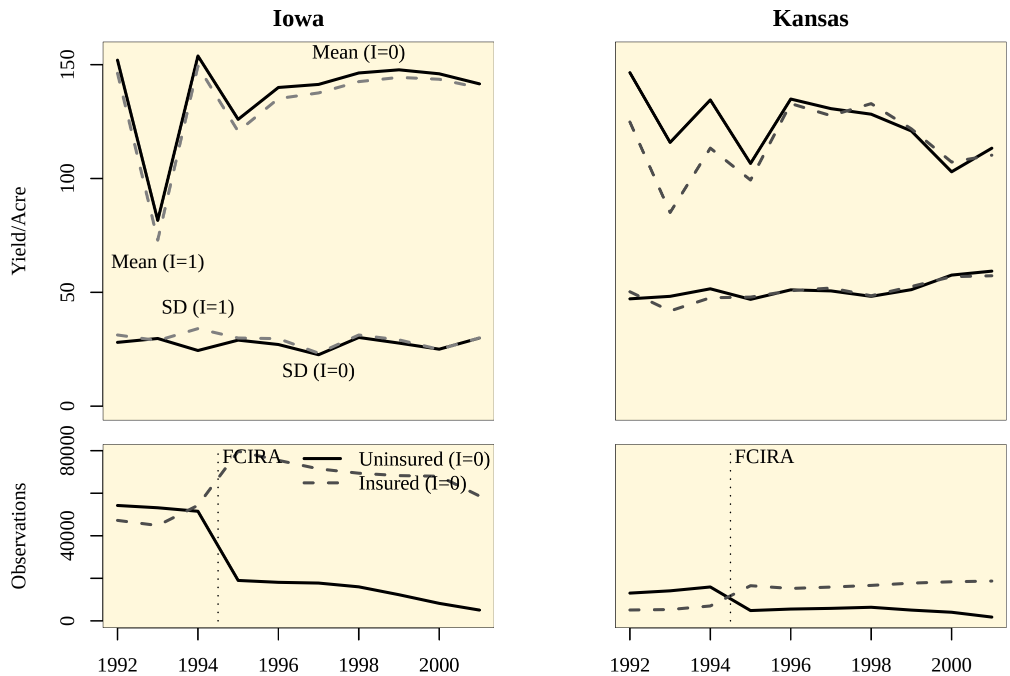

Figure 1. Summary Statistics for Two Representative Corn States

Notes: Average yields and standard deviations are reported. The sample includes the population of insured yields and uninsured yields reported in the crop history of insured yields (all uninsured yields were insured at some later point in time). The vertical FCIRA line indicates the Federal Crop Improvement and Reform Act of 1994. This Act was in effect for 1995 and not in effect in 1994, so the line is drawn between these years. Summary statistics for other states are available from the authors upon request.

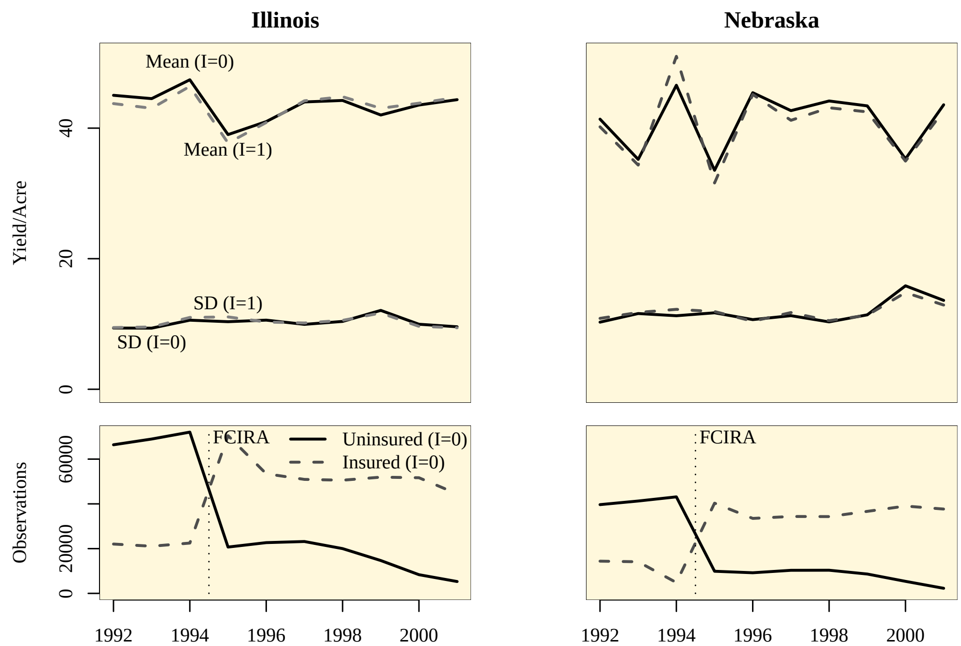

Figure 2. Summary Statistics for Two Representative Soybean States

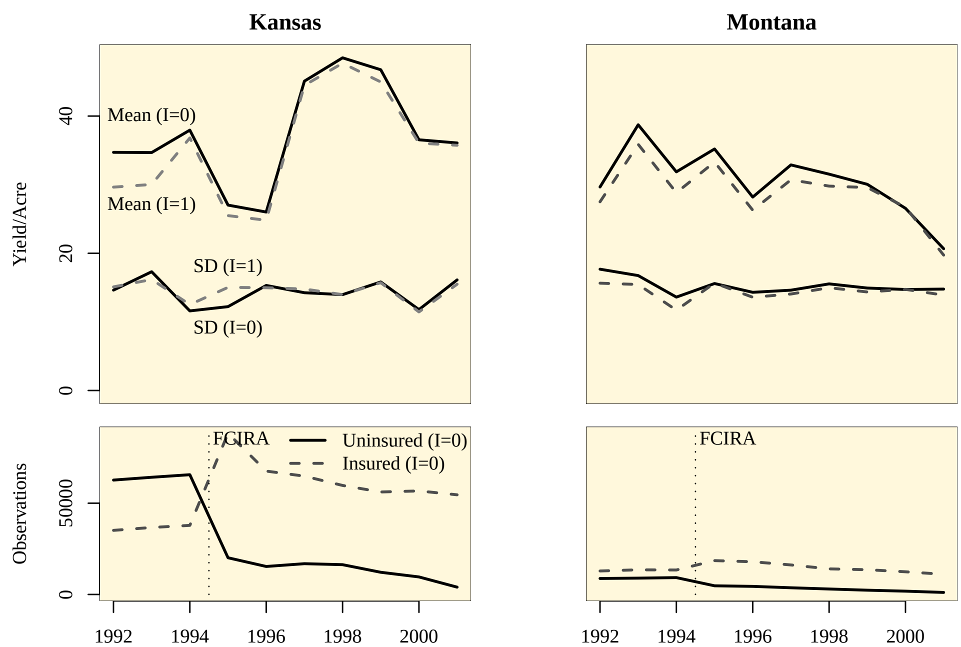

Figure 3. Summary Statistics for Two Representative Wheat States

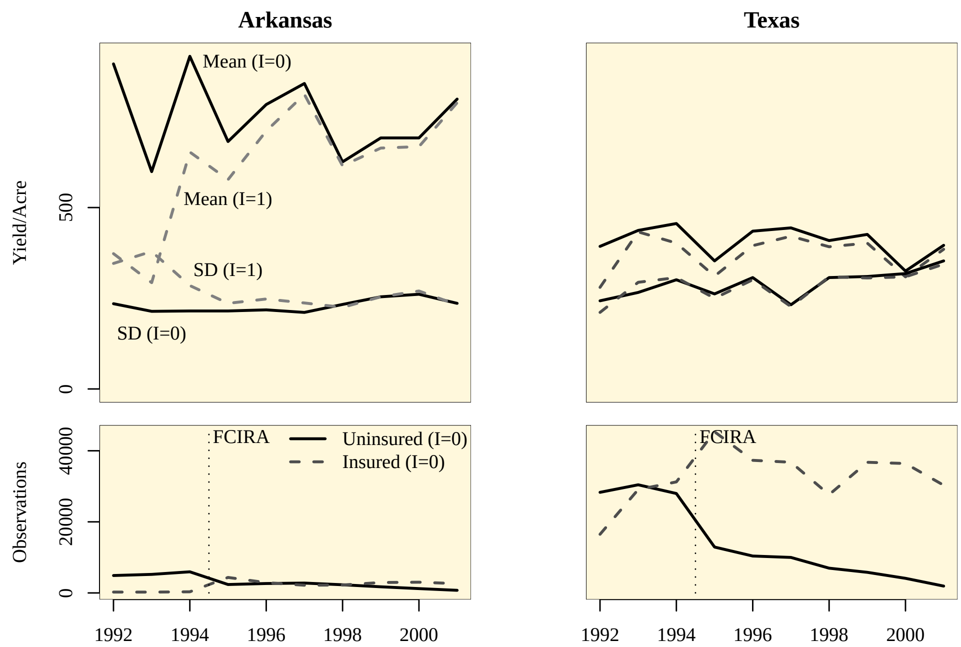

FIGURE 4. SUMMARY STATISTICS FOR TWO REPRESENTATIVE COTTON STATES

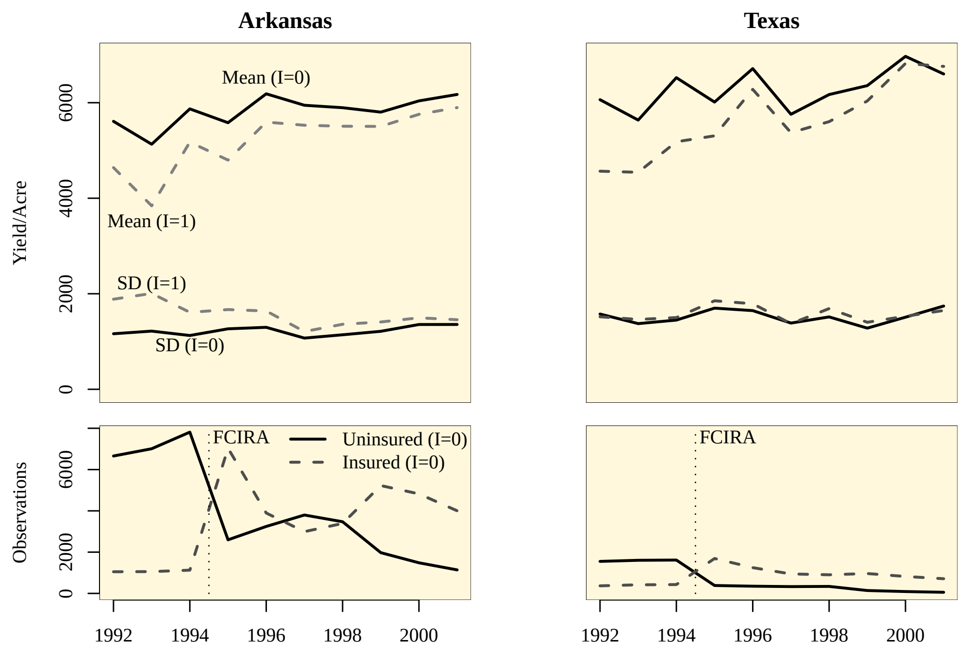

FIGURE 5. SUMMARY STATISTICS FOR TWO REPRESENTATIVE RICE STATES

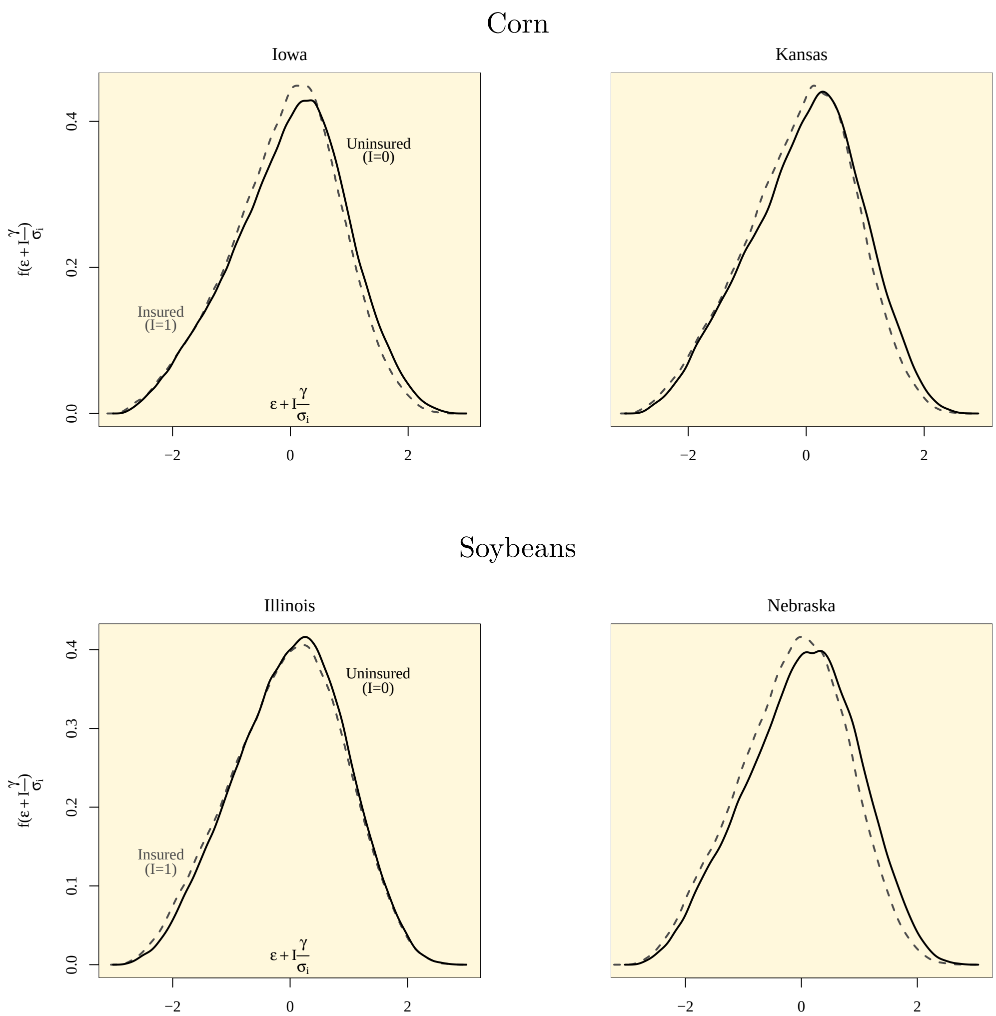

Figure 6. Representative Yield Distribution Shifts for Corn and Soybeans

Notes: The figures show estimated conditional error distributions for two representative states for each crop. The solid black line (Uninsured) shows a kernel density estimate of the farm-specific standardized error distribution for uninsured farms $f(\epsilon_{it}|I = 0)$ and the grey dashed line (Insured) shows the standardized mean shift in yields (γ/σ_i plus estimated error density for insured farms $f(\epsilon_{it} + \gamma/\sigma_i|I = 1)$. To limit measurement error from farm-specific variance estimates (σ_i), only residuals from farms with six or more observations were used in density estimation. Nonparametric kernel densities were estimated using the 'density' function and the default bandwidth selection in the software package 'R' based on Silverman's (1986) rule. Estimates of indemnities paid due to moral hazard, reported in tables 1A-1E, were derived by integrating the difference between these two curves over actual indemnities paid. See text for details.

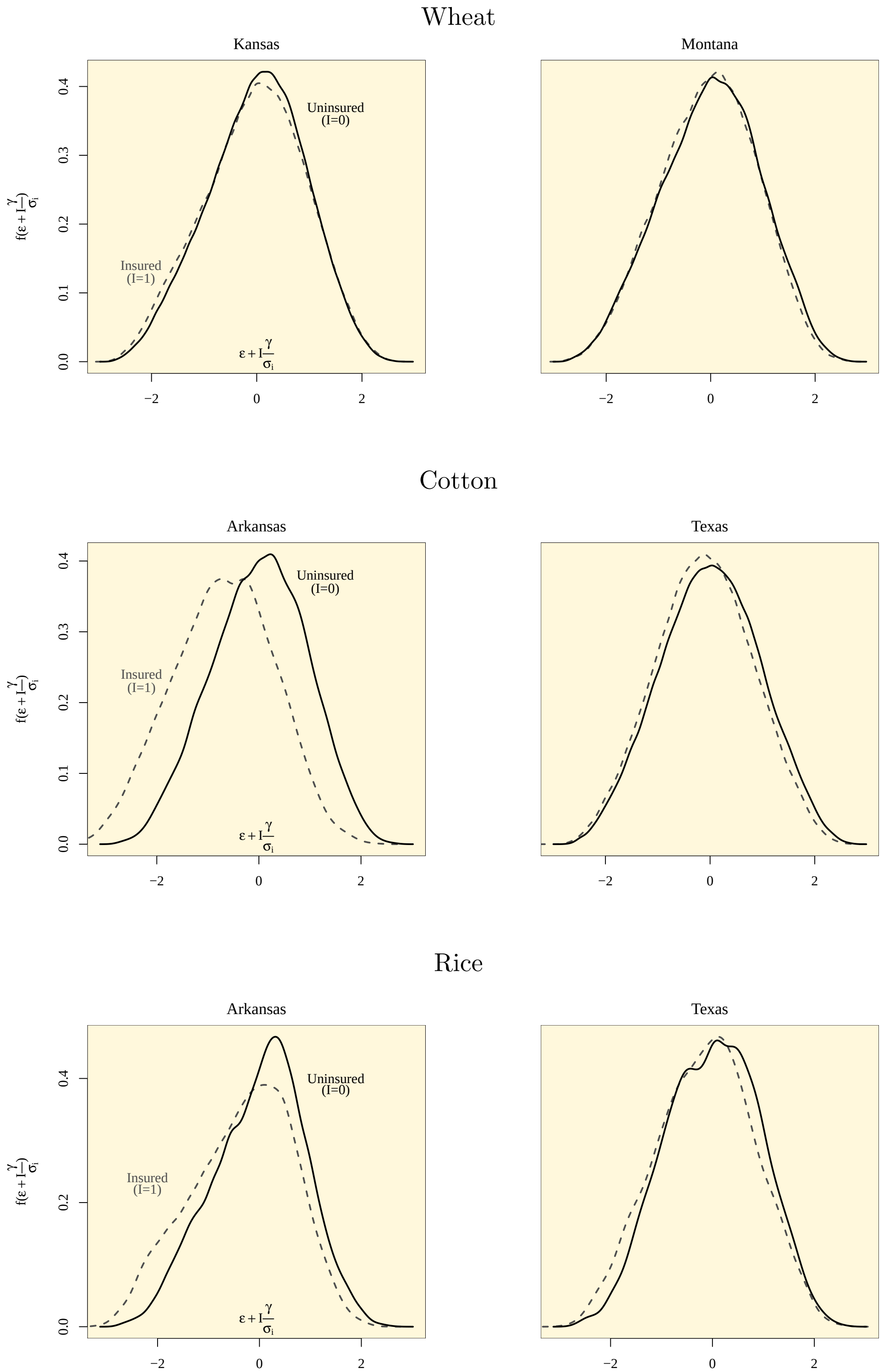

FIGURE 7. REPRESENTATIVE YIELD DISTRIBUTION SHIFTS FOR WHEAT, COTTON AND RICE

Notes: See notes in figure 6.

Table 1—Results for Corn

State (1)	Mean Yield (1992-2001) (2)	$\hat{\gamma}$ (3)	SE of $\hat{\gamma}$ (4)	T-Stat (5)	R^2 (6)	N (7)	*Indemnities Paid Due To Moral Hazard* Share (8)	Total (9)
	(bu/ac)	(bu/ac)	(bu/ac)				(%)	(dollars)
IA	134	-1.22	0.07	-17.03	0.61	861,941	0.47	2,407,733
IN	134	-3.6	0.15	-24.27	0.39	236,696	2.03	2,617,621
IL	141	-0.96	0.08	-12.3	0.43	746,758	0.3	856,769
KS	121	-2.16	0.18	-11.82	0.39	191,315	1.55	736,422
MT	51	7.98	2.06	3.87	0.16	2,420	-7.43	21,503
ND	62	0.4	0.28	1.43	0.48	45,662	-0.23	128,190
NE	122	-0.87	0.1	-9.02	0.39	596,704	0.4	751,072
OH	128	-3.5	0.19	-18.94	0.39	170,031	1.28	1,221,669
TX	105	-1.28	0.25	-5.15	0.44	108,534	0.75	484,182
							Total:	8,003,492

Notes: Separate regressions were estimated for each state. Columns 3-5 report the estimated value, standard error, and t-statistic associated with γ in equation 1–the mean shift in yield associated with insurance. Column 6 reports the R^2 of the regression after farmer-practice fixed effects have been removed from the data (it includes variance explained by county-by-year fixed effects). Column 7 reports the total number of observations used for estimating the error distributions (this excludes farms with fewer than 3 observations). The estimated share of indemnities due to moral hazard (8) accounts for the shift in the error distribution (estimated non-parametrically) around the mean. The estimated budgetary loss is column 8 multiplied by total indemnities paid in the sample. Note that the sample excludes a few outliers so total indemnities paid are slightly less than actual amounts paid. See the text for details.

Table 2—Results for Soybeans

State (1)	Mean Yield (1992-2001) (2)	$\hat{\gamma}$ (3)	SE of $\hat{\gamma}$ (4)	T-Stat (5)	R^2 (6)	N (7)	*Indemnities Paid Due To Moral Hazard* Share (8)	Total (9)
	(bu/ac)	(bu/ac)	(bu/ac)				(%)	(dollars)
IA	44	-0.15	0.02	-6.87	0.53	1,223,312	0.16	262,349
IN	43	-1.27	0.05	-26.01	0.27	235,807	2.26	1,517,840
IL	43	-0.25	0.02	-9.95	0.23	732,649	0.30	227,048
KS	31	-0.54	0.06	-9.74	0.47	214,531	0.40	1,985,422
ND	29	-0.26	0.09	-2.9	0.3	65,610	2.02	163,517
NE	40	-0.79	0.04	-19.88	0.36	416,344	1.62	1,496,462
OH	41	-0.91	0.06	-15.75	0.38	178,193	1.50	973,738
TX	27	-0.87	0.28	-3.07	0.46	13,148	1.65	119,484
							Total:	7,230,042

Notes: See notes to table 1.

Table 3—Results for Wheat

State (1)	Mean Yield (1992-2001) (2)	$\hat{\gamma}$ (3)	SE of $\hat{\gamma}$ (4)	T-Stat (5)	R^2 (6)	N (7)	*Indemnities Paid Due To Moral Hazard* Share (8)	Total (9)
	(bu/ac)	(bu/ac)	(bu/ac)				(%)	(dollars)
IA	34	-0.99	1.19	-0.83	0.63	1,101	1.81	7,301
IN	54	-1.87	0.22	-8.58	0.47	40,987	1.89	142,643
IL	50	-1.33	0.14	-9.22	0.49	99,990	1.26	256,694
KS	36	-0.28	0.04	-7.61	0.51	805,691	0.40	1,449,675
MT	30	-0.27	0.07	-3.78	0.4	178,525	0.71	1,941,474.00
ND	31	-0.74	0.04	-20.03	0.41	513,646	0.81	6,276,735
NE	36	-0.62	0.08	-7.71	0.32	183,768	0.91	658,914
OH	57	-1.07	0.14	-7.92	0.5	85,366	1.15	98,604
TX	27	-0.99	0.07	-13.56	0.37	199,494	1.13	2,388,344
							Total	13,220,384

Notes: See notes to table 1

Table 4—Results for Cotton

State (1)	Mean Yield (1992-2001) (2)	$\hat{\gamma}$ (3)	SE of $\hat{\gamma}$ (4)	T-Stat (5)	R^2 (6)	N (7)	*Indemnities Paid Due To Moral Hazard* Share (8)	Total (9)
	(bu/ac)	(bu/ac)	(bu/ac)				(%)	(dollars)
AR	731	-79.82	5.47	-14.59	0.43	47,536	6.38	1,358,044
CA	1,208	-56.7	9.51	-5.96	0.37	16,825	3.29	1,227,915
TX	388	-10.73	0.68	-15.88	0.39	441,059	1.64	21,822,102
							Total	24,408,061

Notes: See notes to table 1

TABLE 5—RESULTS FOR RICE

State	Mean Yield (1992-2001)	$\hat{\gamma}$	SE of $\hat{\gamma}$	T-Stat	R^2	N	*Indemnities Paid Due To Moral Hazard* Share	Total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	(bu/ac)	(bu/ac)	(bu/ac)				(%)	(dollars)
5,555	-234.65	15.56	-15.08	0.12	69,246	2.32	910,987	
7,528	-26.64	40.09	-0.66	0.24	18,878	-0.90	-51,637	
5,917	-109.7	30.33	-3.62	0.22	14,170	1.01	186,173	
						Total	799,653	

Notes: See notes to table 1

Appendix: Sensitivity to Observations Per Farm

Our data have between three and 10 observations for each farmer-practice i. For purposes of density estimation, we divide the residuals associated with each farmer-practice by an estimated farm-practice specific standard deviation, s_i. Because the estimated standard deviation is based on relatively few observations, it may cause the estimated density function of the standardized errors to be biased too wide.

To investigate the severity of this potential problem we conducted a Monte Carlo simulation. In this simulation we generated farm-specific errors for 10,000 farms, each with three to 10 observations and each with a unique farm-specific variance drawn randomly from a uniform distribution between 25 and 100. In all cases the errors were drawn pseudo-randomly from a normal distribution with mean zero. We then estimated farm specific variances using the simulated data and constructed $\hat{\epsilon}$ like we do with the real data. The correlation between the simulated $\hat{\epsilon}$ and the true ϵ for farms with just 3 observations was relatively weak (0.70), but we found that this correlation grew quickly with larger numbers of observations. With 4 observations per farm the correlation increased to 0.85, for 5 observations it grew to 0.91, and for 6 observations it improved to 0.94. When 10 observations per farm are used the correlation is 0.97. Since there is little remaining error when 6 or more observations are used in farm-specific variance estimation, for purposes of density estimation, we restrict the sample to farms having 6 or more observations.

As a cross check, we estimated the share of indemnities due to moral hazard using density estimates for all subsamples split according to the number of observations per farm. We did this for the 10 crop-states (two states for each crop) for which we present density estimates. These results are summarized in Figure A1. The top panel in the figures shows the estimated share of indemnities due to moral hazard if density estimation is done using only farms with X or more observations. A separate line is drawn for each state-crop. The two lower panels show how the total number of observations declines as the samples become more restricted. The sample sizes for the most restricted samples (farms with 10 observations) is between half and one-third the size of the unrestricted samples. Despite significant variation in the overall sample sizes and the potential bias stemming from farm-specific sample sizes, there is little change in the estimated share of indemnities due to moral hazard.

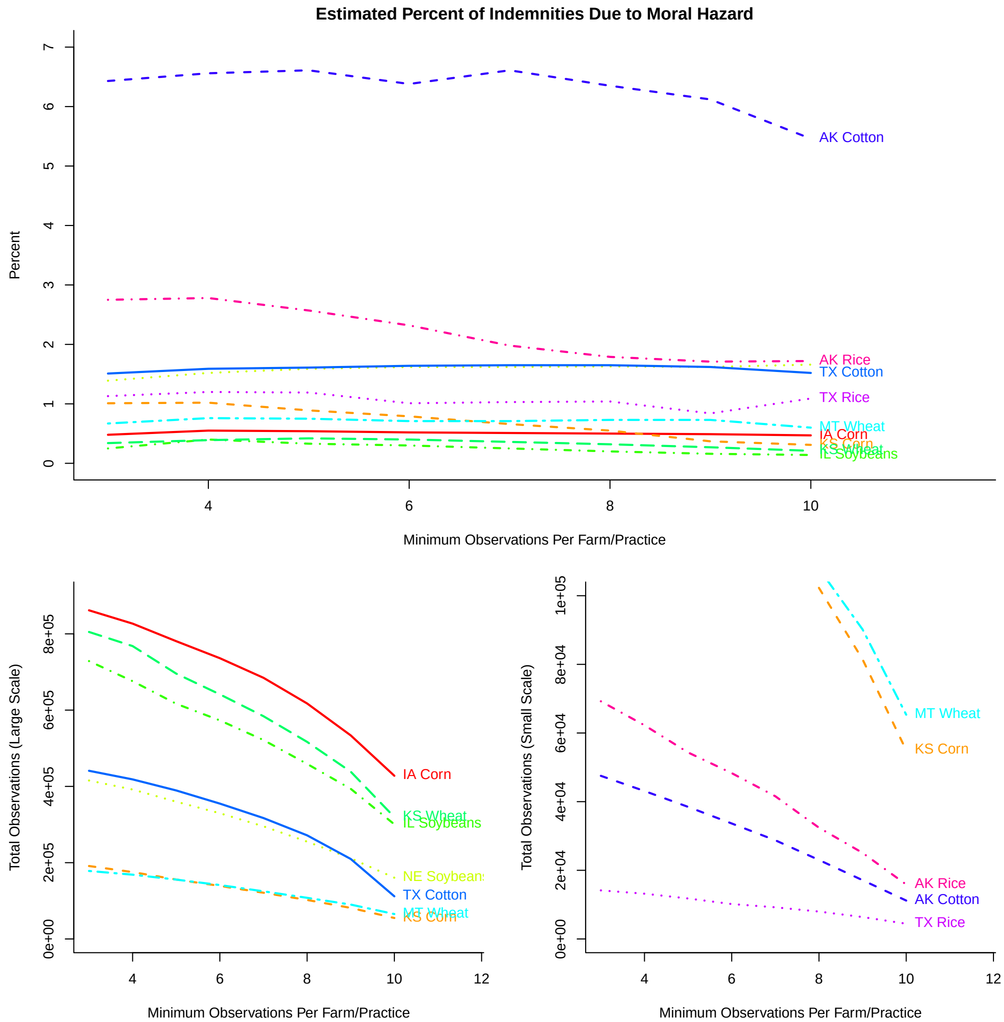

Figure A1. Robustness of Indemnity Cost Estimates to Farm-Practice Sample Size

Notes: The top figure plots the estimated share of indemnities due to moral hazard using different subsamples of the data. Subsamples are selected according to the number of observations used in estimating the farm-specific error varianceσ_i. The bottom two figures show how the total sample size declines as the sample is limited to farms with more observations. Because the number of observations varies so much across states, two plots are presented, the bottom left with y-axis ranging from 0 to 900,000, and the bottom right with a y-axis ranging from 0 to 100,000. Two crops are shown for each state, the same crops and representative states shown in figures 1-7.

Appendix: Data

The Risk Management Agency (RMA) of U.S. Department of Agriculture manages the crop insurance program. To run and monitor the Federal crop insurance program, RMA collects data on all individual crop insurance contracts for all covered crops and maintains the dataset confidentially. Data summarizing coverage, premiums, subsidies and indemnities at the county, state, and national levels are available online at RMAs website. The contract-level data used in this study includes millions of individual observations for a wide range of crops. For example, for the state of Iowa alone, RMA collected almost 240,000 individual yield histories and data on more than 200,000 individual insurance policies that producers enrolled in for 15 different types of crops for the 1992 crop year the first year we include in our analysis. These numbers grew to over 400,000 individual yield histories and data on more than 345,000 individual policies selected by producers for 14 types of crops for the 2001 crop year, the last year of data used in this study.

B1. Merging Data Types

Data used in this study combine four datasets, labeled by RMA as Type10, Type11, Type15, and Type21. Type10, which is called the Policy Record, contains a set of identifying variables for every farmer who obtains insurance. Type11 dataset, called the Acreage Record, holds data concerning the policy coverage purchased by the producer. The Type15 dataset, or Yield Record, records yield histories for each farmer, stretching back up to ten years. Finally, Type21 data, called the Loss Line, contains information on indemnities (if any) received by producers. The variables we used from each of these datasets can be found in table 1 while table 2 contains the number of raw observations for each crop year and data type.

Before merging the four datasets, we dropped duplicate entries (which occurred primarily in the Type10 dataset), and created uniform units of measurement when necessary (e.g., reported acres was originally recorded in the hundredths of acres, which we converted to acres). We next merged the Type 10 and Type 11 datasets and Type 10 and Type 15 datasets by matching observations according to reporting organization, insurance company, state, policy number, and crop year. Since Type 10 held the unique farmer identifier and was merged with the Type 11 and Type 15 files, we could then merge Type 11 and Type 15 datasets together for each individual, crop, and practice in each county by matching the same variables as before, along with the following additional variables: the unique farm identifying number, the crop code (the crop enrolled), the plan code (type of insurance plan), the county, the unit number, the type code (a code that identifies the crop type, class, or variety), the practice code (irrigated or not), and a coverage flag (whether the policy was for catastrophic coverage or more substantial buy-up insurance).

For the Type 21 dataset we summed indemnities by insurance plan, policy number, and coverage level in each year, crop, crop type, and practice in each state and county on each unit for each company and organization delivering the crop insurance. This aggregates indemnities across different insurable units in each farm, which we had to do because we are unable to match outcome yields (from the subsequent contractsee below) with specific insurable units. While individual indemnities for a particular crop and unit may be negative (perhaps an adjustment), the sum of the indemnities in a particular crop year should not be negative. Therefore, we eliminated any observations with an overall negative indemnity for a particular crop.

We then merged indemnities with the merged Type 10, 11, and 15 data using the merge variables described above. We replicated this process for each state and year, leaving us with a data set for each state and each crop year (1992 through 2002). Each observation in this dataset contained information on the type and level of policy enrolled in for the current crop year, the relevant yield history, and any indemnities that might have been paid out for the current year for each unit on the farm for each crop type and practice in each county. See table 3 to get a sense of the total raw number of observations by year, state, and crop.

B2. Panel Construction

Next, we constructed panel datasets for each state and crop from the state-by-year datasets. This process accomplishes two key tasks: (1) it allows us to identify the years in which farmers entered or exited the insurance program or changed insurance levels from minimal catastrophic coverage to more substantial buy up coverage, or vice versa, and (2) because crop year t Type 15 history data only includes yield outcomes for the 10 crop years leading up to crop year t (i.e., the most recent yield outcome is for crop year t-1), the panel dataset construction allows us to match Type 15 yield history outcome data from the crop year t+1 to Type 11 crop year t insurance policy information.

To accomplish these tasks, we first pool all the years for a particular state. If any values for the crop codes, crop years, identification numbers, coverage levels, total premiums, and number of insured acres were missing, or if either the identification number or coverage level was equal to zero, the observations were deleted.

We then created a dataset that, for each individual in each year, sums total premiums paid for insurance, the number of acres insured, and the indemnities (if any). This process aggregates over insurable units within a farm, crop and practice, like we did for indemnities (Type 21) before merging. Then, because each observation has a history of yields and acres insured (for up to 10 years), we reshape these data so that each year of historical yield is assigned a separate individual-by-year observation. Since an individual who continues to enroll in crop insurance will have a history file for each contract year, several of the histories will overlap. For example, if an individual enrolled in crop insurance for a corn crop from 1992 through 2000, then the individual would have 9 years worth of observations in the dataset, and each of those 9 years would have up to 10 years worth of histories associated with them. The 1993 data would contain 1992 yields and acres enrolled as history, 1994 would contain 1993 as well as the 1992 yields and acres enrolled as part of its history, and so forth. We eliminate these duplicate yield outcomes when creating the final panel data set.

Another data issue arises because within each year of the panel data, there is the potential for an individual to have multiple insurable units of a particular crop, creating multiple entries for each crop for an individual in a county in a year. Often, this results from operators dividing up units from one year to the next (producers can also merge units), making it impossible to track histories of specific units. Furthermore, it is not clear that assigned unit numbers remain the same across contract years, inhibiting tracking of specific units even if units are not split or merged. Therefore, for every individual, we generated two separate datasets one which averaged the yield history (weighted by the number of acres in each field) and a second that summed the number of acres of all fields of the crop. This generates a single acre-weighted yield and number of acres insured for each individuals crop-practice within a county. If any yields or acres insured are missing or equal to zero, we eliminated the observation. We also eliminated observations with implausibly high yields (greater than 10,000). If a farmer has fields in multiple counties, units in different counties were treated as separate observations and were ultimately assigned separate farmer fixed effects.

We also create an insurance indicator variable equal to zero if the individual either carried no insurance or only enrolled in catastrophic coverage and equal to one otherwise. For example, if an individual had insurance in 1993, but was not in the dataset in 1992, then any historical data found in the years 1992 (or earlier) would coincide with no insurance. For years after 1993, we have information on whether or not the individual had a policy and what coverage level was assigned to that policy.

To restrict the dataset to the time frame where entry or exit from the program can be observed, we kept those years between 1992 and 2001 (since we only had contract data through 2002, this only provides us with historical outcome yields for years up through 2001).

Note that the methods described above result in a larger number of individual producers than actually enrolled in the insurance program. An individual could have multiple crops insured with crop insurance. Additionally, farmers could utilize a different practice (irrigate or not) on their crops, or have the same crop and practice in multiple counties. To simplify our task and allow for appropriate comparisons later in our analysis, we treated each identification number-state-county-crop-practice combination as a separate individual and each identification number-state-county-crop-practice-year as a separate observation. For example, we created a history for identification number 1 growing corn without irrigating in state 1, county 1 and treated it as a separate individual from identification number 1 growing corn with irrigation in state 1, county 1 and from identification number 1 growing corn without irrigating in state 1, county 2, and also separately from identification number 1 growing wheat without irrigating in state 1, county 1, etc. Even though identification number 1 in state 1, county 1 is likely the same individual (since we believe the identification number to be unique to the individual), we separate this hypothetical individual into four individuals.

To prepare the data for the econometric analysis, then we eliminated any individuals who only had one observation since we needed more than one observation over time for identification. We ended up with a panel dataset for each state (see tables 4 and 5 for number of observations and average values of main variables).

B3. Data Checking

Finally to check our constructed panel datasets, we created totals for the number of acres insured, premiums, and indemnities paid out and compared these totals with publicly available RMA data posted on their website in their Summary of Business files for Iowa corn, soybeans, and wheat and California cotton and rice for the years 1993, 1995, 1997, 1999, and 2001 (i.e. every other year in the dataset we constructed). Results suggest that, after cleaning the data and preparing it for analysis, our estimates are very similar to those reported by RMA (table 6).

Table B1—Variables in Each Data Type

Type 10 *Policy Record*	**Type 11** *Acreage Record*	**Type 15** *Yield Record*	**Type 21** *Loss Line*
Crop Year	Crop Year	Crop Year (t)	Crop Year
State	State	State	State
Policy Issuing Company	Policy Issuing Company	Policy Issuing Company	Policy Issuing Company
Insurance Provider	Insurance Provider	Insurance Provider	Insurance Provider
Policy Number	Policy Number	Policy Number	Policy Number
Producer ID			
	County	County	County
	Crop	Crop	Crop
	Plant Type	Plant Type	Plant Type
	Production Practice	Production Practice	Production Practice
	Insured Unit ID	Insured Unit ID	Insured Unit ID
	Type of Insurance	Type of Insurance	Type of Insurance
	Coverage Level		Coverage Level
	Number of Acres in Unit		
	Share of Acres Insured in Unit		
	Amount of Insurance Purchased		
	RMA Approved Yield		
	Total Premiums		
	Total Subsidies		
	Guarantee per Acre		
	Coverage Flag (CAT vs Buy-up)		
		Up to 10 years of history:	
		Crop Year	
		Crop Yield	
		Yield Type (actual or proxy)	
		Acres Planted	
			Indemnities

Notes: The raw data include a separate file for each state and contract year. Contract years were merged to link current insurance and indemnity information to actual yields reported in the history file (Type 15) of contracts in subsequent years.

Table B2—: Observations in Each File

Year	**Type 10** *Policy Record*	**Type 11** *Acreage Record*	**Type 15** *Yield Record*	**Type 21** *Loss Line*
		Arkansas		
1992	7,414	11,829	18,934	4,619
1993	6,341	11,477	18,241	5,191
1994	9,872	15,278	22,007	3,921
1995	59,528	90,150	115,788	6,315
1996	50,995	77,377	78,508	3,464
1997	31,139	44,747	46,031	3,571
1998	29,755	39,168	44,153	7,304
1999	33,729	52,301	56,656	9,907
2000	37,200	57,436	66,157	12,227
2001	38,146	57,768	68,936	7,924
		California		
1992	7,573	11,853	8,296	888
1993	5,870	9,429	7,482	1,424
1994	8,247	12,287	9,939	819
1995	37,069	45,528	48,142	6,497
1996	34,604	40,317	44,425	2,510
1997	32,260	36,366	38,182	1,713
1998	29,836	39,998	40,307	8,755
1999	34,847	48,301	48,156	5,382
2000	36,794	49,692	52,228	4,007
2001	41,445	49,478	51,624	6,407
		Iowa		
1992	106,070	207,271	239,645	8,533
1993	96,744	198,235	225,790	106,917
1994	130,318	252,143	289,905	6,123
1995	208,631	395,044	429,027	43,853
1996	207,835	372,932	424,093	25,906
1997	173,794	335,091	386,019	7,239
1998	170,867	333,498	412,415	30,474
1999	166,938	339,436	384,413	32,303
2000	171,854	345,328	403,403	37,047
2001	166,287	345,198	406,609	61,830
		Illinois		
1992	69,930	162,069	162,513	15,031
1993	69,178	161,220	165,551	14,765
1994	75,497	167,868	167,947	8,585
1995	221,514	467,793	479,447	43,637
1996	189,660	404,428	404,740	51,616

Continued on next page

Table B2 – continued from previous page

Year	**Type 10** *Policy Record*	**Type 11** *Acreage Record*	**Type 15** *Yield Record*	**Type 21** *Loss Line*
1997	142,333	316,632	322,691	13,171
1998	140,317	314,319	347,060	36,531
1999	142,819	326,382	352,862	39,230
2000	151,409	344,907	390,680	33,725
2001	144,098	332,297	386,371	27,651
		Indiana		
1992	22,934	61,091	68,127	6,304
1993	24,223	61,514	75,124	5,129
1994	24,290	59,931	75,869	4,223
1995	88,471	174,775	202,966	18,559
1996	74,989	149,189	162,426	26,548
1997	47,271	107,646	131,593	14,355
1998	47,667	110,402	144,658	18,426
1999	50,265	12,124	158,122	24,382
2000	55,361	137,081	184,151	17,812
2001	53,951	133,905	185,420	9,745
		Kansas		
1992	84,946	184,556	226,572	38,765
1993	74,939	182,566	216,914	43,969
1994	94,078	197,982	267,938	12,952
1995	229,253	403,018	500,421	82,948
1996	232,371	419,756	527,707	99,403
1997	183,337	332,001	468,214	16,453
1998	175,280	315,742	476,472	20,071
1999	175,146	328,122	485,918	42,829
2000	189,251	334,602	540,941	88,311
2001	195,892	363,424	593,702	84,020
		Montana		
1992	22,863	60,988	89,321	19,210
1993	21,048	59,247	106,744	4,842
1994	22,198	55,959	101,016	5,574
1995	36,829	80,898	158,803	7,098
1996	34,128	80,548	159,972	13,764
1997	29,348	69,270	153,654	5,960
1998	30,429	66,758	151,307	11,703
1999	29,185	69,911	136,065	13,768
2000	31,906	76,931	147,306	28,389
2001	28,954	85,024	144,652	47,881
		Nebraska		
1992	70,622	144,818	195,691	25,757

Continued on next page

Table B2 – continued from previous page

Year	**Type 10** *Policy Record*	**Type 11** *Acreage Record*	**Type 15** *Yield Record*	**Type 21** *Loss Line*
1993	65,823	143,534	179,747	30,963
1994	82,122	160,784	218,917	9,718
1995	169,171	319,975	388,743	47,738
1996	173,416	300,859	399,393	23,191
1997	145,189	267,318	364,681	18,178
1998	141,844	258,197	390,186	14,496
1999	137,846	264,508	360,971	21,491
2000	148,468	279,987	408,035	81,295
2001	147,980	283,385	422,665	43,790
		North Dakota		
1992	104,223	236,957	469,765	21,268
1993	84,134	216,516	407,929	53,926
1994	94,577	209,727	415,609	29,231
1995	154,288	328,359	556,686	72,839
1996	151,345	297,701	604,761	41,375
1997	140,184	272,048	614,477	80,022
1998	143,073	252,455	641,117	41,719
1999	148,237	283,310	643,603	120,036
2000	157,344	268,724	722,983	72,361
2001	158,065	276,963	688,670	88,137
		Ohio		
1992	17,595	33,947	41,643	5,200
1993	16,894	35,150	45,194	6,709
1994	19,228	35,667	47,248	3,021
1995	78,909	133,538	141,343	12,011
1996	63,017	123,261	123,940	24,444
1997	45,542	87,722	96,728	8,659
1998	38,818	68,878	95,706	5,763
1999	38,627	75,936	102,355	15,860
2000	41,911	87,434	118,765	14,992
2001	41,845	88,836	126,129	17,164
		Texas		
1992	61,602	108,624	114,791	42,693
1993	60,350	128,731	132,671	33,575
1994	79,136	135,607	149,652	43,396
1995	157,198	249,834	282,560	84,661
1996	157,576	255,825	299,209	104,145
1997	145,352	233,874	275,326	34,809
1998	142,522	226,483	295,882	106,425
1999	162,314	241,012	299,375	63,333

Continued on next page

Table B2 – continued from previous page

Year	Type 10 *Policy Record*	Type 11 *Acreage Record*	Type 15 *Yield Record*	Type 21 *Loss Line*
2000	175,733	242,445	356,007	119,716
2001	176,947	243,188	365,743	102,419

Continued on next page

Table B3—: Observations by Crop, State and Year

Year	Corn	Soybeans	Wheat	Rice	Cotton
			Arkansas		
1992	308	6,796	4,392	3,950	1,199
1993	336	5,287	5,566	3,694	1,153
1994	435	8,772	4,073	4,319	1,986
1995	1,831	45,813	18,133	21,927	15,525
1996	3,054	38,317	20,408	15,681	11,726
1997	1,346	23,222	8,765	9,280	8,064
1998	1,384	17,551	6,668	8,232	6,925
1999	1,170	24,012	7,426	12,228	8,956
2000	2,034	28,783	8,897	12,676	10,212
2001	2,526	29,993	9,135	13,809	11,438
			California		
1992	8	0	293	58	101
1993	18	0	296	204	267
1994	13	0	286	96	164
1995	2,717	0	4,404	5,806	4,003
1996	2,357	0	4,870	4,503	3,057
1997	1,453	0	3,144	3,298	2,232
1998	1,103	0	2,192	2,409	2,154
1999	1,263	0	2,660	2,704	2,050
2000	1,529	0	2,955	3,046	2,614
2001	1,444	0	3,515	2,986	3,049
			Iowa		
1992	137,617	107,119	409	0	0
1993	130,903	101,756	441	0	0
1994	159,521	138,805	571	0	0
1995	223,266	223,917	650	0	0
1996	223,011	220,470	908	0	0
1997	196,203	199,968	681	0	0
1998	200,793	211,135	569	0	0
1999	203,571	190,231	390	0	0

Continued on next page

Table B3 – continued from previous page

Year	Corn	Soybeans	Wheat	Rice	Cotton
2000	212,815	197,699	526	0	0
2001	216,514	205,921	452	0	0
			Illinois		
1992	100,667	71,921	6,133	0	0
1993	96,258	68,685	14,450	0	0
1994	99,936	73,937	10,156	0	0
1995	231,792	238,762	46,932	0	0
1996	208,515	214,359	45,429	0	0
1997	165,495	163,419	27,779	0	0
1998	166,660	170,387	27,139	0	0
1999	176,428	179,223	23,398	0	0
2000	193,257	195,370	25,182	0	0
2001	190,830	195,148	23,643	0	0
			Indiana		
1992	39,355	30,347	2,153	0	0
1993	37,704	29,394	8,139	0	0
1994	38,444	28,938	5,577	0	0
1995	95,183	93,971	19,354	0	0
1996	85,682	86,327	18,482	0	0
1997	63,701	61,053	10,914	0	0
1998	68,214	66,644	10,210	0	0
1999	76,900	75,339	9,797	0	0
2000	88,587	89,232	10,355	0	0
2001	88,620	90,415	9,734	0	0
			Kansas		
1992	17,947	32,956	132,123	0	51
1993	17,620	30,184	130,843	0	74
1994	24,479	50,099	139,159	0	52
1995	50,753	71,302	250,218	0	66
1996	55,143	69,818	269,726	0	58
1997	52,001	62,221	230,868	0	46
1998	57,609	65,151	220,420	0	39
1999	64,819	75,525	221,794	0	499
2000	75,976	86,578	230,906	0	815
2001	87,277	102,288	249,489	0	1,185
			Montana		
1992	283	0	69,690	0	0
1993	230	2	79,770	0	0
1994	278	0	71,753	0	0
1995	1,131	0	106,672	0	0
1996	817	0	112,762	0	0

Continued on next page

Table B3 – continued from previous page

Year	**Corn**	**Soybeans**	**Wheat**	**Rice**	**Cotton**
1997	667	0	106,773	0	0
1998	634	0	97,825	0	0
1999	612	0	100,167	0	0
2000	681	0	98,981	0	0
2001	733	0	101,188	0	0
			Nebraska		
1992	67,402	53,561	43,032	0	0
1993	65,918	45,645	43,640	0	0
1994	77,347	59,620	43,824	0	0
1995	159,351	108,938	62,231	0	0
1996	161,812	106,345	67,486	0	0
1997	147,746	98,835	58,475	0	0
1998	149,468	101,383	52,912	0	0
1999	156,811	110,566	49,000	0	0
2000	169,648	126,751	55,681	0	0
2001	175,768	136,758	57,986	0	0
			North Dakota		
1992	19,998	15,413	231,934	0	0
1993	14,020	12,492	214,839	0	0
1994	15,808	17,198	202,796	0	0
1995	19,180	20,108	257,464	0	0
1996	18,576	25,426	312,775	0	0
1997	18,045	36,074	295,814	0	0
1998	20,787	44,595	263,395	0	0
1999	22,277	34,795	304,243	0	0
2000	26,795	42,116	307,815	0	0
2001	30,129	47,769	258,191	0	0
			Ohio		
1992	20,135	18,291	3,398	0	0
1993	19,494	18,646	6,385	0	0
1994	19,631	18,073	7,485	0	0
1995	58,569	56,978	31,554	0	0
1996	55,835	61,157	33,312	0	0
1997	44,696	45,711	20,422	0	0
1998	37,263	37,659	15,579	0	0
1999	41,817	42,562	16,132	0	0
2000	48,428	51,641	17,568	0	0
2001	50,525	54,898	17,810	0	0
			Texas		
1992	7,841	714	22,100	888	62,427
1993	9,192	846	27,883	1,066	74,374

Continued on next page

Table B3 – continued from previous page

Year	Corn	Soybeans	Wheat	Rice	Cotton
1994	10,889	1,793	25,609	1,247	79,577
1995	32,659	4,921	61,576	3,398	112,708
1996	32,841	4,691	63,253	3,371	112,062
1997	30,231	4,390	62,665	3,326	107,100
1998	33,822	5,615	53,093	2,272	107,114
1999	33,292	7,413	63,741	2,160	110,832
2000	41,319	10,069	76,573	2,214	124,802
2001	41,965	10,497	82,409	2,301	130,729

Continued on next page

Table B4—Observations in Each Constructed Panel

State	Corn	Soybeans	Wheat	Rice	Cotton
AR	11,293	188,876	70,935	106,311	68,984
CA	9,723	0	20,884	26,492	24,259
IA	1,497,380	2,101,202	3,258	0	0
IL	1,151,546	1,095,700	169,672	0	0
IN	381,024	370,814	69,493	0	0
KS	312,536	433,262	1,342,825	0	1,834
MT	4,304	0	344,251	0	0
NE	947,887	775,859	339,616	0	0
ND	92,943	140,051	1,046,034	0	0
OH	261,506	267,681	131,725	0	0
TX	180,974	30,343	362,699	23,469	792,481

Table B5—Variable means by year

	Reported Acres	Yield	Total Premium	Insured Acres	Indemnity
1982	177	134	–	–	–
1983	127	104	–	–	–
1984	177	123	–	–	–
1985	170	134	–	–	–
1986	157	135	–	–	–
1987	132	135	–	–	–
1988	147	90	–	–	–
1989	160	120	–	–	–
1990	180	125	–	–	–
1991	172	120	–	–	–
1992	184	149	922	116	2,844
1993	172	78	882	111	6,590
1994	188	152	1,018	122	1,749
1995	181	122	738	114	4,579
1996	191	136	1,324	124	2,793
1997	181	138	1,157	122	2,995
1998	184	143	1,330	128	5,857
1999	178	145	1,507	130	3,695
2000	181	144	1,874	135	5,683
2001	168	140	2,134	137	5,630

Notes: All data are derived from the contract files from 1992-2002. Values for yield and reported acres for years prior to 1991 come from crop history files (Type 15). We drop 2002 because we have no yield outcome data for this year, only insurance contract information.

Table B6—Selected comparisons of our totals and USDA's published totals

	Insured Acres		**Total Premiums**		**Indemnities**	
	Ours	*RMA*	*Ours*	*RMA*	*Ours*	*RMA*
Corn (IA)						
1993	5,424,967	5,425,075	43,087,602	43,087,602	213,897,381	213,897,381
1995	11,044,041	10,732,178	71,797,002	70,023,641	68,726,170	68,726,170
1997	9,565,559	9,573,095	90,410,487	90,458,170	8,550,576	8,552,085
1999	9,701,187	9,701,455	112,243,230	112,243,230	36,298,488	36,298,977
2001	9,797,346	9,797,822	152,400,191	152,400,541	89,083,521	102,124,008
Soybean (IA)						
1993	3,205,207	3,205,340	16,463,756	16,463,756	62,937,117	62,955,696
1995	8,528,513	8,386,360	30,887,669	30,480,132	14,675,311	14,675,311
1997	8,258,053	8,263,884	45,465,375	45,487,143	4,867,271	4,867,271
1999	8,847,873	8,848,111	54,527,237	54,527,237	21,767,940	21,767,940
2001	9,333,184	9,333,219	73,494,280	73,486,266	41,563,499	48,872,892
Wheat (IA)						
1993	5,569	5,574	41,515	41,515	274,361	274,361
1995	9,881	9,893	74,035	74,035	53,533	53,533
1997	6,830	6,837	88,219	88,219	137,621	137,621
1999	4,411	4,416	52,785	52,785	38,981	38,981
2001	5,566	5,569	85,552	85,552	57,500	62,232
Rice (CA)						
1993	14,003	14,002	78,635	78,635	487	487
1995	467,992	469,788	1,004,729	1,008,668	151,777	151,777
1997	242,946	245,007	982,302	989,762	11,594	11,594
1999	372,840	374,464	2,037,254	2,045,980	1,158,632	1,158,632
2001	357,723	363,764	2,144,482	2,177,446	2,144,156	2,363,184
Cotton (CA)						
1993	39,125	39,126	532,170	532,170	695,270	695,270
1995	1,045,496	1,045,495	6,450,714	6,450,714	2,908,109	2,908,109
1997	564,235	566,439	5,507,274	5,540,391	6,881,618	6,881,618
1999	489,442	489,452	4,542,982	4,542,982	1,108,367	1,108,367
2001	546,509	546,452	8,194,010	8,103,838	14,126,867	17,597,448