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): North Carolina State University, Box 8109, Raleigh NC 27695, michael_roberts@ncsu.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 often hinge on whether it is moral hazard or adverse selection that 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 and sharing of risk that private markets might not otherwise achieve. If both information problems are acute, then the case for public policy requires that moral hazard be sufficiently rectifiable using a schedule of deductibles, co-payments, and exclusions that is simple enough such that they can be feasibly implemented in a cost-effective manner. Although recent debate about health insurance and the Patient Protection and Affordable Care Act have brought these tensions between moral hazard and adverse selection to the fore, the same essential considerations underly all public insurance programs. The critical challenge of policy evaluation is therefore measuring the incidence of moral hazard separate adverse selection. Separating these effects empirically is difficult because in many ways the two problems are observationally similar: bad outcomes tend to be more prevalent among the insured as compared to the uninsured.

In this article we investigate this issue for a lesser known but pervasive public insurance program: 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. That structure 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 significantl 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 enrolled over 200 million acres—nearly two thirds of all U.S. cropland. In nominal dollars, total liability insured under the program steadily increased from almost \$6 billion in 1981, to nearly \$24 billion in 1995, to over \$40 billion in 2003, to almost \$90 billion by 2008. Indemnities have also increased, from about \$1.2 billion in 1989 to almost \$2.4 billion in 1999 to over \$8.6 billion by 2008.[1]

In agriculture, we observe that farmers with insurance have lower yields, but we cannot necessarily tell 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. Similar observational equivalence of moral hazard and adverse selection is present in most insurance environments.

In earlier years of the newer vintage of the program, insurance products mainly covered a farmer-selected portion of "expected" yield, taken to equal the average of up to 10 years of actual production history (APH). Initially, only major field crops could qualify. Yield outcomes below a contracted percentage of expected yield were indemnified at a contracted share of the FCIC price (typically equal or near the futures price around planting). Other, less popular insurance products included group insurance products that indemnified county-level yields rather than individual farmer yields. Group policies are obviously less susceptible to moral hazard or adverse selection problems but also insure less risk. RMA determines premiums based on a farmer's average verified yield history. But conditional on average yield, RMA assumes all fields within a county have the same yield variance for purposes premium determination.[2] There are various adjustments to premiums depending on whether farmers insure multiple field separately, whether yield history is verifiable, if a farmer plants too late in the growing season,

[1]Much of the more recent increases in indemnities stem from increases in commodity prices rather than increased quantities insured. The data used in this study does not cover this recent period.

[2]This is one source of adverse selection that may be partially avoidable (Just, Calvin and Quiggin 1999).

among other factors.

By the late 1990s additional insurance products and crops were added to the program, with a suite of revenue insurance products gaining popularity. These insurance products insured a contracted portion of revenue per acre: expected yield times the future's price. These insurance products were attractive to both farmers and farm lenders because they assured a given cash flow, which could otherwise prove difficult given the large and oftentimes dependent variability of both prices and yields. Coble and Knight (2002) provide a lucid review of the history of the crop insurance program from roughly 1980 to 2000, which are most critical for our analysis here.

Farmers with crop insurance do not bear the full financial consequences of crop failure and so have less incentive to take costly actions to prevent losses. The resulting change in behavior, called moral hazard, occurs because insurance companies cannot observe and thus contract and verify farmers' production decisions. Because moral hazard may increase the probability and severity of crop failure, understanding the extent of moral hazard is important for insurance policy design. Moral hazard could be mitigated by explicitly limiting insurance coverage through larger deductibles or by enforcing stronger links between premiums and past indemnity claims. In the case of federally subsidized crop insurance, policies might be fine-tuned with precise information about the incidence of moral hazard under past insurance programs. These basic questions about the incidence of moral hazard have counterparts in nearly all private and public insurance contracts, including auto, life, employment, health care, and bank deposit insurance.[3]

In addition to moral hazard, policymakers must account for the fact that insurers have asymmetric information about the characteristics of the insured that

[3]Examples include: 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)

can result in incomplete or even non-existent private insurance markets due to adverse selection. While a combination of mandates and subsidies can entice or force risk pooling in hidden-type environments, thereby solving the problem of adverse selection, a potential unintended consequence is that the insured, protected from downside risk, could alter their hidden actions to be less efficient—perhaps extremely so. Thus, solving the problem of adverse selection may exacerbate the problem of moral hazard. Because adverse selection and moral hazard may have different and sometimes contrary policy prescriptions, debate about appropriate insurance-related policies is often premised on beliefs about the relative importance of these two information problems.

Earlier efforts separating moral hazard and adverse selection include Abbring et al. (2003) and Abbring, Chiappori and Pinquet (2003), who exploit past claims and induced changes in insurance contract terms to identify moral hazard, and Finkelstein and Poterba (2006) who exploit "unused observables" to identify adverse selection separate from moral hazard. 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.

The key contribution of this study is that, for the case of federally subsidized crop insurance, we are able to exploit a large and unique panel data set to derive precise quantitative estimates of the incidence and budgetary cost of moral hazard in a way that plausibly distinguishes these effects from those of adverse selection. We derive separate estimates for each of five crops and each of three to nine states, depending on the crop. We also estimate the share of indemnities paid due to moral hazard for each crop and state.

To separate moral hazard from adverse selection, we examine the link between crop insurance and yield outcomes for individual farmers. Rather than comparing farmers with insurance to those without it, we consider yields of farmers cycling into and out of the program in comparison to farmers in the same county not cycling into or out of the insurance program. 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. The estimates are developed 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 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. Results indicate how the distribution of yields conditional on fixed effects shifts with insurance enrollment. This provides an estimate of the shift in yield distributions due to moral hazard for each crop and state. To estimate the share of indemnities due to moral hazard, we integrate the difference in estimated yield distributions over actual indemnities paid between 1992 and 2001.

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 enrollment in the insurance program. Thus, for each farmer's crop and practice (irrigated or not), one observes yield outcomes in years with and without insurance coverage.

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 plausible 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 is highly plausible that different farmers would take different amounts of time to learn whether enrolling in the crop insurance programs is 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.

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 adverse selections remain, one may view our estimates as an upper bound on the incidence of moral hazard.

I. The Federal Crop Insurance Program

Previous research on the 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 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. Empirical Model

We estimate a separate model for each crop and state. For each model, we develop an index i that spans all farmer and practice combinations.[4] 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.

$$Y_{it} = \alpha_i + \gamma I_{it} + w_{ct} + u_{it} \tag{1}$$

$$\epsilon_{it} = u_{it}/\sigma_i \tag{2}$$

[4]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.

$$\epsilon_{it} \sim F(\epsilon | I_{it}) \tag{3}$$

In 1, 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 2 and 3 give some limited structure for the error in 1. 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 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 1 through 3. 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. Estimating Indemnities Due to Moral Hazard

Indemnity payments are a function of yield outcomes and, depending on the insurance contract, other factors like prices.[5] 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,

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

which include both random components like the weather, fixed factors like farmer ability, soil characteristics, and climate, and potentially complex interactions of these factors.

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:

$$\int P(Y_{it}|\theta_{it}, I_{it} = 1)g(Y_{it}|\theta_{it}, I_{it} = 1)dY_{it} \tag{4}$$

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

$$\begin{aligned} E[IMH] &= \int P(Y_{it}|\theta_{it}, I_{it} = 1)g(Y_{it}|\theta_{it}, I_{it} = 1)dY_{it} - \\ &\qquad \int P(Y_{it}|\theta_{it}, I_{it} = 1)g(Y_{it}|\theta_{it}, I_{it} = 0)dY_{it} \qquad (5) \\ &= \int P(Y_{it}|\theta_{it}, I_{it} = 1)\left[g(Y_{it}|\theta_{it}, I_{it} = 1) - g(Y_{it}|\theta_{it}, I_{it} = 0)\right]dY_{it} \qquad (6) \end{aligned}$$

The first term in equation 5 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:

$$Y_{it}|\theta_{it}, I_{it} = \alpha_i + \gamma I_{it} + w_{ct} + u_{it} \tag{7}$$

where

$$u_{it}/\sigma_i \sim F(\epsilon|I_{it}) \qquad dF/d\epsilon = f(\epsilon|I_{it}) \tag{8}$$

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:

$$Y_{it} = a_i + \gamma + w_{ct} + \sigma_i \epsilon_{it}, \tag{9}$$

and

$$Y_{it} = a_i + w_{ct} + \sigma_i \epsilon'_{it}, \tag{10}$$

which implies

$$\epsilon'_{it} = \epsilon_{it} + \frac{\gamma}{\sigma_i} \tag{11}$$

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,

$$g(Y_{it}|\theta_{it}, I_{it} = 1) = f(\epsilon_{it}|I_{it} = 1) \tag{12}$$

$$g(Y_{it}|\theta_{it}, I_{it} = 0) = f(\epsilon_{it} + \frac{\gamma}{\sigma_i}|I_{it} = 0) \tag{13}$$

Substituting these expressions into 5 gives expected indemnities due to moral hazard:

$$\int P(Y_{it}|\theta_{it}, I_{it} = 1) \left[f(\epsilon_{it}|I_{it} = 1) - f(\epsilon_{it} + \frac{\gamma}{\sigma_i}|I_{it} = 0) \right] dY_{it} \tag{14}$$

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:

$$IMH = \sum_i \sum_t P_{it} \left[f(\epsilon_{it}|I_{it} = 1) - f(\epsilon_{it} + \frac{\gamma}{\sigma_i}|I_{it} = 0) \right] \tag{15}$$

The expression in 15 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 (1 - 3).

IV. 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.[6] 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:

[6] 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.

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.[7] 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 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.[8] 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

[7]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.

[8]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.

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" ($I_{it} = 0$).[9]

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 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.

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

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.

V. 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).[10]

[10]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 $s_i = \sqrt{\sum_i \frac{u_{it}^2}{n_i - 1}}$.

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."[11]

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

[11] 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.

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.

VI. 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.[12] 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

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

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.

VII. 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.

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.

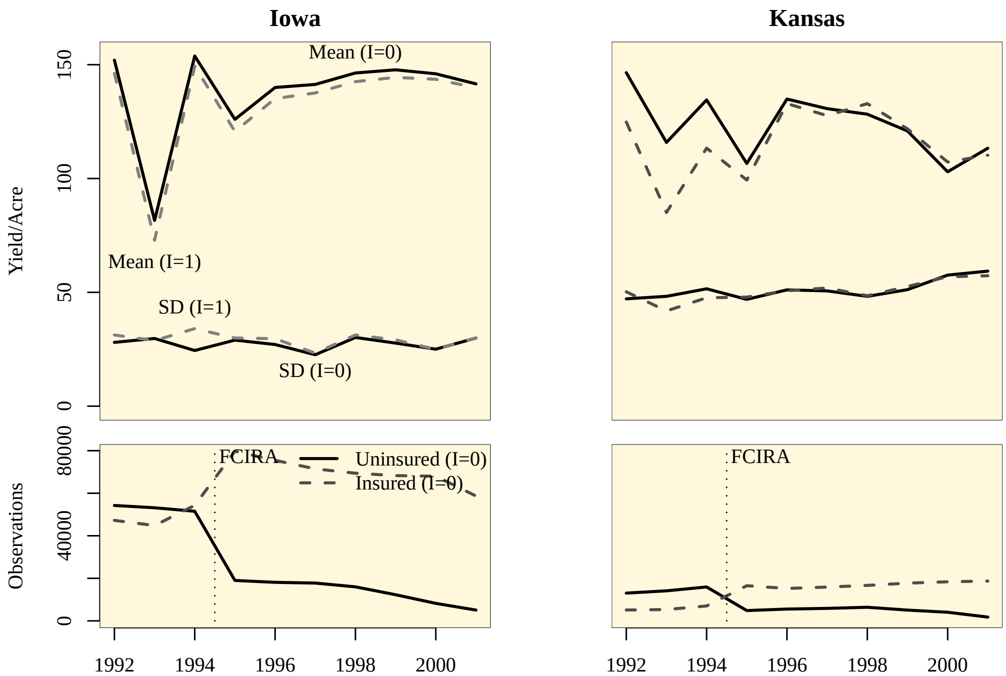

Figure 1. Summary Statistics for Two Representative Corn States

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.

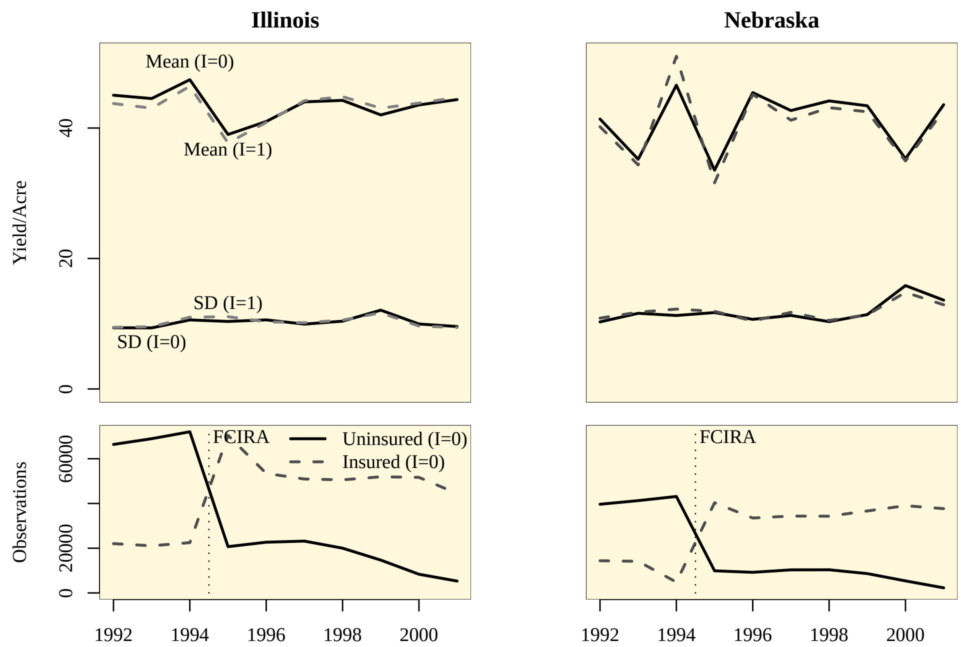

Figure 2. Summary Statistics for Two Representative Soybean States

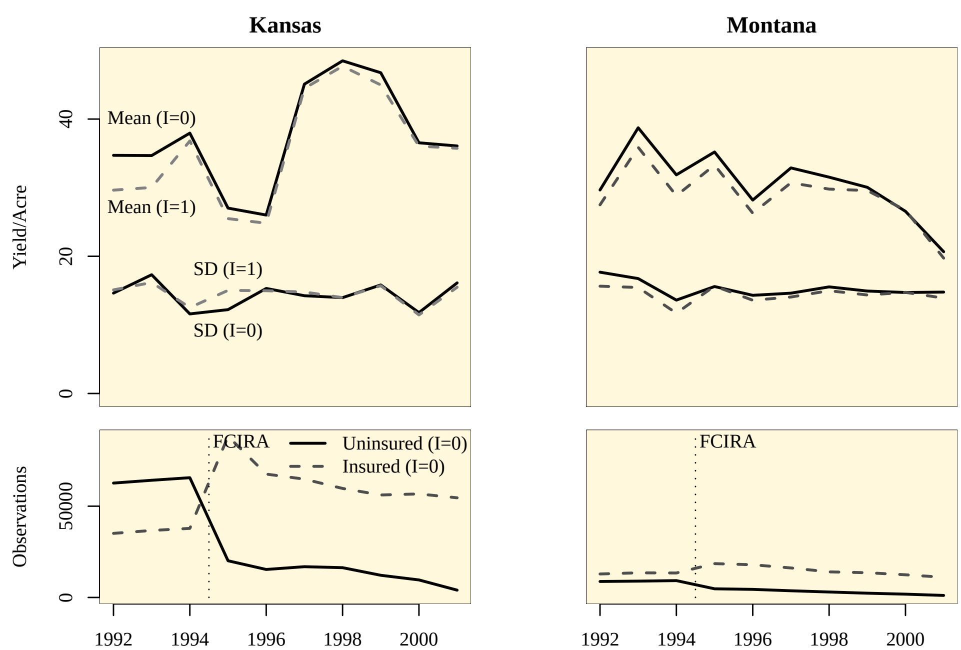

Figure 3. Summary Statistics for Two Representative Wheat States

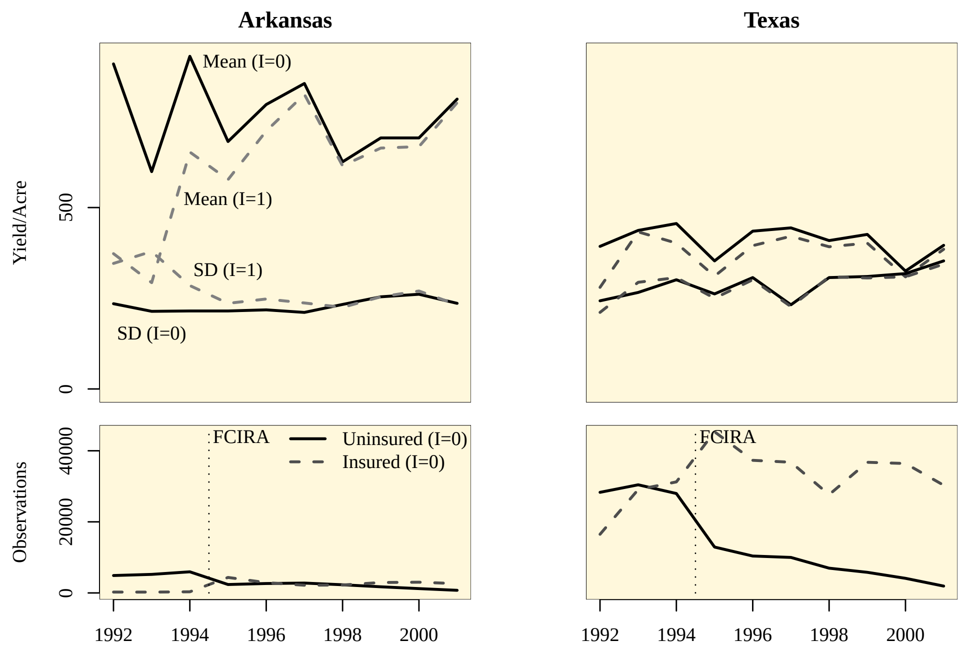

Figure 4. Summary Statistics for Two Representative Cotton States

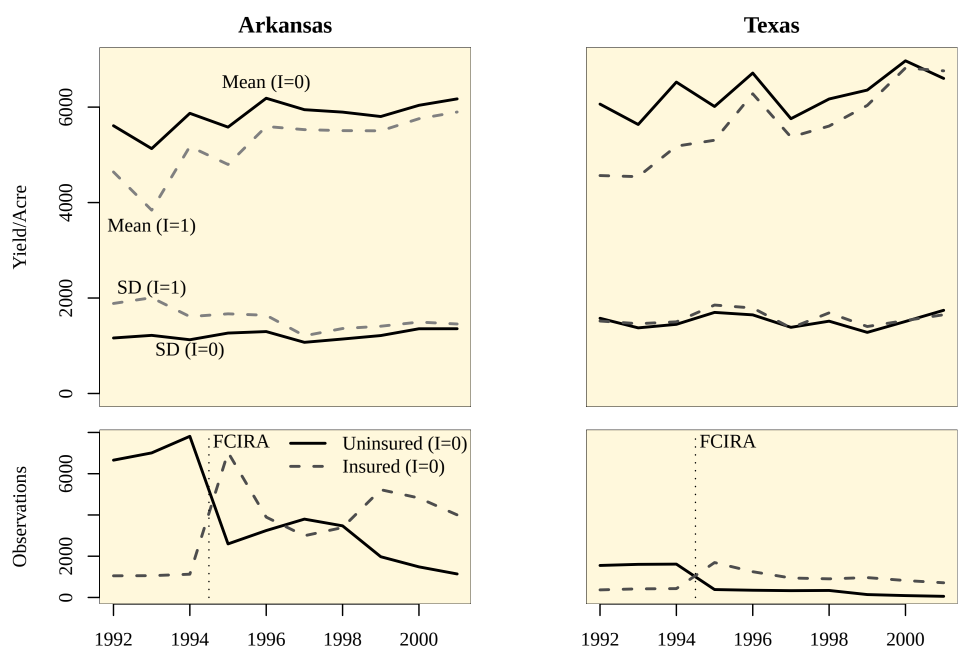

Figure 5. Summary Statistics for Two Representative Rice States

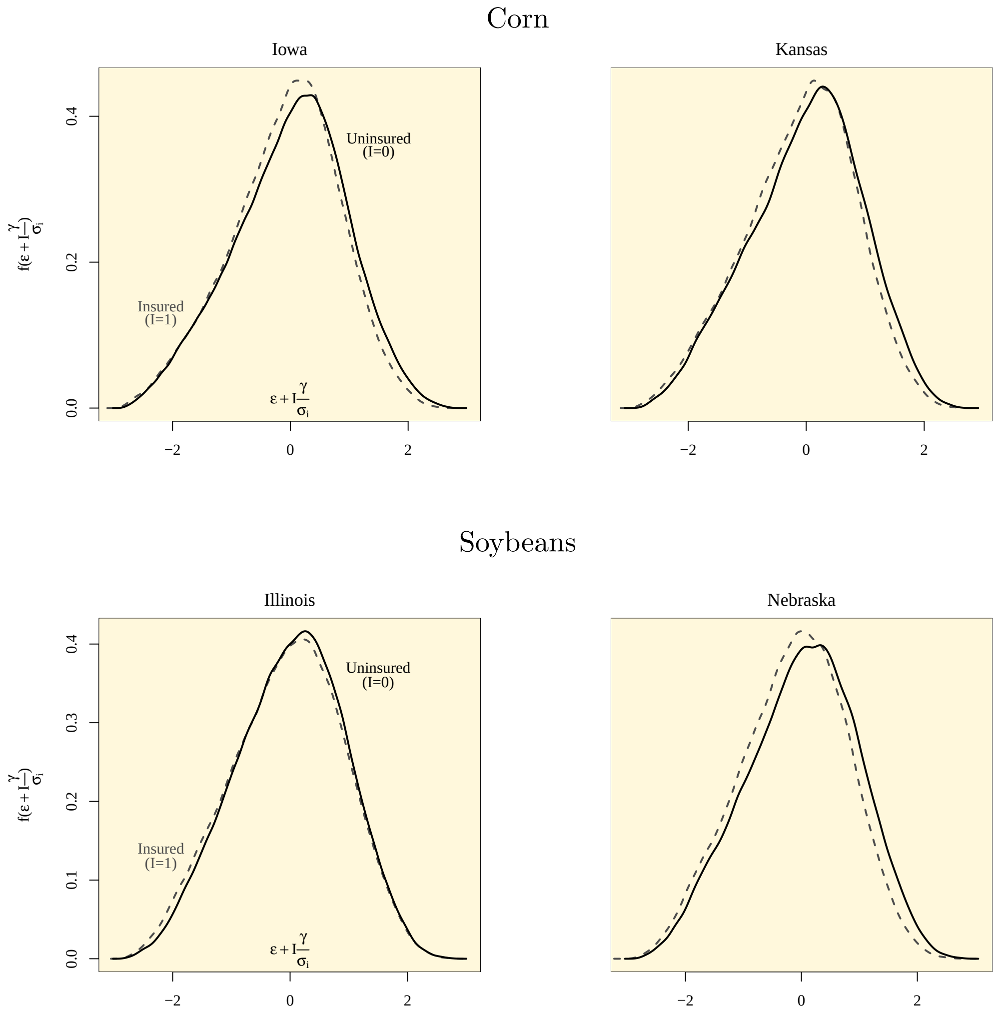

FIGURE 6. REPRESENTATIVE YIELD DISTRIBUTION SHIFTS FOR CORN AND SOYBEANS

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.

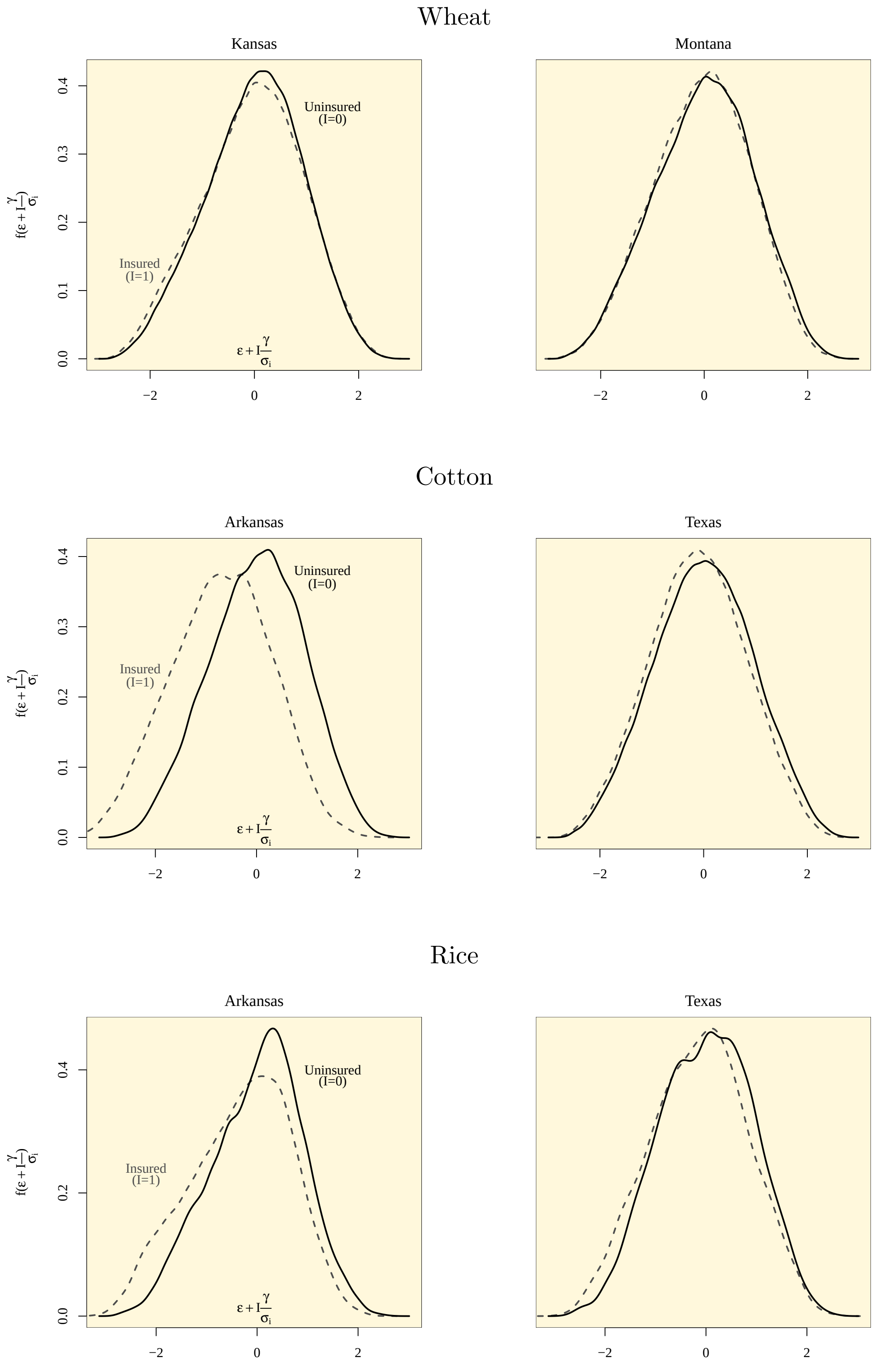

Figure 7. Representative Yield Distribution Shifts for Wheat, Cotton and Rice

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

							Indemnities Paid Due To Moral Hazard	
State	Mean Yield (1992-2001)	$\hat{\gamma}$	SE of $\hat{\gamma}$	T-Stat	R^2	N	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: Data Development

The data used in this study were derived from administrative files of USDA's Risk Management Agency (RMA). The raw files are organized by state, year and "type:" Type10, Type11, Type15, and Type21. Type10 files are called Policy Records and contains a unique producer identification number for each farmer and the policy issuing company (PIC) and reporting organization (the code identifying the Standard Reinsurance Agreement holder, also known as the Approved Insurance Provider, or AIP), the state within which the farm is situated that is enrolling in the crop insurance policy, the policy number, and the crop year of the policy.

Type11 data, called the Acreage Record, contains the indentifying variables contained in the Type 10 data (with the exception of the unique producer identification number) as well as coverage level information: what crop is being insured, what plant type (if applicable), what practice was used to grow the crop (irrigated or not), the county the crop is grown in, the unit (part of field) the crop was grown on, how much insurance was purchased, what share of the total acres in the field was insured, how many acres were reported as being in the field, what type of insurance was purchased, what coverage level was adopted, the RMA approved yield, total premiums, subsidies, the guarantee per acre, and a coverage flag denoting whether coverage was catastrophic or additional (also known as buy-up)..

Meanwhile, Type15 data, the Yield Record, also contains the identifying variables contained in the Type10 and Type11 data (organization and company providing insurance, state and county the farm resides in, the policy number, crop year, crop code, practice, plan code, and unit number), as well as the yield histories for each farmer, stretching back up to ten years for each observation. For each year of history, the Type15 data contains the crop year, the yield type (which can be actual or a variety of assigned yields), the yield for the crop for that crop year, and the number of acres upon which the crop was grown that obtained the yield for the crop. Finally, Type21 data, termed the Loss Line, contained the same identifiers as Type 15 described above as well as the information on indemnities (if any) received by farmers.

Our first step was to merge the four datasets together for each state in each year spanning 1992 through 2002. To begin, it was possible for there to be duplicate entries in the Type10 data. In order to merge the data correctly, we needed to eliminate any duplicates from Type10 data. Additionally, Type 11 data contained a variable called reported acres denoting the number of acres in the field that was being insured. This variable was reported in the hundredths of acres, so to get it to acres, we divided by 100. Multiplying this variable by the share of acres insured gave us the number of acres insured in the field.

After cleaning, we began by merging Type 10 and Type 11 data and Type 10 and Type 15 data together. We did so by matching the reporting organization, the insurance company, the state, the policy number, and the crop year. Since Type 10 held a unique identifier and was now merged in with the Type 11 and Type 15 files, we then merged the 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 unique identifying number, the crop code (the crop enrolled), the plan code (type of insurance plan), the county, the unit number (a unique number assigned for each tract of land qualifying as an individual unit as defined in the policy), 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 enrolled in was for catastrophic coverage or additional buy-up insurance) call these variables the merge variables.

We next cleaned the Type 21 data (the indemnities) which had indemnities entered as a character and had a comma in the data. We cleaned the data, converting indemnities to a numeric value, and then summed the indemnities by insurance plan, policy number, and coverage level in each crop year for each crop, crop type, and crop practice in each state and county, on each field unit for each company and organization delivering the crop insurance. While individual indemnities for a particular crop in a particular crop year in a particular unit within a field may be negative (perhaps an adjustment), the sum of the indemnities in a particular crop year should not be negative. Therefore, we eliminated any individuals who obtained an overall negative indemnity for the crop year for a particular crop.

We then merged indemnities with the dataset that contained data from the Types 10, 11, and 15 using the merge variables. We did this for each state and year, leaving us with a data set for each state and each crop year (from 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.

At this point, as a quick check to see whether or not we were on the right track, for each state, crop, and crop year, we created totals for the number of acres insured, premiums, subsidies, and indemnities paid out and compared these totals with publicly available RMA data posted on their website in their Summary of Business files. Our totals matched theirs, so we continued (we checked again at the end of our panel data process the results of which lie in our table at the end of the appendix).

We then began constructing our panel datasets (one for each state) from these various state-year datasets. The first step involved pooling all the years together for a particular state and cleaning the data with respect to the crop codes, crop years, identification numbers, coverage levels, total premiums, and number of insured acres; if any of these three variables was missing, or if either the identification number or coverage level was equal to zero, they were deleted.

We next created a dataset that, for each individual in every year, sums up the total premiums paid for insurance, the number of acres insured, and the indemnities (if any) for each crop and practice in each county. Then, because each observation has a history of yields and acres insured (for up to 10 years) associated with it, we pull each year of history out to create a separate observation for each year in the history and combine them all back together to create a panel dataset where an individual will now have an observation for each year that either a policy or a history exists. Since an individual who continues to enroll in crop insurance will have a history each year, many of the histories will overlap in this dataset we created. For example, suppose an individual enrolled in crop insurance for a corn crop from 1992 through 2000. The individual would then have 9 years worth of observations in our dataset, and each of those 9 years would have up to 10 years worth of histories associated with them. And many of these histories would overlap1993 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. When pulling out the histories, therefore, and re-entering them into a dataset to create a panel dataset that includes those yield histories, we generate a large number of duplicate observations. We therefore eliminate any duplicates to create a panel data set.

Within each year of the panel data, there is the potential for an individual to have multiple fields 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 their fields from one year to the next (producers can also merge fields), making it impossible to track histories effectively. 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 then merged the summed indemnities, premiums, and number of acres insured with the summed acres and mean yield datasets, generating a panel dataset with this information for each individual, crop, practice, in each county, over time.

Exploring the coverage levels, we noticed that there were a series of coverage levels that appeared to be off by a factor of 10 (for example, we found coverage levels equal to 0.05, 0.055, 0.06, etc., as opposed to actual coverage levels offered of 0.5, 0.55, 0.60). We corrected these coverage levels in the data. Additionally, if yields, number of insured acres, premium totals, number of reported acres, or identification numbers were missing at this point, we deleted the observation. Finally, some yields were incorrectly reported. We eliminated any observation with a yield greater than 10,000.

This dataset then allows us to create an insurance indicator variable. 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, and any years following that, we would have information on whether or not they had a policy and what coverage level was assigned to that policy, allowing us to assign an insurance indicator (equal to zero if the individual either carried no insurance or only enrolled in catastrophic coverage and one otherwise).

Finally, to restrict the dataset to the time frame we were interested in exploring, we kept those years between 1992 and 2001 (since we only had data through 2002, this would only provide us with yields for years up to, and including, 2001). At this point, again, we checked our data to see whether insured acres, total premiums, total subsidies, and total indemnities at the state level matched RMA public records. Results lie in the section below titled Data Checking.

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. To simplify our task and allow for proper 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.

Upon constructing our panel dataset from the RMA data, we needed to get it in shape for our econometric analysis. First, we eliminated any individuals who only had one observation since we needed more than one observation over time for identification. Additionally, while we had an identification number that was unique to the individual, we did not have one unique to the individual-practice (where, if you recall, the practice referred to whether or not the crop was irrigated). We therefore constructed such a variable, completing the preparation of the RMA data for our econometric analysis.

Data checking: Upon creating the dataset, we checked to make sure that we were properly using the data to create the panel datasets for each state. To check, we summed up the total level of premiums and acres insured across all producers in Iowa and Illinois for corn, soybeans, and wheat and in California for cotton and rice for the years 1997 through 2002. The following table shows that our panel dataset, based on the RMA administrative data, comes very close for each of these years.

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 B1. 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.

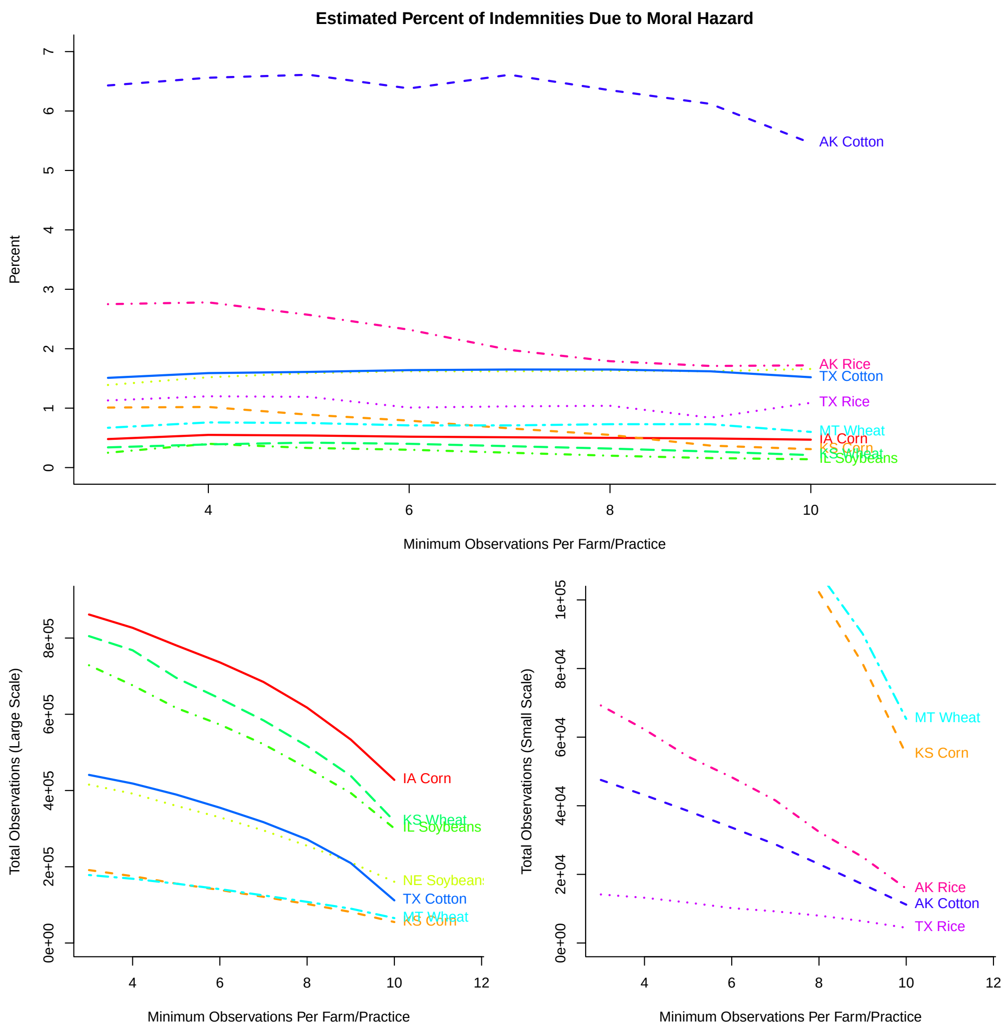

Figure B1. Robustness of Indemnity Cost Estimates to Farm-Practice Sample Size

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.