© *The Journal of Risk and Insurance*, 2007, Vol. 74, No. 4, 761-776

On the Revelation of Private Information in the U.S. Crop Insurance Program

Alan Ker
A. Tolga Ergun

Abstract

The crop insurance program is a prominent facet of U.S. farm policy. The participation of private insurance companies as intermediaries is justified on the basis of efficiency gains. These gains may arise from either decreased transaction costs through better established delivery channels and/or the revelation of private information. We find empirical evidence suggesting that private information is revealed by insurance companies via their reinsurance decisions. However, it is unlikely that such information will be incorporated into subsequent premium rates by the government.

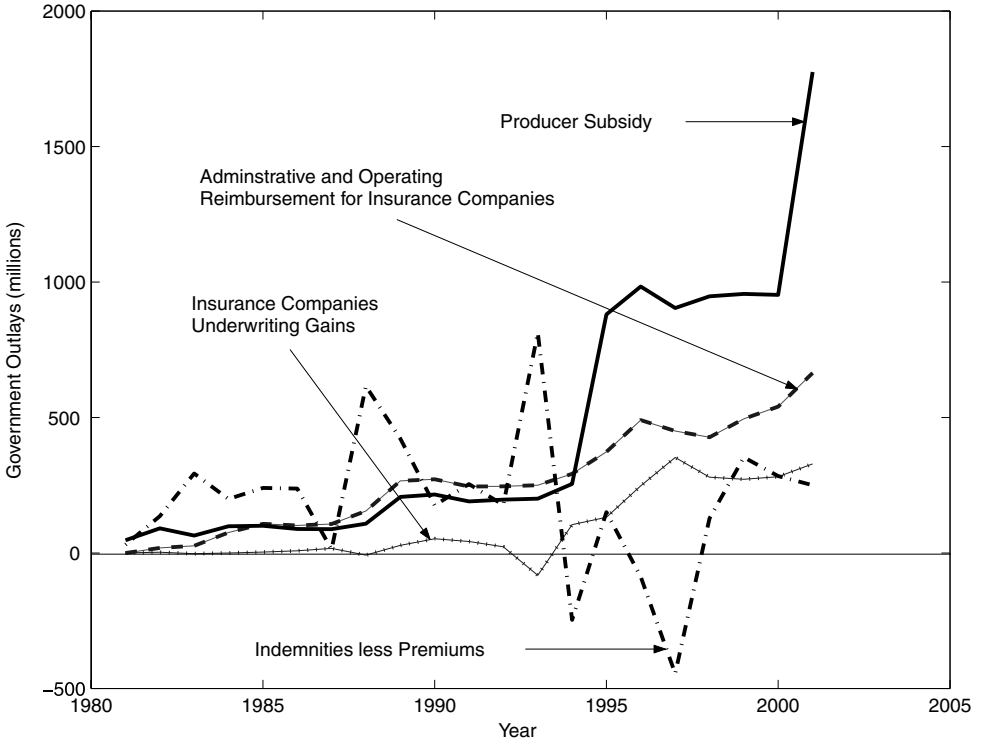
Introduction

Federally regulated crop insurance programs have been a prominent part of U.S. agricultural policy since the 1930s. In 2004, the estimated number of crop insurance policies exceeded 1.24 million with total liabilities exceeding $45 billion. Traditional crop insurance schemes offer farmers the opportunity to insure against yield losses resulting from nearly all risks, including such things as drought, fire, flood, hail, and pests. A variety of crop insurance plans and a number of new pilot programs are currently under development.

In the crop insurance program three economic interests are served: the federal government through the United States Department of Agriculture's Risk Management Agency (RMA), the producers or farmers, and the private insurance companies. In 1980, insurance companies were solicited by the federal government to increase farmer participation. Intermediaries are often used in public policy if efficiency gains are expected. In the crop insurance program, efficiency gains could be expected through two avenues. First, the better established delivery channels of insurance companies could reach a greater number of producers at a given cost. Second, the exploitation of private information (if it exists with the insurance companies) can increase the accuracy of

Alan P. Ker is at the Department of Agricultural and Resource Economics, University of Arizona, and A. Tolga Ergun is at the Department of Economics, Suffolk University. The authors can be contacted via e-mail: aker@ag.arizona.edu and tergun@suffolk.edu.

FIGURE 1
Government Outlays for U.S. Crop Insurance Program



premium rates, thereby decreasing adverse selection losses.[1] In this article, we empirically test if insurance companies reveal private information about risk profiles to the RMA via their reinsurance decisions. Not only is this of economic interest, it is a timely empirical question given the sizeable public funds needed to operate the crop insurance program and that a significant share of those funds—rivaling that of producers—resides with the insurance companies (see Figure 1).

We use semiparametric as well as parametric methods to determine whether a set of policies returns a profit or not using a data set aggregated to the crop–county–year combination. We consider models with and without explanatory variables depicting the allocation decisions of insurance companies. The use of semiparametric methods, which avoid strong distributional assumptions, proves useful as we reject the parametric method.

The remainder of the article proceeds as follows. The second section "Insurance Companies and the Standard Reinsurance Agreement" provides a terse review of the involvement of insurance companies in the U.S. crop insurance program. The third section "Data and Methodology" discusses the data and outlines the econometric

[1] Although the revelation of private information is of economic interest, this rationale was not considered by legislators in any deliberations leading to the passage of the Federal Crop Insurance Act of 1980; the Act that established private sector delivery.

methods. The fourth section "Estimation Results" presents the results, while the final section focuses on the corresponding policy implications.

INSURANCE COMPANIES AND THE STANDARD REINSURANCE AGREEMENT

There is a surprisingly small literature on the role of insurance companies in the U.S. crop insurance program (see Miranda and Glauber, 1997; Ker, 2001). Figure 1 illustrates the breakdown of government program outlays into producer subsidies, indemnities less premiums, administrative and operating reimbursement for insurance companies, and underwriting gains/losses accrued by insurance companies.[2] There are a number of interesting features: (1) producer subsidies increased dramatically in 1995 (a result of the 1994 Federal Crop Insurance Act) and again in 2001 (a result of the 2000 Agricultural Risk Protection Act (ARPA)), (2) indemnities less premiums are quite volatile, (3) insurance companies' administrative and operating expenses have risen with increases in total premiums, and (4) underwriting gains accruing to insurance companies have increased dramatically since 1994. Given the significant underwriting gains realized by insurance companies, it is of economic interest to determine if they reveal private information through their reinsurance decisions.

The involvement of insurance companies in the U.S. crop insurance program is defined by the Standard Reinsurance Agreement (SRA). Insurance companies sell policies and conduct claim adjustments. In return, the RMA compensates them for the corresponding administrative and operating expenses. The underwriting gains/losses are shared asymmetrically, between the insurance companies and the RMA. Both the provisions by which the underwriting gains and losses are shared and the reimbursement for administrative and operating expenses are set out in the SRA.

Section II.A.2 of the 2005 SRA states that an insurance company "... must offer and market all plans of insurance for all crops in any State where actuarial documents are available in which it writes an eligible crop insurance contract and must accept and approve all applications from all eligible producers." An eligible farmer will not be denied access to an available, federally subsidized, crop insurance product. Therefore, an insurance company conducting business in a state cannot discriminate among farmers, crops, or insurance products in that state. This is unusual in that the responsibility for pricing the crop policies lies with the RMA but the insurance companies must accept some liability for each policy they write and cannot choose which policy they will or will not write.

Two mechanisms are provided to entice insurance companies to participate. First, given that insurance companies do not set premium rates, there is a mechanism in the SRA by which they can cede the majority of the liability of an undesirable policy. In a private market, the insurance company would not write a policy deemed undesirable. Second, given that RMA premium rates do not reflect a return to the insurance company's capital, the SRA provides asymmetric sharing of underwriting gains/losses. Essentially, the SRA provides two mechanisms that emulate a private market from

[2] Underwriting gains/losses are defined as total premiums less total indemnities. However, because insurance companies and the government share the underwriting gains/losses asymmetrically by state and certain insurance programs, it is possible that the insurance program as a whole experiences an underwriting loss while the insurance companies experience an underwriting gain. Similarly, the reverse is also possible.

the perspective of the insurance company. In so doing, it also provides a vehicle by which an insurance company uses its information regarding farmer risk profiles to transfer unwanted policies to the RMA.

Insurance companies must place each policy into one of three funds: assigned risk, developmental, or commercial. For each state in which the insurance company does business, there is a separate assigned risk fund, developmental fund, and commercial fund. The structure of risk sharing is identical but the parameters that dictate the amount of sharing vary greatly across funds. For each fund k, the underwriting gain/loss the insurance company retains (denoted Ω_{IC}^k) is equal to the total underwriting gain/loss for the fund (denoted Ω^k) multiplied by two parameters (denoted μ_1^k and μ_2^k). Formally,

$$\Omega_{IC}^k = \Omega^k \cdot \mu_1^k \cdot \mu_2^k.$$

The underwriting gain/loss retained by the RMA (denoted Ω_{RMA}^k) by default is

$$\Omega_{RMA}^k = \Omega^k \cdot \left(1 - \mu_1^k \cdot \mu_2^k\right).$$

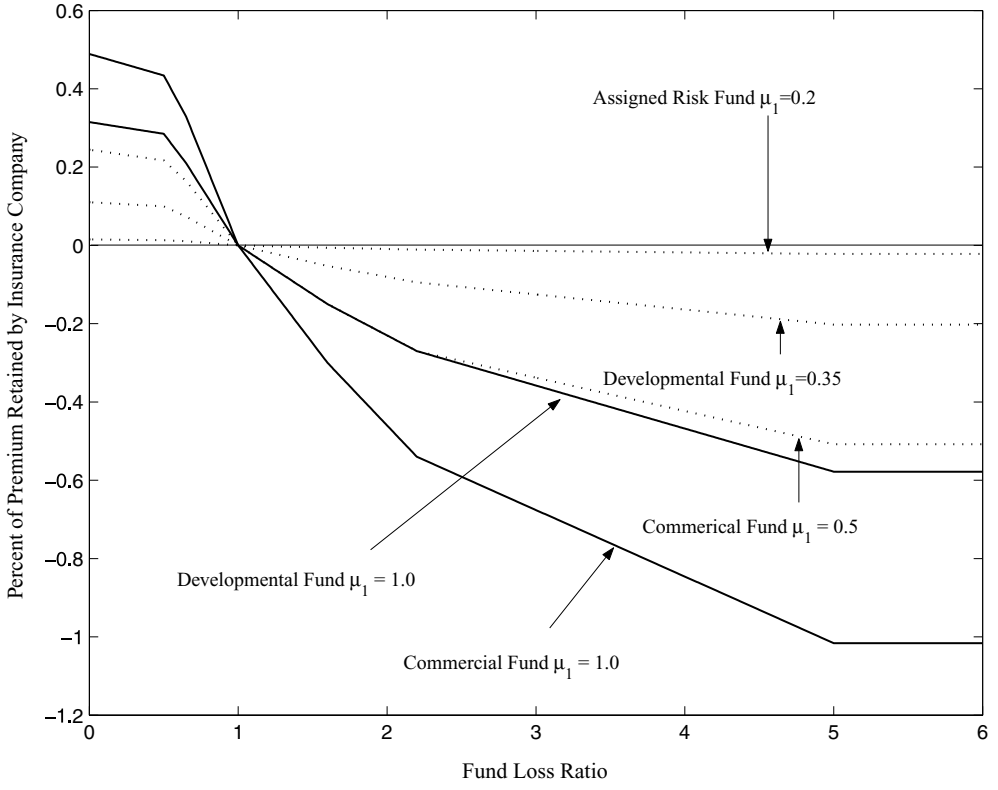
The first parameter, μ_1^k, is fixed at 0.2 for the assigned risk fund but represents an *ex ante* choice variable for the insurance company with respect to the commercial and developmental funds. For the development fund $\mu_1^k \in [0.35, 1.0]$, while for the commercial fund $\mu_1^k \in [0.5, 1.0]$. The insurance company must choose μ_1^k by July 1 of the preceding crop year.

The second parameter, μ_2^k, is not a fixed scalar, but a function of the fund loss ratio. Figure 2 illustrates the relationship between the fund loss ratio and the percentage of premiums retained by the insurance company. For example, if the percentage of premiums retained is -20 percent and the total premiums were \$1 million, the insurance company would incur a loss of \$200,000. The fund loss ratio is defined as the ratio of total indemnities to total premiums.

Note the differences between the percentage of premiums retained for each of the three funds. Consider, for example, if the assigned risk fund has \$2 million in premiums and \$3 million in indemnities. The loss ratio would be 1.5 and the underwriting loss would be \$1 million. For the assigned risk fund, the insurance company would be liable for 0.92 percent or only \$9,200 of the \$1 million underwriting loss. Given total premiums of \$2 million the percentage of premiums retained by the insurance company would be only -0.46 percent. If, on the other hand, this underwriting loss occurred in the commercial fund with $\mu_1 = 1$, the insurance company would be liable for 46 percent or \$460,000 of the \$1 million underwriting loss, resulting in a percentage of premiums retained of -23 percent. Consider a second example: if premiums were \$2 million and indemnities were only \$1 million, the loss ratio would be 0.5 and the underwriting gain would be \$1 million. For the assigned risk fund, the insurance company would retain 2.64 percent (\$26,400) of the underwriting gain and, as such, the percentage of premiums retained would be 1.32 percent. If, on the other hand, this underwriting gain occurred in the commercial fund with $\mu_1 = 1$, the insurance company would retain 86.8 percent (\$868,000) of the underwriting gain and, as such, the percentage of premiums retained would be 43.4 percent.

It is apparent from these examples that policies that a profit-maximizing insurance company expects to be profitable would be placed in the commercial fund where they share a high percentage of any underwriting gains and losses. Conversely, policies that

FIGURE 2
Percentage of Premium Retained by Insurance Company Relative to Fund Loss Ratio



a profit-maximizing insurance company expects to be unprofitable would be placed in the assigned risk fund where they share a low percentage of any underwriting gains and losses.

The expected profit-maximizing optimal reinsurance of policies across the three funds is extremely complicated. Given that there exist three possible funds for which any policy may be allocated, and, assuming N policies, there are 3^N possible reinsurance allocations. For example, if $N = 500$ there exist 3.636E + 238 possible reinsurance allocations, all of which need to be evaluated in terms of expected profit. Not only is it untenable for the insurance company to undertake this, but to do so requires an estimate of the joint density of yields for the N policies—impossible, given the scarce data.[3] Fortunately, our empirical analysis only requires that insurance companies allocate policies that they expect to be *relatively* more profitable to the commercial fund as opposed to the assigned risk fund. It is apparent from the above examples that this requirement is consistent with profit-maximizing behavior.

[3] The reader is directed to Ker and McGowan (2000) for a detailed investigation of a profit-maximizing insurance company's optimal allocation strategy of policies across the three types of funds.

Two final points regarding the SRA require discussion. First, there exist separate developmental and commercial funds for "catastrophic policies," "revenue policies," and "other policies" that comprise multiple peril crop insurance policies and Group Risk Plan policies (Group Risk Plan policies make up a negligible fraction of the total policies). We focus our attention on the three fund allocations for the "other policies" because insurance companies have significantly less experience and historical information with the "revenue policies" and "catastrophic policies" and thus their reinsurance decisions may not be as efficient. Also note, that while these funds (except assigned risk) are not aggregated across types of policies, they are aggregated across crops. Second, insurance companies face a constraint, at the state level, on the maximum percentage of premiums in their book of business that can be placed in the assigned risk fund. These maximums vary quite significantly by state. While this may inhibit the insurance companies' ability to cede unwanted policies, by choosing $\mu_1 = 0.35$ for the developmental fund they can make it resemble the assigned risk fund (see Figure 2) and there are no such percentages of premium restrictions for the development fund.

DATA AND METHODOLOGY

Recall that we wish to test whether relevant private information is revealed in the reinsurance decisions of insurance companies. This hypothesis can be tested by predicting whether policies are profitable or not using two models. The first model uses public information as explanatory variables. The second model nests the first and includes the additional variables representing the reinsurance decisions of the insurance companies. Specifically, we test whether the percentage of correct predictions increases significantly with the inclusion of these reinsurance variables.

Our dependent variable indicates whether a set of policies returned a profit or not. If premiums are greater than indemnities we define $y = 1$. Conversely, if premiums are less than indemnities $y = 0$. Our first model is

$$y = F(v\beta) + \epsilon$$

where v embodies information available to the RMA such as historical loss ratio, crop dummies, state maximums on the assigned risk fund, and liability changes. $F(\cdot)$ is termed the link function and $v\beta$ is termed the index. Our second model is

$$y = F(v\beta + \text{ reinsurance variables} * \gamma) + \epsilon,$$

where the set of explanatory variables now includes the reinsurance decisions.[4]

The data comprise the premiums, indemnities, liability, and number of policies in each of the three funds by crop–county–year combination. We have data on corn,

[4] Our dependent variable is based on whether a set of policies returned a profit or not rather than the level of profit. Consider the reinsurance decision of the insurance company. Whether a policy is expected to be marginally or significantly above a specific profit level, it is reinsured with the commercial fund. All that RMA can ascertain about a policy that has been allocated to the commercial fund is that expected profit is above a specific and unknown level. Therefore, our dependent variable is restricted to whether a set of policies returned a profit or not. However, we did repeat the analysis using the loss ratio and the results remained unchanged.

cotton, soybeans, and wheat for the reinsurance years 1998, 1999, 2000, and 2001. We remove combinations with less than $500,000 in liability leaving 7,602 crop–county–year combinations.

Two caveats regarding our data need noting. First, our data are aggregated to the county level; we do not have policy-specific reinsurance decisions. Second, our data are aggregated across insurance companies. While we would prefer policy and company-specific reinsurance decisions and requested such, we were only able to obtain aggregated data from RMA. This lack of precision will reduce the power of our tests.

The explanatory variables used in our analysis are crop dummies for cotton, soybeans, and wheat, historical loss ratio, ratio of current liability to the previous year liability (denoted liability ratio), the maximum percentage of premiums allowed in the assigned risk fund for that state (denoted state risk), percentage of premiums placed in the commercial fund, and the percentage of premiums placed in the assigned risk fund.[5] We do not include the percentage of premiums placed in the developmental fund since that would result in a singularity problem as the sum of the three percentages in the three funds equals one for each crop–county combination.

Econometric Methodology

For estimation, we consider the parametric probit model alolng with the semiparametric single-index model estimator of Ichimura (1993). Single-index models for binary data have the general form

$$P(y = 1 \mid v) = F(v\beta),$$

where F is an unknown function (not necessarily a distribution function), $v \equiv (1, x)$, x is a $1 \times q$ vector of explanatory variables, and β is a $(q + 1) \times 1$ vector of unknowns. If F is the normal (logistic) distribution function, we have the probit (logit) model. If it is the identity function, we have the linear probability model. If the (normal or logistic) distributional assumption is not correct, the maximum likelihood estimates of coefficients and probability estimates will be inconsistent (see Ruud, 1983, for an exception on the slope coefficient estimates). Choice of a probit or a logit model, almost a standard in the literature, is usually based on estimation convenience rather than any justification of distributional assumptions. These sometimes unrealistic assumptions may lead to erroneous results and implications. Furthermore, since these models are used with cross-sectional data, heteroskedasticity is usually a concern. Unlike linear models where one only loses efficiency, the maximum likelihood estimators of probit and logit models are inconsistent if the error distribution is heteroskedastic (see Yatchew and Griliches, 1985). Single-index models, on the other hand, can accommodate certain forms of heteroskedasticity (general but known form and unknown form if the distribution of the error term depends on x only through the index, i.e., the index restriction).[6]

[5] The historical loss ratio is calculated using data from 1981 to the year prior to the crop year. That is, the historical loss ratio for policies in crop year 1999 comprises data from 1981 to 1998.

[6] See the maximum score estimator of Manski (1975) and its smoothed version by Horowitz (1992) for estimators that can accommodate arbitrary forms of heteroskedasticity although at the cost of a rate of convergence slower than $\sqrt{n}$.

Optimization-based estimation methods have been developed for single-index models without making distributional assumptions and thus avoiding misspecification. These include Ichimura (1993) and Klein and Spady (1993). The first of these estimators is based on minimizing a nonlinear least squares loss function and the latter is based on maximizing a profile likelihood function. The latter estimator is developed specifically for binary-choice model estimation. Ichimura and Klein and Spady show $\sqrt{n}$ convergence and asymptotic normality of their estimators and give a consistent covariance estimator. Since the estimators (and results) are almost identical we only present the results from the Ichimura estimator.

Note that we need a location-scale normalization for identification purposes in single-index models. Since the link function F is assumed to be completely unknown, the intercept term cannot be identified as is subsumed in the definition of F. Also, a scale normalization is needed for the same reason that it is imposed in parametric models (assuming the error term has unit variance). This scale normalization in the semiparametric models can be achieved by setting the coefficient of one continuous regressor equal to a constant.[7]

The semiparametric least squares (SLS) estimator of Ichimura (1993) minimizes

$$\frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{F}(x_i b)]^2,$$

where $\hat{F}$ is the nonparametric estimator for the unknown link function and b is the β vector after location-scale normalization is imposed, i.e., $b \equiv (c, \beta_2, \ldots, \beta_q)^T$, where c is a constant, assuming that the first regressor has a continuous distribution. Ichimura denotes this model as SLS and shows that $\hat{\tilde{b}}$ is consistent and $\sqrt{n}(\hat{\tilde{b}} - \tilde{b}_0) \xrightarrow{d} N(0, \Omega_{SLS})$, where $\tilde{b}$ is b without its first component, and gives a consistent estimator of $\Omega_{SLS} = \Gamma^{-1}\Sigma\Gamma^{-1}$. Γ and Σ can be consistently estimated by

$$\hat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}_i^T\hat{F}'(x_i\hat{b})^2,$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}_i^T\hat{F}'(x_i\hat{b})^2[y_i - \hat{F}(x_i\hat{b})]^2,$$

where $\tilde{x}_i \equiv (x_{2i}, \ldots, x_{qi})$, $\hat{b} \equiv (c, \hat{\tilde{b}}^T)^T$, and $\hat{F}'$ is the derivative of $\hat{F}$. For $\hat{F}$, he uses the Nadaraya–Watson estimator

$$\hat{F}(x_i b) = \sum_{j\neq i} y_j K\left(\frac{x_i b - x_j b}{h}\right) \Big/ \sum_{j\neq i} K\left(\frac{x_i b - x_j b}{h}\right), \tag{1}$$

where K is the kernel function (usually a symmetric density function) and $h = h(n)$ is the smoothing parameter such that $h \to 0$ as $n \to \infty$.[8]

[7] An alternative scale normalization would be $||\beta|| = 1$ where $||\cdot||$ is the Euclidean norm.

[8] Note that in these semiparametric estimators, asymptotic theory requires trimming those observations for which the index is arbitrarily close to the boundary of its support. For the

In parametric probit and logit models, one maximizes the loglikelihood function

$$\sum_{i=1}^{n}(y_i \log[F(v_i\beta)] + (1 - y_i)\log[1 - F(v_i\beta)]), \tag{2}$$

where F is assumed to be the normal or logistic distribution function.

In single-index models, the asymptotic distribution of the normalized and centered estimator does not depend on the smoothing parameter, so asymptotically, any sequence of smoothing parameters is going to give the same estimate as long as it satisfies certain conditions.[9] For this reason, in semiparametric single-index models, selection of the smoothing parameter has not been well studied. One exception is Härdle, Hall, and Ichimura (1993) who show the SLS estimator of Ichimura (1993) can be expanded as $A(b) + B(h)$ and can be minimized simultaneously with respect to both b and h. This is like separately minimizing $A(b)$ with respect to b and $B(h)$ with respect to h. The end result is a $\sqrt{n}$-consistent estimator of b and an asymptotically optimal estimator of h, in the sense that $\hat{h}/h_0 \to 1$ as $n \to \infty$ where h_0 is the optimal bandwidth for estimating F when b is known and is proportional to $n^{-1/5}$ as usual in nonparametrics (see Härdle, Hall, and Ichimura 1993, for technical details). We apply this idea to the SLS objective function and hence we optimize with respect to both b and h. To our knowledge, this is the first article that uses this idea in practice other than the original Härdle, Hall, and Ichimura (1993) article. Note that in estimating F, we are excluding observation i so that we are "cross-validating" the objective functions. In the estimations we use a normal density function truncated at plus and minus three standard deviations as the kernel.

ESTIMATION RESULTS

To test our hypothesis, we randomly partition our sample into an estimation sample (3,801 observations) and a prediction sample (3,801 observations). We evaluate our hypothesis using out-of-sample methods rather than within-sample methods because insurance companies must make reinsurance decisions out-of-sample and out-of-sample tests minimize spurious results from over-fitting the data (particularly for semiparametric methods which, if applied inappropriately, can be made to over-fit the data). We also conducted three tests for the appropriateness of the probit model and all rejected it (see the Appendix for details and test results).

The estimation results and predictive performances for the models without and with the reinsurance variables are located in Table 1 (standard errors are in parentheses). For the semiparametric estimator we restrict the intercept to 0 and the parameter estimate on the historical loss ratio to the probit estimate as is commonly done.[10]

Ichimura estimator, knowledge of the distribution of the index is required, which is unknown in practice. Other applied papers (Horowitz, 1993; Gerfin, 1996; Fernandez and Rodriguez-Poo, 1997) do not consider trimming. As Horowitz (1993, p. 53) explains "This amounts to assuming that the support of [index] is larger than that observed in the data."

[9] But in finite samples the performance of the estimators can be very sensitive to the choice of this smoothing parameter.

[10] This parameter can be set to any finite constant.

TABLE 1
Estimation Results and Predictive Performance

Parameter Estimate	Probit Without Reinsurance Variables	Ichimura Without Reinsurance Variables	Probit With Reinsurance Variables	Ichimura With Reinsurance Variables
Intercept	1.6902*	0.0000**	1.3813*	0.0000**
	(0.0683)	n/a	(0.1426)	n/a
$dummy_{cotton}$	−0.2374*	−5.1377*	−0.2129*	−0.1836*
	(0.0881)	(0.0754)	(0.0885)	(0.0631)
$dummy_{soybeans}$	−0.1077	3.1030*	−0.1083	−0.0728
	(0.0587)	(0.0630)	(0.0589)	(0.0404)
$dummy_{wheat}$	−0.1067	−2.7825*	−0.0880	−0.1330*
	(0.0715)	(0.0911)	(0.0718)	(0.0527)
Liability ratio	−0.0020	−0.0687*	−0.0028	−0.0022
	(0.0078)	(0.0067)	(0.0078)	(0.0095)
State risk	−1.6868*	−6.8218*	−1.6507*	−2.9153*
	(0.1533)	(0.1150)	(0.1545)	(0.1102)
Commercial	n/a	n/a	0.2854*	0.1448*
	n/a	n/a	(0.1236)	(0.0445)
Assigned	n/a	n/a	−0.3957*	−0.7199*
	n/a	n/a	(0.2141)	(0.1147)
Historical LR	−0.2521*	−0.2521**	−0.1738*	−0.1738**
	(0.0475)	n/a	(0.0523)	n/a
h	n/a	0.1098	n/a	0.1025
log L	−1817.1	n/a	−1810.3	n/a
Predictive performance	74.66%	77.84%	75.24%	79.63%

*Significant at 95 percent confidence level. **Restriction necessitated by estimation procedure.

We have no expectations about the signs of the crop dummy variables although we do have expectations about the signs of the other parameter estimates. First, the sign of liability ratio is negative as expected. If liability increases (decreases) significantly from one year to the next, this may suggest that producers perceive their return to that insurance policy to have increased (decreased), and thus the expected return for the insurance company may decrease (increase). The parameter estimate on state risk is negative (as expected) and significant. This indicates, quite interestingly, that policies in those states with higher bounds on the percentage of premiums allowed in the assigned risk fund are less likely to be profitable. The parameter estimates on the historical loss ratio in the probit models are negative and significant as expected; the higher the loss ratio, the less likely the policies are profitable. The parameter on the percentage of premiums in the commercial fund is positive as expected. This suggests that policies the insurance company places in the commercial fund are more likely to be profitable. This is statistically significant in both the probit and semiparametric models. Finally, the parameter on the assigned variable is negative as expected suggesting that policies the insurance company places in the assigned risk fund are less likely to be profitable.

Our null hypothesis is that no private information is revealed in the reinsurance decisions. To test this we compare the percentage of policies correctly predicted with

TABLE 2
Private Information Test Results

Test	Test Statistic	Standard Error
model_f less model_{nf} using probit	0.0058	0.00169
model_f less model_{nf} using Ichimura	0.0179	0.00576

and without the reinsurance explanatory variables.[11] Our test may be formally written as

$$H_o\text{: } \rho_f - \rho_{nf} \leq 0 \text{ versus } H_a\text{: } \rho_f - \rho_{nf} > 0,$$

where ρ_f corresponds to the percentage of correct predictions from the model that includes the two reinsurance variables, while ρ_{nf} corresponds to the percentage of correct predictions from the model that does not include the reinsurance variables. Table 2 summarizes the empirical tests. For convenience, we denote model_f to represent the models including the two reinsurance variables and model_{nf} to represent the models excluding the two reinsurance variables. Standard errors are calculated by bootstrapping the prediction sample and recovering the difference in the percentage of correct predictions (500 bootstraps are used).

The out-of-sample tests show that predictive performance increases significantly when the reinsurance variables are included, indicating that there exists relevant private information revealed through the allocation decisions of the insurance companies.[12] This coincides with the in-sample results that suggested that the explanatory variables depicting the reinsurance decisions were significant in explaining profitable and nonprofitable sets of policies. Therefore, we reject the null that no relevant private information is revealed in the reinsurance data.

CONCLUSIONS AND POLICY IMPLICATIONS

Although the crop insurance program has garnered significant attention in the academic literature, surprisingly little has focused on the involvement of insurance companies. However, the public rents obtained by the insurance companies in return for their involvement are close to rivaling those obtained by producers (see Figure 1). Consequently, more research is needed, both theoretically and empirically, focusing on the involvement of insurance companies as intermediaries.

This article considered whether insurance companies reveal private information through their reinsurance decisions. We conducted out-of-sample tests and showed that the insurance companies do possess statistically significant private information

[11] RMA faces legal and political constraints in setting rates and subsequently they do not reflect all public information. We were unable to obtain unconstrained or target rates from RMA in hopes of including the ratio of constrained to unconstrained rates as an explanatory variable in our regressions. However, the target rates are generally derived from the historical loss ratio that is included in our analyses. Therefore, we are jointly testing the significance of private information and the residual rate-setting constraints not captured by the historical loss ratio variable.

[12] A LR test confirms this result for the probit models.

that may warrant their involvement in the crop insurance program in addition to program delivery efficiencies. Although RMA can adjust their rates over time, they face many political and legal constraints in doing so. Therefore, it is unknown whether RMA could in fact adjust their rates to reflect the revelation of private information. A review of the rating methodologies for all RMA insurance products reveals that the reinsurance behavior of insurance companies is not currently part of any RMA rate-setting formulas.

The policy implications of our results do not call into question the use of insurance companies as intermediaries in the U.S. crop insurance program. The reality is that the program is simply too large to operate without private insurance companies. The results may call into question whether RMA should share the underwriting gains/losses with insurance companies.[13] If risk sharing were eliminated, the administrative and operating expense reimbursement may need to be increased to maintain the participation of insurance companies that do so with the expectation of realizing underwriting gains. Removal of these without compensation could dramatically alter the level of insurance company involvement and hence the delivery of the program. Certainly, there is much room for future research on the role of insurance companies in the U.S. crop insurance program.

APPENDIX

We undertake three tests for the parametric probit model. The first is the so-called HH test of Horowitz and Härdle (1994). This test is motivated by conditional moment tests. Horowitz and Härdle replace the parametric alternative model with a semiparametric one. The advantage of this test relative to tests with arbitrary nonparametric alternatives is that as long as only the shape of the link function, and not the single-index structure, is the issue, the HH test will be more powerful as the latter tests suffer from the curse of dimensionality.[14] The HH test, on the other hand, assumes that the conditional expectation of the dependent variable depends on the regressors only through the index, not only in the null but in the alternative as well, and thus avoids the curse of dimensionality. The HH test statistic is

$$T_n = \sqrt{h} \sum_{i=1}^{n} w(\hat{z}_i)[y_i - F(\hat{z}_i)][\hat{F}(\hat{z}_i) - F(\hat{z}_i)],$$

where $\hat{z}_i = v_i \hat{\beta}_{probit}$ is the estimated index from the parametric probit model, w is a nonnegative weight function that can be chosen to be an indicator variable of an interval that contains 95–99 percent of $\hat{z}$, and F is the normal distribution function. For $\hat{F}$, Horowitz and Härdle (1994) use the jackknife-like method of Schucany and Sommers (1977) to achieve asymptotic unbiasedness. Formally,

[13] The only economic rationale would be to ensure that the government and insurance companies are incentive compatible with respect to fraud. However, given the existence of the assigned risk and developmental funds that enable insurance companies to cede the vast majority of the liability of unwanted policies, this goal is not necessarily attainable under the current SRA.

[14] As the number of regressors increases, estimation precision declines rapidly. This phenomenon is known as the curse of dimensionality.

$$\hat{F}(l) = [\hat{F}_h(l) - (h/s)^r \hat{F}_s(l)]/[1 - (h/s)^r]$$

and

$$\hat{F}_t(l) = \sum_{j \neq i} y_j K\left(\frac{l - \hat{z}_j}{t}\right) \Big/ \sum_{j \neq i} K\left(\frac{l - \hat{z}_j}{t}\right) \quad \text{for} \quad t = h, s,$$

where $h = cn^{-1/(2r+1)}$, $s = c'n^{-\delta/(2r+1)}$ with $c, c' > 0, 0 < \delta < 1$, and K is a kernel of order $r \geq 2$. Horowitz and Härdle show that T_n is asymptotically distributed as $N(0, \sigma_T^2)$ where

$$\sigma_T^2 = 2C_K \int_{-\infty}^{\infty} w(l)^2 [\sigma^2(l)]^2 \, dl. \tag{A1}$$

In (A1), $C_K = \int_{-\infty}^{\infty} K(u)^2 \, du$ and $\sigma^2(l) = \text{Var}(y \mid z = l)$. We conducted this test for model$_{nf}$ and model$_f$ using a standard normal density as the kernel ($r = 2$); w was taken to be the indicator variable that equals 1 on an interval containing 98 percent of $\hat{z}$ and 0 elsewhere. There is no optimal way of choosing h and s. Following Härdle,

FIGURE A1
Test of Probit for the Data Not Including Fund Allocations

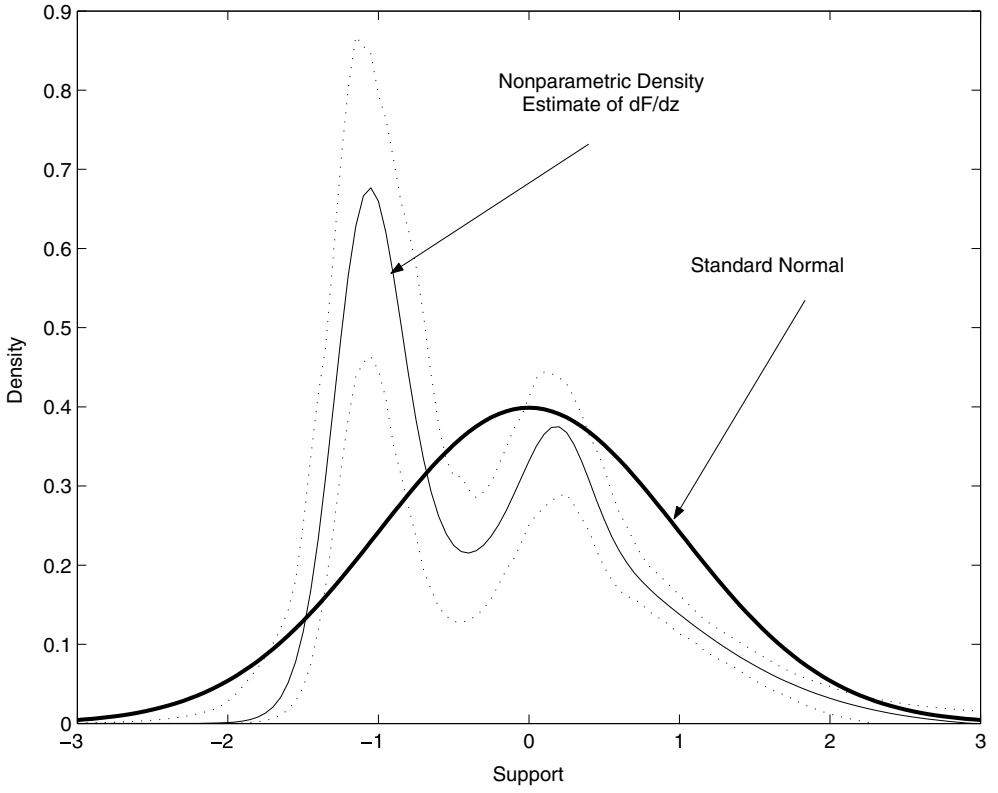
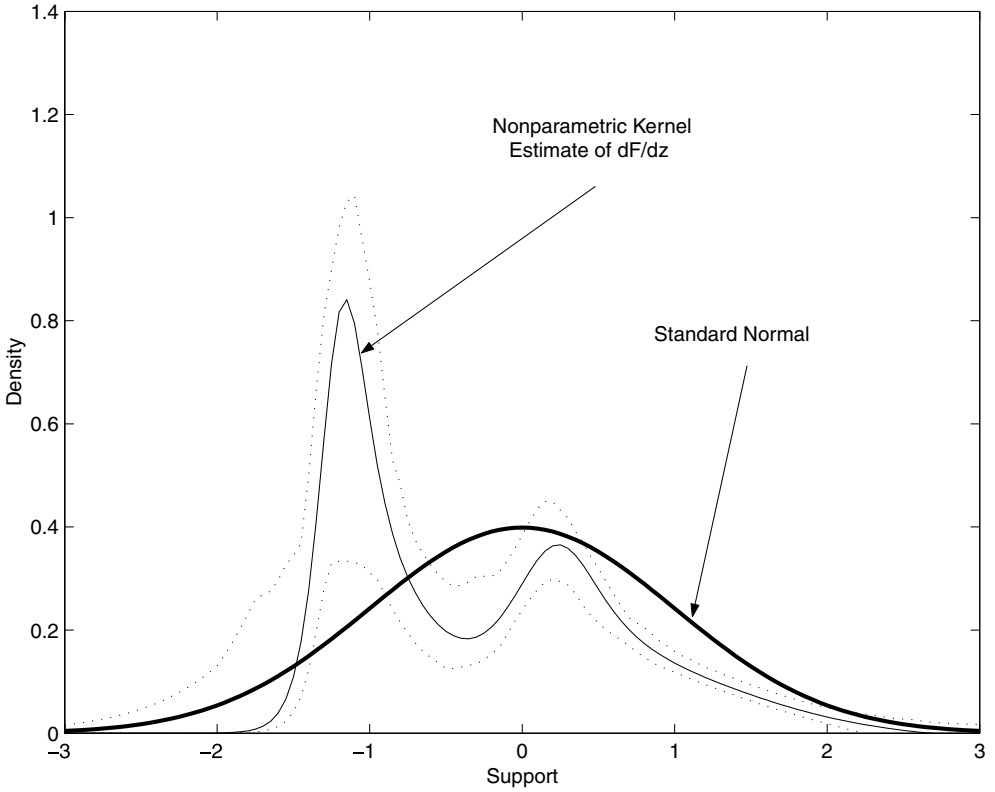


FIGURE A2
Test of the Probit for the Data Including Fund Allocations



Mammen, Proença (2000), we determine s according to $s = hn^{(1-\delta)/5}$ with $\delta = 0.1$.[15] For h, we used several values that were found after a graphical inspection of $\hat{F}$. Based on those values, $T_n/\hat{\sigma}_T$ was in the range 6.66–7.63 for model$_{nf}$ and 6.81–7.27 for model$_f$. Thus, we reject the probit model.

The second test calculates the difference in the predictive performance of the semiparametric method versus the probit for the two models. For models without reinsurance variables, the difference in the percentage correctly predicted is 3.18 percent with a standard error of 0.557 percent. For models with the reinsurance variables, the difference in the percentage correctly predicted is 4.39 percent percent with a standard error of 0.504 percent. Standard errors are calculated by bootstrapping the prediction sample and recovering the difference in the percentage of correct predictions (500 bootstraps are used). These test results strongly reject the probit model in favor of the semiparametric method.

[15] Härdle, Mammen, Proença (2000) suggest using bootstrap methods instead of a normal approximation to calculate critical values and show that bootstrap yields better approximations to the critical values in a simulation study with $n = 200$. We, however, feel more comfortable using normal approximation as we have a relatively large sample ($n = 3{,}801$).

A third test follows the graphical approach of Horowitz (1998, p. 53). Figures A1 and A2 show the nonparametric kernel estimates of $dF/d\hat{z}$, pointwise 95 percent bootstrap confidence interval, and the normal density function. Note that for a probit model, $dF/d\hat{z}$ would be the normal density function. In these nonparametric estimations, we used the standard normal density as our kernel. For bandwidth selection, we initially tried cross-validation for derivative estimation (see Härdle, 1990, pp. 160–161). Numerical minimization of this objective function was not successful for the most part so after experimenting with cross validation, we chose the bandwidths accordingly. The derivative of the link functions is clearly left skewed and hence cannot be accommodated by the symmetric normal density. Pointwise confidence intervals are represented by the dotted lines. The derivatives are bimodal, which suggests that the true data-generating processes may possibly be a mixture of two populations. Using a parametric probit model clearly misses these features of the data.

REFERENCES

Fernandez, A. I., and J. M. Rodriguez-Poo, 1997, Estimation and Specification Testing in Female Labor Participation Models: Parametric and Semiparametric Methods, *Econometric Reviews*, 16: 229-247.

Gerfin, M., 1996, Parametric and Semiparametric Estimation of the Binary Response Model of Labour Market Participation, *Journal of Applied Econometrics*, 11: 321-339.

Härdle, W., 1990, *Applied Nonparametric Regression* (Cambridge, UK: Cambridge University Press).

Härdle, W., P. Hall, and H. Ichimura, 1993, Optimal Smoothing in Single-Index Models, *The Annals of Statistics*, 21: 157-178.

Härdle, W., E. Mammen, and I. Proença, 2000, A Bootstrap Test for Single Index Models, Unpublished manuscript, Humboldt Universitaet, Berlin.

Horowitz, J. L., 1992, A Smoothed Maximum Score Estimator for the Binary Response Model, *Econometrica*, 60: 505-531.

Horowitz, J. L., 1993, Semiparametric Estimation of a Work-trip Mode Choice Model, *Journal of Econometrics*, 58: 49-70.

Horowitz, J. L., 1998, *Semiparametric Methods in Econometrics* (Dordrecht: Springer-Verlag).

Horowitz, J. L., and W. Härdle, 1994, Testing a Parametric Model Against a Semiparametric Alternative, *Econometric Theory*, 10: 821-848.

Ichimura, H., 1993, Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models, *Journal of Econometrics*, 58: 71-120.

Ker, A., 2001, Private Insurance Company Involvement in the U.S. Crop Insurance Program, *Canadian Journal of Agricultural Economics*, 49: 557-566.

Ker, A., and P. McGowan, 2000, Weather-Based Adverse Selection and the U.S. Crop Insurance Program: The Private Insurance Company Perspective, *Journal of Agricultural and Resource Economics*, 25: 386-410.

Klein, R. W., and R. H. Spady, 1993, An Efficient Semiparametric Estimator For Binary Response Models, *Econometrica*, 61: 387-421.

Manski, C. F., 1975, The Maximum Score Estimation of the Stochastic Utility Model of Choice, *Journal of Econometrics*, 3: 205-228.

Miranda, M., and J. Glauber, 1997, Systematic Risk, Reinsurance, and the Failure of Crop Insurance Markets, *American Journal of Agricultural Economics*, 79: 206-215.

Ruud, P., 1983, Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models, *Econometrica*, 51: 225-228.

Schucany, W. R., and J. P. Sommers, 1977, Improvements of Kernel Type Density Estimators, *Journal of the American Statistical Association*, 72: 420-423.

Yatchew, A., and Z. Griliches, 1985, Specification Error in Probit Models, *Review of Economics and Statistics*, 67: 134-139.