Empirical Model and Hypotheses Testing

The econometric analysis is carried out in two stages. First, equation 12 is estimated as a Generalized Polytomous Logit (GPL) function to handle the discrete choice of insurance products, measured on a nominal scale (Greene, 1990; Kennedy, 1992; Long, 1997; Stokes et al., 1998) and second, a three-stage least-squares model is specified to analyze premium rates and the choice of coverage levels.

Generalized Polytomous Logit Model

The probability that a farmer will choose one of the *m* alternative insurance products, Φ_i, from a set of choices, Φ, is given by:

$$(18)\quad prob\left(\Phi_i \middle| \Phi\right) = \frac{\exp\left[U\left(\Phi_i\right)\right]}{\sum_{j=1}^{m}\exp\left[U\left(\Phi_j\right)\right]} = \frac{\exp\left(x_i\beta\right)}{\sum_{j=1}^{m}\exp\left(x_j\beta\right)}$$

where $U(\Phi_i)$ is the utility for alternative Φ_i, x_i is a vector of variables that affect the choice of the insurance product, and β is a vector of parameters. The probability that a farmer will choose a particular product is given by the probability that the utility of that product is greater than the utility from any other available alternative (utility maximization approach). Insurance products available to farmers include CAT, APH, GRP, CRC, and RA.[7] The explanatory variables are risk type, willingness to pay for insurance, and cost of insurance (see table 3).

Since the response variable, choice of insurance product, has no inherent ordering, we estimate equation 18 as a generalized polytomous logit function. The logit of the response variable is formed as a ratio of the probability of choosing a product over the probability of choosing the reference product:

$$(19)\quad logit_{hijk} = log[\frac{\eta_{hijk}}{\eta_{hijr}}]$$

where k = 1, 2, ..., (r-1) indexes the choice of insurance products, r is the reference choice or the choice used as the basis for comparison, h, i, and j reference the explanatory variables, and η_{hijk}, which represents equation 9, is the probability of the k^{th} choice. Specifically, hijk is given by:

$$(20)\quad \eta_{hijk} = \frac{e^{\alpha_k + x_{hji}\beta}}{1 + e^{\alpha_k + x_{hji}\beta}}.$$

A logit of the response variables under consideration is formed for the probability of each product over the reference product. For example, the generalized logits for a four-level nominal response (where the producer chooses among four different insurance products) can be specified as follows:

$$logit_{hij1} = log[\frac{\eta_{hijk1}}{\eta_{hij4}}]$$

$$(21)\quad logit_{hij2} = log[\frac{\eta_{hijk2}}{\eta_{hij4}}]$$

$$logit_{hij3} = log[\frac{\eta_{hijk3}}{\eta_{hij4}}]$$

where product 4 is the reference choice. The model that applies to all logits *simultaneously*, for every combination of the explanatory variables, in a matrix form, is:

$$(22)\quad \mathrm{logit}_{hijk} = \alpha_k + x_{hij}\,\beta_k$$

where k indexes the choice of the product. The matrix x_{hij} is the set of explanatory variable values for the hij^{th} group. This model accounts for each response by estimating separate intercept parameters (α_k) and different sets of regression parameters (β_k) for all explanatory variables. That is, in the GPL model specification, we estimate simultaneously as a panel multiple sets of parameters for both the intercept and the explanatory variables.[8]

We estimate two GPL models using equation 22. Model 1 is a GPL specification with product choices GRP, CRC, and RA with APH as the reference choice. Model 2 is also a GPL specification with product choices APH, GRP, CRC, or RA with CAT are the reference choice. The reason for estimating model 2 is to use the completely subsidized contract as the reference choice.[9] In both models, however, farmers make a choice from a portfolio of yield and revenue insurance products.

Interpretation of the GPL parameter estimates is not very straightforward because the dependent variable has no inherent ordering. To facilitate interpretation of the model parameters, we estimate probabilities and odds ratios. The predicted probability that a particular product is chosen is a function of the estimated model parameters given in equation (22). Odds ratios are obtained from the predicted probabilities (Stokes et al., 1998). For example, to obtain the odds of choosing product k by a high-risk farmer relative to a low-risk farmer, we compute:

$$(23)\quad Odds\ Ratio = \frac{e^{(\alpha_k + x_{hij}\beta)}}{e^{(\alpha_k + x_{lij}\beta)}}, \qquad h \neq l,$$

where *h* and *l* are reference risk types. The odds ratio is a multiplicative coefficient, which means that positive effects are greater than 1, while negative effects are between 0 and 1. Determining the effect of the odds of the event not occurring involves taking the inverse of the effect of the odds of the event occurring (Long, 1997).

Explanatory variables used in the GPL regression model are: (i) probability of yield or revenue falling below the guaranteed level to represent risk type, (ii) level of income or size of operation to represent the willingness to pay for insurance, and (iii) premium per dollar of liability to represent the cost of insurance. Risk type of a farm (RISK) is measured in terms of the probability of yield or revenue falling below the guaranteed level. For yield insurance products, CAT, APH, and GRP, RISK is the probability of *yield* falling below the guaranteed level (Y^p), while for revenue insurance products, CRC and RA, RISK is the probability of *revenue* falling below the guaranteed level (R^p).

The probability of yield falling below the guaranteed level is estimated for *each farm* based on 10 years of yield records, the chosen guaranteed level, and assuming a normal distribution of yield.[10] The probability of revenue falling below the guaranteed level is estimated for *each farm* based on 10 years of yield records and the marketing year average prices, the chosen guaranteed level, and assuming revenues are normally distributed. This measure of risk accounts for both the mean and variance of yield or revenue (Skees and Reed, 1986; Just et al., 1999).

In this study, we use predicted probability of yield or revenue falling below the guaranteed level to measure risk. This measure of risk (Y^p or R^p) is a function of observable variables, including past yield or revenue histories and chosen guaranteed level, and thus provides a robust measure of an individual's risk.[11]

Since neither farm income nor net worth data were available, the level of income that represents willingness to pay is proxied by accumulated savings. Conceptually, income indicates the liquidity position of the farmer, which is an important determinant of the willingness to pay for an insurance contract (Makki and Miranda, 1999). The farmer's level of income, which is proportional to the size of the operation, also indicates the amount of income at risk, along with the operators' ability to pay for insurance or to self-insure against the risk of loss. In our analysis, income is estimated for each farmer as $M = \lambda \Sigma A_t Y_t P_t$, $\forall$ t = 1, 2, ... , 10, where M is the income level, A_t is the number of acres insured in time t, Y_t is the yield per acre in time t, P_t is the marketing-year average price in time t, and λ is the proportion of gross revenue saved in each year. The parameter, λ, is assumed to be equal to 0.10 or 10 percent of gross revenue (Holbrook and Stafford, 1971).[12]

The cost of insurance, captured by premium per dollar of liability, is calculated as total premium (including subsidy) divided by total liability (RATE). Liability represents the maximum potential indemnity or value of the insurance contract if a producer loses the entire crop. This measure of insurance cost facilitates comparison across different insurance contracts. Premiums are subsidized by the Federal Crop Insurance Corporation up to 42 percent (Makki and Somwaru, 1999), but we use total premium in this study.

We adopt the CATMOD procedure in SAS to estimate the GPL model. This procedure is recommended when the dependent variable has several nominal responses without any inherent ordering (Stokes et al., 1998). The CATMOD procedure forms a separate group for each distinct combination of the explanatory variable values. For continuous explanatory variables with many distinct values, the procedure would create a larger number of combinations, rendering the results impossible to interpret. To overcome this limitation, we group each of the explanatory variables into three categories, low, medium, and high.

We group the explanatory variables using their mean and standard deviation. For example, the estimated mean and standard deviation for Y^p were 0.17 and 0.11, respectively, for Iowa corn producers. Farmers with Y^p near the mean (± 1 standard deviation or 0.06

$< Y^p \leq 0.28$) are categorized as medium-risk. Farmers with $Y^p \leq 0.06$ (mean minus one standard deviation) are categorized as low-risk, while those farmers with $Y^p > 0.28$ (mean plus one standard deviation) are categorized as high-risk. Similarly, the estimated mean and standard deviation for R^p were 0.24 and 0.12, respectively. Farmers are categorized as low-risk if $R^p \leq 0.12$, while $0.12 < R^p \leq 0.36$ indicates medium-risk and $R^p > 0.36$ is high-risk. Other variables, income level, and premium rate, are categorized into the three classes using similar procedures.

Three-Stage Least-Squares Model

We specify a simultaneous equation system to analyze premium rates and choice of coverage level:

(24) $\pi_i = x_i \beta + e_i$

(25) $\theta_i = x_i \beta + u_i$

where π_i is premium per dollar of liability (including the subsidy), θ_i is the coverage level chosen by the farmer, x_i is a matrix of explanatory variables, β is a vector of parameters, while e_i and u_i are error terms. The set of explanatory variables included in equation 24 include risk type, coverage level, practice, ownership share, and yield span, while explanatory variables in equation 25 include risk type, level of income, premium rate, practice, ownership share, and yield span (see table 3).

Variables representing risk type, income level, and premium rate are as defined earlier in equation 22, except that they are not grouped. Farm practice—i.e., whether or not a farm is irrigated—is included because irrigation has the potential to reduce yield risks and may provide the incentive to buy higher coverage levels. For the econometric analysis, practice is set equal to 1 for irrigated farms and to 0 for non-irrigated farms.

Ownership share, which is the percentage share of the crop owned by the insured, could potentially influence the choice of an insurance contract. However, the direction of the effect on the level of coverage purchased is indeterminate. A positive effect implies that as the share of ownership increases, farmers are more likely to purchase higher coverage contracts. This is plausible because full ownership could mean greater dependence on farm income for livelihood. On the other hand, a negative effect is also possible, as tenant farmers are usually more leveraged and thus may be subjected to insurance requirements from lenders (Gardner and Kramer, 1986; Goodwin, 1993; Wu, 1999). Given these conflicting effects, the issue of whether ownership share is positively or negatively associated with the insurance purchase decision must be resolved empirically.

USDA's Risk Management Agency (RMA) uses the "yield span" concept to categorize farms into different *classes* (table 3). The yield spanning approach classifies farmers' yields into nine discrete risk categories (R01 through R09) based on the ratio of a farmer's yield to the average county yield. According to the yield span concept, category R01 includes the lowest average yields while category R09 includes the highest average yields. Yield span category R05 includes all those farms whose yields are expected to be equal to the county's expected yield. Yield span ranges are derived from historical county loss experience and are calibrated to the expected county yield reported by the National Agricultural Statistical Service (NASS).

Equations 24 and 25 are estimated simultaneously using the three-stage least-squares procedure. Because error terms are correlated, the farmer's decision choice of coverage levels and the premium rates require a simultaneous equation system approach.[13] The procedure is applied to each insurance product separately.

The purpose of estimating the coverage level and premium rate as a system is to analyze, *ex-post*, the relationship between the producers' choice of coverage levels and the premium rates at which they are offered. Past studies of crop insurance participation have often treated premium rates as exogenous (Coble et al., 1996; Goodwin, 1993). Although the premium rates for different coverage levels are known *ex-ante*, in this analysis we treat them simultaneously as an endogenous choice to gain insight into farmers' decision making processes and the factors affecting those decisions. This is particularly important for an analysis of markets affected by asymmetric information problems. As past yield histories and other farm and farmer risk characteristics are not easily available, the best way to address farmers' attitudes is by observing the choice(s) made by the farmers themselves. Thus, analysis of premium rates and coverage levels can enhance our understanding of farmers' behavior in the crop insurance market. Furthermore, if farmers effectively signal their risk type, through the choice of premium-coverage level, then such information is useful in assessing potential losses and setting premium rates commensurate with risk.

Table 3—Variable description

Variable name	Variable definition
Insurance plans	Alternative insurance plans or products that include Catastrophic Coverage (CAT), Actual Production History Insurance (APH), Group Risk Plan (GRP), Crop Revenue Coverage (CRC), Revenue Assurance (RA), and Income Protection (IP).
Coverage level	Alternative coverage levels that range from 50% to 85% in an interval of 5%.
Premium	Per-acre premium paid in dollars to purchase insurance (includes subsidy).
Rate	Rate is the premium per dollar of liability (premium/liability).
Loss ratio	Loss ratio = Indemnity/Premium.
Loss-cost ratio	Loss-cost ratio = Indemnity/liability.
Risk type	Probability of yield or revenue falling below the guaranteed level, estimated for each farm based on 10 years of yield records and using the corresponding year market average price.
Loss frequency	Ex post observation of whether a farmer filed a claim, also known as loss frequency; set equal to one for those who filed a claim and set equal to zero otherwise.
Yield span	A yield-spanning process creates nine discrete categories (R01 through R09) of yields. Category R01 is associated with the lowest average yields, while category R09 is associated with the highest average. The yield-span ranges are derived from historical county loss experience and are calibrated to the expected NASS county yield. Rates for each category are inversely proportional to the farm's expected yield. Thus, farms in relative expected yield categories 1-4 are charged premium rates which are higher than the base county rate. Conversely, farms in relative expected yield categories 6-9 are charged lower premiums than the base county rate.
Farm income	Income is estimated for each farmer as follows: $M = \lambda \sum A_t Y_t P_t , \ \forall \ t = 1, 2, ... , 10,$ where M is the income level, A_t is the number of acres, Y_t is the yield per acre, P_t is the State average price, and is the proportion of income saved, which is assumed to be 0.10.
Expected indemnity	Expected indemnity, E(I), is estimated for each farmer as follows: E(I) (per acre) from a typical yield insurance contract: $E(I) = MAX(0, Y^g - Y)P^g$, E(I) (per acre) from a typical revenue insurance contract: $E(I) = MAX(0, Y^g P^g - Y P^m)$, where Y^g is the guaranteed yield, Y is the actual farm yield, P^g is the guaranteed price, P^m is the market price at harvest time.
Farm practice	Farm practice, which indicates whether a farm is irrigated, is set equal to one for irrigated farms and zero for non-irrigated farms.
Ownership share	Ownership share is the percentage share of the crop owned by the insured.

Hypotheses Testing

Testing for separating equilibrium. The model is tested for the existence of a separating equilibrium by assessing the signs and the statistical significance of the variable RISK in equations 22 and 25. Significant coefficients for RISK in equations 22 and 25 would indicate a separating equilibrium, implying that low-risk and high-risk farmers purchase different contracts. For example, a significant positive coefficient for RISK in equation 25 would indicate that low-risk types purchase contracts with lower coverage, while high-risk types purchase higher coverage contracts. On the other hand, a non-significant coefficient would indicate a pooling equilibrium, implying that all risk types purchase the same contract.

Testing for the effects of farm income. We expect the choice of insurance contracts to be related to income in a manner consistent with the decreasing risk-aversion hypothesis. This is equivalent to asserting that farmers with higher income retain the risk of some losses. One expects that high-income farmers would be more likely to choose the lower coverage contracts, as they are able to self-insure and manage variations in income within their operations better than would farmers with lower income.

Testing for the effects of cost of insurance. Premium rates are conditioned on insurance product, coverage level, irrigated versus non-irrigated production, and RMA's yield span classification. Assuming low-risk types buy lower coverage levels, a positive correlation between coverage level and the premium rate in equation 24 implies that insurers compensate low-risk types accordingly. The statistical significance of the premium rate in equations 22 and 25 has implications for public subsidization of the risk insurance programs.

Testing for market signaling. A nonlinear relationship in the coverage-premium schedule indicates the presence of signaling in the insurance market.[14] The nonlinearity of the coverage- premium schedule is tested by introducing three dummy variables into the system representing three coverage levels, 55 percent, 65 percent, and 75 percent. Over the range of coverage levels in our sample, nonlinearity would be present if the marginal premiums at the various coverage levels are significantly different. Assuming that farmers make informed decisions, a farmer's selection of an insurance contract reveals information, although imperfectly, about the riskiness of his or her operations. This information could potentially be used to decrease the adverse effects of asymmetric information in the insurance market.

Testing for adverse selection. We test for adverse selection using a two step procedure. First, we test for the independence of the choice of insurance contract and the risk using non-parametric methods. If the choices are correlated with risk, then agents indeed have a better knowledge of their risk (Chiappori and Salanie, 2000). Rejection of independence would suggest that there is evidence of adverse selection in the crop insurance market.

Parametric methods used by Puelz and Snow (1994) or Dionne, Gourieroux, and Vanasse (1998), for instance, rely on a fairly large number of exogenous variables and restrictive functional forms. Hence, the results from parametric methods would be biased. Non-parametric methods are, on the other hand, less restrictive and account for more complicated non-linear relationships between variables (Chiappori and Salanie, 2000). The two non-parametric tests performed are the Kruskal-Wallis χ^2 test and the Kolmogorov-Smirnov test. The Kruskal-Wallis test statistic is given by:

$$(26)\quad H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{T_i^2}{n_i} - 3(N+1)$$

(where N is the sample size, T_i is the rank assigned to the ith group, and n_i is the number of groups in the sample. The test statistic H approximately follows a chi-squared distribution with k-1 degrees of freedom. See Milton and Arnold (1990) for more details on the Kruskal-Wallis test. The Kolmogorov-Smirnov test statistic is given by:

$$(27)\quad K = \sqrt{M}\,\sup_x \left|F_M(x) - F(x)\right|$$

where $F_M(x)$ is the empirical cdf and F(x) is the cdf of a $\chi^2(1)$. Under conditional independence, the test statistic K converges to a distribution that is tabulated in statistics textbooks (Chiappori and Salanie, 2000).

The second step involves comparing the actual and competitive premiums across different risk types. In an efficient market, the competitive premium is equal to the expected indemnity (Puelz and Snow, 1994; Rothschild and Stiglitz, 1976). The expected indemnity E(I) is calculated for each insurance contract separately. For example, E(I) for a typical CRC contract is:

$$(28)\quad E(I) = \frac{1}{n}\sum_{t=1}^{n} MAX[0,\ (\theta Y^e\ \max(P^g,\ P^m)\ -\ Y^t P^m)]\ ,$$

where n is the number of periods for which yield records are available, θ is the coverage level, y^e is the expended yield, y^t is the actual yield in year t, Pg is the guaranteed price (or elected price), and P^m is the market price. We use 10 years (1987-96) of actual yield history for each farm and corresponding annual market prices. Coverage level θ and guaranteed price P^g were chosen by farmers in 1997. We adjust the guaranteed yield for the growth rate in yield to make it comparable with yield in period t. The market price, however, did not exhibit any trend during the 1987-96 period. The calculated E(I) captures farm risk characteristics by accounting for alternative yield and price possibilities.[15]

Under a full-information equilibrium, the difference between actual and competitive premium rates should be zero for all risk types. Under asymmetric information, however, one would expect differences to exist between actual and competitive premium rates, as the accurate determination of individual farmers' risk is either not possible or prohibitively expensive. We use non-parametric tests and graphical illustrations to demonstrate the differences, if any, between the actual and competitive premium rates. The two non-parametric tests performed are the Kruskal-Wallis χ^2 test and the Kolmogorov-Smirnov test as described above.