A STATISTICAL EXAMINATION OF "YIELD SWITCHING" FRAUD IN THE FEDERAL CROP INSURANCE PROGRAM

Federal crop insurance is a core component of U.S. agricultural policy. From its inception, however, policy makers have struggled with two persistent problems: (1) higher than desired loss ratios and the resulting federal expenditures, and (2) lower than desired producer participation despite federal subsidies as high as 67 percent of some producer's premium costs (Goodwin and Smith). The persistence of higher than desired loss ratios and lower than desired producer participation have commonly been attributed to adverse selection and moral hazard (Chambers; Coble et. al.; Goodwin and Smith; Just, Calvin, and Quiggen; Makki and Somwaru; Skees and Reed; Quiggen, Karagiannis, and Stanton; Vercammen and van Kooten). Fraud has received less attention in the federal crop insurance literature although it is a common theme in the general insurance literature (Artis, Ayuso, and Guillen; Dionne; Brockett, Xia, and Derring; Picard, 1996, 2000). Fraudulent activity may also be increasing loss ratios and leading to higher premium rates[1] and a corresponding decrease in participation rates (USDA, Office of the Inspector General; General Accounting Office).

Estimating the costs of fraudulent conduct is difficult because accurately differentiating between legitimate and illegitimate indemnifications requires extensive on-site investigation. Statistically differentiating between legitimate indemnifications and fraud is a difficult identification problem similar to problems encountered when attempting to differentiate between environmentally induced losses, adverse selection, and moral hazard (Chiappori; Dionne; Quiggin et. al.)[2]. In this paper we present a statistical model that tests for one type of fraudulent conduct - switching reported yields between separately insured tracts of land. The statistical model tests for a specific type of dynamic yield reporting that is consistent with fraudulent conduct but is unlikely to have occurred with honest reporting.

RMA provisions define two categories of separately insurable units - basic units and optional units. A producer may separately insure two or more units as basic units if the producer farms the land but each tract is owned by (and the indemnities shared with) an entity different from the producer. Examples of basic units include land rented under a crop share arrangement with different landlords. Optional units are tracts that satisfy certain spatial requirements, are controlled by one producer, and the producer entity retains a 100 percent share of any indemnifications. Examples include land owned by the producer as well as any tracts under a cash rent arrangement. The interested reader is referred Knight and Coble for additional definitions, details and comparisons between basic and optional units.

The fraud examined in this paper involves switching reported yields between separately insured tracts or units of land under the control of a single producer. The possibility of this type of fraud has been discussed in several governmental studies. Both the USDA inspector general and GAO reports explicitly raise the issue of potential abuses of the optional unit provisions of the current crop insurance program. Producers may exploit optional unit provisions by manipulating "their unit structure to benefit themselves when determining if losses actually occurred" (USDA, Office of the Inspector general).[3] Knight and Coble; Kuhling; and Atwood, Watts, and Shaik found that loss cost ratios (indemnifications divided by liability exposure) were higher for optionally insured units than for basic units with similar characteristics and of similar size. While the higher LCR's are not necessarily indicative of misconduct, RMA's current rate structure recognizes the higher loss cost ratio (LCR) and incorporates a ten percent surcharge for optionally insured units.[4]

Among other problems with optional unit provisions, the USDA Office of the Inspector General report indicated that audits of actual producer losses found producers who were not able to provide adequate records that "showed from which optional units the production actually was harvested". As we demonstrate below, the inability of crop adjusters to accurately tie production to specific tracts of land provides strong incentives for some producers to switch reported production between various tracts of land and generate additional fraudulent indemnifications. Atwood, Watts, and Shaik provide anecdotal support for potential yield switching gains.

In this study we expand Atwood et. al's results by presenting a statistical test of potential yield switching that is applied to reported yield histories from 206,952 producers. Procedures are presented that allow identification of: (1) farms whose reported yields are consistent with yield-switching; (2) an estimate of the prevalence of yield switching in the total population; and (3) an estimate of the aggregate cost of yield switching in the population. The paper concludes with a summary and recommendations for further research.

A Model of Yield Switching

We assume that a producer grows an insurable crop on each of U separately insurable tracts or units of land. Let a_u denote the acres of the crop in unit u (u=1,2,…,U) and $y_{u,t}$ the per-acre yields for unit u as reported to the insurance company in year t. On each unit the producer is allowed to insure a proportion K (from 50 to 85 percent)[5] of the unit's year-t Approved Production History ($APH_{u,t}$) computed as a four-to-ten year moving average of previously reported yields. Assuming the producer reports T years of historical data, unit u's APH at the beginning of year t is computed as:

$$APH_{u,t} = \frac{1}{T}\sum_{i=t-T}^{t-1} y_{u,i} \tag{1}$$

At the end of the year, the producer realizes the actual production and reports yields $y_{u,t}$ to the insurance company[6]. If $y_{u,t}$ is less than $K \cdot APH_{u,t}$ for any u, a crop adjustor visits the farm and attempts to determine if there are any discrepancies in the reported yields. Adjustment procedures may include field examinations, measurements of stored production, examinations of sales receipts, and/or contacting marketing establishments within a reasonable transportation distance. If the total production identified during the adjustment process equals the total reported production ($TRP_t = \sum_u a_u y_{u,t}$), and the adjustor finds no substantial evidence of inaccuracies in reported $y_{u,t}$, the producer receives an indemnity of $a_u(K \cdot APH_{u,t} - y_{u,t})$ on each unit for which $y_{u,t}$ is less than $(K \cdot APH_{u,t})$. In the following we assume that the producer's total reported production TRP_t equals the total production potentially identifiable by an adjustor.

Let $z_{u,t}$ be the actual (and unobserved) per-acre production from unit u that is included in TRP_t giving:[7]

$$TRP_t = \sum_u a_u z_{u,t} = \sum_u a_u y_{u,t} \tag{2}$$

Assume that, at the end of year t, $z_{u,t}$ exceed $(K \cdot APH_{u,t})$ for all u and no legitimate indemnifications are forthcoming. The producer realizes that, while the adjustor can determine $TRP_t = \sum_u a_u z_{u,t}$ with relative accuracy (due to total receipts, inventories, etc.), conclusive verification of differences between $z_{1,t}, z_{2,t}, \ldots, z_{U,t}$ is much more difficult. Assume that the producer is willing to illicitly enhance indemnification by manipulating reported yields. The manipulated reported yields can be expressed as:[8]

$$y_{u,t} = z_{u,t} + \Delta_{u,t} \tag{3}$$

subject to:

$$TRP_t = \sum_u a_u y_{u,t} = \sum_u a_u z_{u,t} \qquad \text{for } t = 1, \ldots, T \tag{4}$$

where $\Delta_{u,t}$ denotes the difference between reported and actual per-acre yields on unit u in year t. If $\Delta_{u,t} > 0$ ($\Delta_{u,t} < 0$), unit u's reported yield in year t has been artificially inflated (deflated) relative to the actual $z_{u,t}$. Expressions (3) and (4) can be rearranged allowing (4) to be replaced with an equivalent set of restrictions:

$$\sum_u a_u \Delta_{u,t} = 0 \qquad \text{for } t = 1, \ldots, T \tag{5}$$

Figure 1 graphically demonstrates the producer's yield switching incentives. The first panel of figure 1 plots year-t total cotton indemnities (in pounds of cotton) as a function of $\Delta_{1,t}$ assuming U = 2, $a_1 = a_2 = 1$, K = .75, $(K \cdot APH_{1,t}) = 900$, $(K \cdot APH_{2,t}) = 750$, $z_{1,t} = 950$ and $z_{2,t} = 850$. With honest reporting $\Delta_{1,t} = 0$, $y_{1,t} = 950$, $y_{2,t} = 850$ and total indemnities $INDEM_t = 0$. If 200 pounds of reported production are "switched" from unit 1 to unit 2, ($\Delta_{1,t} = -200$, $y_{1,t} = 750$ and $y_{2,t} = 1000$), total indemnities equal 150 (900-750) pounds on unit 1 and none on unit 2. If

$\Delta_{1,t} = 200$ ($y_{1,t} = 1150$ and $y_{2,t} = 650$) total indemnities equal 100 (750-650) pounds on unit two and none on unit one. Shifting all reported production from unit 1 to unit 2 ($\Delta_{1,t} = -950$, $y_{1,t} = 0$ and $y_{2,t} = 1800$) generates 900 or $(K \cdot APH_{1,t})$ pounds of indemnities. Shifting all production from unit 2 onto unit 1 ($\Delta_{1,t} = 850$, $y_{1,t} = 1800$ and $y_{2,t} = 0$) generates 750 or $(K \cdot APH_{2,t})$ pounds of total indemnities.

If the probability of detection is zero or remains low for all possible $\Delta_{1,t}$ values, a myopic producer maximizes year-t indemnities with "all-or-nothing" yield switching $\Delta_{1,t} = 950$. If the probability of detection increases with $|\Delta_{1,t}|$, the optimal yield switching decision will often involve "partial" yield switching. The first panel of figure 1 also plots net-indemnities less expected agency imposed fraud fines. The plot assumes that the probability of detection increases linearly from 0 (when $\Delta_{1,t} = 0$) to .5 (when $|\Delta_{1,t}* = 950$) and the fraud fine (if detected) is twice the amount of the fraudulent indemnification. With these assumptions, the myopic optimal $\Delta_{1,t}$ level is -500 (partial yield switching) with illicit indemnities of 450 and a 26.3 % probability of detection. In both cases, the optimal myopic yield switching rule involves switching reported production from the unit with the higher APH (unit 1) to the unit with the lower APH (unit 2).

The expression "myopic" is used in the preceding paragraph because the producer's "optimal" solution did not consider the dynamic effects of year-t yield switching upon future yield switching incentives. Since the APH is a T-year moving average of previously reported yields, $\Delta_{1,t}$ will influence both year t indemnities and APH values in years t+1, t+2, ... t+T.

The second panel of figure 1, plots year t +1 indemnity responses to $\Delta_{1, t+1}$ assuming that $APH_{1,t+1}$ decreases to 750 and $APH_{2,t+1}$ increases to 900. Realized $z_{1, t+1}$ and $z_{2, t+1}$ are assumed to equal 950 and 850 respectively. Once again, the optimal myopic policy involves switching production from the highest APH unit (unit 2) to the lowest APH unit (unit 1). Shifting 850 pounds of reported production from unit 2 to unit 1 ($\Delta_{1, t+1} = 850$) generates a maximal indemnity of 900 if the probability of detection is low. Simulations[9] indicate that the “switch all production from high APH to low APH units” rule dominates honest reporting if the probabilities of detection are low. If the probability of detection increases with the severity of the yield switching, partial yield switching dominates all-or-nothing yield switching.

Dynamically switching yields from higher APH units to lower APH units generates higher indemnity payments and oscillating reported yield patterns caused by a given unit's “receiving” production in some years and “donating” production in other years. Within a given year, some of the farm’s units will have inflated yields while other units will have deflated yields with the patterns being reversed in different years. For a two-unit farm, plots of predicted $y_{u, t}$ are similar to figure 2, which plots the reported yields of an actual two-unit cotton producer. The oscillating and negatively correlated yield pattern in figure 2 was “flagged” as statistically significant using the procedures presented in the following paragraphs. We return to this farm in upcoming examples.

A Statistical Model of Yield Switching

We next develop a statistical model that identifies possible yield switching. We first assume that the conditional expected value of the unobserved actual $z_{u, t}$ is proportional to realized county yields in year t with:

$$z_{u,t} = g_u C_t + e_{u,t} \tag{6}$$

where g_u is a “unit - u” expected yield multiplier, C_t are county or regional average yields in year t, and $e_{u,t}$ is a residual. Substituting expression (6) into (3) gives:

$$y_{u,t} = g_u C_t + \Delta_{u,t} + e_{u,t} \tag{7}$$

subject to:

$$\sum_u a_u \Delta_{u,t} = 0 \tag{8}$$

Systems (7) and (8) can be viewed as testing whether the reported yield series $y_{u,t}$ are statistically different from any yield series $z_{u,t}$, with the same total reported production in year t, that could reasonably be expected from process (6). In the following discussion we will also refer to system (7) using the common notation

$$\underline{y} = X\underline{b} + \underline{e} \tag{9}$$

where $\underline{y}' = (y_{1,1},...y_{1,T}, y_{2,1}...y_{2,T},..., y_{U,1}...y_{U,T})$, X is a corresponding design matrix, $\underline{b}' = (g_1, g_2,..., g_U, \Delta_{1,1}..\Delta_{1,T}, \Delta_{2,1},..., \Delta_{2,T},..., \Delta_{U,1},..., \Delta_{U,T})$ is a vector of parameters, and $\underline{e}$ is a vector of corresponding residuals.

Within each farm in a given county or region, the error structure is assumed to satisfy:

$$E(e_{u,t}^2) = s_t^2$$

$$E(e_{u,s}e_{v,t}) = 0 \text{ for all u and v when } s \neq t \tag{10}$$

$$E(e_{u,s}e_{v,t}) = r\ s_t^2 \text{ for all } u \neq v \text{ when } s = t$$

with the s_t^2 and r values assumed common to all farms with an equivalent number of units in the county-region.[10]

In the absence of restrictions (8), system (7) or (9) is over-parameterized (under-identified) with U + UT parameters and only UT observations. The T restrictions in (8), while linearly independent of each other, can be shown to be insufficient to allow the identifiably of the system. We next present a brief review of the two main statistical approaches to the over-parameterized problem. We then discuss the identifiability problem with restrictions (8) and present an additional set of constraints that, with (8), enable identifiably.

Statistical Approaches to the Over-parameterized Problem

System (7) or (9) is a special case of a more general set of over-parameterized statistical problems in which the $(X'X)$ matrix is singular and the normal equations $(X'X)\underline{b} = X'\underline{y}$ have an infinite number of solutions $\underline{b}$. Examples include the classical ANOVA model and the dummy variable problem discussed by Suits and Greene and Seaks. The statistical difficulty introduced by over-parameterized linear models is that the function $\underline{r}'\underline{\tilde{b}}$ is not uniquely identified for all combinations of $\underline{r}$ and $\underline{\tilde{b}}$. The statistics literature addresses this problem in two ways, each of which utilizes restrictions.

The first approach (Rao; Searle) restricts the set of testable hypotheses $\underline{r}$ to the set of "estimable" $\underline{\tilde{r}}$ for with $\underline{\tilde{r}}'\underline{\tilde{b}}$ is invariant for all normal equation solutions $\underline{\tilde{b}}$. The estimable restrictions approach involves the use of an arbitrary generalized inverse $(X'X)^-$ to find “a” solution $\underline{\tilde{b}} = (X'X)^- X'\underline{y}$ to the normal equations and then performing hypothesis tests on estimable restrictions. See Searle and Chipman for numerous results with respect to the estimable restrictions approach. Two results that we use below are (a) a restriction $\underline{\tilde{r}}'\underline{\tilde{b}}$ is estimable if and only if $\underline{\tilde{r}}'$ lies in the row space of X (or equivalently the column space of $(X'X)$) and (b) only estimable restrictions or linear combinations of estimable and non estimable restrictions can influence the SSE of the regression models.

The second approach to the over-parameterized problem works well with restrictions like (8). This approach involves restricting the $\underline{b}$ -space sufficiently so as to allow the constrained system to become invertible (Searle; Johns; Chipman; Green-Seaks). Examples of this approach include the arbitrary deletion of dummy variables or the imposition of more easily interpretable restrictions (Green and Seaks). Green and Seaks' procedure involves the identification of a sufficient set of non-estimable restrictions to allow invertibility of the lagrangian ($\underline{l}$) augmented (LAUG) system:

$$\begin{bmatrix} (X'X) & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \underline{b} \\ \underline{l} \end{bmatrix} = \begin{bmatrix} X'\underline{y} \\ \underline{d}_R \end{bmatrix} \tag{11}$$

with

$$\begin{bmatrix} \underline{\hat{b}}_R \\ \underline{\hat{l}}_R \end{bmatrix} = \begin{bmatrix} (X'X) & R' \\ R & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'\underline{y} \\ \underline{d}_R \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} X'\underline{y} \\ \underline{d}_R \end{bmatrix} \tag{12}$$

Searle; Johns; Chipman; and Green-Seaks present numerous generalized inverse results with respect to systems (11) and (12) including the results that $(X'X)$ is a generalized inverse of G_{11} and G_{11} is related [11] to the variance-covariance matrix of $\underline{\hat{b}}_R$. If $(X'X)$ is rank deficient of order p and expression (11) is constructed with exactly p non-estimable constraints, G_{11} is also a generalized inverse of $(X'X)$, $G_{22} = 0$, and the lagrangians $\underline{\hat{l}}_R$ equal $\underline{0}$. Additional results with respect to systems (11) and (12) are available from the authors.

Deriving an invertible LAUG matrix requires identifying p linearly independent "X nonestimable" restrictions. In addition, (Searle) the non-estimable restrictions must be simultaneously independent of the rows of X'X and the other rows of R. In the following, we denote restrictions that cannot be expressed as linear combinations of the rows of (X'X) and the other rows of R as being "system non-estimable". By definition, system non-estimable restrictions are linearly independent of the other rows of R and are also X non-estimable. However, individual restrictions may be both linearly independent of other restrictions in R and X non-estimable and yet be system estimable. All X estimable restrictions are system estimable.

The importance in identifying the number of "system estimable" restrictions in the LAUG system is two-fold. First, if (X'X) is rank deficient of order p, we must identify p system non-estimable restrictions if the LAUG system is to be invertible. Secondly, when additional hypothesis testing restrictions are imposed on a LAUG system with a singular (X'X), degrees of freedom adjustments in the hypotheses tests involve differences in the rank of basis sets for the two system's estimable restrictions not necessarily the difference in the total number of restrictions between the two systems. Additional results are available from the authors.

<u>Identifiably Restrictions for the YS Model</u>

The (X'X) matrix of system (11) can be shown to be rank deficient of order U. When the T restrictions of (8) are imposed simultaneously, it is easily verified that: (i) any (T-1) of the restrictions can be written as a linear combination of the rows of X'X and the remaining restriction and (ii) the LAUG system remains rank deficient of order (U-1). As a result, it is necessary to identify U-1 additional system non-estimable restrictions to obtain an invertible LAUG matrix.

An additional set of economically interpretable restrictions is suggested by the dynamic characteristics of yield switching. If the "switch reported production from high APH to lower APH unit" rule is followed dynamically, a given unit u will have $\Delta_{u,t} > 0$ in some years and $\Delta_{u,t} < 0$ in other years as the unit's reported yields are cyclically "inflated" and "deflated" by yield switching. Our second set of restrictions limits a given unit's $\hat{\Delta}_{u,t}$ estimates to levels that are offset by $\hat{\Delta}_{u,s}$ estimates of opposite sign within the sample period. The restrictions can be written as:

$$\sum_{t} \hat{\Delta}_{u,t} = 0 \qquad \text{for } u = 1, \ldots, U\text{-}1 \tag{13}$$

Each of the U-1 restrictions[12] in (13) can be shown to be system non-estimable when imposed simultaneously with the T restrictions of (8) and the resulting LAUG system is invertible.

If the restrictions in (13) are not strictly satisfied within the given sample period, individual $\hat{\Delta}_{u,t}$ estimates can be shown to be biased. However, since restrictions (13) are system non-estimable, (i) the SSE of the LAUG system does not depend upon the right hand sides in (13) and (ii) significance tests for the joint hypotheses that $\hat{\Delta}_{u,t} = 0$ (for all u and t) do not depend upon restrictions (13) being exactly satisfied. Restrictions (13) are, in effect, serving the purpose of the identifiability constraints utilized by Greene and Seaks. We demonstrate that restrictions (13) do not affect the SSE in the appendix.

In the next section, we apply the LAUG yield-switching model to yield data from two insured Southeastern cotton farms including the farm whose yields are plotted in figure 2. We then apply the model to reported yield data from 206,952 insured producers of six crops and estimate the incidence of yield switching in the insured population.

An Empirical Application of the Yield Switching Model

In this section we apply the statistical model to reported yield data obtained from RMA's public web site. The data consists of four-to-ten years of unit level reported yields for all producers who purchased barley, corn, cotton, grain sorghum, soybean, or wheat insurance in the 1998 crop year. The procedures described below were individually applied to all farms and units with at least four years of valid[13] yield data (per unit) on two or more units. The example focuses on the two-unit producer whose yields were plotted in figure 2 and a second producer from the same county. Table 1 presents information employed in testing the farms for yield switching and a summary of the results. Numerical matrices and other results are presented in the appendix. The steps and assumptions involved in the estimation are described below.

Assumptions and Procedures In the Estimation Process.

The regression model for each multiple unit farm can be summarized as:

$$y_{u,t} = g_u C_t + \Delta_{u,t} + e_{u,t} \tag{14}$$

subject to:

$$\sum_u a_u \Delta_{u,t} = 0 \qquad \text{for } t = 1, ..., T \tag{15}$$

and

$$\sum_t \Delta_{u,t} = 0 \qquad \text{for } u = 1, ..., U\text{-}1 \tag{16}$$

with the error structure assumptions (10). The C_t, s_t^2, and r values were estimated from the RMA population of single and multiple unit producers in the producers' county-region. For each given county, contiguous counties were pooled until there were at least 100 single-unit farms in the given county's "county-region"[14]. "County-region" yields C_t were averaged across all reported yields in year t.

Annual residual variances s_t^2 were computed by year using pooled residuals from all single unit producers in the county-region. Single unit farm residuals were obtained by applying a three-stage GLS procedure to the following model[15]:

$$y_t^f = g^f C_t + e_t^f \qquad (17)$$

where f denotes farm f. The county average within-farm unit correlation r was computed by applying GLS (with a diagonal variance matrix using the s_t^2 values from the single unit farms) to estimate a "no-switching" model:

$$y_{u,t}^f = g_u^f C_t + e_{u,t}^f \qquad (18)$$

for each multiple unit farm in the county-region. Within each farm f, pair-wise unit residual correlations $\hat{r}_{u,v}^f$ were computed for all pairs of units u and v. The farm's average 'unit-pair" correlation $\hat{r}^f$ was calculated and the county-region average correlation computed as $r = \frac{1}{F_M}\sum_f \hat{r}^f$ where F_M denotes the number of multiple unit farms in the county region.

Testing the Example Farms for Yield Switching

Table 1 presents four years of $y_{u,t}^f$, C_t, s_t^2, and the residual covariance terms for two producers who reported yields in the same years. The first column of yield data lists the first four years[16] of reported yield data from the producer in figure 2. The second column of yield data lists the same year's reported yields from a different producer in the county[17]. Table 1 also presents the GLS-scaled $\mathrm{SSE}_{\mathrm{SW}}$, the likelihood ratio (LRT) statistics and the p-values derived from the "fully restricted" non yield-switching (NSW) model (18) and the "partially restricted" yield-switching (SW) model (14-16) – (see the appendix for numerical computation examples).

Noting that the fully restricted no-switching model (18) is nested in the partially restricted yield-switching model (14-16), we utilize the **population** variance-covariance structure Σ in a likelihood ratio test based on the difference between $\mathrm{SSE}_{\mathrm{NSW\ and}}$ $\mathrm{SSE}_{\mathrm{SW}}$. If the residual vector $\underline{\mathrm{e}}$ is approximately Normal($\underline{0}$, Σ) with known Σ, the following statistic is approximately $c^2(r)$ where r is the difference in the rank of basis sets for the estimable restrictions in the two models:

$$\mathrm{SSE}_{\mathrm{NSW}} - \mathrm{SSE}_{\mathrm{SW}} = (\underline{\hat{\mathrm{e}}}_{\mathrm{NSW}})'\Sigma^{-1}(\underline{\hat{\mathrm{e}}}_{\mathrm{NSW}}) - (\underline{\hat{\mathrm{e}}}_{\mathrm{SW}})'\Sigma^{-1}(\underline{\hat{\mathrm{e}}}_{\mathrm{SW}}) \sim c^2(r) \qquad (19)$$

Expression (19) can be shown to be equivalent to a likelihood ratio test (LRT) statistic and we refer to the statistic as the LRT statistic.

In practice, expression (19) is often of limited use in that Σ is usually unknown and must be estimated using an individual's sample data. If we had data from only one producer, estimates of Σ would be questionable due to the low degrees of freedom in system (14-16). The usual approach of estimating the variance from the sample residuals and utilizing an F-test is problematic when the denominator degrees of freedom is low relative to the number of parameters being estimated. In the restricted system (14-16) there are only T-1 degrees of freedom regardless of the number of units U. In Monte

Carlo simulations we found that traditional F-tests developed with the individual producer's data had low power and did not reliably identify simulated yield switching.

However, in this study, we have a substantial amount of additional information from the population of producers. Monte Carlo simulations indicate that estimating Σ directly from the population of insured producers results in more powerful tests of yield switching as contrasted to using an F-test if the small sample SSE of the farm is used in the denominator of the F-statistic. Alternatively, if the population residual variance is used in the denominator of the F-statistic, the results are equivalent to the LRT test in (19). The degrees of freedom r for the LRT statistic can be shown to be:

$$r = \mathrm{U\,T} - \mathrm{T} - \mathrm{U} + 1 \tag{20}$$

The first farm's LRT statistic is 46.67 with (8-4-2+1) or 3 degrees of freedom giving a p-value of 0.000. We do not interpret these p-values as exact because normality is questionable and sample sizes are small. However a low p-value is strong evidence against the null hypothesis that the first farm's reported yield pattern occurred randomly, given county yields C_t and variance-covariance structure Σ.

The second farm's unit level yields are more closely related to the county yields and do not exhibit the oscillatory pattern show in figure 2. The farm's full sample LRT statistic is 2.25 with 3 degrees of freedom giving a p-value of 0.522 implying that we cannot reject the hypothesis that the second farm's yield series could have occurred randomly.

The third column of yield data in table 1 is presented to demonstrate a potential Type I error problem that we observed when applying the model to the population of multiple unit producers. The LRT was initially individually applied to the reported yield data from 206,952 multiple unit producers of six crops. We then visually examined plots of reported yields for all farms with LRT p-values $\leq .01$. We noted that a number of the low p-value farms exhibited no visible signs of oscillating yields and had one of more units with a very low yield in one year. To examine the sensitivity of the LRT statistic to isolated low yield events[18], we constructed multiple Monte Carlo simulations with various error distributions and correlation structures[19].

If the Monte Carlo error structure is normal with no yield switching, the empirical distribution of the LRT p-values is approximately uniformly distributed over the [0-1] interval. If the correlation structure is improperly specified, the resulting p-values are biased toward 1 (0) if the statistical model's $\hat{r}$ is less than (greater than) the actual r used to generate the data. Monte Carlo samples were also generated with the population's residual s_t^2 and r varying from producer to producer. The simulations indicate utilizing population average s_t^2 and r values produces more robust Type I - Type II error rates as contrasted to estimating small sample farm-specific s_t^2 and r values.

With "long-tailed" distributions or with simulated low probability "hail events", many honest simulated producers whose simulated yields contained an isolated low yield event had low LRT p-values that resulted in Type 1 errors. To examine the incidence of Type 1 – Type 2 error rates, we also constructed simulations in which ten-percent of the simulated population switched yields. While the LRT test identified a significant proportion of the yield-switching group, the test also improperly categorized a number of the "honest-but-hailed" producers as potential yield-switchers.

Given that many of the Type I errors were caused by a single isolated low yield event, we examined the applicability of a jack-knifed test similar to the intersection-union tests discussed by Berger and Hsu. The following jack-knife test was applied to each simulated producer. For each year t = 1,2, …T, LRT_t and p-$value_t$ were estimated while deleting year t's data from the regressions. Let J-MAX = $\max_t$ (p-$value_t$). A "small" J-MAX indicates that the results are not sensitive to events in any single year[20] and is evidence against the null hypothesis of no yield switching. If a producer's J-MAX exceeds some critical value the low LRT p-value was caused by an event that occurred in one year[21]. In the Monte Carlo simulations, setting the critical value for J-MAX at 0.10 gave robust Type I-Type II error results with roughly equal numbers of Type I –Type II error counts. Given that we desire to estimate the incidence of yield switching in the population, we use 0.10 as the critical value for J-MAX.

The lower schedule in Table 1 presents the results of applying the jack-knife procedure to the two example farms. For the first farm, the LRT p-values remain low (0.000) when data from each of years 1, 2, 3, and 4 are sequentially deleted from the analysis. The maximum of these values (J-MAX = 0.000) is below 0.10 and we conclude that the first farm's low LRT p-value was not caused by a single year's yield events. The second farm's higher LRT p-value (0.522) and J-MAX (0.613) both indicate that we cannot reject the hypothesis that the second farm's yield variability is random.

The third set of yield data demonstrates the LRT model's sensitivity to a single low yield event. The third "farm's" yields are equivalent to the second farm's yields with the exception that $y_{1,2}$ has been reduced from 690 pounds to 250 pounds per acre. With the reduction of only one year's yields, the farm's LRT statistic increases from 2.06 to 11.4 reducing the p-value from .522 to .010. The jack-knifed p-values are 0.014, 0.330, 0.031, and .003 when years one, two, three and four are respectively deleted from the analysis giving a J-MAX of 0.330. Since J-MAX exceeds 0.10 we conclude that the original low p-value resulted from a single year event and not from a consistent cyclical pattern of yield switching.

Estimates of Yield Switching Incidence and Costs in the General Population

The results of applying the yield switching tests to 206,952 of barley, cotton, grain sorghum, corn, soybeans and wheat are summarized in tables 2 and 3. Due to space limitations, each table only presents results from states with 500 or more multiple unit producers reporting at least four years of non-replicated reported yields on each unit. Only years for which yields were reported on two or more units were included in the farm's analysis. The tables also present total values aggregated over all states with 100 or more multiple unit producers.

The first five columns list the state, the total number of screened farms, the number of farms "flagged" with a 1 % likelihood ratio test (LRT), the number of farms with J-MAX less than 0.10, and the proportion of farms flagged with the jack-knifed J-MAX statistic. As previously indicated, the J-MAX statistic flags substantially fewer farms than the 1% LRT as a result of the LRT's sensitivity to single year events. The proportion of farms flagged with the J-MAX statistic varies by state and crop with values as low as 2.3 percent in Nebraska grain sorghum to as high as 13.7 percent in Alabama cotton. The estimated incidence rate varies by crop from 3.4 percent in grain sorghum to 6.1 percent in cotton. The overall incidence rate across all six crops is estimated at 10256 of 206,952 producers or about 5 percent.

Figure 3 plots the reported yields from six producers flagged as statistically suspicious. The plots demonstrate the yield switching model's predicted oscillating patterns of high reported yields on some units with concurrently low yields on other units with the patterns being reversed in subsequent years. We visually examined the plots of all "J-MAX flagged" farms and usually observed visually apparent oscillatory patterns of concurrently high and low yields.

The statistical model identified farms with suspicious yield patterns that were unlikely to have occurred randomly given the local yield events reflected in C_t and Σ_C. Behaviorally, if the low J-MAX producers were enhancing indemnities by yield switching, a Rothschild-Stiglitz type model predicts that the low J-MAX producers will elect higher insurance coverage. To test this prediction, we pooled the producers by state and regressed their individual coverage elections[22] on J-MAX. The fifth column of tables 2 and 3 present the estimated regression parameter from the pooled regressions. For the larger states, the parameter estimates are usually negative and statistically significant indicating that lower J-MAX producers tend to elect higher coverage. Across all 67 state-crop combinations examined, 58 regression parameter estimates were negative with 35 significant at the 10% level or lower. Two of the nine positive estimates were statistically significant at the 10% level. The regression results support the hypothesis that the lower J-MAX producers are electing higher coverage, possibly in anticipation of higher indemnities from yield switching.

Estimates of the Costs of Yield Switching

The final two columns in tables 2 and 3 present our estimates of program costs of yield switching in the 1993-1997 period. We first downloaded 1993-1997 historical state-crop level liability and indemnification amounts from the RMA's online summary of business. However, RMA's summary of business does not provide coverage level and single-multiple unit breakdowns of the data. Using procedures similar to those described in Atwood, et.al., we simulated 1993-1997 coverage level, multiple unit, and flagged-nonflagged liability and loss shares using RMA's 1998 online yield histories[23]. Applying the estimated shares to the online -summary of business totals allowed us to estimate the total amount of liability for flagged versus non-flagged farms by coverage level. Multiplying this amount by the difference in simulated loss cost ratios (indemnities divided by liability exposure) of the flagged versus non-flagged farms and aggregating across coverage levels generates the values in tables 2 and 3.

The estimated five-year costs of detected yield switching totaled $ 121,720,000 for the states and crops examined. For individual crops, the estimated costs are $1,975,000 for barley, $ 16,116,000 in cotton, $556,000 for grain sorghum, $ 43,671,000 in corn, $29,774,000 for soybeans, and $ 29,619,000 for wheat. During this period estimated yield switching costs amount to about three percent of total indemnifications. This amount varies by crop (from .4 percent in grain sorghum to over five percent in soybeans) and by state (over eight percent in several soybean states).

The estimated costs of yield switching in tables 2 and 3 are lower than RMA's current surcharge (10%) associated with insuring optional units. However, the incidence and costs of yield switching as reported in tables 1 and 2 should be interpreted cautiously. A reviewer questioned why the incidence rates would differ in different locations. One possibility is that detection of yield switching fraud is more statistically difficult in areas, such as the high plains, with higher underlying yield variability. In these areas, estimated variances s_t^2 are proportionally higher - possibly resulting in fewer farms being flagged as statistically suspicious. In these areas it is likely that we have underestimated the incidence of yield switching.

Unpublished research by one of the authors (for the RMA) has also shown that different forms of apparent abuse are more common in some geographic regions than in others and also differ across time. For example, when examined intertemporally, it appears that once a given type of questionable conduct has gone undetected in a given year, the local incidence of the conduct increases in later years. The pattern of “learned” yield-switching conduct is especially evident in some of the southeastern cotton states where we estimate a higher yield-switching incidence. We suspect that, in other regions, different methods of “indemnification enhancement” are the method of choice resulting in lower estimates of yield switching incidence. It is probable that both yield switching and other types of fraud are more prevalent than indicated in tables 1 and 2. In this study we were not able to examine the extent of yield switching between singly insured units controlled by related parties nor were we able to examine the extent of yield switching for the large number of producers who fail to submit a complete and accurate historical yield history.

SUMMARY AND CONCLUSIONS

This paper demonstrated that multiple unit crop insurance provisions provide incentives for some producers to consider switching yields between separately insured tracts of land so as to generate larger insurance indemnifications. Statistical procedures were developed to identify producers whose reported yield patterns are consistent with the predictions of a model of yield-switching model. Estimates of yield switching incidence in the population range from two to over thirteen percent of the population in some states. In some states, yield-switching indemnifications were estimated to amount to as much as eight percent of total indemnifications received by all multiple unit producers.

There are problems with the interpretation of the results of this study. The statistical procedures used are not intended by the authors to be viewed as proof of fraudulent conduct on the part of any individual or group of insurees. Proof of illicit conduct requires detailed and careful on-site investigation. However, we believe that the procedures presented in this paper can provide a useful screening device by more efficiently focusing the efforts of crop adjustors and agency compliance personnel on the smaller subset of the population who are more likely to have participated in suspicious conduct.

The procedures presented in this paper do, however, provide evidence that an abnormally high number of producers report yields that are consistent with yield switching. As there are undoubtedly other types of undetected fraudulent activities present in the insured population, the results presented in this paper could be viewed as lower bound estimates of the prevalence and cost of crop insurance fraud.

ENDNOTES

[1] The RMA periodically adjusts premium rates to incorporate recent loss information. As a result fraudulent indemnifications eventually lead to higher net premiums for producers as well as higher taxpayer subsidization costs.

[2] Chiappori and Dionne state that differentiating between moral hazard and adverse selection is difficult but may be facilitated with the identification of moral hazard conduct consistent with testable predictions from dynamic economic models. In following paragraphs, we present an intuitive discussion of a dynamic economic model that predicts cyclical patterns of reported yields generated by yield switching fraud. The model is similar to Vercammen and van Kooten's model of moral hazard cycles. Space prevents a detailed discussion of the model. A more thorough discussion of the model and it's predictions is available from the authors

[3] The report estimated that taxpayer savings might have been as high as $300 million in 1991 if producers had been required to jointly insure all separately insured tracts of land.

[4] It is expected that the LCR's of larger jointly insured multiple unit farms would be smaller than if the units were insured independently as optional units (see Kuhling). However, both Knight and Coble and Kuhling found higher LCR's for individual optionally insured units as contrasted to the LCR's of similar sized basic units insured by smaller producers. The reasons why LCR's of individual optional units should be higher than the LCR's of similar smaller farm basic units are not clear. One possibility is that some optional unit farms are abusing optional unit provisions. If the RMA increases premium charges sufficiently to cover the average cost of yield switching or other abuses of optional unit provisions, one might argue that the problem has been adequately addressed. However such an argument ignores two important issues. The first is that fraud increases premiums for all honest producers. Thus honest producers who desire to separately insure their units are forced to subsidize dishonest producers through higher premiums. The second issue is that the higher optional unit premiums receive public subsidization at the same rate as producers who insure under a single ownership structure.

[5] The proportion K is selected at the beginning of the year and is common across all U units.

[6] The producer may commit a different type of fraud by hiding production from the adjustor. Identifying this type of fraud in beyond the scope of this paper. This paper examines the potential to enhance indemnifications by the manipulation of the manner in which total reported production ($\sum_u a_u \; y_{u,t}$) is reported. All following discussions involve attempting to identify whether certain patterns in reported yields $y_{u,t}$ are consistent with yield switching.

[7] If the producer reports all production to the agency, $z_{u,t}$ is the realized per acre yield from unit u in year t. If the producer hides production from the adjuster (see footnote 6), $z_{u,t}$ is the amount of production from unit u that is included in the total reported production.

[8] Atwood, Watts and Shaik term this practice “yield switching”. Discussions with producers indicate that production is sometimes physically shifted between storage facilities on different tracts of land giving rise to a local producer expression “bin arbitrage”.

[9] The simulations are available from the authors. Stochastic dynamic programming (SDP) can be used to derive globally optimal solutions if the APH state evolution is modeled as an AR(1) process (as in VerCammen and van Kooten) rather than the MA(T) process. With SDP, the optimal yield switching policies involve switching from high to low APH units with the severity of the yield switching decreasing if the probability of detection or the associated fines increase with the magnitude of yield switching.

[10] For each state, we estimated the amount by which the average unit residual correlation decreases with the number of units in the farm. This estimated reduction in correlation was incorporated into the error structure.

[11] When we use s_t^2 for yearly variances and $s_{u,v}$ for within-year unit covariances, we actually use $(X'\Sigma^{-1}X)$ and $(X'\Sigma^{-1}\underline{y})$ in system (12).

[12] We include only U-1 restrictions since using U restrictions in (13) with the T restrictions in (8) results in row dependencies in the resulting R matrix.

[13] We included only units with four or more years of non-replicated yield data. A number of producers reported four or more years of identical yields on two or more of their units. If a unit's yields were replicated, only the first unit's yield data was used in the analysis. A large number of producers did not

report at least four years of data on any of their separately insured units. RMA allows these producers to establish coverage using county-based T-yields.
All units reporting T-yields were excluded from the analysis.

[14] We assume that single unit farms do not commit yield-switching fraud and use them as a "control group" to estimate s_t^2 in the absence of yield switching. If there were fewer than 100 single unit farms in a given county, the geographically closest counties were pooled into the county's "county-region" until the "county-region" had 100 or more single unit farms. If a given county had 100 or more single unit farms, the county's "county-region" consists only of the given county.

[15] The 3-stage GLS procedures involved (i) individually estimating (17) will OLS for each single unit farm in the county-region, (ii) pooling residuals across all single unit farms (iii) calculating s_t^2 as the variance of year t residuals from the population of single unit farms reporting yields in year t, (iv) individually re-estimating (18) with GLS for all single unit farms in the county-region, (v) pooling the GLS residuals across all single unit farms and re-calculating s_t^2, and (vi) repeating steps (iv) and (v) for one additional stage.

[16] The number of years in the example has been reduced from figure 2 due to space limitations in the tables.

[17] The third column of yield data is derived from the second producer's yield data and is used to illustrate a potential Type-I error problem with the yield-switching model.

[18] The GLS model is less susceptible to low yield events if such events also occurred on a sufficient number of single unit farms so as to influence s_t^2. The following procedures were developed to address isolated low yield events that had negligible effects upon s_t^2.

[19] The Fortran code and the results from the Monte Carlo simulations are available from the authors.

[20] We were originally concerned with the implications of the jack-knife procedure with respect to restrictions (13). If restrictions (13) are satisfied over the time period of the entire sample, they will undoubtedly be violated in the jackknifed model when one of the years is deleted from the analysis. However the non-estimability status of restrictions (13) imply that the violating the restrictions does not affect the jackknifed model's SSE. Tests on the additional hypothesis restrictions that all jack-knifed $\hat{\Delta}_{u,t}$ estimates are simultaneously equal to zero are not affected by the underlying data's violation of (13).

[21] **The J-MAX "critical value" of 0.10 used in this paper is a level that gave acceptable Type I - Type II error rates in the Monte Carlo simulations**. The "critical value" for J-MAX is a statistic and **should not** be interpreted as a traditional p-value.

[22] Governmental subsidy policies complicate coverage election regression tests due to their distorting effects upon producer coverage elections. For the 1998 crop year, proportional subsidies were highest at the 65 percent coverage election and the majority of insured producers elected 65 percent coverage. The presence of different proportional subsidies by coverage level is likely to introduce "stickiness" in producer coverage choices thus reducing measurable producer responses as predicted by Rothschild-Stiglitz type models.

[23] Additional details are available from the authors.

FIGURE 1: TWO YEARS OF YIELD SWITCHING INCENTIVES WITH AND WITHOUT FRAUD PENALTIES

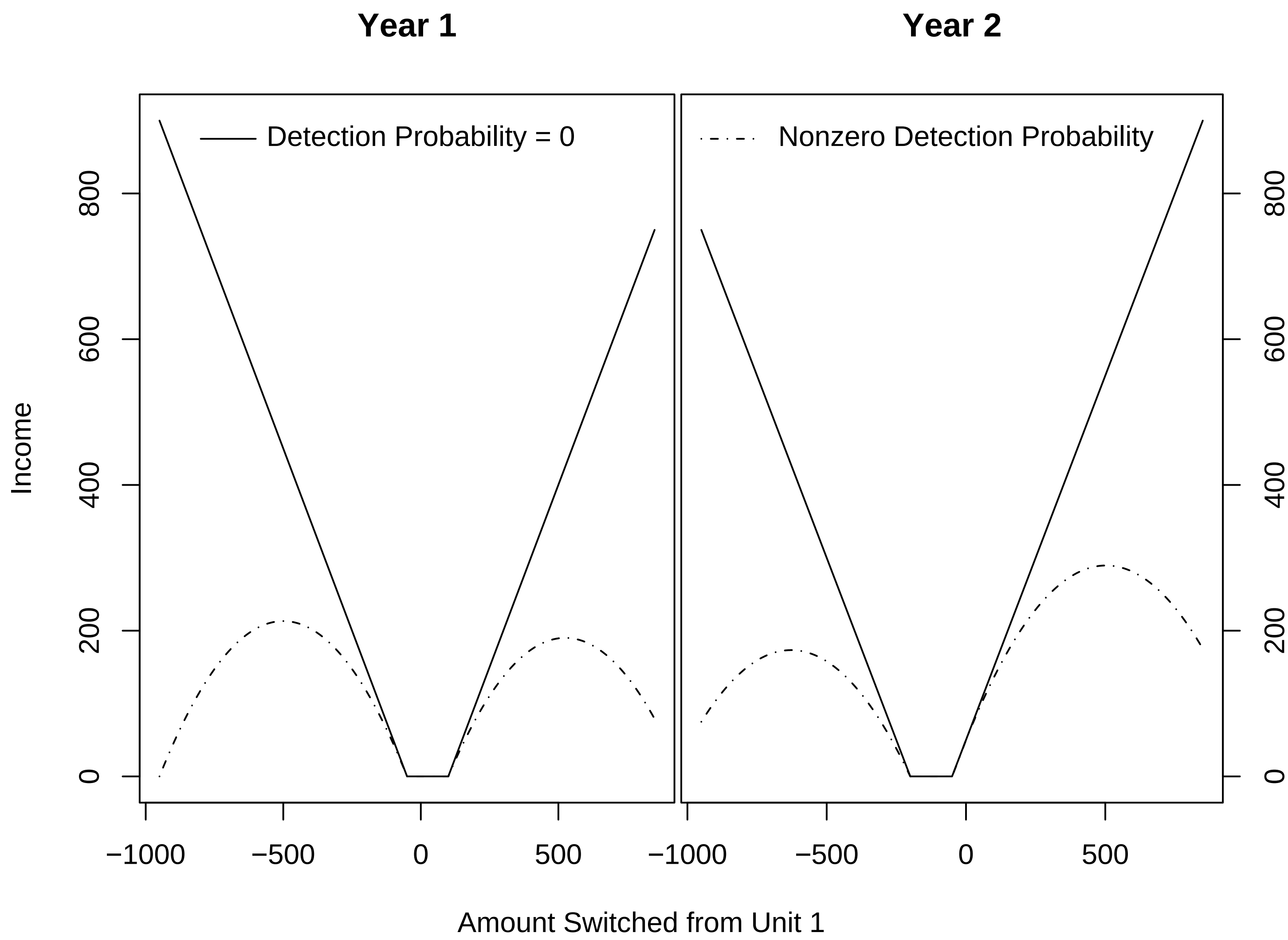

FIGURE 2 : PLOT OF REPORTED YIELDS FROM A TWO UNIT COTTON FARMER

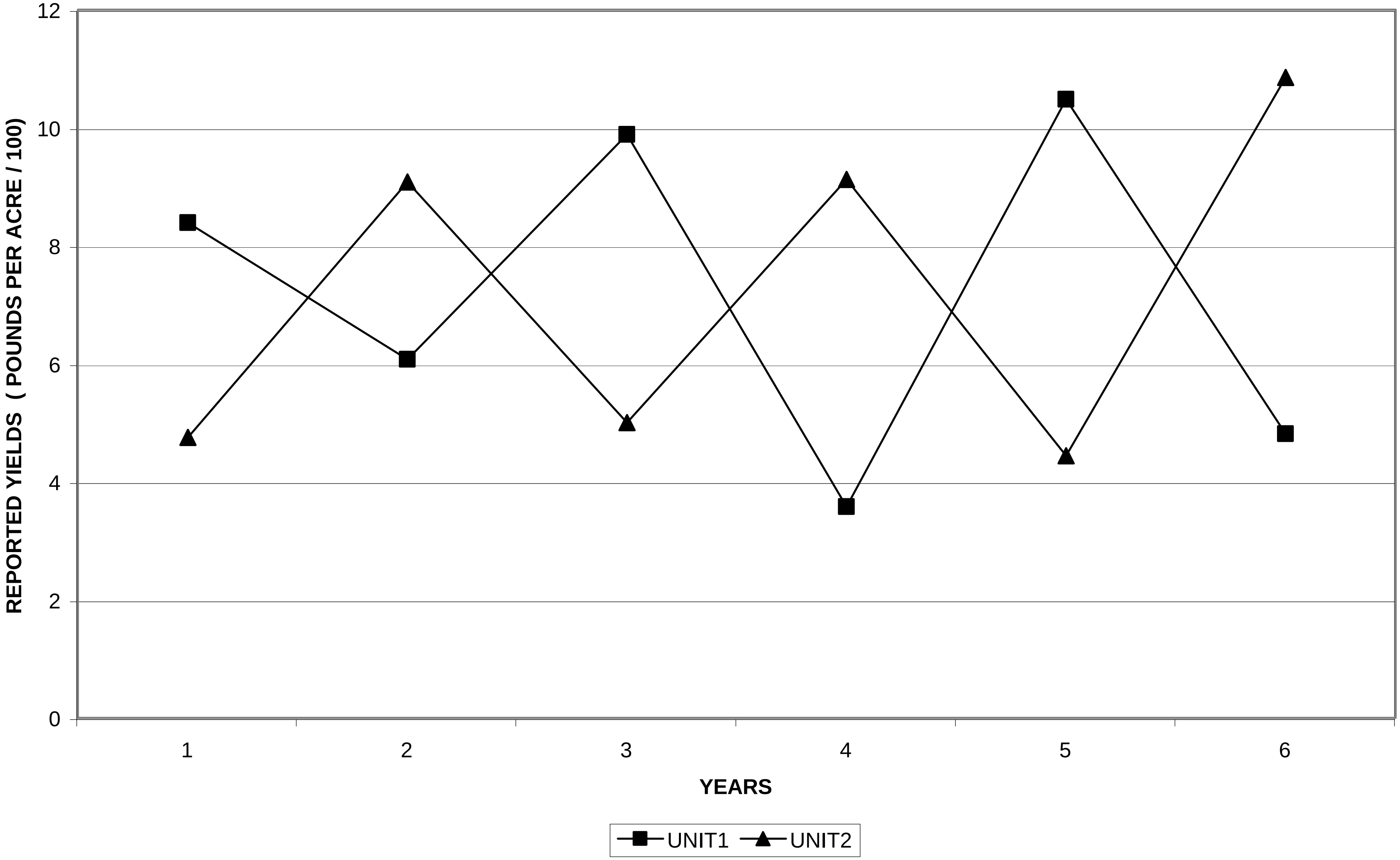

Figure 3: Example Plots of Farms With Small J-MAX Values

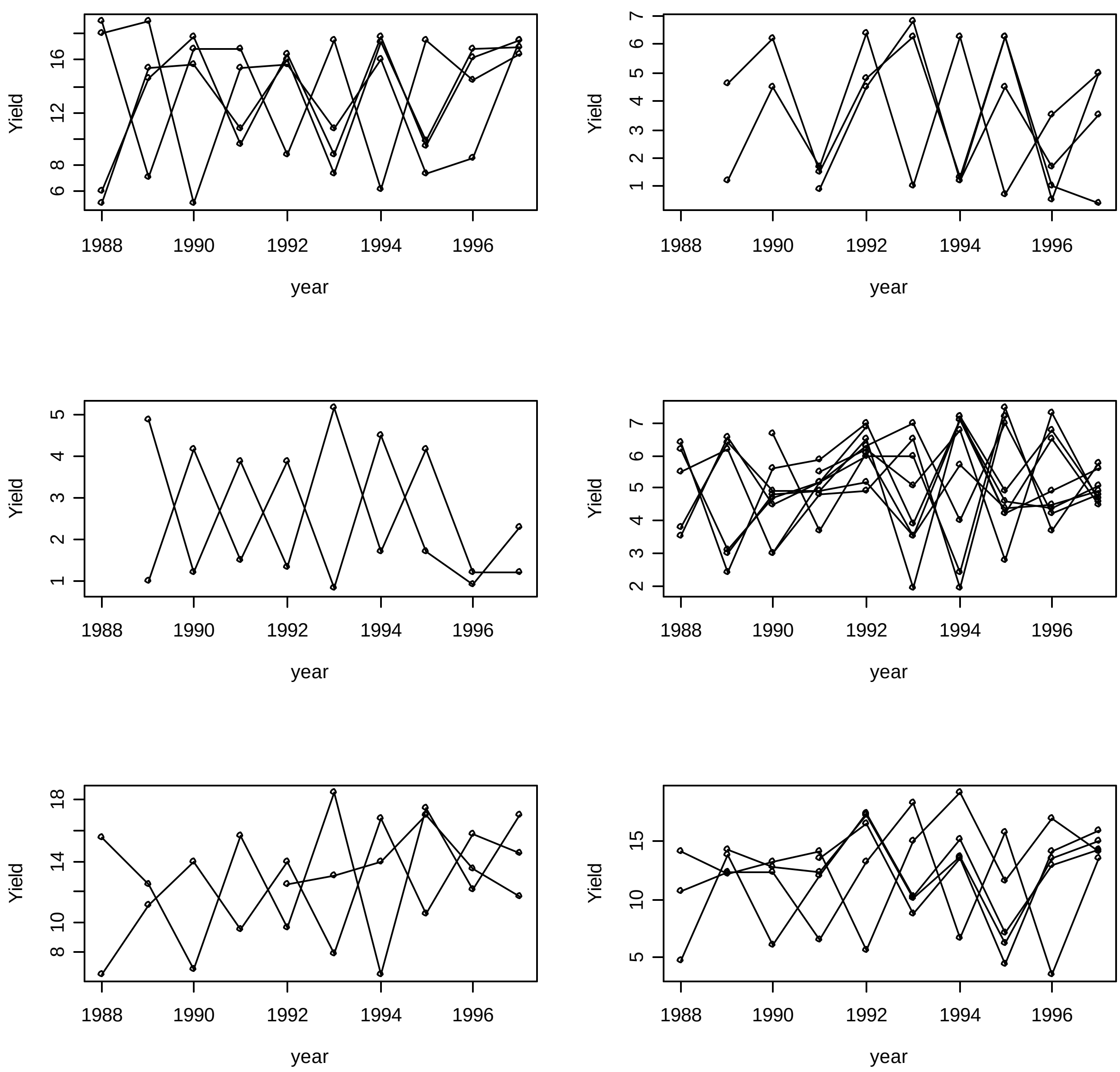

TABLE 1 : YIELD DATA, SIGMA MATRIX, SSE, LRT STATISTICS AND P-VALUES FOR EXAMPLE FARMS

	FARM YIELDS (100 POUNDS)			COUNTY YIELDS (100 POUNDS)	SIGMA MATRIX UNIT-YR							
UNIT-YR	FARM1	FARM2	FARM3		1-1	1-2	1-3	1-4	2-1	2-2	2-3	2-4
1-1	6.1	3.8	3.8	5	1.45	0	0	0	0.58	0	0	0
1-2	9.9	6.9	2.5	6.3	0	1.11	0	0	0	0.44	0	0
1-3	3.6	2.6	2.6	4.2	0	0	1.58	0	0	0	0.63	0
1-4	10.5	5.7	5.7	7.3	0	0	0	2.76	0	0	0	1.1
2-1	9.1	2.4	2.4	5	0.58	0	0	0	1.45	0	0	0
2-2	5	5.9	5.9	6.3	0	0.44	0	0	0	1.11	0	0
2-3	9.1	0.4	0.4	4.2	0	0	0.63	0	0	0	1.58	0
2-4	4.4	6.2	6.2	7.3	0	0	0	1.1	0	0	0	2.76

YEARS EXCLUDED	NSW MODEL(18) SCALED SSE FARM1	FARM2	FARM3	SW MODEL (14-16) SCALED SSE FARM1	FARM2	FARM3	LRT STATISTIC FARM1	FARM2	FARM3	LRT P-VALUES FARM1	FARM2	FARM3
FULL SAMPLE LRT RESULTS												
NONE [a]	51.24	8.57	13.56	4.57	6.32	2.16	46.67	2.25	11.4	0.000	0.522	0.010
JACK KNIFE TEST RESULTS												
1 [b]	36.76	7.38	10.66	2.45	5.32	2.13	34.31	2.06	8.53	0.000	0.357	0.014
2	34.98	4.35	4.35	4.05	2.13	2.13	30.93	2.22	2.22	0.000	0.330	0.330
3	24.79	3.69	7.56	3.39	2.71	0.58	21.4	0.98	6.98	0.000	0.613	0.031
4	41.45	7.27	12.53	2.53	6.31	1.18	38.92	0.96	11.35	0.000	0.619	0.003
									J-MAX =	0.000	0.619	0.330

[a] STATISTICS IN THIS ROW HAVE (8-2-4+1) OR 3 DEGREES OF FREEDOM

[b] STATISTICS WITH YEARS 1, 2, 3, OR 4 DELETED HAVE (6-2-3+1) OR 2 DEGREES OF FREEDOM

TABLE 2: SUMMARY RESULTS OF APPLYING THE STATISTICAL MODEL TO PRODUCERS OF BARLEY, COTTON, GRAIN SORGHUM, AND CORN

STATE	TOTAL # FARMS [1]	# FARMS 1% LRT P-VALUE	# FARMS J-MAX (< 0.10)	PROPORTION 10% PVALMAX FARM	REGRESS CVG ON J-MAX [2]	ESTIMATED YLD SWITCHING COSTS [3]	PROPORTION OF TOTAL INDEMNITIES
				BARLEY			
MT	609	61	39	0.064	-1.53 ***	219	0.025
ND	1800	177	88	0.049	-1.03**	1734	0.032
ALL [4]	2554	246	128	0.050		$1,975	0.026
				COTTON			
AL	526	97	72	0.137	-0.939	2268	0.029
TX	9739	972	572	0.059	-0.8471***	2156	0.004
ALL [4]	11376	1173	693	0.061		$16,116	0.019
				GR. SORGHUM			
KS	4460	228	131	0.029	-0.48***	321	0.013
NE	1418	61	33	0.023	-0.128	154	0.010
TX	1884	173	93	0.049	-1.37***	0	0.000
ALL [4]	7873	478	264	0.034		$566	0.004
				CORN			
CO	767	89	43	0.056	-1.12*	1015	0.048
IA	18946	2252	1258	0.066	-1.16***	10299	0.034
IL	11956	1182	677	0.057	-0.406***	3876	0.041
IN	4008	367	207	0.052	-0.702**	4517	0.064
KS	2589	217	88	0.034	-0.616**	1114	0.037
MN	9046	947	394	0.044	-0.688***	4241	0.020
MO	2091	202	120	0.057	1.61***	1421	0.027
NE	13474	1214	490	0.036	-0.944***	6969	0.058
OH	1975	127	52	0.026	-1.28***	1814	0.047
TX	1163	119	55	0.047	-0.646	1563	0.012
WI	1505	163	75	0.050	-1.23***	2512	0.050
SD	5858	464	222	0.038	-2.01***	1253	0.019
ALL [4]	75826	7583	3800	0.050		$43,671	0.034

[1] DUE TO SPACE LIMITATIONS THIS TABLE LISTS ONLY THOSE STATES FOR WHICH THERE WERE AT LEAST 500 MULTIPLE UNIT PRODUCERS WITH FOUR OR MORE YEARS OF VALID YIELD DATA.

[2] * DENOTES 10% SIGNIFICANCE, ** DENOTES 5% SIGNIFICANCE AND *** DENOTES 1% SIGNIFICANCE

[3] ESTIMATED COSTS IN $1,000 UNITS

TABLE 3: SUMMARY RESULTS OF APPLYING THE STATISTICAL MODEL TO PRODUCERS OF SOYBEANS AND WHEAT

STATE	TOTAL # FARMS [1]	# FARMS 1% LRT P-VALUE	# FARMS J-MAX (< 0.10)	PROPORTION 10% PVALMAX FARM	REGRESS CVG ON J-MAX [2]	ESTIMATED YLD SWITCHING COSTS [3]	PROPORTION OF TOTAL INDEMNITIES
				SOYBEANS			
IA	13618	1456	748	0.055	-0.976***	5,006	0.056
IL	8692	858	476	0.055	-0.146	1,948	0.070
IN	3006	275	141	0.047	-1.777***	2,291	0.093
KS	2722	163	84	0.031	-0.327	627	0.035
MN	9200	998	493	0.054	-0.804***	5,967	0.044
MO	2601	275	157	0.060	0.448	2,285	0.063
ND	718	72	46	0.064	-1.77***	1,005	0.043
NE	6452	489	244	0.038	-0.246	1,345	0.045
OH	2132	159	72	0.034	-0.476	1,585	0.081
SD	4988	376	169	0.034	-0.692***	4,179	0.051
ALL [4]	55596	5266	2710	0.049		$29,774	0.053
				WHEAT			
CO	2038	118	81	0.040	-0.25	427	0.010
IL	561	39	28	0.050	0.167	581	0.034
KS	16367	1362	783	0.048	-0.126	2,971	0.015
MN	1843	178	97	0.053	-1.12**	2,397	0.016
MT	4205	352	191	0.045	-0.599**	617	0.010
ND	14085	1469	802	0.057	-0.609***	13,804	0.049
NE	3331	188	90	0.027	-0.076	484	0.013
OH	540	39	11	0.020	-2.488***	12	0.002
OK	3749	348	209	0.056	-1.07***	2,810	0.023
SD	2286	187	105	0.046	-1.085***	5,036	0.051
TX	2092	225	144	0.069	-1.333***	0	0.000
WA	1121	84	54	0.048	-1.026*	26	0.003
ALL [4]	53727	4692	2661	0.050		$29,619	0.025

[1] DUE TO SPACE LIMITATIONS THIS TABLE LISTS ONLY THOSE STATES FOR WHICH THERE WERE AT LEAST 500 MULTIPLE UNIT PRODUCERS WITH FOUR OR MORE YEARS OF VALID YIELD DATA.

[2] * DENOTES 10% SIGNIFICANCE, ** DENOTES 5% SIGNIFICANCE AND *** DENOTES 1% SIGNIFICANCE

[3] ESTIMATED COSTS IN $1,000 UNITS

[4] THE TOTAL ROW INCLUDES THE SUMMARY RESULTS FROM ALL STATES WITH AT LEAST 100 MULTIPLE UNIT FARMS WITH VALID YIELD DATA.

APPENDIX

Table A1 presents the yield vectors and the X, $X'\Sigma^{-1}X$, $X'\Sigma^{-1}\underline{y}$, and R matrices from the example farms. FARM 1 and FARM 2 are reported yields from two actual cotton farms from a county in the southeastern U.S. FARM 3 is a construct with all yields (except $y_{1,t}$ = 2.5) set equal to those of FARM 2. Each farm has the same X and $X'\Sigma^{-1}X$. Column headers GAM $_u$, D $_{u,t}$, and LAMDA $_j$ respectively denote g_u, $\Delta_{u,t}$, and l_j from the main text.

Table A2 presents the LAUG^{-1} matrix partitioned as $LAUG^{-1} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}$ The lower section of table A2 lists the right-hand-side (RHS) vectors of the restricted normal equations (($X'\Sigma^{-1}\underline{y}$) appended with $\underline{d}_R$ from restrictions R $\underline{b} = \underline{d}_R$) and the corresponding solution vectors for each farm. The solution vectors are computed as:

$$\begin{bmatrix} \hat{\underline{b}} \\ \hline \hat{\underline{l}} \end{bmatrix} = LAUG^{-1} \begin{bmatrix} X'\Sigma^{-1}\underline{y} \\ \underline{d}_R \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} X'\Sigma^{-1}\underline{y} \\ \underline{d}_R \end{bmatrix} \qquad \text{(A-1)}$$

For the main body's yield switching tests (the lower left section of Table A2) $\underline{d}_R = \underline{0}$.

The lower right section of table A2 demonstrates that changing the RHS of the non-estimable restriction (r5) from 0 to 5 changes the estimates of $\hat{g}_u$ and $\hat{\Delta}_{u,t}$ but does not affect the lagrangians $\hat{\underline{l}}$ or the farm's SSE_{SW} (4.57). The non-zero lagrangians associated with restrictions r1-r4 indicate that changing any one or more of the first four d_R^j values in $\underline{d}_R$ may affect SSE_{SW}. Below we briefly present results indicating that each the first four restrictions in R can be expressed as a linear combination of a basis set of three linearly independent estimable functions and one non-estimable function. As such, the basis set for the estimable restrictions implied by R is of rank three.

Since changing the RHS on restriction (r5) changes individual $\hat{g}_u$ and $\hat{\Delta}_{u,t}$ estimates and does not change the predicted $\underline{\hat{y}}$ vector, we cannot have confidence in individual $\hat{\Delta}_{u,t}$ estimates unless we have sufficient external evidence to believe the original summing-across-years restriction is true. As we have no such evidence for any farm in our data set, individual $\hat{\Delta}_{u,t}$ remain non-estimable. However, this does not prevent us from testing whether the complete set of $\hat{\Delta}_{u,t}$ estimates are simultaneously zero (i.e. there is no yield switching on the farm).

Results available from the authors demonstrate that G_{22} contains a significant amount of information with respect to the estimability status of the restrictions in R. These results include: (i) even if the LAUG matrix is singular, a full row rank R implies G_{22} is symmetric and invariant for all generalized inverses of LAUG, (ii) if any row in G_{22} is zero, the corresponding lagrangian is zero and it's corresponding restriction is system non-estimable, and (iii) the rank of G_{22} equals the rank of a basis set for the estimable restriction space implied by R. The dimension less the rank of G_{22} equals the number of implicitly non-estimable restrictions in R.

In this example, the rank of the 5x5 matrix G_{22} is 3 implying that there are three intrinsically estimable and two intrinsically non-estimable restrictions in R. The number of intrinsically estimable restrictions for the main body’s system (14)-(16) can be shown to equal T-1. In the following we denote G_{22} from the partially restricted system (14)-(16) as G_{22}^{0}.

Testing the hypothesis that a farm is not yield switching involves testing whether all $\hat{\Delta}_{u,t}$ simultaneously equal zero. This can be accomplished with several methods. One method is to explicitly exclude all $\hat{\Delta}_{u,t}$ from the regression (the main body's expression (18)). A second, less numerically stable procedure is to construct a larger LAUG system with each $\hat{\Delta}_{u,t}$ explicitly restricted to zero. Although we do not recommend this method in practice, it has useful analytical properties. The G_{22} matrix of the larger LAUG system, denoted G_{22}^{*}, can be shown to be of rank $U\,T - U$ (8-2 or 6 in the example). The appropriate degrees of freedom for the LRT tests can be shown to equal the difference in the ranks of G_{22}^{*} and G_{22}^{0} or $UT - U - T + 1$. For this example, the degrees of freedom are (8 – 2 – 4 + 1) or 3.

Other results when additional hypothesis testing restrictions are imposed on a partially restricted LAUG system are available from the authors.

TABLE A1: DATA AND DESIGN MATRICES FOR THE EXAMPLE FARMS

DESIGN MATRIX / **FARM YIELDS**

UNIT-YR	GAM_1	GAM_2	D_{11}	D_{12}	D_{13}	D_{14}	D_{21}	D_{22}	D_{23}	D_{24}	FARM1	FARM2	FARM3
1-1	5	0	1	0	0	0	0	0	0	0	6.1	3.8	3.8
1-2	6.3	0	0	1	0	0	0	0	0	0	9.9	6.9	2.5
1-3	4.2	0	0	0	1	0	0	0	0	0	3.6	2.6	2.6
1-4	7.3	0	0	0	0	1	0	0	0	0	10.5	5.7	5.7
2-1	0	5	0	0	0	0	1	0	0	0	9.1	2.4	2.4
2-2	0	6.3	0	0	0	0	0	1	0	0	5	5.9	5.9
2-3	0	4.2	0	0	0	0	0	0	1	0	9.1	0.4	0.4
2-4	0	7.3	0	0	0	0	0	0	0	1	4.4	6.2	6.2

X'*SIGINV*X / **X'*SIGINV*Y**

GAM_1	GAM_2	D_{11}	D_{12}	D_{13}	D_{14}	D_{21}	D_{22}	D_{23}	D_{24}	FARM1	FARM2	FARM3
99.177	-39.468	4.105	6.734	3.161	3.144	-1.642	-2.669	-1.260	-1.253	90.828	60.240	30.611
-39.468	99.177	-1.642	-2.669	-1.260	-1.253	4.105	6.734	3.161	3.144	59.486	35.263	47.008
4.105	-1.642	0.821	0	0	0	-0.328	0	0	0	2.020	2.332	2.332
6.734	-2.669	0	1.069	0	0	0	-0.424	0	0	8.463	4.875	0.172
3.161	-1.260	0	0	0.753	0	0	0	-0.300	0	-0.021	1.837	1.837
3.144	-1.253	0	0	0	0.431	0	0	0	-0.172	3.767	1.391	1.391
-1.642	4.105	-0.328	0	0	0	0.821	0	0	0	5.468	0.722	0.722
-2.669	6.734	0	-0.424	0	0	0	1.069	0	0	1.150	3.383	5.247
-1.260	3.161	0	0	-0.300	0	0	0	0.753	0	5.768	-0.479	-0.479
-1.253	3.144	0	0	0	-0.172	0	0	0	0.431	0.093	1.692	1.692

R MATRIX / **RESTRICTIONS RHS**

RESTRICTION NUMBER	GAM_1	GAM_2	D_{11}	D_{12}	D_{13}	D_{14}	D_{21}	D_{22}	D_{23}	D_{24}	RHS		
r1	0	0	1	0	0	0	1	0	0	0	0	0	0
r2	0	0	0	1	0	0	0	1	0	0	0	0	0
r3	0	0	0	0	1	0	0	0	1	0	0	0	0
r4	0	0	0	0	0	1	0	0	0	1	0	0	0
r5	0	0	1	1	1	1	0	0	0	0	0	0	0

TABLE A2: LAUG-INVERSE MATRIX AND SOLUTIONS FOR THE EXAMPLE FARMS

LAUG-INVERSE MATRIX

G11 | G12

GAM_1	GAM_2	D_{11}	D_{12}	D_{13}	D_{14}	D_{21}	D_{22}	D_{23}	D_{24}	$LAMDA_1$	$LAMDA_2$	$LAMDA_3$	$LAMDA_4$	$LAMDA_5$
0.012	0.004	-0.001	-0.01	0.004	0.007	0.001	0.01	-0.004	-0.007	0.001	-0.012	0.006	0.006	-0.044
0.004	0.012	0.001	0.01	-0.004	-0.007	-0.001	-0.01	0.004	0.007	-0.043	-0.056	-0.038	-0.038	0.044
-0.001	0.001	0.344	-0.068	-0.1	-0.176	-0.344	0.068	0.1	0.176	0.39	-0.11	-0.11	-0.11	0.219
-0.01	0.01	-0.068	0.308	-0.087	-0.153	0.068	-0.308	0.087	0.153	-0.138	0.362	-0.138	-0.138	0.276
0.004	-0.004	-0.1	-0.087	0.37	-0.183	0.1	0.087	-0.37	0.183	-0.092	-0.092	0.408	-0.092	0.184
0.007	-0.007	-0.176	-0.153	-0.183	0.511	0.176	0.153	0.183	-0.511	-0.16	-0.16	-0.16	0.34	0.32
0.001	-0.001	-0.344	0.068	0.1	0.176	0.344	-0.068	-0.1	-0.176	0.61	0.11	0.11	0.11	-0.219
0.01	-0.01	0.068	-0.308	0.087	0.153	-0.068	0.308	-0.087	-0.153	0.138	0.638	0.138	0.138	-0.276
-0.004	0.004	0.1	0.087	-0.37	0.183	-0.1	-0.087	0.37	-0.183	0.092	0.092	0.592	0.092	-0.184
-0.007	0.007	0.176	0.153	0.183	-0.511	-0.176	-0.153	-0.183	0.511	0.16	0.16	0.16	0.66	-0.32
0.001	-0.043	0.39	-0.138	-0.092	-0.16	0.61	0.138	0.092	0.16	-0.196	0.084	0.039	0.039	0
-0.012	-0.056	-0.11	0.362	-0.092	-0.16	0.11	0.638	0.092	0.16	0.084	-0.184	0.065	0.064	0
0.006	-0.038	-0.11	-0.138	0.408	-0.16	0.11	0.138	0.592	0.16	0.039	0.065	-0.196	0.03	0
0.006	-0.038	-0.11	-0.138	-0.092	0.34	0.11	0.138	0.092	0.66	0.039	0.064	0.03	-0.1	0
-0.044	0.044	0.219	0.276	0.184	0.32	-0.219	-0.276	-0.184	-0.32	0	0	0	0	0

G21 | G22

GLS NORMAL EQUATION RHS VALUES AND RESULTS RESTRICTIONS (15 & 16)

X'*EPSINV*Y

	FARM1	FARM2	FARM3
	90.828	60.240	30.611
	59.486	35.263	47.008
	2.020	2.332	2.332
	8.463	4.875	0.172
	-0.021	1.837	1.837
	3.767	1.391	1.391
	5.468	0.722	0.722
	1.150	3.383	5.247
	5.768	-0.479	-0.479
	0.093	1.692	1.692
r1	0.000	0.000	0.000
r2	0.000	0.000	0.000
r3	0.000	0.000	0.000
r4	0.000	0.000	0.000
r5	0.000	0.000	0.000

SOLUTION VECTORS

	FARM1	FARM2	FARM3
GAM_1	1.31	0.89	0.64
GAM_2	1.2	0.71	0.66
D_{11}	-1.77	0.25	0.73
D_{12}	2.1	-0.07	-1.66
D_{13}	-2.98	0.72	1.13
D_{14}	2.65	-0.91	-0.2
D_{21}	1.77	-0.25	-0.73
D_{22}	-2.1	0.07	1.66
D_{23}	2.98	-0.72	-1.13
D_{24}	-2.65	0.91	0.2
$LAMDA_1$	0.64	-0.44	-0.07
$LAMDA_2$	-0.31	0.88	0.07
$LAMDA_3$	0.48	-0.84	-0.56
$LAMDA_4$	-0.45	0.03	0.31
$LAMDA_5$	0	0	0
SSE-SW	**4.57**	**6.32**	**2.16**

EFFECT OF CHANGING RHS ON EQ (16)

X'*EPSINV*Y

FARM1
90.828
59.486
2.020
8.463
-0.021
3.767
5.468
1.150
5.768
0.093
0.000
0.000
0.000
0.000
5.000

SOLUTION

	FARM1
GAM_1	1.09
GAM_2	1.42
D_{11}	-0.68
D_{12}	3.49
D_{13}	-2.06
D_{14}	4.25
D_{21}	0.68
D_{22}	-3.49
D_{23}	2.06
D_{24}	-4.25
$LAMDA_1$	0.64
$LAMDA_2$	-0.31
$LAMDA_3$	0.48
$LAMDA_4$	-0.45
$LAMDA_5$	0
SSE-SW	**4.57**

References

Artis,M., M. Ayuso, and M. Guille'n. "Modeling Different Types of Automobile Insurance Fraud Behavior in the Spanish Market." *Insurance: Math. and Econ.* 24(1999):67-81.

Atwood, J. A., M. J. Watts, and S. Shaik. " An Examination of Potential Fraud in Crop Insurance Due to Yield Switching between Optional Units in US Cotton". Staff Paper #2003-1. Department of Agricultural Economics and Economics Staff Paper Series. Montana State University. Bozeman, MT (2003)

Berger, R.L. and J.C. Hsu. "Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets." *Statistical Science.* 11(1996): 283-319.

Brockett, P. L., X. Xia, and R. Derrig. "Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud." *J. Risk and Insurance.* 65(1998):245-74.

Chambers, R.G. "Insurability and Moral Hazard in Agricultural Insurance Markets." *Amer. J. Agr. Econ.* 71(1989): 604-16.

Chiappori, P. "Econometric Models of Insurance under Asymmetric Information". *Handbook of Insurance: G. Dionne, Ed.* Boston, MA: Kluwer Academic Publishers, 2000. (pp. 365-393)

Chipman, J. S. "On Least Squares with Insufficient Observations." *J. Amer. Stat. Assoc.* 59(1964):1078-1111.

Coble, Keith., T.O. Knight, R.D.Pope, and J.R.Williams. "An Expected Indemnity Approach to the Measurement of Moral Hazard in Crop Insurance." *Amer. J. Agr. Econ.* 79(1997): 216-226.

Dionne, G. "The Empirical Measure of Information Problems with Emphasis on Insurance Fraud". *Handbook of Insurance: G. Dionne, Ed.* Boston, MA: Kluwer Academic Publishers, 2000. (pp. 395-419)

Goodwin, B. K. and V. H. Smith. *The Economics of Crop Insurance and Disaster Aid.* Washington D.C.: AEI Press, 1995.

Greene, W. H. and T. G. Seaks. "The Restricted Least Squares Estimator: A Pedagogical Note." *Review. Econ. and Stat.* 73(1991):563-67.

John, P. W. M., "Pseudo-inverses in the Analysis of Variance." *Annals Math. Stat.* 35(1964):895-96.

Just, R. E., L. Calvin, and J. Quiggen. "Adverse Selection in Crop Insurance: Actuarial and Asymmetric Information Incentives." Am. J. Agr. Econ. 81(1999):834-49.

Knight, T. O., and K. H. Coble. "Actuarial Effects of Unit Structure in the U.S. Actual Production History Crop Insurance Program." *Journal of Agricultural and Applied Economics*. 31 (1999): 519-535.

Kuhling, J. "The Effects of Optional Units on Crop Insurance Indemnity Payments". *Master Thesis*, Montana State University, January, 2002.

Makki, S. and A Somwaru. "Evidence of Adverse Selection in Crop Insurance Markets." *J. Risk and Insurance.* 68(2001): 685-708.

Picard, P. "Auditing Claims in the Insurance Market with Fraud: The Credibility Issus." J. Public Econ. 63(1996):27-56.

Picard, P. "Economic Analysis of Insurance Fraud". *Handbook of Insurance: G. Dionne, Ed.* Boston, MA: Kluwer Academic Publishers, 2000 (pp 315-362)

Quiggin, J., G. Karagiannis, and J. Stanton. "Crop Insurance and Crop Production: An Empirical Study of Moral Hazard and Adverse Selection." *Australian J. Agr. Econ.* 37(1993): 95-113.

Rao, C. R. "A note on a Generalized Inverse Matrix with Applications to Problems in Mathematical Statistics." *J. Royal Stat. Soc.* 24(1962):152-58.

Searle, S. R. *Linear Models.* New York, NY: John Wiley and Sons, 1971

Searle, S. R. *Matrix Algebra Useful for Statistics.* New York, NY: John Wiley and Sons, 1982.

Skees, J., and M. Reed. "Rate Making for Farm Level Crop Insurance: Implications for Adverse Selection." *Amer. J. Agr. Econ.* 68(1986): 653-659.

Smith, V.H., and B.K. Goodwin. "Crop Insurance, Moral Hazard and Agricultural Chemical Use." *Amer. J. Agr. Econ.*. 78(1996): 428-38.

Suits, D. B. "Dummy Variables: Mechanics V. Interpretation." Review of Economics and Statistics. 66(1984): 177-180.

U. S. Department of Agriculture, Office of the Inspector General. "Report to the Secretary on Federal Crop Insurance Reform Revised as of April 19, 1999, Audit Report No. 05801-2-At, 1999.

U. S. General Accounting Office. "Crop Insurance: USDA Needs a Better Estimate of Improper Payments to Strengthen Controls over Claims, GAO/RCED-99-266, September, 1999.

Vercammen, J., and G. Cornelis van Kooten. "Moral Hazard Cycles in Individual-Coverage Crop Insurance." *Amer. J. Agr. Econ.*. 76(1994): 250-61.