Unbalanced Nested Component Error Model for Estimating Pest Damage Functions

Juan Yang[1]

and

Paul D. Mitchell[2]

May 2004

Abstract prepared for presentation as a Select Paper at the American Agricultural Economics Association Annual Meetings in Denver, CO

Short Abstract

A recently developed nested error component model for unbalanced panel data is used to estimate insect damage functions. The model estimates the separate random effects for location and year on the variability of yield loss and has smaller standard errors for the regression coefficient than the comparable OLS model.

[1]Ph.D. student in Department of Agricultural Economics, Tom Slick Senior Graduate Fellow at Texas A&M University, 333 Blocker Building, College Station, TX 77843-2124, voice (979) 845-5946, and email yangjuan@tamu.edu

[2]Assistant Professor, Department of Agricultural Economics, Texas A&M University, 333 Blocker Building, College Station, TX 77843-2124, voice (979) 845-6322, fax (979) 862-1563, and e-mail p-mitchell@tamu.edu.

A wide variety of pest problems have been analyzed using the tools of economics. Examples include assessments of the economic impact of an invasive species, estimates of the value of transgenic crops, evaluations of pest eradication programs, impact assessments of pesticide bans, development of economic thresholds for IPM, and designing optimal strategies for managing insect resistance to pesticides (Carlson, Sappie and Hammig; Hurley, Babcock, and Hellmich; Mitchell, Gray and Steffey; Hurley, Mitchell, and Rice; Perrings, Williamson, and Dalmazzone).

Economic analyses of pest issues can use yield and pesticide application data to estimate production functions or pesticide demand using duality based methods (Lichtenberg and Zilberman; Saha, Shumway and Havennar). However, problems can occur if pest population data and/or pest damage data are not available (Norwood and Marra). Experimental data are another commonly utilized source of data to directly estimate a pest damage function that predicts yield loss as a function of pest population densities or measures of plant damage by the pest (Mitchell Gray and Steffey; Hurley, Mitchell, and Rice). This paper focuses on the use of experimental data to estimate pest damage functions.

Unbalanced panels are a common problem when estimating damage functions with experimental data. Field experiments typically use multiple replicates, but the experiments are often conducted in different locations, for more than one year, with different hybrids, different pesticides, or different management regimes (e.g. crop rotation or tillage). However, in field experiments, replicates are often lost, and locations, hybrids, pesticides, and management regimes change over the years of the project, so that the number of replications for each possible grouping variable (year, hybrid, locations, pesticide, etc.) for analysis using panel data methods is not equal. Recent advances in panel data methods have included development of estimators for unbalanced panels for nested random effects models. The purpose of this paper is to describe and illustrate the application of some of these unbalanced nested panel data estimators to estimate pest damage functions, and then demonstrate the statistical and economic weaknesses of using ordinary least squares (OLS) or some of the simpler analysis of variance (ANOVA) estimators.

This paper uses the nested error component model recently developed by Baltagi, Song, and Jung to estimate an insect damage function with unbalanced panel data. With unbalanced data, the OLS estimates of regression coefficients are still unbiased and consistent, but their standard errors are biased, which may lead to incorrect conclusions concerning their significance. The unbalanced nested composed error model improves the accuracy of the estimated standard errors for the regression coefficients and allows use of panel data methods to estimate the random effects of factors such as location and year on the estimated distribution of yield loss due to insect damage. Also, since the standard OLS regression model uses a single error term, it attributes all variability in yield loss from the pest, regardless of the source. The component error model uses a component error term to estimate the effect of location and year on the variability in yield loss due to the pest, separate from variability due to experimental errors, measurement errors, and similar effects.

Thus the unbalanced nested component error model has two important advantages. First, it allows the data to be unbalanced and accounts for obvious nesting structures, both of which commonly occur with agricultural field data. Second, it estimates the specific random effects attributable to experimental error and the nesting variables such as year, location, and hybrid. The first advantage improves the analysis of the pest's effect on mean yield loss, while the second advantage improves the analysis of the pest's effect on the variance of yield loss.

This second advantage of using the unbalanced, nested component error model is important and often missed by economists using panel data methods. Baltagi, Song, and Jung (p. 358) note that "[s]tatisticians and biometricians are more interested in the estimates of the variance components, … [e]conometricians, on the other hand, are ore interested in the regression coefficients." Because farmers and agricultural economists are interested in assessing risk and changes in risk, we believe that agricultural economists are interested in estimating both the regression coefficients and the variance components and hence should benefit from applying unbalanced nested component error model to analyze data from field experiments.

This paper first presents a general unbalanced nested random effects panel data model, and then describes four different estimators for the regression coefficients and error components based on the work of Baltagi, Song, and Jung. Next, an empirical application is presented that involves estimating a western corn rootworm damage function and the value of a new transgenic corn that controls this important corn pest.

Unbalanced Nested Component Error Model

Grouping variables for panel data analysis of field data from pest experiments are usually clear. For example, data can be grouped by year, location, crop, hybrid, pesticide, and similar. If the data can be grouped by more than one such index, the data are nested. For this description of the unbalanced nested error component model, we assume that the grouping variables are year t = 1 to T and location i = 1 to L, since these fit the data used for the empirical application. The unbalancedness of the data is reflected in the last index, the replication r = 1 to R_t $\forall t$, implying that for each year t, the number of replicates is R_t.[1]

Replication is part of standard experimental methods, but field experiments often do not have the same number of replications across years and locations. Observations are lost because of weather events, accidents, and similar factors, as well as changes in the availability of funding, land, chemicals, and other experimentally controlled factors determining the number of replicates, so that the data become unbalanced.

Finally, the total number of observations for this model is $N = L\sum_{t=1}^{T} R_t$. Note that the unbalancedness could be expressed equivalently as r = 1 to R_i $\forall i$, implying that for each location i, the number of replicates is R_i. In this case, $N = T\sum_{t=1}^{T} R_i = L\sum_{t=1}^{T} R_t$.

The standard OLS regression model for estimating a pest damage function is:

$$y_{tir} = x_{tir}'b + u_{tir}, \tag{1}$$

where y is yield loss, x is a K x 1 vector of regressors (e.g., pest population densities, pest damage measures), b is a K x 1 vector of regression coefficients to estimate, and u is the error term, assumed to be independent and identically distributed (iid) with mean zero and variance s_u^2. The OLS model aggregates all experimental errors into the single error term u and estimates its variance s_u^2.

[1] Baltagi, Song, and Jung develop estimation methods that allow unbalancedness in the index i as well (i.e., i = 1 to L_t for all t). However, we do not explore this extension here, since it does not occur for our data, but it is fairly straightforward.

The nested error component model is the same, except that it uses a component error term $u_{tir} = m_t + n_{ti} + e_{tir}$,

(2) $$y_{tir} = x_{tir}'b + m_t + n_{ti} + e_{tir}.$$

In this case, m_t is the t^{th} unobservable random year effect, n_{ti} is the unobservable nested random effect of the i^{th} location within t^{th} year, and e_{tir} is the random disturbance. Each component of the error term is assumed to be iid, with zero mean and respective variances s_m^2, s_n^2, and s_e^2. The nested error component model estimates the three variance components, but only s_e^2 is attributed to experimental errors.

Equation (2) is a random effects model because the fixed effect (within) estimator performs poorly when the ratio of either component error variance to the experimental error variance (s_m^2 / s_e^2, s_n^2 / s_e^2) is small. Moreover, Baltagi, Song, and Jung find that random effects analysis of variance (ANOVA) estimators perform well for estimating regression coefficients, and that random effects maximum likelihood methods perform best for estimating variance components and standard errors of regression coefficients. Therefore, we describe different ANOVA estimators and maximum likelihood estimation of the parameter vector b and the variance components s_m^2, s_n^2, and s_e^2, but first we reformulate the model for presentation.

In matrix notation, the standard OLS regression model in equation (1) is

(3) $$y = Xb + u,$$

where y is a N x 1 vector of yield losses, X is a N x K matrix of regressors, and u is a N x 1 vector of disturbances. Similarly, write the component error term for the nested error component model in equation (2) as:

(4) $u = Z_m m + Z_n n + e$,

where m is a T x 1 vector of year effects, n is TL x 1 vector of location effects for each year, and e is a N x 1 vector of errors for each replication within each year and location, i.e., $m¢ = (m_1, \ldots, m_T)$, $n¢ = (n_{11}, \ldots, n_{1T}, \ldots, n_{L1}, \ldots, n_{LT})$, and $e' = (e_{111}, \ldots, e_{1LR_1}, \ldots, e_{TLR_T})$. Also, $Z_m = diag(l_L \otimes l_{R_t})$ and $Z_n = diag(I_L \otimes l_{R_t})$, where l_L and l_{R_t} are L x 1 and R_t x 1 vectors of ones, I_L is a L x L identity matrix, $\otimes$ denotes the Kronecker product, and $diag(l_L \otimes l_{R_t})$ implies $diag(l_L \otimes l_{R_1}, \ldots, l_L \otimes l_{R_T})$.

With this reformulation, the disturbance variance-covariance matrix E(uu') is Ω = $s_m{}^2 Z_m Z_m' + s_v{}^2 Z_v Z_v' + s_e{}^2 diag(I_2 \otimes I_{R_t})$, or

(5) $\Omega = diag[s_m{}^2 (J_L \otimes J_{R_t}) + s_v{}^2 (I_L \otimes J_{R_t}) + s_e{}^2 (I_L \otimes I_{R_t})]$,

where $J_L = l_L l_L¢$ and $J_{R_t} = l_{R_t} l_{R_t}'$ are matrices with all elements equal to one and respective dimensions of L x L and R_t x R_t. Ω is a block diagonal matrix with the t^{th} block given by:

(6) $\Lambda_t = s_m{}^2 (J_L \otimes J_{R_t}) + s_V{}^2 (I_L \otimes J_{R_t}) + s_e{}^2 (I_L \otimes I_{R_t}) \ \forall\ t = 1$ to T.

Following Wansbeek and Kapteyn, decompose Λ_t as follows:

(7) $\Lambda_t = LR_t s_m{}^2 (\bar{J}_L \otimes \bar{J}_{R_t}) + R_t s_n{}^2 (I_L \otimes \bar{J}_{R_t}) + s_e{}^2 (I_L \otimes I_{R_t})$,

where $\bar{J}_L = J_L / L$, and $\bar{J}_{R_t} = J_{R_t} / R_t$. Substituting $E_L = I_L - \bar{J}_L$ and $E_{R_t} = I_{R_t} - \bar{J}_{R_t}$ into equation (7) and combining equivalent terms gives the following decomposition for Λ_t:

(8) $\Lambda_t = l_{1t} Q_{1t} + l_{2t} Q_{2t} + l_{3t} Q_{3t}$,

where $Q_{1t} = I_L \otimes E_{R_t}$, $Q_{2t} = E_L \otimes \bar{J}_{R_t}$, and $Q_{3t} = \bar{J}_L \otimes \bar{J}_{R_t}$ and

(9) $l_{1t} = s_e{}^2$, $l_{2t} = R_t s_v{}^2 + s_e{}^2$, $l_{3t} = LR_t s_m{}^2 + R_t s_v{}^2 + s_e{}^2$.

Furthermore, following Baltagi, Song, and Jung,

(10) $$Q_1 = diag(I_L \otimes E_{R_t}),\ Q_2 = diag(E_L \otimes \bar{J}_{R_t}),\ Q_3 = diag(\bar{J}_L \otimes \bar{J}_{R_t}).$$

The advantage of this decomposition is that

(11) $$\Lambda_t^{\ P} = l_{1t}^{\ P} Q_{1t} + l_{2t}^{\ P} Q_{2t} + l_{3t}^{\ P} Q_{3t},$$

where P is an arbitrary scalar, so that finally we can write:

(12) $$\Omega^{-1} = diag[\Lambda_t^{\ -1}] = diag[\,l_{1t}^{\ -1} Q_{1t} + l_{2t}^{\ -1} Q_{2t} + l_{3t}^{\ -1} Q_{3t}].$$

The OLS estimator $\hat{b}_{OLS} = (X'X)^{-1} X'y$ is still unbiased and consistent in unbalanced nested panel regression if the variance components are positive, but its standard errors are biased. For notation, define OLS residuals as $\hat{u}_{OLS} = y - X\hat{b}_{OLS}$.

The within (fixed effects) estimator can be obtained by pre-multiplying equation (3) by $Q_1 = diag(I_2 \otimes E_{R_t})$ and then applying OLS. Pre-multiplying by Q_1 removes m_t and n_{ti} whether they are fixed or random effects, since $Q_1 u = Q(Z_m m + Z_n n + e) = Qe$, so that $\tilde{b}_{wtn}$, the $K - 1$ vector of within coefficient estimates excluding the intercept, is

(13) $$\tilde{b}_{wtn} = (X_s' Q_1 X_s)^{-1} X_s' Q_1 y,$$

where X_s denotes the N x $K - 1$ matrix of regressors excluding the intercept. The within intercept estimate is $\tilde{a}_{wtn} = \bar{y} - \bar{X}_s \tilde{b}_{wtn}$, where the bar indicates averaging, and the within residuals are $\tilde{u}_{wtn} = y - \tilde{a}_{wtn} l_N - X_s \tilde{b}_{wtn}$, where l_N is a N x 1 vector of ones (Amemiya).

Unbalanced Nested Component Error Model Estimators

Following Baltagi, Yong, and Jung, this section reports the derivation of different estimators for the regression coefficients and the variance components, i.e., the parameter vector b and s_m^2, s_n^2, and s_e^2. First, three ANOVA estimators are reported, then maximum likelihood estimation. The ANOVA estimators are derived by equating sums of squared residuals to their expectations and then solving for the variance components. The estimators differ because each uses different residuals. Because each extends balanced panel ANOVA estimators to the unbalanced case, they are termed modified estimators. Regression coefficients are then estimated by GLS with these variance components estimates inserted into the variance-covariance matrix Ω. Maximum likelihood here assumes normality for the error components m_t, n_{ti}, and e_{tir} in equation (2). Nevertheless, solving first order conditions requires an interative numerical procedure. Baltagi, Song, and Jung report Monte Carlo results showing that these simple ANOVA estimators for the regression coefficients compare well with the more complicated maximum likelihood estimates, but perform poorly for estimating the variance components when the unbalancedness is severe.

Modified Wansbeek and Kapteyn Estimators

The modified Wansbeek and Kapteyn (WK) estimator uses the within residuals for the Q_1, Q_2, and Q_3 diagonal matrixes defined by equation (10), specifically

(15) $$q_1 = \tilde{u}_{wtn}'Q_1\tilde{u}_{wtn},\ q_2 = \tilde{u}_{wtn}'Q_2\tilde{u}_{wtn},\ q_3 = \tilde{u}_{wtn}'Q_3\tilde{u}_{wtn}.$$

Defining $R = \sum_{t=1}^{T} R_t$, the respective expected values of q_1, q_2, and q_3 are:

(16a) $$\mathrm{E}(q_1) = (N - TL - K + 1)s_e^2,$$

(16b) $$\mathrm{E}(q_2) = [TL - T + tr\{(X_s'Q_1X_s)^{-1}X_s'Q_2X_s\}]s_e^2 + (N - R)s_v^2,$$

(16c) $$\mathrm{E}(q_3) = [T - 1 + tr\{(X_s'Q_1X_s)^{-1}X_s'Q_3X_s\} - tr\{(X_s'Q_1X_s')^{-1}X_s'\bar{J}_mX_s\}]s_e^2$$

$$+[R - L\sum_t R_t^{\,2}/N]\sigma_v^{\,2} + [N - L^2 \textstyle\sum_t R_t^{\,2}/N]\sigma_m^{\,2}\ .$$

Equating the q_i in equation (15) to their expected values in equation (16) and solving for the variance components gives the following modified WK estimators:

(17a) $\sigma_e^{\,2} = \tilde{u}_{wtn}{}'Q_1\tilde{u}_{wtn}/(N - TL - K + 1)$

(17b) $\sigma_v^{\,2} = \left(\tilde{u}_{wtn}{}'Q_2\tilde{u}_{wtn} - [TL - T + tr\{(X_s{}'Q_1X_s)^{-1}(X_s{}'Q_2X_s)\}\sigma_e^{\,2}]\right)/(N-R)$

(17c) $\sigma_m^{\,2} = \left(\tilde{u}_{wtn}{}'Q_3\tilde{u}_{wtn} - [T - 1 + tr\{(X_s{}'Q_1X_s)^{-1}X_s{}'Q_3X_s\}\right.$

$$\left. - tr\{(X_s{}'Q_1X_s)^{-1}X_s{}'\bar{J}_mX_s\}]\sigma_e^{\,2} - [R - L\textstyle\sum_t R_t^{\,2}/N]\sigma_v^{\,2}\right)/\left(M - L^2\textstyle\sum_t R_t^{\,2}/N\right).$$

These variance components can be used with equation (12) to obtain Ω^{-1}, as well as

(18) $$\sigma_e\Omega^{-1/2} = diag[\left(\sigma_e^{\,2}/\lambda_{1t}\right)^{0.5}Q_{1t} + \left(\sigma_e^{\,2}/\lambda_{2t}\right)^{0.5}Q_{2t} + \left(\sigma_e^{\,2}/\lambda_{3t}\right)^{0.5}Q_{3t}]$$

$$= diag[I_L \otimes I_{R_t}] - diag[q_{1t}(I_L \otimes \bar{J}_{R_t})] - diag[q_{2t}(\bar{J}_L \otimes \bar{J}_{R_t}),$$

where $q_{1t} = 1 - \sigma_e/(\lambda_{2t})^{0.5}$, $q_{2t} = \sigma_e/(\lambda_{2t} - \lambda_{3t})^{0.5}$. To apply feasible GLS, multiply equation (3) by $\sigma_e\Omega^{-1/2}$ and run OLS on this transformed model. The variance of the estimated coefficients follows the GLS rule, so that $\mathrm{var}(\hat{\beta}_{GLS}) = (X'\Omega^{-1}X)^{-1}$.

Modified Swamy and Arora Estimators

The modified Swamy and Arora (SA) estimator uses three regressions to obtain residuals. Specifically, multiply equation (3) by Q_1 and run OLS to obtain residuals $\tilde{u}_1$. In the same manner, multiply by Q_2 to obtain residuals $\tilde{u}_2$ and multiply by Q_3 to obtain residuals $\tilde{u}_3$. Let $\tilde{q}_1 = \tilde{u}_1Q_1\tilde{u}_1$, $\tilde{q}_2 = \tilde{u}_2Q_2\tilde{u}_2$, and $\tilde{q}_3 = \tilde{u}_3Q_3\tilde{u}_3$. It can be shown that $\tilde{q}_1$ is the same as q_1 in equation (15), so that the expected value of $\tilde{q}_1$ is the same as in equation (16). The other expected values are:

(19a) $\quad \mathrm{E}(\tilde{q}_2) = (N - TL - K + 1)s_e^{\,2} + [N - R - tr\{(X_s{}'Z_vZ_v{}'Q_2X_s)(X_s{}'Q_2X_s)^{-1}\}]s_v^{\,2}$,

(19b) $\quad \mathrm{E}(\tilde{q}_3) = (TL - K)s_e^{\,2} + [R - tr\{(X_s{}'Z_vZ_v{}'Q_3X_s)(X_s{}'Q_3X_s)^{-1}\}]s_v^{\,2}$

$$+ [N - tr\{(X'Z_mZ_m{}'X)(X'Q_3X)^{-1}\}]s_m^{\,2}.$$

Equating the $\tilde{q}_i$ to their respective expected values and solving for the variance components gives the following modified SA estimators:

(20a) $\quad \tilde{s}_e^{\,2} = \tilde{u}_{wtn}{}'Q_1\tilde{u}_{wtn}/(N - TL - K + 1)$,

(20b) $\quad \tilde{s}_v^{\,2} = \dfrac{\tilde{u}_2{}'Q_2\tilde{u}_2 - (TL - T - K + 1)s_e^{\,2}}{N - R - tr\{(X_s{}'Z_vZ_v{}'Q_2X_s)(X_s{}'Q_2X_s)^{-1}\}}$,

(20c) $\quad \tilde{s}_m^{\,2} = \dfrac{\tilde{u}_3{}'Q_3\tilde{u}_3 - (T - K)s_e^{\,2} - [R - tr\{[X'Z_vZ_v{}'Q_3X)(X'Q_3X)^{-1}\}]\tilde{s}_v^{\,2}}{N - tr\{(X'Z_mZ_m{}'X)(X'Q_3X)^{-1}\}}$.

The same GLS procedure as for the WK estimators gives the regression coefficients.

Henderson and Fuller and Battese Estimators

Based on the extension of Henderson by Fuller and Battese (HFB), this estimator also uses three different residuals. First, use the within residuals to obtain $\tilde{q}_1^{\,*} = \tilde{u}_{wtn}{}'\tilde{u}_{wtn}$. Second, multiply equation (3) by (Q_1 + Q_2), run OLS, and collect residuals to obtain $\tilde{q}_2^{\,*} = \tilde{u}_2^{*}{}'\tilde{u}_2^{*}$. Third, use the standard OLS residuals to obtain $\tilde{q}_3^{*} = \hat{u}_{OLS}{}'\hat{u}_{OLS}$. The expected value of $\tilde{q}_1^{\,*}$ is the same as for the WK and SA estimators, while the expected values of $\tilde{q}_2^{\,*}$ and $\tilde{q}_3^{\,*}$ are:

(22a) $\quad \mathrm{E}(\tilde{q}_2^{\,*}) = s_e^{\,2}(N - T - K + 1)$

$$+ s_v^{\,2}\{N - R - tr[(X_s{}'Z_vZ_v{}'Q_2X_s)(X_s{}'(Q_1 + Q_2)X_s)^{-1}]\},$$

(22b) $\quad \mathrm{E}(\tilde{q}_3^{\,*}) = s_e^{\,2}(N - K) + s_v^{\,2}\{N - tr[(X'Z_vZ_v{}'X)(X'X)^{-1}]\}$

$$+ s_m^{\,2}\{N - tr[(X'Z_m Z_m'X)(X'X)^{-1}]\}\,.$$

Equating $\tilde{q}_1^{\,*}$, $\tilde{q}_2^{\,*}$ and $\tilde{q}_3^{\,*}$ to their expected values and solving for the variance components gives the following HFB estimators:

(23a) $\tilde{s}_e^{\,2} = \tilde{u}_{wtn}'Q_1\tilde{u}_{wtn}/(N - TL - K + 1)$,

(23b) $$\tilde{s}_v^{\,2} = \frac{\tilde{u}_2{*}'\tilde{u}_2{*} - (N - T - K + 1)\tilde{s}_e^{\,2}}{N - R - tr\{(X_s'Z_v Z_v'Q_2 X_s)[X_s'(Q_1 + Q_2)X_s]^{-1}\}},$$

(23c) $$\tilde{s}_m^{\,2} = \frac{\hat{u}_{OLS}'\hat{u}_{OLS} - (T - K)s_e^{\,2} - \{N - tr[(X'diag(I_L \otimes J_{R_t})X)(X'X)^{-1}]\}\tilde{s}_v^{\,2}}{N - tr\{(X'Z_m Z_m'X)(X'X)^{-1}\}}.$$

The GLS procedure used for the WK and SA estimators gives the regression coefficients.

Maximum Likelihood Estimation

Define $r_1 = s_m^{\,2}/s_e^{\,2}$, $r_2 = s_v^{\,2}/s_e^{\,2}$, and $\Omega = s_e^{\,2}\Sigma$. Rearranging equation (12) with these definitions gives $\Sigma = \Omega / s_e^{\,2} = diag[Q_{1t} + (R_t r_2 + 1)Q_{2t} + (2R_t r_1 + R_t r_2 + 1)Q_{3t}]$. Because Σ has the same arbitrary scalar as Ω,

(24) $$\Sigma^{-1} = diag[Q_{1t} + \frac{1}{(R_t r_2 + 1)}Q_{2t} + \frac{1}{(2R_t r_1 + R_t r_2 + 1)}Q_{3t}]\,.$$

After removing constants, the log-likelihood function is (Baltagi, Song, and Jung):

(25) $$\ln L(\cdot) = -\frac{N}{2}\ln s_e^{\,2} - \frac{1}{2}\sum\nolimits_t (LR_t r_1 + R_t r_2 + 1)$$

$$-\frac{L-1}{2}\sum\nolimits_t \ln(R_t r_2 + 1) - \frac{1}{2}u'\Sigma^{-1}u / 2s_e^{\,2}\,.$$

Solving the first order conditions for b and $s_e^{\,2}$ gives the following closed form solutions as functions of r_1 and r_2:

(26a) $$\hat{b}_{ML} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y\,,$$

(26b) $$\hat{s}_e^{\,2} = (y - X\hat{b}_{ML})'\Sigma^{-1}(y - X\hat{b}_{ML})/N\,.$$

The first order conditions for r_1 and r_2 give the following implicit definitions for r_1 and r_2, given b and $s_e^{\,2}$:

(27a) $$\frac{\partial \ln L(\cdot)}{\partial r_1} = -\frac{1}{2}tr(Z_m'\Sigma^{-1}Z_m) + \frac{1}{2s_e^{\,2}}(y - Xb)'\Sigma^{-1}Z_mZ_m'\Sigma^{-1}(y - Xb) = 0,$$

(27b) $$\frac{\partial \ln L(\cdot)}{\partial r_2} = -\frac{1}{2}tr(Z_v'\Sigma^{-1}Z_v) + \frac{1}{2s_e^{\,2}}(y - Xb)'\Sigma^{-1}Z_vZ_v'\Sigma^{-1}(y - Xb) = 0.$$

Because no analytical solution exists for equation (27), solving these first order conditions (26) and (27) requires a numerical iteration procedure. We summarize the Fisher scoring procedure described by Baltagi, Song, and Jung.

Beginning with initial values of $\hat{r}_1$ and $\hat{r}_2$ (the WK, SA, or HFB estimates are obvious choices, but other values can be used), calculate updated values as follows:

(28) $$\begin{bmatrix}\hat{r}_1\\ \hat{r}_2\end{bmatrix}_{j+1} = \begin{bmatrix}\hat{r}_1\\ \hat{r}_2\end{bmatrix}_{j} + \begin{bmatrix} \mathrm{E}\left[-\dfrac{\partial^2 \ln L(\cdot)}{\partial^2 r_1^{\,2}}\right] & \mathrm{E}\left[-\dfrac{\partial^2 \ln L(\cdot)}{\partial r_1 \partial r_2}\right] \\ \mathrm{E}\left[-\dfrac{\partial^2 \ln L(\cdot)}{\partial r_1 \partial r_2}\right] & \mathrm{E}\left[-\dfrac{\partial^2 \ln L(\cdot)}{\partial r_2^{\,2}}\right] \end{bmatrix}_j^{-1} \begin{bmatrix} \dfrac{\partial \ln L(\cdot)}{\partial r_1} \\ \dfrac{\partial \ln L(\cdot)}{\partial r_2} \end{bmatrix}_j .$$

Here the subscript j denotes the j^{th} iteration. Equation (27) gives the elements of the gradient vector for r_1 and r_2, using equation (26) to calculate $\hat{b}_{ML}$ and $\hat{s}_e^{\,2}$. The elements of the information matrix can be obtained as follows:

(29a) $$\mathrm{E}\left[-\frac{\partial^2 \ln L(\cdot)}{\partial^2 r_1^{\,2}}\right] = \frac{1}{2}\sum_t \frac{(2R_t)^2}{(1 + r_2R_t + 2r_1R_t)^2}\,,$$

(29b) $$\mathrm{E}\left[-\frac{\partial^2 \ln L(\cdot)}{\partial r_1 r_2}\right] = \frac{1}{2}\sum_t \frac{2R_t^{\,2}}{(1 + r_2R_t + 2r_1R_t)^2}\,,$$

(29c) $$E\left[-\frac{\partial^2 \ln L(\cdot)}{\partial^2 r_2^{\,2}}\right]=\frac{1}{2}\sum_t \frac{R_t^{\,2}}{(1+r_2 R_t)^2}+\frac{1}{2}\sum_t \frac{R_t^{\,2}}{(1+r_2 R_t+2r_1 R_t)^2}.$$

Iteration continues until the values of $\hat{r}_1$ and $\hat{r}_2$ converge, then the associated $\hat{b}_{ML}$ and $\hat{s}_e^{\,2}$ can be determined. The information matrix allows calculation of standard errors.

Empirical Application

As an empirical illustration of the differences between these different estimators for the unbalanced nested component error model, we estimate a pest damage function for the western corn rootworm. To illustrate the economic significance of these differences, we then use each estimator to assess the farmer value of the new Bt corn for corn rootworm.

Corn rootworms, a group of related insect species, are among the most economically important pests of corn in the United States, with yield losses and control costs estimated to exceed $1 billion annually (Metcalf). The most problematic species are typically the western and the northern corn rootworm, though other species are important in some areas. Corn rootworm larvae hatch in the soil during the spring and feed almost exclusively on corn roots. Larvae emerge from the soil as adults in summer and adult females lay eggs in the soil in the late summer to continue the cycle (Levine and Oloumi-Sadeghi 1991).

Larval feeding causes yield loss by disrupting several plant functions and making corn plants more likely to lodge (Gray and Steffey). Because corn rootworms typically lay eggs only in existing corn fields, crop rotation is an effective and widely used control strategy in much of the Corn Belt. For non-rotated corn, soil insecticides applied at planting to control larvae and aerial applications in summer to control adults are the most common control strategies (Gray, Steffey, and Oloumi-Sadeghi; USDA-NASS).

In recent years, two corn rootworm species have developed resistance to crop rotation as a control strategy. The western corn rootworm soybean variant lays eggs in corn and in other crops, especially soybeans (Levine and Oloumi-Sadeghi 1996; O'Neal, Gray, and Smyth; Levine et al. 2002). Where a corn-soybean rotation is common, eggs laid in soybean fields hatch in corn fields the next spring and larvae cause yield loss. The soybean variant first appeared in east-central Illinois and northwestern Indiana in the mid-1990's and has spread through the eastern Corn Belt (Onstad et al.). Northern corn rootworm have evolved extended diapause as an adaptation to two-year corn rotations (Krysan, Jackson, and Lew). Extended diapause eggs hatch after two winters, so that where a corn-soybean rotation is common, eggs laid in corn hatch when corn is again planted in a field. Extended diapause occurs in varying levels wherever northern corn rootworm are found, but is most prevalent in Iowa, Minnesota, and South Dakota.

Bt corn active against western and northern corn rootworm larvae was registered for sale during the 2003 crop year (CITATION). As with Bt corn active against European corn borer and other lepidopteran pests, only limited seed was available during the initial years of product sales. Because of the widespread prevalence of economic damage from corn rootworm and the success of other Bt corn products, sales are expected to grow. Additional demand is expected since a stacked variety of Bt corn active against both corn rootworm and lepidopteran pest has been registered and sales will occur during the 2004 crop year. However, as with all new technologies, its value is somewhat uncertain, especially during initial years of its availability. The value of Bt corn for controlling European corn borer and other lepidopteran pests has been studied (Hyde et al.; Hurley, Mitchell, and Rice). As a result, we estimate on the value of Bt corn active against corn rootworm as an illustration of the economic differences that result when using each of the previously described panel data estimators.

Estimation Data and Results

Data for estimation were from three years (1994-1996) of field experiments conducted in two locations in Illinois (Urbana and DeKalb) concerning the effect of corn rootworm on corn yield and the effectiveness of soil insecticides for controlling damage (Gray and Steffey). Whole plot treatments were 6-10 replicates each for several commonly grown hybrids. Sub-plot treatments were two rows treated with the soil insecticide Counter® (terbufos) and two untreated rows. Collected data included machine-harvested yield for each sub-plot and the average root rating for five plants in each sub-plot. The final data are 574 observations of the soil insecticide yield (Y_t) and average root rating (A_t) and the untreated control yield (Y_c) and average root rating (A_c).

Root ratings are commonly used to assess corn injury from corn rootworm because accurately measuring larval densities is difficult—the tiny larvae live underground and hundreds can infest a single plant. The root rating is an index of corn root injury based on the number of corn root nodes exhibiting feeding scars or completely destroyed by corn rootworm larval feeding. Though other root rating scales exist, the most widely used when the experiments were conducted was the 1 to 6 scale of Hills and Peters. The larger the root rating, the greater the damage—a 1 indicates no corn rootworm feeding injury and a 6 indicates three or more root nodes completely destroyed.

Following Mitchell, Gray, and Steffey, who used most of these same data for their analysis, the dependent variable for estimation is proportional yield loss $y = (Y_t - Y_c)/Y_t$ and the independent variable is the squared root rating difference $x = A_c - A_t$, which is always positive.[2] In terms of the unbalanced nested component error model in equation (2), the grouping variables are year t = 1 to 3 (so T = 3) and location i = 1 to 2 (so L = 2).[3] The number of replicates each year is 108, 113, and 56 for 1994 to 1996 respectively, so the unbalanced pattern is significant. Also, since only one regressor is used, X in equation (2) is the vector x and the parameter vector $\boldsymbol{b}$ consists of an intercept $\boldsymbol{b}_0$ and slope $\boldsymbol{b}_1$.

Table 1 reports estimation results for all previously described estimators. According to the standard errors, the OLS estimates for the intercept and the slope parameters are significant. However, since OLS ignores the year and location effects, the corresponding standard errors are biased, so that this conclusion concerning significance may be incorrect. Table 1 also shows the differences in the regression coefficients that exist between the OLS estimates and the other estimators. For example, the unbalanced nested panel data estimators all indicate that the intercept is not significant, opposite the conclusion based on the OLS estimate. An insignificant intercept makes biological sense, since when the squared root rating difference is zero, no difference between the damage measures exists and so the plots should have the same expected yield, which implies a zero intercept. In terms of the slope parameter, the unbalanced nested panel data estimators all (with the exception of the SA estimator) imply a value between 0.0.1 and 0.015, substantially less than the OLS estimate. This difference is substantial and implies a much greater proportional yield loss with the OLS estimate than with the unbalanced nested panel data estimators for the same squared root rating difference.

[2] Unlike Mitchell, Gray, and Steffey, who only used the data for one location (Urbana), we found the squared root rating difference provided a better fit when the data from DeKalb was also included.
[3] Because preliminary data analysis found no significant hybrid effect (also reported by Gray and Steffey), it was dropped from the nesting structure.

In terms of the estimated variability in proportional yield loss, the estimators vary substantially. Since the unbalanced nested panel data models assume independent errors for each component, the total variance of proportional yield loss is the sum of the three error component variances. Hence, relative to the OLS estimate of 0.036, only the maximum likelihood estimate of 0.299 is lower. The WK estimate of 0.039 is comparable, while the other estimators are much larger—0.0483 and 0.0770 for the SA and HFB estimators respectively. However, all panel data estimators all agree that the contribution of experimental errors to proportional yield loss is 0.022, the same as the within (fixed effects) estimators. Our results concerning the differences between the ANOVA estimators and the maximum likelihood estimate are consistent with the findings of Baltagi, Song, and Jung, since the unbalancedness of our panel is substantial. They conclude that the ANOVA estimators compare well with maximum likelihood estimators for the regression coefficients, but perform poorly for estimating the variance components when the unbalancedness is severe.

Conclusion

Our analysis concludes at this point, but more work is needed. To better understand the economic implications of the differences among these estimators, we will build a model of per acre farmer returns in order to estimate the value of the new Bt corn active against corn rootworm. We intend to build a hierarchical model similar to that developed by Mitchell, Gray and Steffey. Because they differ not only in terms of the mean effect of corn rootworm damage (the regression coefficients), but also in the terms of the variance effect (the error components), we expect substantial economic differences among the estimators.

However, before developing this empirical economic model, we want to better assess and finalize the unbalanced nested panel data model. Two issues remain to be addressed. First, because the logic of pest damage implies that a zero intercept is reasonable, and the panel data models support this conclusion, we want to impose this restriction on the estimation before conducting the economic analysis. Mitchell, Gray and Steffey impose this restriction on their analysis as well. Second, wew want to better asses the use of the squared root rating difference as the regressor. Mitchell, Gray and Steffey used many of the same data and found that the linear model fit better. We found that when the squared difference was included, the linear term was insignificant, and so we dropped the linear term. Perhaps when we drop the intercept as Mitchell, Gray and Steffey recommend, our analysis will be consistent with theirs.

References

Amemiya, T. "The Estimation of Variance in a Variance Components Model." *International Economic Review* 12(1971):1-13.

Antweiler, W. "Nested Random Effects Estimation in Unbalanced Panel Data." *Journal of Econometrics* 101(2001):295-313.

Baltagi, B. H., S. H. Song, and B. C. Jung. "The Unbalanced Nested Error Component Regression Model." *Journal of Econometrics* 101(2001):357-381.

Baltagi, B. H. *Econometric Analysis of Panel Data,* 2nd ed. New York: Wiley 2001.

Carlson, G. A., G. Sappie, and M. Hammig. "Economic Returns to Boll Weevil Eradication." Agricultural Economic Report No. 621, U.S. Department of Agriculture, Economic Research Service, Washington, DC, September 1989.

Fuller, W. A., and G. E. Battese. "Transformations for Estimation of Linear Models with Nested Error Structure." *Journal of the American Statistical Association* 68(1973):626-632.

Gray, M. E., and K. L. Steffey. "Corn Rootworm (Coleoptera: Chrysomelidae) Larval Injury and Root Compensation of 12 Maize Hybrids: An Assessment of the Economic Injury Index." *Journal of Economic Entomology* 91(1998):723-740.

Gray, M.E., K.L. Steffey, and H. Oloumi-Sadeghi. "Participatory On-Farm Research in Illinois Cornfields: An Evaluation of Established Soil Insecticide Rates and Prevalence of Corn Rootworm (Coleoptera: Chrysomelidae) Injury." *Journal of Economic Entomology* 86(1993):1473-1482.

Harville, D. A. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association* 72(1977):320-340.

Hills, T. M., and D. C. Peters. "A Method of Evaluating Postplanting Insecticide Treatments for Control of Western Corn Rootworm Larvae." *Journal of Economic Entomology* 64(1971):764-765.

Hurley, T. M., B. A. Babcock, R. L. Hellmich. "Bt Corn and Insect Resistance: An Economic Assessment of Refuges." *Journal of Agricultural and Resource Economics* 26(July 2001):176-194.

Hurley, T. M., P. D. Mitchell, and M. E. Rice. "Risk and the Value of Bt Corn." *American Journal of Agricultural Economics* 86(2004): 345-358.

Hyde, J., M. A. Martin, P. V. Preckel, and C. R. Edwards. "The Economics of Bt Corn: Valuing Protection from the European Corn Borer." *Review of Agricultural Economics* 21(Fall/Winter 1999):442-454.

Krysan, J. L., J. J. Jackson, and A. C. Lew. "Field Termination of Egg Diapause in *Diabrotica* with New Evidence of Extended Diapause in *D. barberi* (Coleoptera: Chrysomelidae)." *Environmental Entomology* 13(1984):1237-1240.

Levine, E., and H. Oloumi-Sadeghi. "Management of Diabroticite Rootworms in Corn." *Annual Review of Entomology* 36(1991):229-255.

Levine, E., and H. Oloumi-Sadeghi. "Western Corn Rootworm (Coleoptera: Chrysomelidae) Larval Injury to Corn Grown for Seed Production Following Soybeans Grown for Seed Production." *Journal of Economic Entomology* 89(1996):1010-1016.

Levine, E., J. L. Spencer, S. A. Isard, D. W. Onstad, and M. E. Gray. "Adaptation of the Western Corn Rootworm to Crop Rotation: Evolution of a New Strain in Response to a Management Practice." *American Entomologist* 48(Summer 2002):64-107.

Lichtenberg, E. and D. Zilberman. "The Econometrics of Damage Control: Why Specification Matters." *American Journal of Agricultural Economics* 68(May 1986): 261-273.

Metcalf, R. L. "Forward." *Methods for the Study of Pest Diabrotica*. J. L. Krysan and T. A. Miller, eds. New York: Springer-Verlag, 1986.

Mitchell, P. D., M. E. Gray, and K. L. Steffey. "A Composed-Error Model for Estimating Pest-Damage Functions and the Impact of the Western Corn Rootworm Soybean Variant in Illinois." *American Journal of Agricultural Economics* 86(2004):332-344.

Norwood, F. B., and M. C. Marra. "Pesticide Productivity: Of Bugs and Biases." *Journal of Agricultural and Resource Economics* 28(2003):596-510.

O'Neal, M.E., M.E. Gray, and C.A. Smyth. "Population Characteristics of a Western Corn Rootworm (Coleoptera: Chrysomelidae) Strain in East-Central Illinois Corn and Soybean Fields." *Journal of Economic Entomology* 92(1999):1301-1310.

O'Neal, M.E., M.E. Gray, S. Ratcliffe, and K.L. Steffey. "Predicting Western Corn Rootworm (Coleoptera: Chrysomelidae) Larval Injury to Rotated Corn with Pherocon AM Traps in Soybeans." *Journal of Economic Entomology* 94(2001):98-105.

Onstad, D.W., M. Joselyn, S. Isard, E. Levine, J. Spencer, L. Bledsoe, C. Edwards, C. Di Fonzo, and H. Wilson. "Modeling the Spread of Western Corn Rootworm

(Coeloptera: Chrysomelidae) Populations Adapting to Soybean-Corn Rotation." *Environmental Entomology* 28(1999):188-194.

Perrings, C., M. Williamson, and S. Dalmazzone, eds. *The Economics of Biological Invasions*. Cheltenham, UK: Edward Elgar, 2000.

Saha, A., C. R. Shumway, and A. Havennar. "The Economics and Econometrics of Damage Control." *American Journal of Agricultural Economics* 79(August 1997):773-785.

Swamy, V. B. and A. A. Arora. "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models." *Econometrica* 40(1972):261-275.

United States Department of Agriculture, National Agricultural Statistics Service (USDA-NASS). Agricultural Chemical Usage: 2001 Field Crops Summary. Washington, DC. jan.mannlib.cornell.edu/reports/nassr/other/pcu-bb/#field.

Wansbeek, T., and A. Kapteyn. "A Simple Way to Obtain the Spectral Decomposition of Variance Components Models for Balanced Data." *Communications in Statistics – Theory and Method* 11(1982):2105-2111.

Wansbeek, T., and A. Kapteyn. "A Note on Spectral Decomposition and Maximum Likelihood Estimation of ANOVA Models with Balanced Data." *Statistics and Probabilities Letters* 1(1983):213-215.

Wansbeek, T., and A. Kapteyn. "Estimation of the Error Components Model with Incomplete Panels." *Journal of Econometrics* 41(1989):341-361.

Table 1. Parameter estimates for western corn rootworm damage function for different unbalanced nested component error model estimators (standard errors in parenthesis).

Estimator	Intercept b_0	Slope b_1	Year Effect s_m^2	Location Effect s_v^2	Experimental Error s_e^2
Ordinary Least Squares (OLS)	0.087 (0.013)	0.029 (0.002)	--	--	0.036
Within (Fixed Effect) (WTN)	0.160 (--)	0.013 (0.003)	--	--	0.022
Wansbeek and Kapteyn (WK)	0.140 (0.071)	0.015 (0.003)	0.012	0.005	0.022
Swamy and Arora (SA)	0.113 (0.093)	0.022 (0.002)	0.026	0.0003	0.022
Henderson and Fuller and Battese (HFB)	0.146 (0.130)	0.014 (0.003)	0.046	0.009	0.022
Maximum Likelihood (ML)	0.145 (0.266)	0.015 (0.017)	0.0009	0.007	0.022