TESTS AND ESTIMATORS OF MULTIPLICATIVE MODELS FOR VARIETY TRIALS

P. L. Cornelius

J. Crossa

M. S. Seyedsadr

Recommended Citation

TESTS AND ESTIMATORS OF MULTIPLICATIVE MODELS FOR VARIETY TRIALS

P. L. Cornelius[1], J. Crossa[2], and M. S. Seyedsadr[3]

[1]Dept. of Agronomy, University of Kentucky,
Lexington, KY 40546-0091

[2]International Maize and Wheat Improvement Center (CIMMMYT),
Apdo. 6-641, 06600 Mexico, D.F., Mexico

[3]Biostatistics and Data Management, Bristol-Myers Squibb Co.,
5 Research Parkway, Wallingford, CT 06492-7600

Abstract. Some recently obtained results on cross validation, hypothesis test and estimation procedures for multiplicative models applied to multi-site crop variety trials are presented. The PRESS statistic is more sensitive to overfitting and choice of model form than data-splitting cross-validation. Because of their extreme liberality, Gollob F-tests should not be used to test multiplicative terms. F_{GH} tests effectively control Type I error, but are conservative for tests of terms for which the previous term is small. "Simulation tests" have greater power than F_{GH} tests, but still effectively control Type I error rates. Simulation results and cross validation in two examples suggest that BLUP-style shrinkage estimators of multiplicative terms produce fitted models with predictive value at least as good as the best truncated models and would eliminate the need for cross validation as a criterion for model choice. Shrinkage estimators of multiplicative models were better than BLUPs computed under the assumption of random unpatterened interaction in one example and were at least as good in the second example. Both were much better than empirical cell means in both examples. It is suggested that variety performance estimates derived from shrinkage estimators of multiplicative models should replace empirical cell means routinely reported in experiment station crop variety trial bulletins.

Key words: Multiplicative models, crop variety trials, PRESS statistics, simulation test, shrinkage estimators, genotype $\times$ environment interaction.

1 INTRODUCTION

In multi-site crop variety trials, multiplicative models (MMs) may be useful for studying yield response patterns and/or obtaining estimates of realized varietal response levels in specific environments which provide better estimates than are given by the empirical cell means. We suppose that there are g varieties evaluated with n replicates in each of e environments (sites). The empirical mean yield of the ith variety in the jth environment will be denoted $\bar{y}_{ij.}$. Potentially useful MM forms are:

Completely multiplicative model (COMM),

$$\bar{y}_{ij.} = \sum_k^t \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij.} \, ;$$

Shifted multiplicative model (SHMM),

$$\bar{y}_{ij.} = \beta + \sum_k^t \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij.} \, ;$$

Genotypes (varieties) regression model (GREG),

$$\bar{y}_{ij.} = \mu_i + \sum_k^t \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij.} \, ;$$

Environments regression model (EREG),

$$\bar{y}_{ij.} = \mu_j + \sum_k^t \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij.} \, ;$$

Additive main effects and multiplicative interaction (AMMI),

$$\bar{y}_{ij.} = \mu + \tau_i + \delta_j + \sum_k^t \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\epsilon}_{ij.} \, .$$

The λ_k are such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_t > 0$; α_{ik} and γ_{jk} satisfy orthonomality constraints $\sum_i \alpha_{ik} \alpha_{il} = \sum_j \gamma_{jk} \gamma_{jl} = 1$ if $k = l$, zero otherwise. The cell means of errors $\bar{\epsilon}_{ij.}$ are assumed $NID(0, \sigma^2/n)$. It will be useful in the sequel to define $\theta_k = n^{1/2} \lambda_k / \sigma$. A suffix appended to the acronym for a MM form will indicate the number of multiplicative terms, e.g., COMM1, COMM2, etc., and, for generality, COMMt.

Perhaps the earliest use of a multiplicative model for analysis of a crop variety trial was Fisher and Mackenzie's (1923) use of COMM1 for the analysis of a factorial arrangement of 12 potato varieties and six manurial treatments. Cornelius (1978) suggested COMM, GREG, EREG and AMMI as useful models for the analysis of variety trials arranged in lattice designs with another treatment factor imposed on the lattice replicates. SHMM has been used as a basis for methods

for grouping varieties or environments into groups without variety rank changes (Cornelius, et al., 1992; Crossa et al., 1993; Crossa and Cornelius, 1993; Cornelius et al., 1993).

Least squares estimates of multiplicative parameters in any of the models is by singular value decompostion (SVD) of the matrix $\boldsymbol{Z} = [z_{ij}]$ where z_{ij} is the residual of $\bar{y}_{ij.}$ after subtracting the estimates of the additive terms (Gabriel, 1978). Estimates of the additive terms in GREG, EREG and AMMI are the usual estimates, $\hat{\mu}_i = \bar{y}_{i..}$, $\hat{\mu}_j = \bar{y}_{.j.}$, $\hat{\mu} = \bar{y}_{...}$, $\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$ and $\hat{\delta}_j = \bar{y}_{.j.} - \bar{y}_{...}$. In SHMM, $\hat{\beta} = \bar{y}_{...} - \sum_k^t \hat{\lambda}_k \hat{\bar{\alpha}}_k \hat{\bar{\gamma}}_k$, where $\hat{\bar{\alpha}}_k = g^{-1} \sum_i \hat{\alpha}_{ik}$ and $\hat{\bar{\gamma}}_k = e^{-1} \sum_j \hat{\gamma}_{jk}$, which makes the matrix to be decomposed dependent on the result of the SVD. Newton-Raphson and EM-type iterative algorithms for SHMM estimation are described by Seyedsadr and Cornelius (1992).

GREG1 is essentially the equivalent of the widely used "stability analysis" model (Finlay and Wilkinson, 1963; Eberhart and Russell, 1966), but reported stability analyses have generally not used the least squares estimates of Gabriel, but, instead, have been done as variety regressions on the observed environment main effects (the $\hat{\delta}_j$ of AMMI) as described by Mandel (1961).

Our main consideration in this paper will be estimation (or prediction) of the true yield response in the cells. For this purpose, Gauch (1988) and Gauch and Zobel (1988) recommended using a truncated AMMI model, i.e., $t < p = \text{rank}(\boldsymbol{Z})$, retaining only as many multiplicative terms as could be shown by cross-validation to improve the model's predictive value. The data are split (randomly) into n_m replicates of each variety for modelling and n_v replicates for validation. AMMI1, AMMI2, etc., are fitted to the modelling data and their mean squared errors of prediction (expressed as its square root and denoted RMS PD) determined. Generally, ten such random data splittings are done and the results averaged. The model with smallest RMS PD is assumed optimal and fitted to the entire data set to obtain improved estimates. It is argued that this "predictive" assessment is a better strategy for model choice than "postdictive" assessment based on statistical tests (Gollob, 1968) of hypotheses H_{0k}: $\lambda_k = 0$, computed essentially like analysis of variance F-tests for linear models. In all published examples (e.g., Gauch, 1988; Gauch and Zobel, 1988; Crossa et al., 1990; Nachit et al., 1992; van Oosterom et al., 1993) such tests have found more significant terms than could be shown to be predictively useful.

In this paper we present results of our investigations of alternatives to current practice. In Section 2 we consider the "PRESS" statistic (Allen, 1971). In Section 3 we present simulation results showing the Gollob F-test to be extremely liberal and suggest procedures which do effectively control Type I error rates. Finally, in Section 4, we suggest shrinkage estimates of multiplicative models as improvements over truncated models.

2 CROSS VALIDATION

The objective of the cross validation is to choose the model which will most effectively combine the direct information given by the empirical mean of all obser-

vations in a cell with indirect information which can be extracted from the other cells to give the best estimate of the true response in that cell. The best estimator of the direct information should be the mean of all the observations in the cell. The PRESS statistic, obtained as $\sum\sum(\bar{y}_{ij} - \hat{y}_{[ij]})^2$ where $\hat{y}_{[ij]}$ is the prediction of $\bar{y}_{ij.}$ when the (ij)th cell is omitted, uses all of the information available from the other cells, and should most effectively identify the model which extracts the most indirect information from the other cells. However, when the model is fitted to the entire data set, the observed cell mean will contribute to the estimate only to the extent that it contributes to the effects retained in the chosen model (one cannot depend on getting an optimally weighted average of the two pieces of information).

The first and third authors have developed efficient Newton-Raphson algorithms for computing the PRESS statistic in MMs in complete two-way classifications. For the CIMMYT EVT16B data set (nine varieties, 20 environments; for the cell means, see Cornelius et al., 1992), Table 1 shows RMS PD from data splitting ($n_m = 3, n_v = 1$) and an adjusted RMS PD(PRESS) obtained as $[\text{PRESS}/ge + 3s^2/4]^{1/2}$, where s^2 is the pooled within site error mean square. The term in s^2 is an adjustment for the difference between variance of the validation data (cell means as opposed to individual observations) done to make results comparable to RMS PD from 3-1 data splitting. The adjustment may be underestimated, because it does not compensate for sacrificed control of differences among blocks in blocked designs which occurs with random data-splitting cross validation.

These data suggest that PRESS is much more sensitive than data splitting to an overfitted model, will more definitively differentiate between model forms and will not be less parsimonious. The model with smallest PRESS (SHMM1) predicted data in a deleted cell better than they were predicted by three replicates of data with all cells present. Many overfitted models gave extremely unreliable prediction of a missing (i.e., deleted) cell, a situation which can only get worse if there are multiple missing cells. If a MM is to be used to impute missing cells (Gauch and Zobel, 1990), one should be extremely careful to avoid overfitting. To be able to compute PRESS for incomplete data sets would be highly desirable. To develop an algorithm to do so is a future objective.

3 TESTS OF SIGNIFICANCE OF MULTIPLICATIVE TERMS

Tests of hypotheses H_{0k}: $\lambda_k = 0$ are based on sequential sums of squares "explained" by the multiplicative terms. For AMMI, GREG, EREG and COMM, sequential sums of squares are n times the eigenvalues $[\hat{\lambda}_k^2, k = 1, \ldots, \text{rank}(\boldsymbol{Z})]$ of $\boldsymbol{Z'Z}$ (or of $\boldsymbol{ZZ'}$) which are distributed as eigenvalues of a p-variate Wishart matrix with q df where p and q are the smaller and larger, respectively, of elements of the following couples: for AMMI, $(g - 1, e - 1)$; GREG, $(g, e - 1)$; EREG, $(g - 1, e)$; COMM, (g, e) (Johnson and Graybill, 1972). In SHMM, sequential sums of squares are given by $SS_k = n[\sum_{m=k}^p \hat{\lambda}_{m(k-1)}^2 - \sum_{m=k+1}^p \hat{\lambda}_{m(k)}^2]$ where $\hat{\lambda}_{m(k)}$ is the mth largest eigenvalue of $\boldsymbol{Z}$ when the fitted model contains k multiplicative terms.

Let s^2/n be an independent estimate of the variance of a cell mean with f df. The Gollob (1968) approximate F-test assumes $n\hat{\lambda}_k^2/\sigma^2$ is distributed as chi-square with df $= p + q + 1 - 2k$ (the Gollob analog for SHMM has df the same as COMM except that there is a loss of one df from the last term). Since eigenvalues of Wishart matrices do not have chi-square distributions, the assumption does not hold. Computer simulation of AMMI for nine varieties and 20 sites has shown Gollob tests intended to be at 0.05-level to be extremely liberal, e.g., Type I error rate of 66.2% for testing H_{01} and 17.6% for H_{02} when $\theta_1 = \theta_2 = 0$ (Cornelius, 1993), and 34.9% for H_{03} if $(\theta_1, \theta_2) = (10, 5)$ (Table 2). Clearly, the Gollob test cannot be recommended and its use should be discontinued.

Tests which effectively control Type I error rates are the F_{GH1}, F_{GH2} and "simulation" tests. Cornelius et al. (1992) developed F_{GH1} and F_{GH2} as modifications of tests which they denote as F_1 and F_2. F_2 is an approximation developed by Johnson (1976) for $k = 1$, viz., $F_2 = n\hat{\lambda}_k^2/u_{1k}s^2$, approximately distributed under H_0 as central F with $2u_{1k}^2/u_{2k}^2$ and f df, where $u_{1k} = E(n\hat{\lambda}_k^2/\sigma^2)$ and $u_{2k}^2 = V(n\hat{\lambda}_k^2/\sigma^2)$ given all $\lambda_k = 0$. To obtain F_1, put $X = n\hat{\lambda}_k^2/fs^2$ and find beta distribution parameters a and b such that under H_{0k}, to a two-moment approximation, $X + 1$ is distributed as the reciprocal of a beta random variable. Solutions for a and b are $a = 1 + (f-2)q_1/q_2$ and $b = u_{1k}q_1/q_2$ where $q_1 = u_{2k}^2 + u_{1k}^2 + (f-4)u_{1k}$ and $q_2 = (f-2)u_{2k}^2 + 2u_{1k}^2$. Then $F_1 = aX/b$ is distributed approximately as F with $2b$ and $2a$ df. In practice, P-values computed for F_1 and F_2 always agree very closely. For SHMM, SS_k is substituted for $n\hat{\lambda}_k^2$ in these test procedures.

Recent computer simulation results (unpublished) show that F_1 and F_2 computed under the complete null hypothesis (all $\lambda_k = 0$) are liberal tests for $k > 1$ in SHMM analysis, and this is probably also true for other MM forms. Theoretical results of Goodman and Haberman (1990), Marasinghe (1985) and Schott (1986) concerning properties of estimates, residuals and eigenvalues as one or more $\lambda_k \to \infty$ (with the other $\lambda_k = 0$) suggest, as conservative procedures, that for testing terms beyond the first, F_1 and F_2 be modified so as to make them equivalent to tests of the first term in a $(g - k + 1) \times (e - k + 1)$ table. We denote such tests as F_{GH1} and F_{GH2}. A worked example is given by Cornelius (1993).

Functions which may be used in computer programs to approximate values of u_{11} and u_{21} for the F_{GH} tests for $p - k + 1 \le 19$, $q - k + 1 \le 99$, are given by Seyedsadr and Cornelius (1991) for SHMM and Cornelius (1980) for the other MM forms. Tables of such values may be found in Mandel (1971) for model forms other than SHMM.

Cornelius (1993) found intended 0.05-level F_{GH} tests of H_{0k} have Type I error rates close to 0.05 if $\theta_{k-1} \ge 10$, but become conservative as θ_{k-1} decreases. For example, the Type I error rate for the F_{GH2} test of H_{03} in simulated AMMI analyses of 9×20 tables with $(\theta_1, \theta_2) = (10, 5)$ was 0.019 (Table 2).

Simulation and iterated simulation tests (Cornelius, 1993) use computer simulation to find improved values for u_{1k} and u_{2k} to use in F_2-type tests. Define u_{1k}^*, u_{1k} and u_{2k} such that $E(n\hat{\lambda}_k^2) = u_{1k}^*\sigma^2 + n\lambda_k^2$, $E(n\hat{\lambda}_k^2|\lambda_m = 0, m \ge k) = u_{1k}\sigma^2$, $V(n\hat{\lambda}_k^2|\lambda_m = 0, m \ge k) = u_{2k}^2\sigma^4$. The scheme is as follows.

1. Begin with an initial value of u_{1k}^* for $k = 1, \ldots, p$; we use Gollob's df for this.

2. Compute $F_k^* = n\hat{\lambda}_k^2 / u_{1k}^* s^2$.

3. Compute $\tilde{\lambda}_k = \hat{\lambda}_k[(F_k^* - 1)/F_k^*]^{1/2}$ provided $F_k^* > 1$, else $\tilde{\lambda}_k = 0$.

4. By simulation, determine u_{1k} and u_{2k} under the assumption $\theta_m = \tilde{\theta}_m = n^{1/2}\tilde{\lambda}_m/s$ if $m < k$, else $\theta_m = 0$, and u_{1k}^* under the assumption $\theta_k = \hat{\theta}_k$ for all k.

5. To test H_{0k}: $\lambda_k = 0$, use $F_k = n\hat{\lambda}_k^2 / u_{1k}s^2$.

If the $\tilde{\lambda}$ fail to give a non-increasing sequence, pool $\hat{\lambda}_k$ through $\hat{\lambda}_{k+m}$ by computing $F_{k \cdot m}^* = n \sum \hat{\lambda}_k^2 / \sum u_{1k}^* s^2$, where the sums are over the terms whose sums of squares are to be pooled, and put $\tilde{\lambda}_k = \tilde{\lambda}_{k+1} = \ldots = \tilde{\lambda}_{k+m} = [(m+1)^{-1} \sum \hat{\lambda}_k^2 (F_{k \cdot m}^* - 1)/F_{k \cdot m}^*]^{1/2}$.

Cornelius (1993) described a computational strategy for the simulation in Step 4 which exploits the fact that $\hat{\theta} = n^{1/2}\hat{\lambda}_k/\sigma$ can be simulated as the kth largest singular value of a $q \times p$ matrix $\boldsymbol{R} = [\text{diag}(\theta_k, k = 1, \ldots, p), \boldsymbol{0}_{p \times (q-p)}]' + \boldsymbol{E}$ where $\boldsymbol{E}$ is a matrix of random standard normal deviates (Gaussian "noise"). Thus, one can begin by generating $\boldsymbol{R} = [r_{ij}] = \boldsymbol{E}$, obtain its SVD to give a simulated result under the scenario that all $\lambda_k = 0$, add $\tilde{\theta}_1$ to r_{11}, do the SVD again to obtain a result under the scenario $\theta_1 = \tilde{\theta}_1$ (i.e., $\lambda_1 = \tilde{\lambda}_1$), then add $\tilde{\theta}_2$ to r_{22}, again obtain the SVD, etc. Once this has been done for all $\tilde{\theta}_k > 0$ the SVD gives simulated $\hat{\theta}_k$ values under the scenario that $\theta_k = \tilde{\theta}_k$ for all k.

After completing the tests at Step 5, the iterated simulation test is done by returning to Step 2 using the u_{1k}^* values from the simulation. We recommend 1000 simulations, but, in this paper, we present simulated results of simulation tests which used only 100 simulations for each test of 1000 simulated tables tested. Type I error rates and power of test do not change much with increases in number of simulations per test beyond 100, but increasing the number of simulations per test will make individual applications of the tests less subject to simulation error. Simulation tests have not been developed for SHMM.

The simulation and iterated simulation tests are less conservative and have greater power than the F_{GH2} test without seriously sacrificing Type I error rates (Table 2; see also Cornelius, 1993). Simulation results for a case where all $\theta_k = 5$ in 9×20 tables indicate that one might actually detect all eight terms if the simulation tests are used.

Since it will generally be more parsimonious than the Gollob test, the F_{GH2} test is a viable alternative to cross validation as a criterion for model choice. For AMMI with balanced data, using data-splitting cross validation, the expected mean squared error of prediction is equal to the sum of (1) the variances of the estimates of the mean and additive main effects, (2) the mean squared error of the estimated interaction and (3) the error variance in the validation data. A change in the number of terms retained will change only the second component. Thus, the predictively

best truncated AMMI model is the one with the smallest expected interaction mean squared error (IMSE). (More generally, for AMMI, GREG, EREG or COMM, the predictively best truncated model is the one with the smallest mean squared error of the estimated multiplicative effects.) Simulation estimates of E(IMSE) in AMMI models fitted to 9×20 tables with various interaction patterns (Table 3) indicate that using a 0.05-level F_{GH2} test as criterion for model choice will result in an E(IMSE) only slightly larger than E(IMSE) for the best truncated model. Neither cross validation nor the F_{GH2} test will always identify the predictively best truncated model obtainable from the full data set, but evidently F_{GH2} will usually choose the best or something not very much worse. In view of the simplicity of the F_{GH2} criterion, it is highly questionable whether cross validation is worth the effort.

4 SHRINKAGE ESTIMATORS OF MULTIPLICATIVE MODELS

The estimates of the multiplicative portion of the model in truncated models may be written as $\sum_k^p S_k \hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk}$, i.e., the kth term is multiplied by a "shrinkage factor" S_k which is either 0 or 1. Table 3 includes E(IMSE) when S_k is instead chosen as $\max[(F_k^* - 1)/F_k^*, 0]$ where $F_k^* = n\hat{\lambda}_k^2/u_{1k}^* s^2$. If the u_{1k}^* value used is correct, then S_k is an estimate of $n\lambda_k^2/(u_{1k}^*\sigma^2 + n\lambda_k^2)$. Results are shown for three choices of F_k^*. F_{GH2}, as previously defined, uses an approximate value of u_{1k} instead of u_{1k}^*. F_{sim} is F_k^* using u_{1k}^* estimated from simulation using the shrinkage estimates $S_k \hat{\lambda}_k$ as the supposed true λ_k for the simulation, with the initial S_k obtained using Gollob's F (F_{sim} is not the F_k or F_k^* which results from the previously described simulation test). Theorems of Goodman and Haberman (1990) suggest that Gollob's df should be a good choice for u_{1k}^* for λ_k which are large, but not for λ_k which are small. Interestingly, the only case shown (Table 3) where shrinkage estimation did not give E(IMSE) smaller than the best truncated model is where the only nonnull λ_k was $\lambda_1 = 10\sigma^2/n$. Here, Gollob's df should be a good value for u_{11}^*, but a poor value of u_{12}^*. Shrinkage using F_{GH2} was superior to Gollob's F in cases with no more than two nonnull terms, but Gollob's F was better otherwise. In all cases, shrinkage using F_{sim} gave E(IMSE) close to the better choice of F_{GH2} or Gollob's F.

The shrinkage factors S_k were constructed by analogy to shrinkage factors involved in empirical best linear unbiased predictors (BLUPs) in random effects and mixed models (Henderson, 1984) and to empirical Bayes estimators in linear models with normal priors (Lindley and Smith, 1972). In particular, the empirical BLUP of a cell mean under a two-way random effects model with a balanced data set is easily derived in the form $S_G(\bar{y}_{i.} - \bar{y}_{..}) + S_E(\bar{y}_{.j} - \bar{y}_{..}) + S_I(\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$, where the shrinkage factors are of the same form as we are using, viz., $S_X = \max[(F_X - 1)/F_X, 0]$ for $X = G, E$ or I and F_G, F_E and F_I are F-statistics for varieties, environments and interaction, each with the pooled error s^2 as denominator. We claim no optimal properties for S_k of this form for multiplicative models, but these results suggest that they will almost always result in better predictors than truncated models without shrinkage.

Computation of shrunken $\hat{\lambda}_k$ using F_{sim} is shown in Table 4 for the EVT16B data. The pooling of $\hat{\lambda}_3$ and $\hat{\lambda}_4$ is done by computing

$$
\begin{aligned}
\hat{\lambda}_{3\&4} &= [(6514651 + 5383288)/2]^{1/2} = 2439.05\,, \\
F^*_{3\&4} &= 2(2439.05)^2/[(27.30 + 20.06)(150712)] = 1.667\,, \\
S_{3\&4} &= 0.667/1.667 = 0.4001, \\
S_{3\&4}\hat{\lambda}_{3\&4} &= (0.4001)(2439.05) = 976\,.
\end{aligned}
$$

We used 3-1 data-splitting cross validation to validate the shrinkage estimation approach with real data. Shrinkage estimates of multiplicative terms in GREG, EREG and COMM were constructed in the same way as in AMMI. SHMM shrinkage is described in the Appendix. Table 5 shows the RMS PD values for the EVT16B data and another data set (EVT14B, also corn, with $g = 8, e = 47$). BLUP cell means refers to empirical BLUPs of the cell means under a random (unpatterned) interaction model. Clearly, empirical BLUPs are superior to the ordinary cell means in both data sets. Truncated multiplicative models were superior to empirical BLUPs in EVT16B, but not in EVT14B. Shrinkage with u^*_{1k} values obtained by simulation was equal to or better than either truncated models or empirical BLUPs in both data sets.

Improvement of shrinkage estimates over truncated models should be better than Table 5 indicates because, in practice, the shrinkage factors will be computed from the complete data, not from a modelling subset, whereas the truncated model still must be chosen by cross validation (unless chosen by F_{GH} tests). With shrinkage estimates, cross validation is unnecessary, except possibly to choose a model form. Results suggest that choice of model form is not of great importance.

SUMMARY

Either truncated or shrunken MMs can provide much better estimates of realized response levels of crop varieties in environments where tested than are given by the empirical cell means. Good truncated MMs may be chosen by cross validation, but simulation results suggest that a slightly conservative test procedure such as F_{GH1} or F_{GH2} as a criterion will result in only a small loss of efficiency. However, simulation results also indicate that shrinkage estimates of the type proposed here should be superior to truncated MMs. Cross validation of the EVT16B and EVT14B examples did not show as much advantage for shrinkage estimates, but opportunity to compute shrinkage factors from the complete data set should result in more advantage than can be demonstrated by data splitting cross validation. In cross validation, shrinkage estimates of MMs were superior to BLUPs of cell means using a two-way random effects model in EVT16B and were equally good in EVT14B.

State agricultural experiment stations or affiliated entities routinely conduct crop variety trials in their respective states. Results of this study strongly suggest

that yield performance estimates obtained by shrinkage estimates of MMs should replace the empirical means now routinely reported in published bulletins.

Obviously, the potential use of MMs for two-way studies is not limited to trials of crop varieties over a set of sites. Other possibilities immediately come to mind. For example, consider performance of a set of corn hybrids consisting of g new inbred lines, each crossed with e elite lines already in commercial use. Also, consider "on farm" trials where, say, g crop management practices are experimentally evaluated on each of e farms. The potential beneficial applications of shrinkage estimation of MMs would appear to be virtually unlimited.

APPENDIX

Shrinkage of SHMM was done as follows. Suppose for the moment that the left and right singular vectors are known. Then $\hat{\beta}$ and the $\hat{\lambda}_k$ can be obtained as intercept and regression coefficients for the multiple regression of $\text{vec}(\boldsymbol{Y}) = \text{vec}([\bar{y}_{ij.}])$ on $\boldsymbol{X}_k = \boldsymbol{\alpha}_k \otimes \boldsymbol{\gamma}_k$. Then the partial sum of squares owing to regression on $\boldsymbol{X}_k$ is $nq_k\hat{\lambda}_k^2$ where $q_k = (1 - ge\sum\bar{\alpha}_m^2\bar{\gamma}_m^2)/[1 - ge(\sum\bar{\alpha}_m^2\bar{\gamma}_m^2 - \bar{\alpha}_k^2\bar{\gamma}_k^2)]$. This suggests the following iterative scheme which is similar to a generalized ridge regression (Hoerl and Kennard, 1970) with iterative re-estimation of the left and right singular vectors from which the $\boldsymbol{X}_k$ are obtained. Initialize $u_{1k}^* = g+e-2k$. Given an initial value of $\hat{\beta}$, obtain the SVD of $\boldsymbol{Z} = [z_{ij}] = [\bar{y}_{ij.} - \hat{\beta}]$. Compute $F_k^* = n\hat{q}_k\hat{\lambda}_k^2/u_{1k}^*s^2$ and $S_k = (F_k^* - 1)/F_k^*$ provided $F_k^* > 1$, else $S_k = 0$. Then the shrunken $\hat{\lambda}_k$ is $S_k\hat{\lambda}_k$ and the new value of

$$\hat{\beta} = (\bar{y}_{...} - \sum_m S_m\hat{\bar{\alpha}}_m^2\hat{\bar{\gamma}}_m^2 \sum_i \sum_j \hat{\alpha}_{im}\hat{\gamma}_{jm}\bar{y}_{ij.})/(1 - ge\sum_m S_m\hat{\bar{\alpha}}_m^2\hat{\bar{\gamma}}_m^2). \qquad (A.1)$$

For a solution based on the initial u_{1k}^*, iterate these equations until they converge. By simulation, obtain new values for $u_{1k}^* = nE[\hat{q}_k\hat{\lambda}_k^2 - q_k\lambda_k^2]/\sigma^2$ where q_k, λ_k, α_{ik} and γ_{jk} are put equal to the previous $\hat{q}_k$, shrunken $\hat{\lambda}_k$, $\hat{\alpha}_{ik}$ and $\hat{\gamma}_{jk}$, respectively. Repeat the above scheme, the solution of which will be defined as the F_{sim} shrinkage.

The right hand side of (A.1) is a function only of $\boldsymbol{Y}$, u_{1k}^* for $k = 1, \ldots, \min(g, e)$, and $\hat{\beta}$. If we consider the u_{1k}^* as given constants, we can move $\hat{\beta}$ to the right hand side giving an equation $f(\hat{\beta}) = 0$ to be solved for $\hat{\beta}$. The solution can usually be obtained by a Newton-Raphson algorithm. The mathematical details are extremely lengthy and will be presented in another paper. Once $\hat{\beta}$ is so obtained, computation of $S_k, \hat{\lambda}_k, \hat{\alpha}_{ik}$ and $\hat{\gamma}_{jk}$ (some of which will be involved in the Newton-Raphson iteration) is straightforward.

REFERENCES

Allen, D. M. 1971. Mean square error of prediction as a criterion for selecting variables. *Technometrics* 13:469-495

Cornelius, P. L. 1978. Empirical models for the analysis of unreplicated lattice split-plot cultivar trials. *Crop Sci.* 18:627-633.

Cornelius, P. L. 1980. Functions approximating Mandel's tables for the means and standard deviations of the first three roots of a Wishart matrix. *Technometrics* 22:613-616.

Cornelius, P. L. 1993. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. *Crop Sci.* (In press).

Cornelius, P. L., M. Seyedsadr and J. Crossa. 1992. Using the shifted multiplicative model to search for "separability" in crop cultivar trials. *Theor. Appl. Genet.* 84:161-172.

Cornelius, P. L., D. A. Van Sanford and M. S. Seyedsadr. 1993. Clustering cultivars into groups without rank-change interactions. *Crop Sci.* (In press)

Crossa, J., and P. L. Cornelius. 1993. Recent developments in multiplicative models for cultivar trials. In: Buxton, D. R., et al. (Eds.) *International Crop Science I.* Crop Science Society of American, Madison, WI (In press).

Crossa, J., P. L. Cornelius, M. Seyedsadr and P. Byrne. 1993. A shifted multiplicative model cluster analysis for grouping environments without genotypic rank change. *Theor. Appl. Genet.* 85:577-586.

Crossa, J., H. G. Gauch, Jr., and R. W. Zobel. 1990. Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. *Crop Sci.* 30:493-500.

Eberhart, S. A., and W. A. Russell. 1966. Stability parameters for comparing varieties. *Crop Sci.* 6:36-40.

Fisher, R. A. and W. A. Mackenzie. 1923. Studies in variation II: The manurial response in different potato varieties. *J. Agri. Sci.* 13: 311-320.

Finlay, K. W. and G. N. Wilkinson. 1963. The analysis of adaptation in a plant breeding programme. *Australian J. Agri. Res.* 14:742-754.

Gabriel, K. R. 1978. Least squares approximation of matrices by additive and multiplicative models. *J. Royal Statist. Soc., Ser. B,* 40:186-196.

Gauch, H. G., Jr. 1988. Model selection and validation for yield trials with interaction. *Biometrics* 44:705-715.

Gauch, H. G., Jr., and R. W. Zobel. 1988. Predictive and postdictive success of statistical analyses of yield trials. *Theor. Appl. Genet.* 76:1-40.

Gauch, H. G., Jr., and R. W. Zobel. 1990. Imputing missing yield trial data. *Theor. Appl. Genet.* 79:753-761.

Gollob, H. F. 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* 33:73-115.

Goodman, L. A., and S. J. Haberman. 1990. The analysis of nonadditivity in two-way analysis of variance. *J. Amer. Statist. Assoc.* 85:139-145.

Henderson, C. R. 1984. *Applications of Linear Models in Animal Breeding.* University of Guelph, Guelph, Ontario, Canada.

Hoerl, A. E., and R. W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12:55-67.

Johnson, D. E. and F. A. Graybill. 1972. Analysis of a two-way model with interaction and no replication. *J. Amer. Statist. Assoc.* 67:862-868.

Johnson, D. E. 1976. Some new multiple comparison procedures for the two-way AOV model with interaction. *Biometrics* 32:929-934.

Lindley, D. V. and A. F. M. Smith. 1972. Bayes estimates for the linear model. *J. Royal Statist. Soc., Ser. B,* 34:1-41.

Mandel, J. 1961. Nonadditivity in two-way analysis of variance. *J. Amer. Statist. Assoc.* 56:878-888.

Mandel, J. 1971. A new analysis of variance model for nonadditive data. *Technometrics* 13:1-8.

Marasinghe, M. G. 1985. Asymptotic tests and Monte Carlo studies associated with the multiplicative interaction model. *Comm. Stat. A. Theory and Meth.* 14:2219-2231.

Nachit, M. M., G. Nachit, H. Ketata, H. G. Gauch, Jr., and R. W. Zobel. 1992. Use of AMMI and linear regression models to analyze genotype-environment interaction in durum wheat. *Theor. Appl. Genet.* 83:597-601.

Schott, J. R. 1986. A note on critical values used in stepwise tests for multiplicative components of interaction. *Comm. Stat. A. Theory and Meth.* 15:1561-1570.

Seyedsadr, M. and P. L. Cornelius. 1991. Functions approximating the expectations and standard deviations of sequential sums of squares in the shifted multiplicative model for a two-way table. Univ. Ky. Dept. Statist. Tech. Report 332.

Seyedsadr, M. and P. L. Cornelius. 1992. Shifted multiplicative models for nonadditive two-way tables. *Comm. Stat. B. Simul. and Comp.* 21:807-832.

van Oosterom, E. J., D. Kleijn, S. Ceccarelli and M. M. Nachit. 1993. Genotype-by-environment interactions of barley in the Mediterranean region. *Crop Sci.* 33:669-674.

Table 1: RMS PD (data split 3-1) and adjusted RMS PD(PRESS) for models fitted to EVT16B data.

No. terms	AMMI	GREG	EREG	COMM	SHMM
			Model form		
			Data splitting		
0	980	—	—	—	—
1	<u>915</u>	954	<u>908</u>	962	947
2	934	<u>907</u>	935	<u>911</u>	<u>906</u>
3	951	926	947	930	924
4	955	946	949	951	944
5	963	957	967	957	959

Cell means model: 985

			Adjusted RMS PD (PRESS)		
0	970	—	—	—	—
1	<u>956</u>	939	<u>892</u>	942	925
2	2725	<u>912</u>	9155	<u>935</u>	<u>886</u>
3	30708	994	7071	1557	925
4	19030	1071	14670	2682	2246
5	30540	3251	8165	8688	5094

Table 2: Rejection rates of H_{0k}: $\lambda_k = 0$ at $\alpha = 0.05$ by the Gollob test, F_{GH2} test, simulation test and iterated simulation test in AMMI analyses of 9×20 tables.

Description	Principal Component(k)							
	1	2	3	4	5	6	7	8
$\theta_k = n^{1/2}\lambda_k/\sigma$	10	5	0	0	0	0	0	0
Gollob	1	.957	.349	.041	.002	0	0	0
F_{GH2}	1	.496	.019	0	0	0	0	0
Simulation	1	.520	.046	.006	.003	.008	.010	.020
Iterated sim.	1	.516	.057	.018	.007	.013	.018	.026
$\theta_k = n^{1/2}\lambda_k/\sigma$	5	5	5	5	5	5	5	5
Gollob	1	1	1	.991	.872	.550	.181	.015
F_{GH2}	.998	.985	.899	.687	.382	.168	.058	.015
Simulation	.999	.985	.961	.884	.807	.738	.630	.527
Iterated sim.	.999	.992	.978	.919	.830	.727	.593	.410

Table 3: Simulated E(IMSE) values for AMMI models fitted to 9×20 tables with various interaction patterns

Nonnull θ-values*	Best truncated model	Model chosen by F_{GH2} test[†]	Shrinkage using		
			Gollob F	F_{GH2}	F_{sim}
10	0.167 (1)[‡]	0.181	0.178	0.148	0.149
14, 6	0.342 (2)	0.350	0.275	0.266	0.265
12, 8, 4	0.418 (2)	0.444	0.344	0.353	0.345
12, 10, 8, 4	0.547 (3)	0.565	0.446	0.473	0.451
14, 6, 4, 4, 2	0.562 (1)	0.592	0.404	0.425	0.402
Cell means model: E(IMSE)=0.844 for all sets of θ values.					

*$\theta_k = n^{1/2}\lambda_k/\sigma$.

[†].05−level.

[‡]Number of terms shown in parentheses.

Table 4: Computation of shrunken $\hat{\lambda}_k$ values for prediction using AMMI (EVT16B data, u_{1k}^* obtained by simulation, $s^2/n = 150712$).

$\hat{\lambda}_k^2$	$\hat{\lambda}_k$	u_{1k}^*	F_k^*	S_k	$S_k\hat{\lambda}_k$	3rd and 4th Pooled
35080839	5923	28.34	8.214	0.8783	5202	5202
9427899	3070	32.06	1.951	0.4874	1497	1497
6514651	2552	27.30	1.584	0.3685	941	976
5383288	2320	20.06	1.781	0.4384	1017	976
3092937	1759	18.88	1.087	0.0801	141	141
1725053	1313	13.37	0.856	0	0	0
628268	793	8.73	0.478	0	0	0
573093	757	4.86	0.782	0	0	0

Table 5: RMS PD values for models fitted to two data sets (3-1 data split).

Model form	Best truncated model		Shrinkage with u_{1k}^* obtained by		
	w/o BLUPs*	w/ BLUPs*	Gollob df	Simul- ation	Iterated simulation
			EVT16B data		
AMMI	915 (1)[†]	914	913	911	911
GREG	907 (2)	907	909	906	905
EREG	908 (1)	908	910	908	908
COMM	911 (2)	NA	911	910	910
SHMM	906 (2)	NA	908	905	907
	Cell means: 985		BLUP cell means[‡]: 932		
			EVT14B data		
AMMI	819 (1)	818	806	805	805
GREG	816 (1)	816	804	803	803
EREG	819 (1)	819	809	808	809
COMM	816 (1)	NA	804	803	803
SHMM	816 (1)	NA	811	808	809
	Cell means: 873		BLUP cell means[‡]: 808		

*Without and with BLUPs of additive main effects in AMMI, GREG and EREG. Shrinkage results are also "w/BLUPs".

[†]Number of multiplicative terms retained.

[‡]Empirical BLUPs of cell means under a two-way random effects model.