Parsimonious Estimation of Multiplicative Interaction in Analysis of Variance using Kullback-Leibler Information

Kert Viele and C. Srinivasan

July 9, 1999

Department of Statistics
University of Kentucky
Lexington, KY 40506-0027

Abstract

Many standard methods for modeling interaction in two way ANOVA require mn interaction parameters, where m and n are the number of rows and columns in the table. By viewing the interaction parameters as a matrix and performing a singular value decomposition, one arrives at the Additive Main Effects and Multiplicative Interaction (AMMI) model which is commonly used in agriculture. By using only those interaction components with the largest singular values, one can produce an estimate of interaction that requires far fewer than mn parameters while retaining most of the explanatory power of standard methods.

The central inference problems of estimating the parameters and determining the number of interaction components has been difficult except in "ideal" situations (equal cell sizes, equal variance, etc.). The Bayesian methodology developed in this paper applies for unequal sample sizes and heteroscedastic data, and may be easily generalized to more complicated data structures. We illustrate the proposed methodology with two examples.

Keywords: Bayes Estimates, Singular Value Decomposition, AMMI Model
AMS Classification: 62J10

1 Introduction

Many standard analyses of interaction in the twoway Analysis of Variance (ANOVA) model begin with the relation

$$y_{ijk} = \mu + \alpha_i + \beta_j + \theta_{ij} + \epsilon_{ijk} \tag{1}$$

where $i = 1, \ldots, m$, $j = 1, \ldots, n$, k indexes the observations within each cell, α_i and β_j are the main effects, θ_{ij} are interaction parameters, and ϵ_{ijk} is independent $N(0, \sigma^2)$ random error. The θ_{ij} parameters form a $m \times n$ matrix Θ with mn parameters. Our goal in this paper is to describe a method for analyzing the interaction that is more parsimonious than using mn parameters but retains most of the explanatory power.

Parsimonious modeling of the interaction is achieved by taking the singular value decomposition $\Theta = \Gamma^T \Lambda \Delta$ and retaining only the first few singular values, where Λ is a diagonal $\min(m,n) \times \min(m,n)$ matrix, Γ is a $\min(m,n) \times m$ matrix with orthonormal rows, and Δ is a $\min(m,n) \times n$

matrix with orthonormal rows. If $\Theta = \Gamma^T \Lambda \Delta$, then $\theta_{ij} = \sum_{t=1}^{T} \lambda_t \gamma_{ti} \delta_{tj}$ and thus equation (1) becomes

$$y_{ijk} = \mu + \alpha_i + \beta_j + \sum_t \lambda_t \gamma_{ti} \delta_{tj} + \epsilon_{ijk} \tag{2}$$

with the constraints $\lambda_t > 0$, $\sum_i \gamma_{ti}^2 = \sum_j \delta_{tj}^2 = 1$, and $\sum_i \gamma_{ri}\gamma_{si} = \sum_j \delta_{rj}\delta_{sj} = 0$ for all $r \neq s$ (the γ and δ vectors form orthonormal matrices in the singular value decomposition).

This reparameterization, originally considered in Gollob (1968), has already been used in a variety of contexts, such as genetics (Elkind and Cahaner 1986), infrared spectroscopy (Isakkson and Naes 1988), latent variable models in education (Muthen and Satorra 1989) and analysis of crop yield data (Zobel, Wright, and Gauch 1989, Cornelius, Crossa and Seyedsadr 1995, hereafter CCS). Some related work appears in Tukey (1955), Maransinghe and Johnson (1982), Yochmowitz and Cornell (1978), and Rao (1967). Within agriculture equation (2) is known as the Additive Effects and Multiplicative Interaction (AMMI) model. The central advantage of the AMMI formulation is that T may typically be chosen small and still retain much of the explanatory power of the full interaction model (1). Instead of mn parameters, one uses $T(m+n+1)$ parameters.

The central difficulties in using the AMMI model are estimating the parameters of the model and determining T. Typically the AMMI parameters are estimated by an iterative least squares method (see CCS) where one computes least squares estimate of the main effects and then performs a singular value decomposition on the residuals. Estimation of T is performed by approximate F tests. These methods, however, are insufficiently flexible to handle heteroscedastic data and unequal cell sizes. The contribution of this paper is to provide a set of tools that are sufficiently flexible to be applied to heteroscedastic data and unequal cell sizes.

In section 2.1 we describe the our prior distribution. As the values of the λ_t are particularly relevant in determining T, the number of interaction components, we derive the conditional posterior means of λ_t. With the exception of the γ and δ vectors, posterior computation is straightforward. However, γ and δ are defined as coordinate axes, and thus computation is nontrivial and described in section 2.3.

In section 3 we describe a method for evaluating the fit of the interaction components based on model embeddings from Viele (1996). The goal of the method is to determine the number of components required to produce a reasonable approximation of the data. The proposed method is flexible enough to allow the investigator to determine the functional definition of what is "reasonable". We provide a default measure based on Kullback-Leibler (KL) information and calculations for implementing the KL measure. KL information is fundamental to determining the complexity of a model (van Emden 1971) and hence is used is many model selection criteria such as AIC (Akaike 1973) and ICOMP (Bozdogan and Haughton 1996).

Two examples are shown in section 4, reanalyzing data on a soybean yield experiment from Zobel, Wright, and Gauch (1988) and a tomato softness study from Oman (1991). The tomato dataset with unequal variances demonstrates the flexibility of the Bayesian approach in analyzing data with heteroscedasticity or random effects. The results qualitatively agree with the frequentist analysis in estimating the parameters. The examples also provide additional analyses evaluating the fit of the interaction terms in the model. Further discussion of the method is provided in section 5.

2 Prior Distribution and Computation

2.1 Prior Distribution

The parameters of the model are μ, σ^2, the additive parameters α and β, and the interaction parameters λ, γ, and δ. The purpose of this section is to describe a prior distribution on these parameters that obeys the constraints of the model.

We place a normal prior on μ and a half normal prior $N^+(\mu_\lambda, \sigma_\lambda^2)$ on each λ_t, defined as the distribution proportional to a $N(\mu_\lambda, \sigma_\lambda^2)$ distribution for the nonnegative real numbers and 0 elsewhere. This is a conjugate prior that assures the length parameters will be nonnegative. Either distribution may be made approximately flat by choosing large variances.

The α and β vectors are constrained to sum to 0 to avoid confounding with μ. The prior on the α vector (the prior on β is similar) is the multivariate normal distribution

$$N_m\left(0, \sigma_\alpha^2\left[I - (1/m)11^T\right]\right),$$

which is the conditional distribution of m independent $N(0, \sigma_\alpha^2)$ random variables given the sum is 0 (Schervish 1995, section 8.2.2). Although the covariance matrix is singular, it has the advantages that the generated α_i sum to 0 and all the α_i are exchangeable. In practice, we avoid difficulties with the singular covariance matrix by including only $\alpha_1, \ldots, \alpha_{m-1}$ as parameters and then defining $\alpha_m = -(\alpha_1 + \ldots + \alpha_{m-1})$.

The vector γ_1 (δ_1 is similar) must have zero sum and unit length. The constraint that γ_1 must sum to 0 is equivalent to γ_1 being orthogonal to the 1 vector. These vectors form a hyperplane in $\mathbb{R}^m$. The constraint of unit length corresponds to the vectors on the unit sphere in $\mathbb{R}^m$. The intersection of these subspaces is an $m-2$ dimensional sphere in $\mathbb{R}^m$. In three dimensions, this corresponds to intersecting the sphere $x^2 + y^2 + z^2 = 1$ with the plane defined by $x + y + z = 0$. The intersection of these subspaces is a circle. We place a uniform prior over the possible values of γ_1, which we write SUNI. For the current discussion we assume the subspace is understood. We describe the details of sampling from this distribution in section 2.3.

Additional axes $\gamma_2, \gamma_3, \ldots$ (similarly for $\delta_2, \delta_3, \ldots$) must be orthogonal to all previously generated axes. After γ_1 has been generated, γ_2 is in the intersection of the hyperplane of vectors orthogonal to 1 and γ_1 and having unit length. This subspace is an $m-3$ dimensional sphere in $\mathbb{R}^m$. Additional axes reside in lower dimensional spheres. If the model specifies a full rank interaction, then either the final γ or δ (depending on whether m or n is smaller) is supported on only two vectors which are additive inverses of one another. In two dimensions, the intersection of the unit circle with the line $x + y = 0$ results in two solutions $x = -y = 1/2^{1/2}$ and $-x = y = 1/2^{1/2}$. The conditional prior distribution of each γ_t given the previous axes is a spherical uniform on the correct subspace. For the final axis this results in a Bernoulli(1/2) distribution on the two possible vectors.

The complete scheme is the following set of independent prior distributions.

$$
\begin{array}{rcl}
\mu & \sim & N(\theta_\mu, \sigma_\mu^2) \\
\alpha_1, \ldots, \alpha_{m-1} & \sim & N_{m-1}\left(0, \sigma_\alpha^2\left[I - (1/m)11^T\right]\right) \\
\beta_1, \ldots, \beta_{n-1} & \sim & N_{n-1}\left(0, \sigma_\beta^2\left[I - (1/n)11^T\right]\right) \\
\lambda_t & \sim & N^+(\mu_\lambda, \sigma_\lambda^2) \\
\gamma_t & \sim & \text{SUNI} \\
\delta_t & \sim & \text{SUNI} \\
\sigma^2 & \sim & \text{InvGamma}(a_\sigma, b_\sigma)
\end{array}
\tag{3}
$$

This prior may then be restricted to satisfy whatever assumptions one chooses for identifiability of the interaction components, such as using only a half sphere for γ_t if γ_{t1} is constrained to be nonnegative. Unequal variances across rows may be included in the model by placing separate inverse gamma priors on each σ_i^2.

2.2 Conditional Posterior Means

Though the AMMI model may be formulated in considerable generality, in the following the focus is on the special case where $T = 1$ and the errors ϵ_{ijk} are independent and distributed according to $N(0, \sigma_i^2)$. In addition assume, without loss of generality, that $K = 1$. Thus the model under consideration is

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda\gamma_i\delta_j + \epsilon_{ij}$$

This special case is adequate to address heteroscedasticity in the error (Oman 1991). The following discussion extends to the general case with $T > 1$ or $K > 1$. To insure the parameters are uniquely defined assume $\sum \alpha_i = \sum \beta_j = \sum \gamma_i = \sum \delta_j = 0$, and that γ_1 and δ_1 are nonnegative. This is equivalent to the usual assumptions of $\lambda > 0$ and $\gamma_1 > 0$. The assumption $\lambda \geq 0$ is customary because it facilitates the interpretation of the magnitude of the interaction. However, leaving λ free with a $N(\mu_\lambda, \sigma_\lambda^2)$ prior facilitates theoretical calculations and still allows evaluation of the magnitude of the interaction by $|\lambda|$.

The formulation in equation (3) does not permit closed form expressions for the posterior means (Bayes estimates) of the parameters. However, one can gain considerable insight into the estimators of α, β, and λ by computing the posterior means conditional on the remaining parameters σ_i^2, γ, and δ.

For simplicity, let the prior variances for α, β, and λ be sufficiently large to be approximately uniform. Letting $\bar{y}_{i.} = n^{-1}(y_{i1} + \ldots + y_{in})$, it is easy to see that the joint likelihood of y can be factored into the likelihoods of $\bar{y}_{i.}$ and $w_{ij} = y_{ij} - \bar{y}_{i.}$. Thus $\bar{y}_{1.}, \ldots, \bar{y}_{m.}$ are independent with $\bar{y}_{i.} \sim N(\mu + \alpha_i, \sigma_i^2/m)$ and w_{ij} is normal with mean $\beta_j + \lambda\gamma_i\delta_j$.

The posterior mean of $\mu + \alpha_i$, conditional on σ_i^2, is easy to obtain and is the familiar $\bar{y}_{i.}$ when the priors are approximately flat. The posterior means of β_j and λ (denoted $\bar{\beta}_j$ and $\bar{\lambda}$) conditional on σ_i^2, γ, and δ are also easy to derive when the prior variances are large. They are given by the equations

$$\bar{\beta}_j = \left(\sum 2/\sigma_i^2\right)^{-1} \sum (2/\sigma_i^2)(w_{ij} - \bar{\lambda}\gamma_i\delta_j)$$

and

$$\bar{\lambda} = \left(\sum \gamma_i^2/\sigma_i^2\right)^{-1} \sum_i (1/\sigma_i^2) \left(\sum_j w_{ij}\gamma_i\delta_j - \sum_j \bar{\beta}_j \gamma_i \delta_j \right)$$

Solving these equations we get

$$\bar{\lambda}\left(1 - \frac{\left(\sum \gamma_i/\sigma_i^2\right)^2}{\sum \gamma_i^2/\sigma_i^2} \frac{1}{\sum 1/\sigma_i^2}\right) = \left(\sum \gamma_i^2/\sigma_i^2\right)^{-1} \left(\sum_{ij} \frac{\gamma_i \delta_j}{\sigma_i^2} - \frac{\sum \gamma_i/\sigma_i^2}{\sum 1/\sigma_i^2} \sum_{ij} \frac{w_{ij}\delta_j}{\sigma_i^2} \right)$$

Setting $z_i = \sum w_{ij}\delta_j$, this equation reduces to

$$\bar{\lambda} = \frac{\frac{\sum z_i\gamma_i/\sigma_i^2}{\sum 1/\sigma_i^2} - \frac{\sum z_i/\sigma_i^2}{\sum 1/\sigma_i^2}\frac{\sum \gamma_i/\sigma_i^2}{\sum 1/\sigma_i^2}}{\frac{\sum \gamma_i^2/\sigma_i^2}{\sum 1/\sigma_i^2} - \left(\frac{\sum \gamma_i/\sigma_i^2}{\sum 1/\sigma_i^2}\right)^2}$$

The numerator can be interpreted as the covariance between Z and γ with respect to the probability distribution $\sigma_i^{-2}/\sum \sigma_i^{-2}$, while the denominator is the variance of γ with respect to the same measure. The entire quantity is therefore the weighted regression coefficient of Z on γ.

2.3 Sampling from the Spherical Uniform Distributions

The supports of the γ and δ vectors are not easily expressible in terms of the standard basis, so in this section we describe how to sample from the spherical uniform distributions with the correct supports. This method is used in the Gibbs Sampler in section 2.4. We describe the method for γ, the method for δ is identical. We construct a latent variable τ_{γ_t} that has a uniform prior on the the unit sphere in $\mathbb{R}^{m-t}$. We then construct a basis of $\mathbb{R}^m$ where the support of γ_t is easily expressed and transform τ_{γ_t} to the correct subspace to generate γ_t.

To generate from a uniform distribution on a d dimensional unit sphere, generate d $N(0,1)$ random variates and retain their direction. If $\kappa_1, \ldots, \kappa_d$ are the random variables, then define the vector τ by $\tau_i = \kappa_i/(\sum \kappa_i^2)^{1/2}$. The τ vector has unit length. We refer to the distribution of τ as a spherical uniform distribution of dimension d which we write SUNI_d. The vector τ_{γ_1} has the distribution SUNI_{m-1}.

Although τ_{γ_1} has the same dimension as γ_1, it is not supported on the same sphere. If $m = 3$, then τ_{γ_1} would be generated on the circle $x^2 + y^2 = 1$ with $z = 0$ which is not the intersection of $x^2 + y^2 + z^2 = 1$ and $x + y + z = 0$. We transform τ_{γ_1} to γ_1 by constructing a basis of $\mathbb{R}^m$ where one of the basis vectors is the 1 vector. The remaining vectors of the basis are calculated by Gram-Schmidt orthogonalization of the final $m-1$ standard basis vectors to 1. The result is the columns of the $m \times m$ matrix B_{γ_1} with entries

$$b_{rc} = \begin{cases} 1 & c = 1 \\ 0 & c > 1 \text{ and } r \le c-2 \\ -1 & c > 1 \text{ and } r \ge c \\ m-r & c > 1 \text{ and } r = c-1 \end{cases}$$

While difficult to clearly describe in symbols, the pattern should be apparent from the matrix B_{γ_1} for $m = 5$.

$$B_{\gamma_1} = \begin{bmatrix} 1 & 4 & 0 & 0 & 0 \\ 1 & -1 & 3 & 0 & 0 \\ 1 & -1 & -1 & 2 & 0 \\ 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

Define A_{γ_1} to be the matrix with the columns of B_{γ_1} normalized to have unit length, so $a_{rc} = b_{rc}/(b_{1c}^2 + \ldots + b_{mc}^2)^{1/2}$. The columns of A_{γ_1} form an orthonormal basis of $\mathbb{R}^m$, with A_{γ_1} the change of basis matrix from the standard basis. Define η_{γ_1} to be the vector $(0, \tau_{\gamma_1 1}, \ldots, \tau_{\gamma_1 (m-1)})^T$ and let $\gamma_1 = A_{\gamma_1} \eta_{\gamma_1}$. The vector γ_1 has unit length because A_{γ_1} is orthonormal and η_{γ_1} was constructed with unit length. Since the first coordinate of η_{γ_1} is 0 by definition (corresponding to the coefficient of the 1 vector), $\sum \gamma_{1i} = 0$, so γ_1 satisfies the constraints of the model. Furthermore, since A_{γ_1} is orthonormal γ_1 has a spherical uniform distribution on the correct subspace.

A convenient basis for the subspace supporting γ_2 is constructed by the vectors 1, γ_1, and the Gram-Schmidt orthogonalization of the final $m-2$ standard basis vectors. These form the columns of a matrix B_{γ_2}, which are then normalized to generate A_{γ_2}. Generate τ_{γ_2} as SUNI_{m-2}, define $\eta_{\gamma_2} = (0, 0, \tau_{\gamma_2 1}, \ldots, \tau_{\gamma_2 (m-2)})$, and let $\gamma_2 = A_{\gamma_2} \eta_{\gamma_2}$. By the same reasoning that γ_1 obeys the constraints of the model, γ_2 has unit length, zero sum, is orthogonal to γ_1, and is uniformly distributed within the correct subspace. This procedure may be iterated to generate as many γ_t as the model specifies. For example, γ_3 would be generated by constructing a basis by Gram Schmidt orthogonalization of the final $m-3$ standard basis vectors to 1, γ_1, and γ_2, then generating $\tau_{\gamma_3} \sim \mathrm{SUNI}_{m-3}$ and defining $\eta_{\gamma_3} = (0, 0, 0, \tau_{\gamma_3 1}, \ldots, \tau_{\gamma_3 (m-3)})$, and $\gamma_3 = A_{\gamma_3} \eta_{\gamma_3}$.

2.4 Gibbs Sampler

The posterior distribution of the parameters may be estimated through Markov Chain Monte Carlo (MCMC) techniques (Gelfand and Smith 1990, Tierney 1994). The complete conditionals for μ, α_i, β_j, λ_t, and σ^2 may be computed using the conjugate property of the prior distributions. As discussed in section 2.1, we avoid singularities in the covariance matrix of the α and β vectors by using only $\alpha_1, \ldots, \alpha_{m-1}$ and $\beta_1, \ldots, \beta_{n-1}$ as parameters and defining $\alpha_m = -(\alpha_1 + \ldots + \alpha_{m-1})$ and $\beta_n = -(\beta_1 + \ldots + \beta_{n-1})$.

Updating the coordinate axis parameters γ and δ requires use of the Metropolis algorithm (Metropolis et al 1953). Because the γ vectors are fairly rigidly constrained with respect to one another, care must be taken in updating the set. If a full rank interaction is specified, one cannot update the vectors one at a time. The conditional distribution of γ_t given the remaining γ vectors is supported on only two vectors as described in section 2.1, so the chain will not move except to change the signs of the components. A near full rank interaction would not be as severe a problem, although it seems reasonable that a near full rank interaction will produce a chain that moves slowly. Alternatively, one could consider moving all the axes as a group. However, it is unclear if the chain mixes properly. For example, if all the axes are changed by a particular angle at each iteration, the estimated variances of all the axes will be equal since all moves are of equal distance. Clearly the posterior variances may be different for different axes. Axes corresponding to λ_t near 0 have large variances around the sphere, since they provide only small changes to the estimated means.

We therefore chose to update the τ vectors instead of the actual interaction components. The τ parameters have no constraints between them so the actual Metropolis chain should have better mixing properties. The τ vectors may be initialized by subtracting the empirical main effects from the model, performing a singular value decomposition on the residuals, and converting the largest eigenvectors to τ parameters.

To update each τ_{γ_t}, we construct a random walk Metropolis Chain (Tierney 1994) by starting with the current vector $\tau_{\gamma_t}^{cur}$, adding a $N_{m-t}(0, \sigma_\tau^2 I_{m-t})$ random vector, then normalizing the result to generate a candidate $\tau_{\gamma_t}^{can}$ vector. This creates a symmetric random walk so the acceptance probability is the ratio of posterior densities for the current and candidate vectors. This ratio is further simplified since we are using a uniform prior, so the acceptance probability is the ratio of the likelihoods for the observed Y values.

To compute the likelihood ratio, use the method from section 2.3 to compute all the γ_t^{cur} and γ_t^{can} vectors from the τ values (updating τ_{γ_t} changes all the γ_g with $g \geq t$). The chain should be moved from $\tau_{\gamma_t}^{cur}$ to $\tau_{\gamma_t}^{can}$ with probability

$$\min\left(1, \frac{\prod_{ijk} N(\mu + \alpha_i + \beta_j + \sum_t \lambda_t \gamma_{ti}^{can} \delta_{tj}, \sigma^2)(y_{ijk})}{\prod_{ijk} N(\mu + \alpha_i + \beta_j + \sum_t \lambda_t \gamma_{ti}^{cur} \delta_{tj}, \sigma^2)(y_{ijk})}\right) \tag{4}$$

The Metropolis chain may be run for multiple iterations to provide extra mixing for the τ vectors. In the examples in section 4, we ran 30 Metropolis steps within each Gibbs iteration. Should the model specify a full rank interaction, the final interaction component has a Bernoulli complete conditional with probability determined by the likelihood (4).

The step width σ_τ^2 should be chosen carefully. If σ_τ^2 is too large, the chain will produce almost uniformly distributed candidate vectors around the sphere. If the posterior distribution is relatively concentrated, few of these candidate values will be accepted and the chain will converge to the posterior distribution slowly. If σ_τ^2 is too small, then the chain will move slowly, also resulting in slow convergence. A simple method for choosing σ_τ^2 is to run preliminary chains to determine the proportion of candidate values that are accepted, and then increase or decrease σ_τ^2 until the acceptance proportion is reasonable (proportions between 0.4 and 0.9 should all be effective).

After running the Gibbs sampler, one has a large list of samples from the posterior distribution. The sample means may be used to obtain estimates for most of the parameters. The interaction components are an exception. Because they are supported on spheres, sample means are usually neither unit length nor orthogonal. We suggest finding the sample means of the τ vectors, normalizing them to have unit length (this retains the direction), and then transforming them to get estimates of γ and δ.

In implementing the Gibbs Sampler, we did not constrain the parameters to be identifiable as noted in sections 1 and 2.1. Nonidentifiability of the AMMI model occurs because of sign changes in the interaction components. These changes are not easily permitted by the updating scheme, so are extremely unlikely to occur. The chain randomly picks a set of signs and sticks with them.

3 Evaluating Fit of Interaction Components

We approach evaluating fit from the assumption that the model is not exactly correct (although it may be extremely close) and formulate a method for quantifying the inaccuracy. While models with more parameters usually have greater predictive power, if the increase in predictive power is small then more parsimonious models may be preferred. This section describes a method for quantifying the increase in predictive power from using a particular number of interaction components. An alternative approach would be to compute Bayes Factors (Kass and Raftery 1995) for the interaction components, which we discuss briefly in this section.

Carota, Parmigiani, and Polson (1996) (CPP) and Viele (1996) discuss methods of evaluating fit based on model embedding that may be used for the AMMI model. Let F_q be an AMMI model with q components for which we are evaluating fit. We refer to the distribution for a particular cell of the table as $F_q(i,j)$. We evaluate fit by computing the posterior distribution of the model F_p with $p > q$ and then determine if the F_q model adequately predicts the data compared to the model F_p.

Suppose we incorrectly use an F_q model to approximate a particular F_p model with K observations in each cell. The posterior distribution of the parameters for the F_q model will converge to the parameter values that minimize

$$(mn)^{-1}\sum_{ij} KL(F_p(i,j), F_q(i,j)),$$

the averaged KL information across the cells in the table (Berk 1966). Let $\hat{F}_q$ be the minimizing model. The parameters of $\hat{F}_q$ are random variables since the F_p parameters are unknown. We evaluate fit by looking at the posterior distribution of $d(F_p, \hat{F}_q)$ as the F_p model parameters vary across their posterior distribution, where d is a measure of distance between distributions. This method follows Viele (1996). CPP's measure for this example would correspond to looking at the KL information between the prior and posterior distributions of the $q+1$ through p interaction components. This would involve computing the KL information for the γ and δ parameters, which is difficult because of their supports.

Ideally, the investigator has a clear idea about how closely the individual cell means should be estimated by the model, which might then be elicited and used as the distance d. If such a d is unavailable, CPP and Viele (1996) recommend KL information as a default choice. CPP provide a detailed list of references concerning the use of KL information for similar purposes. Letting d be the average of the KL informations across the cells, we find the posterior distribution of

$$(mn)^{-1}\sum_{ij} KL(F_p(i,j), \hat{F}_q(i,j)). \tag{5}$$

To compute the posterior distribution of (5), we use the Gibbs sampler output for an AMMI model with p interaction components. Each iteration is a sample from the posterior distribution of F_p. For each iteration, we evaluate the minimizing parameter values for $\hat{F}_q$ and compute the KL information (5) for those values. The resulting KL informations are a sample from the posterior distribution of (5). Proposition 3.1 describes the details of performing the calculation.

This distribution estimates how well an F_q model predicts data from an F_p model, where the F_p model is unknown. Heuristically, KL information measures how many observations are required before one can tell the F_q model is not exactly correct. If P and D are fixed distributions (no unknown parameters) and B is the Bayes factor for testing between the two models when data is generated from P, then $E[\ln B] = nKL(P, D)$. A KL information of 0 corresponds to both models having exactly the same predictive power. If the distribution of (5) concentrates around small values, then this indicates it would require a large amount of data to determine F_q is not accurate, so the F_q model may be a useful approximation. If the distribution concentrates around larger values, the F_q model is insufficient to predict the data and more axes should be included.

If, for some $a > 0$, the posterior distribution contains negligible probability for values of $KL < a$, this is evidence of a statistically significant axis since $KL = 0$ corresponds to $H_0 : \lambda_{q+1} = \ldots = \lambda_p = 0$. Staying with our idea of testing for "practical significance", a Bayes Factor for comparing $H_0 : KL \leq a$ against the hypothesis $H_1 : KL > a$ may be computed by determining the ratio of the posterior and prior masses of KL values less than a. Negligible posterior probability for $KL < a$ corresponds to the Bayes Factor favoring H_1. If a is small, however, the posterior distribution may still concentrate around a small value, indicating the extra axes may be "statistically significant" but not "substantively meaningful". With large samples, it is quite possible an axis may be found to be significant although if provides minimal predictive power. The proposed KL method allows one to calibrate the increase in predictive power. If one does want a Bayes factor for comparing the point null $H_0 : KL = 0$ to $H_1 : KL > 0$, then a may be decreased to 0 and the Bayes factor may be found using the Savage-Dickey density ratio (Verdinelli and Wasserman 1995).

Since the observed Y have normal distributions, it is useful to calibrate the KL information in terms of normal distributions to determine exactly what values are large or small. A $N(0, 1)$ distribution and a $N(\theta, 1)$ distribution have a KL information of $\theta^2/2$. If the KL information is ϕ, then this is the same KL information as normal distributions $(2\phi)^{1/2}$ standard deviations apart. References for other methods of calibrating the KL information may be found in CPP.

Proposition 3.1 *Let F_p be a fixed AMMI model with p interaction components, where*

$$Y_{ijk} \sim F_p(i, j) = N(\mu + \alpha_i + \beta_j + \sum_{t=1}^{p} \lambda_t \gamma_{ti} \delta_{tj}, \sigma^2)$$

The parameter values for $\hat{F}_q$ are $\hat{\mu} = \mu$, $\hat{\alpha}_i = \alpha_i$, $\hat{\beta}_j = \beta_j$, $\hat{\lambda}_t = \lambda_t$, $\hat{\gamma}_i = \gamma_i$, $\hat{\delta}_j = \delta_j$, and

$$\hat{\sigma}^2 = \sigma^2 + (mn)^{-1} \sum_{t=q+1}^{p} \lambda_t^2$$

The KL information (5) at these parameter values is

$$(1/2) \ln \left(\frac{\sum_{t=q+1}^{p} \lambda_t^2}{mn\sigma^2} + 1 \right)$$

Proof

The parameter values that achieve the infimum in (5) may be determined using Berk's (1966) result. Suppose we observe data from F_p and we assume (correctly or incorrectly) that an F_q model is correct. Berk's result states that for any prior (with certain regularity conditions) on F_q, the posterior distribution will converge (as K approaches infinity) to a point mass at the parameter values that minimize (5). For normal models, the MLE corresponds asymptotically to the posterior mean. We therefore assume data is generated from F_p and compute the limiting value of the MLE to find $\hat{F}_q$. Those asymptotic values may then be used to compute (5). The results are not approximations nor do they require large sample sizes. This is a calculation device to find $\hat{F}_q$.

The MLEs for the μ, α, and β parameters are the empirical row and column effects and are consistent, so $\hat{\mu} = \mu$, $\hat{\alpha}_i = \alpha_i$, and $\hat{\beta}_j = \beta_j$. Conditional on those MLEs, the log likelihood is

$$-mnK \ln \sigma (2\pi)^{1/2} - \frac{1}{2\sigma^2} \sum_{ijk} \left(y_{ijk} - \mu - \alpha_i - \beta_j - \sum_{t=1}^{q} \hat{\lambda}_t \hat{\gamma}_{ti} \hat{\delta}_{tj} \right)^2$$

Rewriting y_{ijk} as $y_{ijk} - \bar{y}_{ij.} + \bar{y}_{ij.}$, it may be seen that maximizing this for $\hat{\lambda}$, $\hat{\gamma}$, and $\hat{\delta}$ is equivalent to maximizing

$$-K\sigma^{-2}\sum_{ij}\left(\bar{y}_{ij.}-\mu-\alpha_i-\beta_j-\sum_{t=1}^{q}\hat{\lambda}_t\hat{\gamma}_{ti}\hat{\delta}_{tj}\right)^2$$

Asymptotically, $\bar{y}_{ij.}$ converges to

$$\mu+\alpha_i+\beta_j+\sum_{t=1}^{p}\lambda_t\gamma_{ti}\delta_{tj},$$

so the goal is to maximize

$$-K\sigma^{-2}\sum_{ij}\left(\sum_{t=1}^{p}\lambda_t\gamma_{ti}\delta_{tj}-\sum_{t=1}^{q}\hat{\lambda}_t\hat{\gamma}_{ti}\hat{\delta}_{tj}\right)^2 \tag{6}$$

Equation (6) is a scalar multiple of the Euclidean norm between two $m \times n$ matrices, both expressed in terms of their singular value decomposition. The Euclidean norm is minimized by choosing $\hat{\lambda}_1,\ldots,\hat{\lambda}_q$ to be the q largest eigenvalues of $\lambda_1,\ldots,\lambda_p$ and choosing $\hat{\gamma}$ and $\hat{\delta}$ to be their corresponding eigenvectors (Horn and Johnson 1985, example 7.4.52), so $\hat{\lambda}_t=\lambda_t$, $\hat{\gamma}_i=\gamma_i$, and $\hat{\delta}_j=\delta_j$. Conditional on the mean MLEs, the MLE of σ^2 is

$$\hat{\sigma}^2=(mnK)^{-1}\sum_{ijk}\left(y_{ijk}-\hat{F}_q(i,j))^2=(mnK)^{-1}\sum_{ijk}(\epsilon_{ijk}+\sum_{t=q+1}^{p}\lambda_t\gamma_{ti}\delta_{tj}\right)^2=$$

$$(mnK)^{-1}\sum_{ijk}\left\{\epsilon_{ijk}^2+2\epsilon_{ijk}\sum_{t=q+1}^{p}\lambda_t\gamma_{ti}\delta_{tj}+\left(\sum_{t=q+1}^{p}\lambda_t\gamma_{ti}\delta_{tj}\right)^2\right\}$$

Because of the unit length and orthogonality properties of the γ and δ vectors

$$\sum_{ij}\left(\sum_{t=q+1}^{p}\lambda_t\gamma_{ti}\delta_{tj}\right)^2=\sum_{ij}\sum_{t=q+1}^{p}\lambda_t^2\gamma_{ti}^2\delta_{tj}^2=\sum_{t=q+1}^{p}\lambda_t^2$$

so

$$\hat{\sigma}^2=(mn)^{-1}\sum_{t=q+1}^{p}\lambda_t^2+(mnK)^{-1}\sum_{ijk}\left(\epsilon_{ijk}^2+2\epsilon_{ijk}\sum_{t=q+1}^{p}\lambda_t\gamma_{ti}\delta_{tj}\right)$$

As the sample size K increases, $\hat{\sigma}^2$ converges to

$$\sigma^2+(mn)^{-1}\sum_{t=q+1}^{p}\lambda_t^2$$

The averaged KL information across the table is

$$(mn)^{-1}\sum_{ij}KL(N(E(F_p(i,j)),\sigma^2),N(E(\hat{F}_q(i,j)),\hat{\sigma}^2))=$$

$$(mn)^{-1} \sum_{ij} \left(\frac{(E(F_p(i,j)) - E(\hat{F}_q(i,j)))^2 + \sigma^2 - \hat{\sigma}^2}{2\hat{\sigma}^2} + (1/2)\ln \frac{\hat{\sigma}^2}{\sigma^2} \right) =$$

$$(2\hat{\sigma}^2 mn)^{-1} \sum_{ij} \left\{ \left(\sum_{t=q+1}^{p} \lambda_t \gamma_{ti} \delta_{tj} \right)^2 - (mn)^{-1} \sum_{t=q+1}^{p} \lambda_t^2 \right\} + (1/2)\ln \left(\frac{\sum_{t=q+1}^{p} \lambda_t^2}{mn\sigma^2} + 1 \right) =$$

$$(1/2)\ln \left(\frac{\sum_{t=q+1}^{p} \lambda_t^2}{mn\sigma^2} + 1 \right)$$

□

If the variances are not equal, the analysis becomes more complicated because the log likelihood is no longer expressed as the Euclidean norm between two matrices. In principle, however, one could compute asymptotic MLEs and determine the KL information, which we do not pursue here.

If one is interested only in examining the fit of the additive model with separate variances for each column, the analysis still proceeds smoothly. Even for unequal variances, the MLEs for the row and column effects converge to α_i and β_j. The conditional MLE for each row variance converges to

$$\hat{\sigma}_i^2 = \sigma_i^2 + n^{-1} \sum_t \lambda_t^2 \gamma_{ti}^2.$$

The KL information is

$$m^{-1} \sum_i (1/2)\ln \left(\frac{\sum_t \lambda_t^2 \gamma_{ti}^2}{n\sigma_i^2} + 1 \right) \tag{7}$$

4 Examples

This section applies the methods in sections 2 and 3 to two previously analyzed datasets, a soybean yield experiment described in Zobel, Wright, and Gauch (1988) (ZWG) and a tomato softness experiment described in Oman (1991). In both examples, the prior distributions of μ, α, β, and λ were made nearly flat by choosing the corresponding variances large. The prior density of σ^2 is $1/\sigma^2$.

4.1 Soybean Yield Data

This dataset consists of approximately 980 observations of soybean yield. The 35 rows are site-years and the 7 columns are genotypes, with the observations measured in kg/ha. Most cells in the table have 4 observations, although a few cells have 2 or 3 observations. We fit an AMMI model with three interaction components to the data. Table 1 contains the data and the posterior means for the parameters. The interaction terms are shown as $\lambda_t^{1/2}\gamma_{ti}$ and $\lambda_t^{1/2}\delta_{tj}$, following ZWG, so that to determine a component's contribution to a cell one can multiply the corresponding interaction terms. The posterior means for the additive effects and interaction terms are almost identical to the least squares estimates given in ZWG, who provide a discussion interpreting the parameter estimates in the context of soybean planting.

The top row of figure 1 shows the estimated posterior distributions of the three λ parameters. The posterior distributions of λ_1 and λ_2 both contain negligible mass near 0, indicating that the first two interaction components probably have nonzero length. The posterior distribution of λ_3, in contrast, does contain mass near 0, indicating that the data do not necessarily support a third axis.

We calibrated the lengths of the axis using the method from section 3. The first plot on the bottom row shows the estimated KL information for using the additive model ($q = 0$) compared to using three interaction components ($p = 3$). As with ZWG, the interaction components play a significant role in prediction. The KL information has a posterior mean of 0.27, corresponding to normal distributions 0.735 standard deviations apart.

The second plot on the bottom row shows the estimated KL information for using a model with 1 component compared with the 3 component model. This plot, as with its counterpart on the top row, has negligible probability near 0, indicating the fit is improved, but it is unclear the improvement is worth keeping after calibration. The posterior mean for the KL information is 0.034 corresponding to normal distributions 0.26 standard deviations apart. Decisions regarding the inclusion of this axis should be made on substantive grounds. Looking at the γ_2 and δ_2 vectors, the largest components correspond to the HODG and S200 genotypes. The third plot on the bottom row shows the KL information for comparing a two axis model with the three axis model. There is considerable posterior probability near 0 with most of the posterior mass less than 0.02, indicating the third axis provides little explanatory power.

The analysis in ZWG finds only the first axis to be significant, which qualitatively agrees with the Bayesian analysis when the magnitude of the second axis is calibrated by the KL information. The KL information provides more information than significance, however, since it gives a way of determining if the axis is significant (by observing the amount of posterior probability near 0) and quantifying the predictive power of the axis.

The Gibbs Sampler also allows us to compute posterior variances for any of the parameters. For example, the posterior standard deviations of the γ and δ components for the first axis are between 0.03 and 0.04, suggesting that a γ_{1i} or δ_{1j} should be considered statistically nonzero if they are greater in absolute value than around 0.08. This includes 14 of the 35 γ_{1i} and 5 of the 7 δ_{1j}. This guides the investigator toward the significant components of the interaction.

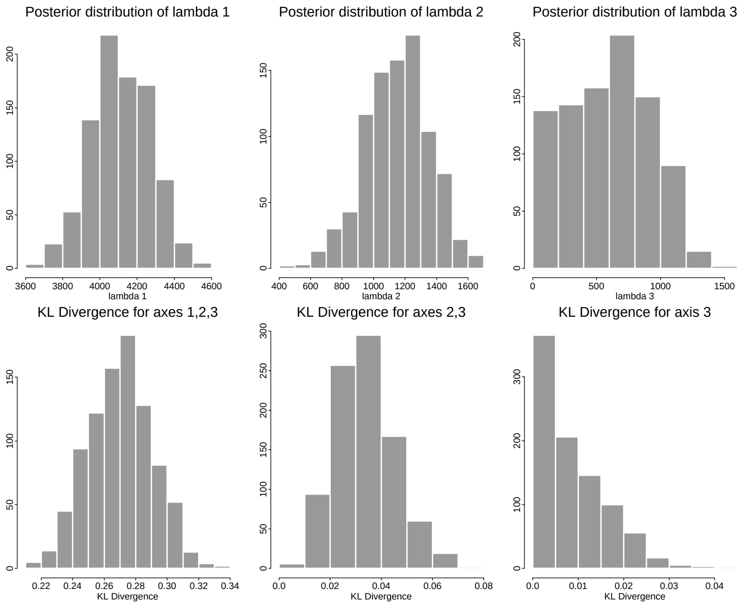
4.2 Tomato Softness Data

This example is based on the dataset from Oman (1991). The data consist of measures of tomato softness as the force required to compress a given tomato 0.5mm. The 21 rows, indexed by j, represent randomly selected mother plants and the 3 columns, indexed by i, represent genotypes. The observed data clearly exhibit heteroscedasticity across the genotypes. Oman analyzed the data with a random multiplicative heteroscedastic model of the form $y_{ijk} = \mu + \alpha_i + b_j(1 + \theta_i) + \epsilon_{ijk}$ with $b_j \sim N(0, \sigma_\beta^2)$ and $\epsilon_{ijk} \sim N(0, \sigma_i^2)$ to reflect the random selection of the mother plants and the heteroscedasticity.

We reanalyzed these data with the AMMI model with one axis. The goal here is to emphasize the flexibility of the Bayesian method regarding unequal variances and random effects. Our model is

$$y_{ijk} \sim N(\mu + \alpha_i + \beta_j + \lambda\gamma_i\delta_j, \sigma_i^2),$$

Figure 1: Evaluating Fit of Interaction Components for Soybean Data



with nearly flat prior distributions as indicated in the beginning of this section, with one exception. To account for the random plant effect, we placed a hierarchical prior distribution on β_j with the prior density on σ_β^2 being nearly proportional to $1/\sigma_\beta^2$. In frequentist analyses, fixed effects affect the distribution of the response through the mean vector while random effects affect the covariance matrix. In the Bayesian approach, fixed effects are estimated using the prior distribution described in equation (3). This prior distribution induces a structure similar to a random effect. The variance of the random effect may be estimated by looking at the posterior distribution of σ_β^2.

The results are shown in Table 2. Our results are similar to Oman's even though the models are slightly different. The posterior distribution of σ_β^2 is shown in Figure 2. The posterior mean and variance of σ_β^2 are 1.15 and 0.45, similar to Oman's results (see Table 4, Oman 1991). The posterior distribution of σ_β^2 is clearly shifted away from 0, indicating that softness varies among the mother plants. Moreover, the posterior means of $\mu + \alpha_i$ are almost identical to Oman's estimates. Our estimates of σ_i^2, though comparable, are slightly smaller than Oman's estimates.

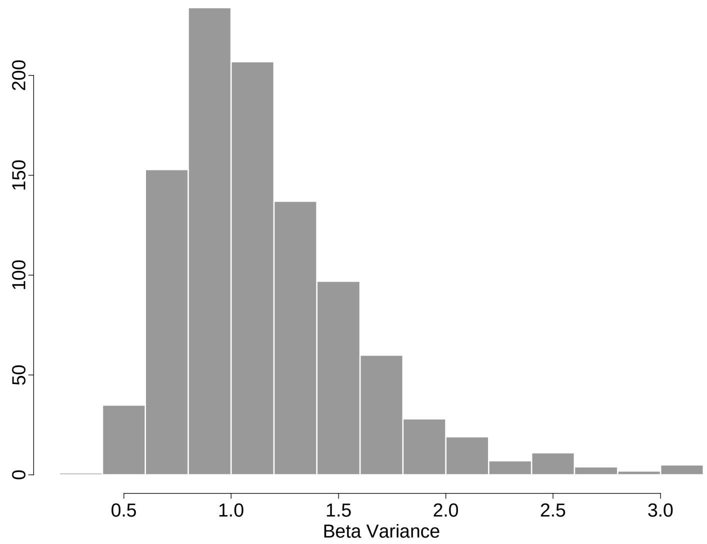
Although multiplicative effects appear to be present from both the posterior distribution of λ and the KL information (7) shown in Figure 3, the overall multiplicative effect is relatively small. The mean KL information is 0.0516 corresponding to normal distributions 0.32 standard deviations apart. This is also consistent with Oman's results, who found a significant, but not large, interaction component in his model formulation.

5 Discussion

Our goal in this paper has been to construct a Bayesian method for analyzing the AMMI model, including both methods for estimating the model parameters and goodness of fit criteria useful for determining the number of interaction components.

One interesting theoretical contrast between the Bayesian and frequentist analyses of the AMMI model occurs for small cell sizes K, which occur often in practice. Many frequentist methods for determining the number of components are based on asymptotics in K (see CCS and references therein), and thus require adjustments for small K. The Bayesian analysis allows exact calculation

Figure 2: Posterior Distribution of σ^2_β



of the posterior distribution, and thus replaces the uncertainty involved in the asymptotic approximation with the uncertainty of determining how much influence the prior has over the posterior distribution. We hope to investigate this issue further in future work.

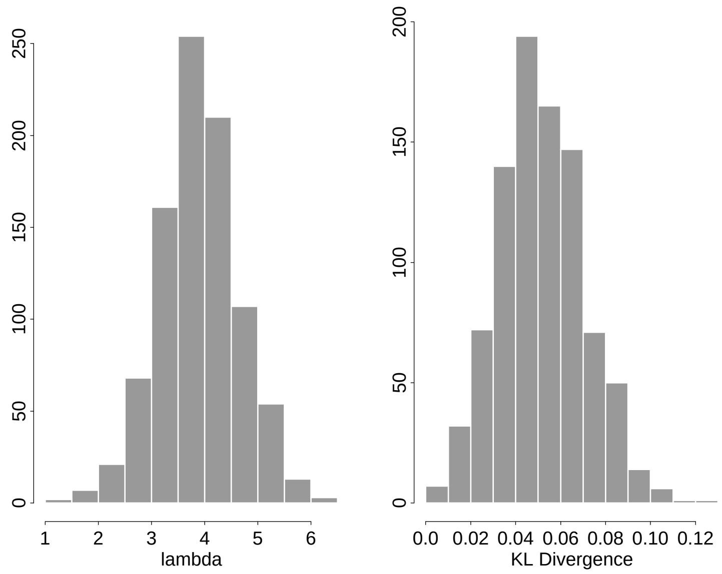
From a practical perspective, the Bayesian analysis based on the Gibbs Sampler easily provides information about the parameters or any functional of the parameters, and is flexible enough to estimate generalizations of the AMMI model such as unequal variances. To the best of our knowledge, no one has attempted a complete frequentist analysis of heteroscedastic data. The expression for $\hat{\lambda}$ suggests a method for beginning such an analysis based on weighted regression. From a Bayesian computational perspective, we have developed a prior for sampling coordinate axes and provided an implementation using MCMC techniques. The proposed method is computationally intensive, although it is unclear if there are intuitively appealing alternatives for estimating the axes directly from the discussion in section 2.4.

The proposed goodness of fit criterion allows fit of the interaction components to be examined in two ways. Using the posterior distribution of either the individual λ or the KL information it is possible to visually determine if any particular axes are statistically significant from 0. This may be formalized to a Bayes factor using the Savage-Dickey density ratio to provide a hypothesis test. The posterior distribution of the KL information also allows interpretation of the magnitude of the distance. It is possible an interaction component may be present, but may be small enough to ignore. Ignoring the component allows a more parsimonious model while keeping the component allows better predictions. The second axis of the soybean data provides an example. A second principal component appears to be present in the data, but its actual effect seems small when calibrated by the KL information. Whether the axis should be retained should be decided on substantive grounds. The KL information therefore provides some calibration in determining whether a "statistically significant" axis is a "substantively meaningful" axis.

Acknowledgment

The authors would like to thank Paul Cornelius for drawing their attention to this problem and

Figure 3: Evaluating Fit of Interaction Components for Tomato Data



providing useful comments on drafts of the paper. The authors would also like to thank Genzhou Liu and two anonymous referees for helpful comments.

6 References

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In *2nd International Symposium on Information Theory* Petrov and Czaki, Editors. Akademiai Kiado, Budapest, 267-281.

Berk, R. (1966). Limiting Behavior of Posterior Distributions when the Model is Incorrect. *Annals of Mathematical Statistics* **37** 51-58

Bozdogan, H. and Haughton, D. (1998) Informational Complexity Criteria for Regression Models. *Computational Statistics and Data Analysis* **28** 57-76

Carota, C., Parmigiani, G., Polson N. (1996). Diagnostic Measures for Model Criticism. *Journal of the American Statistical Association* **91** 753-762

Cornelius, P., Crossa, J., Seyedsadr, M. (1995). Statistical Tests and Estimators of Multiplicative Models for Genotype by Environment Interaction. Chapter 8 of Genotype by Environment Interaction. Kang, M. and Gauch, H. eds. CRC Press. Boca Raton.

Elkind, Y., Cahaner, A. (1986). A Mixed Model for the Effects of Single Genes, Polygenes, and their Interaction on Quantitative Traits, I : The Model and Experimental Design. *Theoretical and Applied Genetics* **72** 377-383.

Gelfand, A., Smith, A. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* **85** 398-409

Gollob, H. (1968). A Statistical Model which Combines Features of Factor Analytic and Analysis of Variance Techniques. *Psychometrika* **33** 73-115

Horn, R., Johnson, C. (1985). *Matrix Analysis.* Cambridge University Press. Cambridge.

Isakkson, T., Naes, T. (1988). The Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. *Applied Spectroscopy* **42** 1273-1284.

Kass, R., Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association* **90** 773-795

Mandel, J. (1969). The Partitioning of Interaction in Analysis of Variance. *Journal of Research of the National Bureau of Standards B* **73** 309-328

Mandel, J. (1971). A New Analysis of Variance Model for Nonadditive Data. *Technometrics* **13** 1-18

Maransinghe, M, Johnson, D. (1982). A Test for Incomplete Additivity in the Multiplicative Interaction Model. *Journal of the American Statistical Association* **77** 869-877.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21** 1087-1092

Methun, B., Satorra, A. (1989). Multilevel Aspects of Varying Parameters in Structural Models. In *Multilevel Analysis of Educational Data.* Bock, R. Ed. 87-99. Academic Press. San Diego.

Oman, S. (1991). Multiplicative Effects in Mixed Model Analysis of Variance. *Biometrika* **78** 729-739

Rao, C. (1967). Least Squares Theory Using an Estimated Dispersion Matrix and its Application to Measurement of Signals. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **1** 355-372.

Schervish, M. (1995). *Theory of Statistics.* Springer Verlag. New York.

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics* **22** 1701-1762

Tukey, J. (1955). Answer to Query 113. *Biometrics* **11** 111-113

van Emden, M. (1971). An Analysis of Complexity. Mathematical Centre Tracts 35, Amsterdam.

Verdinelli, I., Wasserman, L. (1995). Computing Bayes Factors using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association* **90** 614-618

Viele, K. (1996). *Nonparametric Bayesian Methods for Evaluating Fit in Hierarchical Models.* Doctoral Thesis. Department of Statistics. Carnegie Mellon University.

Yochmowitz, M., Cornell, R. (1978). Stepwise Tests for Multiplicative Components of Interaction. *Technometrics* **20**, 79-84.

Zobel, R., Wright, M., Gauch, H. (1988). Statistical Analysis of a Yield Trial. *Agronomy Journal* **80** 388-393

Table 1: Posterior means for the Soybean yield data. The posterior mean of σ^2 is 327^2

Genotype / Env	EVAN	WILK	CHIP	HODG	S200	CORS	WELL	Mean $\mu+\beta_j$	δ_1	δ_2	δ_3
A77	2757	2502	2361	2771	2878	3141	2777	2739	3.45	3.62	5.03
C77	2946	2771	1755	2192	1089	1836	1217	1975	-25.48	-7.15	2.58
V77	1567	1103	2266	2468	2730	2569	2616	2192	21.59	0.06	0.62
V78	1735	1493	1506	2172	2011	2145	1553	1812	3.61	0.27	1.69
A79	3127	2623	2488	3201	3430	2878	2791	2930	2.53	9.14	-5.77
C79	2777	2562	1728	1944	1849	1486	1264	1949	-19.22	8.42	-7.97
G79	2986	2367	2340	3154	2623	3040	2455	2705	-0.06	-4.24	-0.94
R79	1843	1110	1816	2495	2105	1769	2058	1894	9.49	-3.62	-6.79
V79	1083	578	1278	1500	1964	1661	1715	1405	15.26	4.88	1.62
A80	2919	2784	2582	3208	2703	3100	2219	2780	-4.09	-3.17	-1.96
C80	2596	3248	2280	2710	2172	2260	1432	2392	-18.15	2.99	-3.77
G80	3901	3194	3376	4096	3887	4250	3517	3737	4.79	-1.58	0.72
L80	3706	3820	2993	3732	3739	3161	3215	3471	-7.66	10.27	-8.26
D80	1937	1580	1580	2374	1997	2609	1890	1997	4.97	-3.03	5.98
R80	2199	1870	2199	2966	2461	2327	2186	2315	4.64	-4.15	-4.56
V80	2334	2018	1802	1964	1601	2165	1782	1958	-6.81	-3.73	5.93
A81	3033	2609	2636	3013	2831	3611	2959	2953	3.51	-2.81	8.70
C81	3053	3053	1849	2522	1654	2186	1708	2289	-21.39	-2.28	5.29
G81	3322	2892	3208	3840	3383	4028	3235	3409	5.98	-6.25	3.25
L81	2972	2710	2636	2972	2569	2818	2777	2777	-2.84	-2.92	-0.28
D81	2529	1997	2582	3268	2112	2529	2320	2477	1.61	-15.93	-1.50
R81	2038	1385	2347	2798	2616	2966	2757	2417	17.34	-6.11	1.95
V81	3026	3127	2387	2367	2361	2461	2260	2571	-13.35	5.76	5.39
A82	2186	1870	1883	2441	2441	2562	2239	2236	5.24	2.78	3.02
L82	2663	1957	2535	2798	3241	3147	2737	2726	11.26	4.85	-1.38
G82	3652	3295	2724	3712	3901	3921	3322	3492	2.27	9.87	-0.96
V82	2273	1910	1123	1755	1184	1399	1345	1572	-13.45	-0.84	3.71
A83	2582	2125	2018	2313	2192	2058	1970	2182	-4.97	2.39	-5.60
I83	1278	1029	1688	1701	2105	1964	1856	1666	12.76	3.75	-1.51
G83	4499	4015	3329	4620	3564	4062	3625	3947	-9.14	-6.46	2.67
A84	3161	2717	3188	3860	3376	3423	3544	3312	8.45	-5.50	0.00
N84	2993	2603	2630	3221	3369	3201	2703	2950	3.98	6.30	-2.39
C84	3181	2690	2448	3060	2576	2784	2670	2765	-4.72	-2.41	-0.14
I84	1950	1701	2078	2260	2145	2246	1917	2043	3.96	-0.84	-1.58
G84	4015	3329	3329	3961	4277	4015	3692	3790	4.62	7.64	-2.82
Mean $\mu+\alpha_i$	2706	2363	2318	2834	2603	2732	2411	2566			
γ_1	-28.74	-43.23	5.17	3.30	24.07	16.30	23.12		λ_1 4112		
γ_2	0.25	8.29	-9.69	-16.51	25.99	-5.88	-2.46			λ_2 1151	
γ_3	-0.76	2.00	-2.19	-8.20	-14.85	13.90	10.10				λ_3 592

Table 2: Posterior Means for the Tomato Softness Data

	Genotype			$\mu+\beta_j$	δ
	11.609 (7)	14.005 (9)	15.987 (7)	14.905	-0.527
	10.120 (7)	15.615 (8)	18.955 (5)	14.745	0.246
	10.120 (6)	16.839 (10)	19.508 (9)	15.182	0.507
	11.238 (5)	16.940 (3)	17.790 (9)	15.460	0.105
	10.834 (8)	12.963 (8)	17.198 (6)	14.317	-0.569
M	12.442 (6)	15.663 (8)	23.988 (4)	16.239	-0.387
o	11.637 (8)	17.210 (10)	18.262 (8)	15.990	0.182
t	10.292 (5)	14.777 (14)	18.295 (4)	14.568	-0.058
h	10.140 (4)	12.453 (9)	17.824 (5)	13.715	-0.506
e	11.583 (5)	19.256 (10)	21.967 (8)	16.888	0.768
r	11.077 (4)	15.380 (7)	19.479 (8)	15.199	-0.157
	10.810 (5)	17.277 (8)	18.986 (7)	15.600	0.422
P	10.939 (5)	15.083 (7)	19.205 (7)	15.098	-0.092
l	12.415 (5)	17.117 (11)	22.124 (8)	16.684	0.034
a	10.816 (4)	16.620 (14)	19.230 (8)	15.529	0.312
n	9.549 (6)	14.198 (8)	16.804 (7)	13.886	-0.007
t	13.028 (6)	18.432 (10)	23.012 (7)	17.412	0.144
	9.989 (7)	14.755 (13)	16.246 (5)	14.293	0.039
	12.194 (6)	17.304 (9)	20.745 (6)	16.439	0.071
	10.874 (7)	15.307 (12)	16.305 (5)	14.933	-0.058
	11.705 (7)	14.661 (11)	21.055 (6)	15.396	-0.468
$\mu+\alpha_i$	11.118	15.789	19.161	$\mu = 15.356$	
σ_i^2	0.95	4.89	11.91		
γ	-1.104	1.465	-0.360		$\lambda = 3.884$