Handling Outlier in Two-Ways Table by Robust Alternating Regression of FANOVA Models: Towards Robust AMMI Models

Alfian Futuhul Hadi
Jurusan Matematika FMIPA Universitas Jember

ABSTRACT

AMMI (Additive Main Effect Multiplicative Interaction) model for interactions in two-way table provide the major mean for studying stability and adaptability through genotype × environment interaction (GEI), which modeled by full interaction model. Eligibility of AMMI model depends on that assumption of normally independent distributed error with a constant variance. Nowadays, AMMI models have been developed for any condition of MET data which violence the normality, homogeneity assumpion. We can mention in this class of modelling as M-AMMI for mixed AMMI models, G-AMMI for generalized AMMI models. The G-AMMI was handling non-normality i.e categorical response variables using an algorithm of alternating regression. While in handling the non-homogeneity in mix-models sense, one may use a model called factor analytic multiplicative. The development of AMMI models is also to handle any outlier that might be found coincides with non-homogeneity condition of the data. In this paper, we will present of handling outlier in multplicative model by robust approach of alternating regression algorithm.

Keywords: AMMI, G-AMMI, M-AMMI, factor analytic, multiplicative models, alternating regression, robust approach

INTRODUCTION

AMMI (Additive Main Effects and Multiplicative Interaction Analysis) was well described by Gauch (1988, 1992) and Gollob (1968). AMMI may be viewed as a procedure to separate pattern (the G x E interaction) from noise (mean error of treatment mean within trials). This is achieved by PCA, where the first axes (i.e. the axes with the largest eigenvalues) recover most of the pattern, whilst most of the noise ends up in later axes. The pattern can be viewed as the whole G × E effects weighed by an estimate of the pattern-to-noise ratio associated with the respective effect. This pattern-to-noise ratio is a variance component ratio analogue to a repeatability or heritability coefficient (Piepho 1994). Multiplicative models AMMI have been popularised in a fixed model context and a number of applications have been found (Gauch 1988, 1992, Crossa *et al*. 1990). AMMI analysis combines, in a model, additive components for main effects (treatments and environments) and multiplicative components for G × E effects.

AMMI model combines a univariate technique, ANOVA for the main effects and a multivariate technique PCA-principal component analysis, for G × E effects. Crossa (1990) suggests that the use of multivariate techniques permits a better use of information than the traditional regression methods. This models are also give a visually pattern of the main interaction through biplot. The power of multiplicative AMMI model is placed on visualized interactions by using biplot.

The AMMI model represents observations into a systematic component that consists of main effect and interaction effect through multiplication of interactions components, apart from random errors component. Basically, the AMMI analysis combines both additive analysis of variance for the main effect of treatment and analysis of multiple main components uses bilinear modeling for the interaction effect, by using singular value decomposition (SVD) of its interaction matrix (Mattjik & Sumertajaya 2000, Mattjik 2005). Eligibility of AMMI model depends on that assumption of normally independent distributed error with a constant variance. Nowadays, AMMI models have been developed for any condition of MET data which is violences in i.e the normality and homegeneity assumpion. We can mention in this class of medelling as M-AMMI for mixed AMMI models, G-AMMI for generalized AMMI models. The G-AMMI was handling non-normality i.e categorical response variables using an algorithm of alternating regression (Eeuwijk 1995). The G-AMMI with normal data distributited and indentity link function is equal to AMMI models. Some aplications of generalized alternating

regression speificlaly on poisson distribution can be found at Hadi *et al* . (2010).

Handling the non-homogeneity in mix-models sense of AMMI, one may use a model called factor analytic multiplicative (Smith *et al.* 2002). Smith *et al.* 2002 conclude that the factor analitic models is equivalent wtth AMMI mixed models. The development of AMMI models is also to handle any outlier that might be found coincides with non-homogeneity condition of the data. We know that SVD is venurable to the outliers, so that the construction of robust model is a promising subject to be investigated. In this paper, we will present of handling outlier in multiplicative model by robust approach of alternating regression algorithm.

Factor analytic ANOVA (FANOVA) models and its relation to AMMI models

A model concerning the evaluation of several treatments or genotypes in several environments is given by:

$$y_{ij} = \mu + g_i + e_j + ge_{ij} + \varepsilon_{ij}$$

where: μ, *g*, *e*, *ge* and ε are the fixed constant, genotype, environment, genotype x environment interaction and within environment error effects, respectively. The μ and *e* effects can be regarded as fixed and the others as random. In the context of MET data, the factor analysis approach can be used to provide a class of structures for the variance-covariance matrix of g_{ij} (*G*). The model is postulated in terms of the unobservable genotype effects in different environments:

$$ge_{ij} = \sum_{r=1}^{k} \lambda_{jr} f_{ir} + \delta_{ij}$$

where: ge_{ij} : interaction effect of genotype *i* in environment *j*; λ_{jr} : loading for factor *r* in environment *j*; f_{ir} : score for genotype i in factor *r*; δ_{ij}: error representing the lack of fit of the model.

The factor analytic mixed-model is presented according to Smith *et al*. (2002). Applied to *g* genotype effects on *s* environments (Resende & Thompson 2005), the factor analytic model postulates dependence on a set of random hypothetical factors $f_r^{(gx1)}$ (r=1, 2,, . . . $k < s$) . In vector notation, the factor analytic model for these effects is

$$g_s = (\lambda_1 \otimes I_g) f_1 + ... + (\lambda_k \otimes I_g) f_k + \delta$$

where: $\lambda_r^{(sx1)}$: loadings or weights of the factors in environments; $\delta^{(gs\ x1)}$: vector of residuals or lack of fit for the model (also called vector of specific factors).

In a compact way, the model is:

$$ge = \Lambda \otimes I f + \delta_{sg},$$

where:

$$\Lambda^{(sxk)} = [\lambda_1 .... \lambda_k]; f^{(gk\ x1)} = (f_1', \ldots, f_k')'$$

The joint distribution of *f and* δ is given by

$$\begin{pmatrix} f \\ \delta \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_k \otimes I_m & 0 \\ 0 & \psi \otimes I_m \end{pmatrix}\right]$$

where: Ψ=diag(ψ1....ψ. $_p$) ; ψ_i : specific variance for the i^{th} trial. The variance matrix for genotype effects on environments is given by

$$\text{var}[ge_s] = (\Lambda \otimes I_m)\,\text{var}[f](\Lambda' \otimes I_m) + \text{var}(\delta)$$
$$= (\Lambda\Lambda' + \psi) \otimes I_m$$

The model for genotype effects in each environment leads to a model for *G* in which:

$\sigma_{g_{jj}} = \sum_{r=1}^{k} \lambda_{jr}^2 + \psi_j$ genotype variance in environment *j*;

$\sigma_{g_{jj'}} = \sum_{r=1}^{k} \lambda_{jr}\lambda_{j'r}$ genotype covariance between environments *j* and *j*';

$\rho_{g_{jj'}} = \sum_{r=1}^{k} \lambda_{jr}\lambda_{j'r} / \sqrt{(\sum_{r=1}^{k} \lambda_{jr}^2 + \psi_j)(\sum_{r=1}^{k} \lambda_{j'r}^2 + \psi_{j'}}$: genotype correlation between environments *j* and *j*'

The equation for g_s has the form of a (random) regression on *k* environmental covariates λ_1.... λ_k in which all regressions pass through the origin. It may be more appropriate to allow a separate (non-zero) intercept for each genotype. This is equivalent to the model with genotype main effects, g, and a k-factor analytic model for *g* x *e* interaction. Then, the expression for g_s turns to

$$g_s = (1_s \otimes I_g)g + ge = (1_s \otimes I_g)g + (\Lambda \otimes I_g) f + \delta .$$

Vector *g* has mean zero and variance $\sigma_g^2 I$ or $\sigma_g^2 A$, where A is a genetic relationship matrix. The model can be written as

$$g_s = (\sigma_g 1_s \otimes I_g) f_0 + (\Lambda \otimes I_g) f + \delta$$
$$= (\Lambda_g \otimes I_g) f_g + \delta ,$$

where:

$$\Lambda^{s(k+1)} = [\sigma_g \sigma 1_s \Lambda]; f_0 = \frac{g}{\sigma_g}; f_g' = (f_0' f');$$

Thus the model with genotype main effects and a k-factor analytic model for *g* x *e* interactions is a special case of a (*k*+1)-factor analytic genotype effects in each environment, in which the loadings in the first set are

constrained to be equal. The feature that distinguishes equations for g, from standard random multivariate regression problems is that both the covariates and the regression coefficients are unknown and therefore must be estimated from the data.

The AMMI model has become a popular method for analysing MET data. The model is a fixed-effects model with (additive) main effects for genotypes and environments and multiplicative terms for the interaction. The latter are obtained using a singular value decomposition (SVD) of the G× E interactions. Let $\mathbf{U}_{ve}$ denote the $m \times p$ matrix of G × E interactions. In AMMI, $\mathbf{U}_{ve}$ is decomposed as $\mathbf{U}_{ve} = \mathbf{ALB}^{*'}$, where $\mathbf{A}$ and $\mathbf{B}^{*}$ are $m \times t$ and $p \times t$ matrices, such that $\mathbf{A'A} = \mathbf{I}_t = \mathbf{B}^{*'}\mathbf{B}^{*}$, $\mathbf{L} = \text{diag}(l_1 \ldots l_t)$, and t is the rank of $\mathbf{U}_{ve}$.
Defining $\mathbf{B}=\mathbf{B}^{*}\mathbf{L}$, the decomposition can be written as

$$\text{Uve} = \text{AB'} = \sum_{r=1}^{t} a_r b_r'$$

The columns of A (ar of m×1) are called the genotype scores and the columns of B (br of p×1) are the environment loadings. As in factor analysis, the aim of the AMMI approach is to account for structure in the genetic effects using the minimum number, k, of multiplicative terms. Isolation of the first k terms in

$$\text{var}[ge_s] = (\Lambda \otimes I_m)\,\text{var}[f](\Lambda' \otimes I_m) + \text{var}(\delta)$$

Then we have: $\text{Uve} = \sum_{r=1}^{k} a_r b_r' + \sum_{r=k+1}^{t} a_r b_r'$

$$= A_1 B_1' + A_2 B_2'$$

Where A1 and B1 are m × k and p × k matrices, respectively. Thus, in the AMMI model, the G × E interactions are modeled as

$$\text{uve} = (\text{B1} \otimes \text{Im})\text{a} + \text{eg}$$

where a = vec[A1] = ($a_1' \ldots a_1'$) of mk x 1vector and eg of mp x 1 are the residual G × E interactions that remain if not all the t components of the SVD are used. The latter are assumed to be independent with constant variance. There is a clear connection between

$$\text{uve} = (\text{B1} \otimes \text{Im})\text{a} + \text{eg}$$

and the k factor-analytic model for the G × E interactions, namely:

$$\text{uve} = (\Lambda \otimes \text{Im})\text{f} + \delta$$

There is a correspondence between the environment loadings for the two models (B1 and Λ) and the genotype scores (a and f). Thus, the k factor-analytic model is a random-effects analogue of the AMMI model. The model is then a multiplicative model of environment and genotypes coefficients (known as loadings and factorial scores, respectively). Here, the FANOVA models is analogue to AMMI models.

In historical perspectives the FANOVA model is proposed by Gollob in 1968, combines aspects of analysis of variance and factor analysis. Among others, Gabriel (1978) considered models like FANOVA and estimated the unknown parameters using a least squares fit. AMMI then developed afterwards. Cornelius *et al.* 1996 say that Gaugh & Zobel renamed the Gollob FANOVA model as AMMI model.

The RAR estimator

As usual, the n×p data matrix Y contains the individuals (cases, objects) in the rows and the observed variables (characteristics) in the columns. The variables are already standardized to have zero location and unit spread. A factor score is denoted as fil. The ith score vector is given by fi = (fi1, . . . , fik)T, while the jth loading vector is λj = (λj1 , . . . , λjk) T. Both the loading vectors and the score vectors are unknown. Denote by θ = (f1 T, . . . , fn T, λ1 T , . . . , λp T) the vector of all scores and loadings, and let

$$\hat{y}_{ij}(\theta) = \sum_{i=1}^{k} f_{ii}\lambda_{ij} = f_i^T \lambda_j = \lambda_j^T f_i$$

be the fitted value of yij . By choosing θ such that the fitted and the actual values of the data matrix are close together, we define estimates $\hat{f}_i$ the score vectors and $\hat{\lambda}_j$ for the loading vectors. The fitted data matrix $\hat{Y}$ can then be decomposed as $\hat{Y} = \hat{F}\hat{\Lambda}^T$ where the rows of $\hat{F}$ are the estimated scores and the rows of $\hat{\Lambda}^T$ are the estimated loadings.

Observe that the rank of $\hat{Y}$ is at most k < p, while the rank of Y is typically p. The least squares (LS) approach is to minimize the sum of squared residuals:

$$\hat{\theta}_{LS} = \arg\min_{\theta} \sum_{i=1}^{n}\sum_{j=1}^{p} (y_{ij} - \hat{y}_{ij}(\theta))^2$$

The resulting $\hat{Y}$ can be seen as the "best" (in the least squares sense) approximation of the data matrix Y by a rank k matrix.

The Eckart-Young theorem (Gower & Hand 1996) says that this best fit can be obtained by performing a singular value decomposition Y = U*DV* ^{T} of the data matrix. By replacing all singular values in *D* by zero except for the *k* largest ones, one obtains D_k and finally $\hat{Y} = UD_kV^T$. By taking $\hat{F} = \sqrt{n}U$ and $\hat{\Lambda} = VD_k / \sqrt{n}$ we obtain the so-called *Principal Component solution* to the FA problem (cfr. Johnson and Wichern 1998). Moreover, the sample covariance matrix of the estimated score vectors equals $\hat{F}^T\hat{F} = I_k$ which is consistent with the assumption *Cov*(F) = I_k.

It is important to note that the estimates $\hat{F}$ and $\hat{\Lambda}^T$ are only specified up to a linear transformation. Since $\hat{Y} = (\hat{F}T^T)(\hat{\Lambda}T^{-1})^T$ for any non singular *k* by *k* matrix *T*, it follows that $\hat{F}T^T$ and $\hat{\Lambda}T^{-1}$ attain the same value for the objective $\hat{\theta}_{LS}$. However, the fitted values $\hat{Y}$ are uniquely defined. Moreover, if we add the restriction that the estimated covariance matrix of the score vectors needs to be the identity matrix, then the estimates $\hat{F}$ and $\hat{\Lambda}^T$ are specified up to an orthogonal transformation, making the matrix $\hat{\Lambda}\hat{\Lambda}^T$ uniquely defined.

Since the LS criterion gives too much weight to large residuals, a first idea is to use the L_1 criterion (or Least Absolute Deviations criterion) instead, which is known to give a very robust additive fit to two-way tables (Terbeck & Davies 1998). This yields the estimator

$$\hat{\theta}_{L1} = \arg\min_{\theta} \sum_{i=1}^{n}\sum_{j=1}^{p} | y_{ij} - \hat{y}_{ij}(\theta) |$$

For the optimal $\hat{F}$ and $\hat{\Lambda}^T$, it must hold that $\hat{f}_1$ minimizes $\sum_{j=1}^{p} | y_{ij} - f_i^T\hat{\lambda}_j |$ and $\hat{\lambda}_j$ minimizes $\sum_{i=1}^{n} | y_{ij} - f_i^T\hat{\lambda}_j |$

Therefore, instead of minimizing both sums in L_1 criterion at the same time, one fixes an index *j* and scores f_i and selects the λ_j to minimize

$$\sum_{i=1}^{n} | y_{ij} - f_i^T\lambda_j |$$

The above problem is now linear instead of bilinear and can easily be solved with a Least Absolute Deviations regression algorithm. One sees immediately that minimizing $\sum_{i=1}^{n} | y_{ij} - f_i^T\lambda_j |$ consecutively for j = 1, . . . , p corresponds to minimizing L1 criterion for fixed scores. Analogously, for fixed loadings λj, finding the fi minimizing $\sum_{j=1}^{p} | y_{ij} - f_i^T\lambda_j |$ (for each i = 1, . . . , n in turn) corresponds to minimizing L1 criterion when the loadings are given. Alternating $\sum_{i=1}^{n} | y_{ij} - \hat{f}_i^T\lambda_j |$ and $\sum_{j=1}^{p} | y_{ij} - f_i^T\lambda_j |$ leads to an iterative scheme of alternating regressions. Note that the value of the criterion in L1 decreases at each step.

Similar algorithms, but based on alternating classical least squares regressions and generalized linear models, see de Falguerolles & Francis (1992), Gabriel (1998) for generalized bilinear models.

Unfortunately, *L*1 regression is sensitive to leverage points. If outlying score or loading vectors are present, the *L*1 regressions can be heavily influenced by them. By downweighting these leverage points we obtain a weighted *L*1 regression, resulting in the RAR estimator ($\hat{\theta}_{RAR}$) below,

$$\hat{\theta}_{RAR} = \arg\min_{\theta} \sum_{i=1}^{n}\sum_{j=1}^{p} w_i(\theta)v_j(\theta) | y_{ij} - \hat{y}_{ij}(\theta) |$$

One single objective function estimates $\hat{F}$ and $\hat{\Lambda}^T$ simultaneously from the rows and columns of *Y*. The result of $\hat{\theta}_{RAR}$ used Robust Alternating Regressions to compute it. The estimator will not be misled by outlying observations.

The row weights, $w_i(\theta)$ are defined by

$$w_i(\theta) = \min\left(1, \frac{X^2_{k;0,95}}{RD_i^2}\right) \quad \text{for } i = 1, \ldots, n$$

where $\chi^2_{k;0.95}$ is the upper 5% critical value of a chi-squared distribution with *k* degrees of freedom, and

$$RD_i = \sqrt{\left(f_i - T(F)^T C(F)^{-1}\left(f_i - T(F)\right)\right)}$$

for *i* = 1, . . . , *n* are robust distances (Rousseeuw & van Zomeren 1990) computed from the collection of score vectors $F = \{ f_i \mid 1 \le i \le n \}$ *k*-dimensional space. Such weights were used by Simpson *et al.* (1992) and yielded stable results. The robust multivariate location and scatter estimators *T* and *C* are taken as the location and scatter part of the MVE estimator. The MVE estimator was

chosen here since it performs well as an outlier identifier (Becker & Gather 2001). Analogously, the set of column weights v_j is defined using the loading vectors. Note that, since the true loadings and scores are unobserved, w_i and v_j depend on the unknown parameter vector θ.

The RAR algorithm

The RAR estimator can be approximated by an alternating algorithm, as outlined below (Croux *et al*. 2003).

Step 0 is used to obtain invariance with respect to a change of measurement units, the data are first scaled in a robust way:

$$y_{ij} \leftarrow \frac{y_{ij} - med_i\left(y_{ij}\right)}{MAD_i\left(y_{ij}\right)}$$

where MAD stands for the Median Absolute Deviation. Note that orthogonal or affine equivariance properties are not necessary in a factor model. This initial standardization corresponds with a correlation matrix based FA.

Step 1 is starting values. First, a robust principal component analysis (PCA) procedure is performed. The resulting scores are then taken as starting values $\hat{f}_i^{(0)}$ for the factor scores. We use the projection pursuit (PP) based estimator of Li & Chen (1985), implemented as in Croux and Ruiz-Gazen (1996). This PP-based method is fast to compute, can deal with $p > n$, and is highly robust. Moreover, this approach allows one to compute just the first k principal components (the only ones that are needed here), which reduces the computation time even further. Using classical PCA in this first stage would slow down the convergence considerably, and could lead to a nonrobust FA when there are many outliers. Alternatively, one could take several random starting values, which could help to check for a local versus global optimum. But the latter approach will increase computation time significantly. In any case, experiments have shown that the choice of the starting values is not too crucial for finding a good approximation.

Step 2: The iteration process. Now suppose that the iteration process has reached step t (t ≥ 1) of the algorithm, and the $\hat{f}_i^{(t-1)}$ are available.

* First compute the row weights $w_i^{(t)}$, which downweight outliers in the set of estimated score vectors $\left\{ \hat{f}_i^{(t-1)} \mid 1 \le i \le n \right\}$ in $\mathbb{R}^k$.

Then compute $\hat{\lambda}_j^{(t)} = \arg\min_{\lambda \in \mathbb{R}^k} \sum_{i=1}^{n} w_i^{(t)} \mid y_{ij} - \lambda^T \hat{f}_i^{(t-1)} \mid$ for j = 1, . . . , p: In this part of the procedure, one needs to perform an L_1 fit p times (and this will be the case at every iteration step). Note that the loadings are estimated one at a time, which turned out to be more convenient for the implementation of the algorithm. Fortunately, very efficient algorithms for L1 regression exist (Bloomfield & Steiger 1983), so this takes little time. Note that the weights $w_i^{(t)}$, only need to be computed once every iteration step. They require computation of a robust scatter estimator in the factor space, which is usually of a low dimension k.

* We analogously compute column weights $v_j^{(t)}$, which downweight outliers in the set of estimated loading vectors $\left\{ \hat{\lambda}_i^{(t)} \mid 1 \le j \le p \right\}$ in $\mathbb{R}^k$. Then compute $\hat{f}_i^{(t)} = \arg\min_{f \in \mathbb{R}^k} \sum_{j=1}^{p} v_i \mid y_{ij} - f^T \hat{\lambda}_j^{(t)} \mid$ for i = 1, . . ., n.

* The values of the objective function of $\hat{\theta}_{RAR}$ computed for the estimates obtained in step t–1 and step t are compared. If there is no essential difference in the objective function, the iterative process is stopped and we set $\hat{f}_i = \hat{f}_i^{(t)}$ for $1 \le i \le n$ and $\hat{\lambda}_j = \hat{\lambda}_j^{(t)}$ for $1 \le j \le p$. If not, Step 2 is repeated.

Step 3: Orthogonalization. This last step is optional and will not alter the fitted values $\hat{Y} = \hat{F}\hat{\Lambda}^T$. We compute a robust estimator $\hat{\Sigma}_f$ of the covariance matrix of the estimated scores $\left\{ \hat{f}_i \mid 1 \le i \le n \right\}$. Since the scores only have k dimension, where k is small, the matrix $\hat{\Sigma}_f$ can be computed quickly. We compute $\hat{f}_i$ by the reweighted MCD estimator with 25% breakdown value, using the FAST-MCD algorithm of Rousseeuw and van Driessen (1999). The breakdown value 25% for the MCD has been chosen since this combines robustness with efficiency (see e.g. Croux & Haesbroeck 1999). Afterwards we set

$\hat{F} \leftarrow \hat{F}\hat{\Sigma}_f^{-1/2}$ and $\hat{\Lambda} \leftarrow \hat{\Lambda}\hat{\Sigma}_f^{1/2}$

The effect of the above transformation is that the robust covariance matrix of the estimated scores is now an identity matrix, which mimics the model condition $Cov(\hat{F}) = I_k$. Another effect is that the biplot representation of the n cases (see Step 4) will show no correlation structure, as is common practice in the biplot literature (Gower & Hand 1996).

Step 4: Residuals, uniquenesses, biplot. The residuals are obtained as $\hat{\varepsilon}_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \hat{f}_i^T \hat{\lambda}_j$ and can be plotted versus (*i, j*) in the horizontal plane. This residual plot is very useful for detecting outliers. From the residuals the uniquenesses can be estimated as $\hat{\psi}_j = (MAD_j(\hat{\varepsilon}_{ij}))^2$. In the common case k = 2 one can represent the individuals by $(\hat{f}_{i1}, \hat{f}_{i2})$ and the variables by $(\hat{\lambda}_{j1}, \hat{\lambda}_{j2})$ in the same 2D plot, called the biplot.

It also allows to perform alternating regression using other regression estimators, like M-estimators or the highly robust Least Trimmed Squares (LTS) and Least Median of Squares (LMS) estimators. It is even possible to execute the algorithm with the nonrobust Least Squares regression estimator, yielding the same result as the classical approach of Gabriel (1978) based on the singular value decomposition. Alternating regression using the LMS algorithm was already considered by Ukkelberg & Borgen (1993). However, using the LMS yields a very time consuming algorithm. In our experience, the RAR estimator gave the most satisfying factor analysis method with respect to computation time, robustness, and stable convergence of the algorithm. Although no proof of convergence exists, many simulations and examples have shown its good numerical and statistical performance.

The RAR procedure required the choice of several auxiliary robust estimators and a weighting function. Most of these choices are standard, and simulations for other robust choices led to essentially identical results.

RESULTS AND DISCUSSION

Theorical costruction: developing robust AMMI by applying RAR to the FANOVA models

The standard model for a two-way table is the ANOVA model

$$y_{ij} = \mu + g_i + e_j + \delta_{ij}$$

where μ is called the overall mean, g_i represents genotype effect (the row) and e_j environment (the column) effect. In a classical setup, the row and column effects are assumed to have zero mean. The terms δ_{ij} can either be seen as residuals or as interaction terms between rows and columns. That expression is called an additive model. It is however quite possible that the interaction terms δ_{ij} still contain some structure that can be described by a factor model $\delta_{ij} = \sum_{r=1}^{k} \lambda_{jr} f_{ir} + \varepsilon_{ij}$ yielding the overall model

$$y_{ij} = \mu + g_i + e_j + \sum_{r=1}^{k} \lambda_{jr} f_{ir} + \varepsilon_{ij}$$

The first idea would be to proceed sequentially by estimating the additive model first, and afterwards performing a factor analysis on the residuals. But better fits can be obtained by estimating all parameters jointly. For the least squares fit there is no difference between the simultaneous and the sequential approach, but this is no longer true for the robust fits. Therefore we will estimate additive and multiplicative terms simultaneously.

The RAR estimator for the FANOVA model was defined in previuos section. Denote θ as the vector collecting the scores, loadings, row and column effects and the overall effect μ. In order to estimate the $(k + 1)(n + p) + 1$ unknown elements of θ from the np available data, we can use the RAR estimator :

$$\hat{\theta}_{RAR} = \arg\min_{\theta} \sum_{j=1}^{p} \sum_{i=1}^{n} v_j(\theta) w_i(\theta) \mid y_{ij} - \hat{y}_{ij}(\theta) \mid$$

The weights w_i and v_j (defined in previous section) are downweighting outlying scores and loadings in the k-dimensional spaces of scores and loadings. To uniquely identify the parameters in y_{ij}, the function $\hat{\theta}_{RAR}$ will be minimized under the constraints:

$$\underset{i}{\text{med}}(g_i) = \underset{j}{\text{med}}(e_j) = 0 \quad \text{and}$$

$$\underset{i}{\text{med}}(f_{i1}) = \underset{j}{\text{med}}(\lambda_{j1}) = 0 \text{ for } i = 1, 2, \ldots, k$$

Those constrains are consisten with robust approach. The algorithm to compute the RAR estimator in FANOVA models is based on alternating regressions, and is almost identical to the iterative scheme outlined before (see The RAR Algorithm section). The only difference is that, it is nolonger working with regression through the origin, intercepts need to be estimated.

Numerical example: fitting robust AMMI models by robust alternating regression

The data we will use is national research concucted by Indonesian National Consortium of Rice (INCR) in cooperation with Indonesian Centre for Rice Research in 2008. The aims of this research was evaluating the phenotypic performance of rice from the latest generation in the different environment.

There are 11 cultivars evaluated at 20 environments. There are 3 cultivars from BATAN,4 cultivars from ICRR, 2 cultivars from Biogen, and 2 cultivars from IPB, with 3 varieties to compare (Ciherang, Inpari1, Cimelati). It used 3 plots of replication for each cultivar in each environment, sized in 4 m x 5 m. We will used two way table of genotype × environment interaction, the cell was median of 3 plots of replication.

There were no sufer outler in the data, only one outlier identified in the row effect (Figure 2). So it can be predicted that the result will be sligthly different with the leatsquare criterion. From Figure 1, it can be say that L4 have relatively largest variance. The genotype which coordinates near the centre point (0,0) is clustered as a stable genotype, so G3, G8, G4, G12, and G14 are relatively more stable than other genotypes.

When there were no sufer outlier then the robust methods give the slightly different result from the least square ones.

Figure 2 shows that only one observation that suspected to be an outlier in column effect. And its downweigthted alraeady as shown in Tabel 1, that was in the L4. The residual plot (Figure 3) shows, that it is randomly distributed by row and column way.

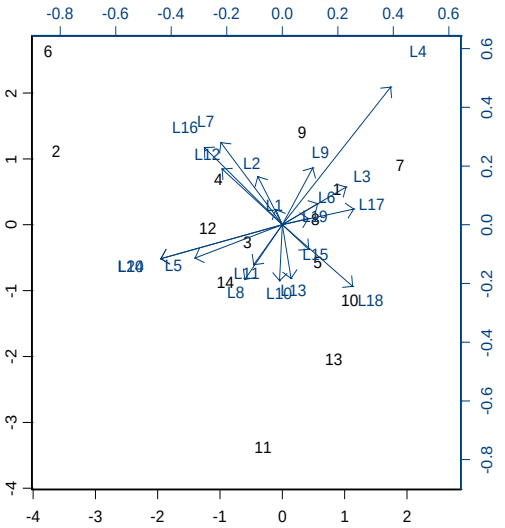

Figure 1. Biplot of G × E interaction of robust FANOVA models . (factor 1 49.43%, cumulative 2 factor 53.76%).

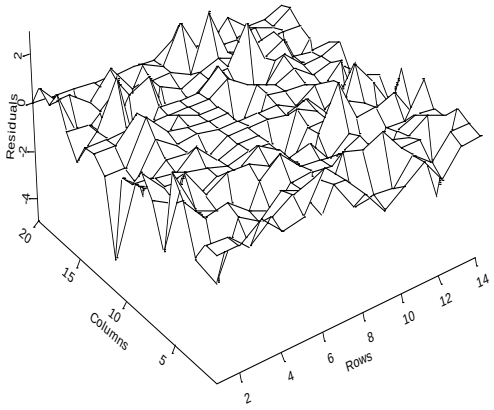

Figure 3. Residual plots of robust FANOVA models.

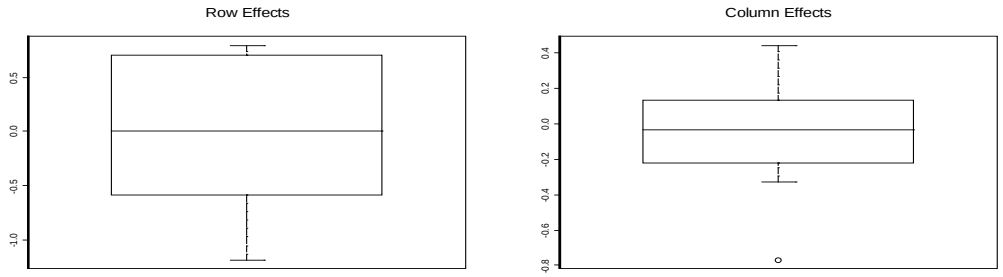

Figure 2. The boxplot of row (environment) and column (genotype) effect in INCR 2008 data.

Tabel 1. Row and column downweigth to the outlier(s).

Row Weight:														
G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Column Weight:														
L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20
1	1	1	0.698925	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

CONCLUSION

FANOVA model of Gollob combines aspects of analysis of variance and factor analysis and estimated the unknown parameters using a least squares fit. AMMI then developed afterwards. Gaugh & Zobel renamed the Gollob FANOVA model as AMMI model. AMMI model combines a univariate technique (ANOVA) for the main effects and a multivariate technique (PCA and SVD) for G × E effects. While in handling the non-homogeneity in mix-models sense, one may use a model called factor analytic multiplicative. Factor analitic models is equivalent with AMMI mixed model. SVD is venurable to the outliers, so that the construction of robust model is a promising subject to be investigated. In this paper, we will present of handling outlier in multiplicative model by robust approach of alternating regression algorithm.

Basically, the idea of constructing robust FANOVA model was to proceed sequentially by estimating the additive effect by robust model first, and afterwards performing a robust factor analysis on the residuals. But this is no longer true for the robust fits. Therefore we will estimate additive and multiplicative terms simultaneously by applying Robust Alternating Regression in FANOVA model. RAR FANOVA model is are downweighting outlying scores and loadings in the *k*-dimensional spaces of scores and loadings, and uniquely identify the parameters in y_{ij}, the function $\hat{\theta}_{RAR}$ will be minimized under the constraints that are consisten with robust approach of the median of parameters. When there were no sufer outlier, then the robust methods give the slightly different result from the least square ones.

Acknowledgment

1. This article is part of research funded by the Ministry of National Education, The Republic of Indonesia, Directorate Generale of Higher Education at the HIBAH PENELITIAN PASCASARJANA, 2008-2010
2. Thanks to Dr. Ir. Aan Darajat, Senior Researcher at Indonesian Center for Rice Research, for the data and the discussions.
3. Also to Prof. Peter Filzmoser. Department of Statistics and Probability Theory. Vienna University of Technology for the S-PLUS scripts and the articles.

REFERENCES

Becker C & Gather U. 2001. The largest nonidentifiable outlier: A comparison of multivariate simultaneous outlier identification rules. *Computational Statistics and Data Analysis*. **3**6:119-127.

Bloomfield P & Steiger WL. 1983. *Least Absolute Deviations: Theory, Applications, and Algorithms*. Boston. Mass, Birkh¨auser.

Cornelius PL, Crossa MS & Seyedsadr. 1996. *Statistical Test and Estimators of Multiplicative Models for GEI. In Genotype by Environment Interaction*, Manjit S. Kang, & HG. Gauch (eds), CRC Press. http://books.google.co.id/books?id=NPqiC-mvRTUC&lpg=PP1&pg=PA199 [10 Jully 2010].

Croux C & Haesbroeck G. 1999. Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis*. **71**:161-190.

Croux C & Haesbroeck G. 2000. Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*. **87**:603-618.

Croux C & Ruiz-Gazen A. 1996. A fast algorithm for robust principal components based on projection pursuit. COMPSTAT 1996. *Proceedings in Computational Statistics* (ed. A. Prat). Heidelberg, Physica-Verlag.

Crossa J. 1990. Statistical Analysis of Multilocation Trials. *Adv. Agron*. **44**: 55-85.

Croux C, Filzmoser P, Pison G & Rousseeuw PJ. 2003. Fitting Multiplicative Models by Robust Alternating Regressions. *Statistics & Coomputin.* **13**: 23-36.

de Falguerolles A & Francis B. 1992. Algorithmic approaches for fitting bilinear models. COMPSTAT 1992, *Proceedings in Computational Statistics*. Vol. **1** (eds. Y. Dodge and J. Whittaker), Heidelberg, Physica-Verlag.

Denis JB & Gower JC. 1996. Asymptotic confidence regions for biadditive models: Interpreting genotype-environment interactions. *Applied Statistics*. **45**:479-493.

Gabriel KR. 1978. Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Statistical Society B.* **40**(2):186-196.

Gabriel KR. 1998. Generalized bilinear regression, *Biometrika*. **85**:689-700.

Gabriel KR & Zamir S. 1979. Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics*. **21**:489-498.

Gauch HG. 1988. Model selection and validation for yield trial with interaction. *Biometrics*. **44**:705-716.

Gollob HF. 1968. A statistical model which combines features of factor analytic and analysis of variance technique. *Psychometrika*. **33**:73-116.

Gower J & Hand D. 1996. *Biplots*. New York, Chapman & Hall.

Hadi AF, Mattjik A & Sumertajaya IM. 2010. Generalized AMMI Models For Assessing The Endurance of Soybean to Leaf Pest. *Jurnal Ilmu Dasar*. **11**(2):151-159.

Johnson RA & Wichern DW. 1998. *Applied Multivariate Statistical Analysis*. 4th edition, New Yersey. Prentice Hall.

Li G & Chen Z. 1985. Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. *Journal of the American Statistical Association*. **80**:759-766.

Mattjik AA & Sumertajaya IM. 2000. *Perancangan Percobaan dengan Aplikasi SAS dan Minitab*. IPB Press. Bogor.

Mattjik AA. 2005. *Interaksi Genotipe dan Lingkungan dalam Penyediaan Suumberdaya Unggul. Naskah Orasi Ilmiah Guru Besar Biometrika*. Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor. Bogor.

Resende MDP & Thompson R. 2004. Factor Analytic Multiplicative Mixed Model in The Analysis of Multiple Experiments. *Rev. Mat. Estat. Sao Paulo*. **22**(2).

Rousseeuw PJ & Van Driessen K. 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*. **41**:212-223.

Smith A, Cullis B, Luckett D, Hollamby G & Thompson R. 2002. Exploring genotype-environment data using random effects AMMI models with adjustments for spatial field trend: Part 1-2: Theory-Aplications. In. Kang. M.S. (Ed). Quantitative Genetics. *Genomics and Plant Breeding*. CAB International. UK.

Terbeck W & Davies P. 1998. Interactions and outliers in the two-way analysis of variance. *The Annals of Statistics*. **26**:1279-1305.

Ukkelberg °A & Borgen O. 1993. Outlier detection by robust alternating regressions. *Analytica Chimica Acta*. **277**:489-494.

Van Eeuwijk FA. 1995. Multiplicative Interaction in Generalized Linear Models. *Biometrics*. **51:** 1017–1032.