ASYMPTOTICALLY DISTRIBUTION FREE TESTS IN HETEROSCEDASTIC UNBALANCED HIGH DIMENSIONAL ANOVA

Haiyan Wang and Michael G. Akritas

Kansas State University and The Pennsylvania State University

Abstract: In this paper, we develop the asymptotic theory for hypotheses testing in high-dimensional analysis of variance (HANOVA) when the distributions are completely unspecified. Most results in the literature have been restricted to observations of no more than two-way designs for continuous data. Here we formulate the local alternatives in terms of departures from the null distribution so that the responses can be either continuous or categorical. The asymptotic theory is presented for testing of main factor and interaction effects of up to order three in unbalanced designs with heteroscedastic variances and arbitrary number of factors. The test statistics are based on quadratic forms whose asymptotic theory is derived under non-classical settings where the number of variables is large while the number of replications may be limited. Simulation results show that the present test statistics perform well in both continuous and discrete HANOVA in type I error accuracy, power performance, and computing time. The proposed test is illustrated with a gene expression data analysis of *Arabidopsis thaiana* in response to multiple abiotic stresses.

Key words and phrases: Asymptotic distribution theory of quadratic forms, local alternatives, Neymann-Scott problem, nonparametric hypotheses, projection method.

1. Introduction

As data gathering technology in various fields advances, large amount of data become increasingly available. A typical example arises in the significance analysis of genes using expression data. Numerous experiments were done using a model plant *Arabidopsis thaliana* under different conditions, and the gene expression data are available from various websites. With all these data produced, effective statistical methods are essential for biologists to identify genes involved in many biological processes. Due to the high cost in array experiments and the large number of experimental conditions (such as many biotic, abiotic stresses, and pathogen infections), the number of replicates is typically very small. A particular example we analyze concerns 18 abiotic stress conditions using leaf or root tissue at multiple time points (see Section 3.4). Only

two replications are available and the gene expression data are highly skewed and heteroscedastic. To properly analyze such data, we develop a nonparametric version of high-dimensional analysis of variance (HANOVA) that allows a large number of factors or factor levels with small replications.

The terminology HANOVA was introduced by Fan and Lin (1998) to compare multiple sets of curves. To develop general HANOVA test procedures requires asymptotic techniques under non-classical settings in which the number of variables, p, approaches infinity. In high-dimensional designs, the p variables correspond to the factor-level combinations (also called groups or cells). The challenges posed to non-classical asymptotic settings have been addressed in the literature (cf., Neymann and Scott (1948), Portnoy (1988), Fan (1996) Sarandasa and Altan (1998), Johnstone (2001), Li, Lindsay, and Waterman (2003), Inglota and Ledwina (2005)).

In spite of the considerable amount of research that has been done, inference in many settings with high-dimensional designs is still not adequately developed. The discrete case deserves special attention as results for designs with high-dimensional discrete data are very limited. Generalized linear models (GLM) are the well-accepted methods for discrete data. The response is related to a linear combination of the predictors through a link function. GLM relies heavily on the asymptotic properties of the MLE as the sample size approaches infinity while the number of factor levels remains fixed. For log-linear models of Poisson or multinomial count data, the asymptotic properties of MLE and chi-square statistics from GLM still hold when both the sample size and the number of cells are large (Haberman (1977)). Parametric inferences on other discrete data under the nonclassical settings are not available.

Recently, a few authors have considered asymptotically distribution-free procedures for factorial designs with a large number of factor levels. Boos and Brownie (1995), Akritas and Arnold (2000), Bathke (2002), and Bathke (2004) are mainly focused on showing that the ANOVA F-statistics can still be used under homoscedastic or a special form of heteroscedastic variances when there are a large number of treatments for balanced design. Akritas and Papadatos (2004) considered heteroscedastic one-way layout; Wang and Akritas (2006) gave results for heteroscedastic two-way designs; Wang and Akritas (2004) and Bathke and Harrar (2008) considered methods based on ranks. None of the references above allow a combination of unbalanced design with arbitrary number of factors in the presence of arbitrary heteroscedastic variances. In addition, none of these references provide testing procedures for the main effects or interactions of factors that contain only a small number of levels while other factors have a large number of levels.

Though the results and methods in the aforementioned references are distribution-free under the null hypotheses, they do not provide inference for discrete data under local alternatives under the non-classical setting. For example, Akritas and Papadatos (2004) and Wang and Akritas (2006) used a representation of the observations under the local alternatives as the observations under the null plus an integral of a continuous function. Specifically, both references considered local alternatives of the form $Y_{ij} = X_{ij} + \alpha_i(a)$, where X_{ij} is a sequence of random variables satisfying the null hypothesis and $\alpha_i(a) = a^{3/4} n(a)^{-1/2} \int_{(i-1)/a}^{i/a} g(t) dt$, with $g(t)$ being a continuous function on $[0, 1]$. Such a formulation is not applicable to discrete random variables, such as Bernoulli where data for both X_{ijk} and Y_{ijk} take values 0 or 1.

In this paper, we consider the nonparametric theory of hypothesis testing for general HANOVA allowing unbalanced and heteroscedastic independent data. Since the results under the null hypothesis follow from those under local alternatives, we state our theory for local alternatives directly. We formulate the local alternatives through departures from the distributions under the null. Such formulation of local alternatives is more flexible and general than location alternatives. It pertains, for the first time, to both continuous and discrete data. Moreover, the alternative and null distributions need not belong to the same family of distributions. See Section 2.2 for further discussion. A comprehensive set of asymptotically distribution-free procedures for testing of both main effects and interactions of up to order three is developed.

The asymptotic theory in the current setting is substantially different from the classical one. For example, for testing the hypothesis about a parameter $\boldsymbol{\theta}$ of dimension $d > 1$ against unrestricted alternatives, an asymptotically uniformly most powerful invariant test (AUMPI) under a group of nonsingular linear transformations of the whole vector can be constructed for Pitman's moving alternatives. The parameters under the alternatives approach those under the null hypothesis at a rate related to the sample size increases while the dimension stays fixed. However, the invariance methodology does not extend to hypotheses about an infinite-dimensional parameter of interest, since by considering every finite-dimensional projection of local departures, the only test that satisfies the invariance requirement is the trivial test or its asymptotic equivalents (Choi, Hall, and Schick (1996)). In addition, the methods developed for the traditional setting, in which neither the number of parameters of interest nor the number of nuisance parameters is large, tend to have low power in the current setting. See Fan and Lin (1998) for an illustration of the power loss for likelihood ratio test with Gaussian data. For data from the Gamma distribution as given in our simulation study, the test in Brunner, Dette, and Munk (1997) showed accurate type I error at level 0.01 but only about half the power of the proposed test,

even though their test demonstrated comparable power to the Wald-type test for most of their small sample sizes simulations in the classical setting.

The rest of the paper is organized as follows. Section 2 gives the data representation, the hypotheses, the test statistics, and their asymptotic results. Numerical results, including simulations and an application in stress response gene expression data, are presented in Section 3. Finally the proofs are given in Section 4. Throughout the paper, we use replication, group sample size, or cell sample size interchangeably.

2. Test Statistics and Their Asymptotic Distributions

2.1. Construction of test statistics

We consider testing for no main effects, and no two- and three-way interactions with the possible presence of additional factors. The treatment of these testing problems captures all ideas and techniques that would be required for developing test procedures for still higher-way interactions. In order to achieve as concise a formulation as possible we combine all factors that do not contribute parameters to the testing problem. Thus, for testing of no main effects, it suffices to consider a surrogate two-way design where the second factor encompasses the levels of all other factors. Similarly, for a general testing theory of no two-way interaction effects, it is sufficient to consider a surrogate three-way design where the additional factor encompasses the levels of all other factors. Since we consider testing of no more than three-way interactions, we employ no more than five indices to describe the observations (one index for each of the factors of the surrogate design and one index for the replications).

Score statistics derived from estimating equations are commonly used with both discrete and continuous data. However, the inferences based on score statistics rely heavily on the specified relationship between the mean and variance function. When the specified relationship holds and the model has only a small number of fixed effects, score statistics are efficient. However, when the number of fixed effects increases to infinity, as is the case in this paper, the number of estimating equations approaches infinity. With only a small number of replications, the estimated parameters from these large number of estimating equations are not consistent.

Another class of statistics, rank statistics, have certain advantage such as robustness to outliers. In the present context, however, this advantage comes at the expense of efficiency. For example, when only a small proportion (such as 5%) of observations come from distributions with higher mean values and the rest of the observations have common mean, the simulations in Section 3.1 suggest that the rank statistics have low power.

In this paper we propose two different types of statistics based on the original observations: a Wald-type statistic, applicable to hypotheses regarding a

small number of parameters, and an F-type statistic, applicable to hypotheses regarding a large number of parameters.

The proposed F-type statistics differ from the classical statistics in two ways.

- Use of unweighted means instead of weighted ones. For example, in an ANOVA with three factors, we observe X_{ijkm}, $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, c$, $m = 1, \ldots, n_{ijk}$. The unweighted mean for each i is $\widetilde{X}_{i...} = (bc)^{-1} \sum_{j,k} \overline{X}_{ijk.}$, while the weighted mean is $\overline{X}_{i...} = (\sum_{j,k} n_{ijk})^{-1} \sum_{j,k} \sum_{m=1}^{n_{ijk}} X_{ijkm}$. Similarly, the overall unweighted mean is $\widetilde{X}_{....} = (abc)^{-1} \sum_{i,j,k} \overline{X}_{ijk.}$, while the overall weighted mean is $\overline{X}_{....} = (\sum_{i,j,k} n_{ijk})^{-1} \sum_{i,j,k,m} X_{ijkm}$. The idea of using unweighted means is related to the statistics used by Yates (1934) for unbalanced homoscedastic two-way designs; see also (Sahai and Ageel (2000, pp.220-222)).

- Summation over the replications is not included in our quadratic forms for the numerators of F, and the denominators are adjusted so that the numerator and denominator have the same expectation under corresponding null hypothesis. For example, in the classical three-way ANOVA, the numerator of F for testing of no main row effect is $MST = (a-1)^{-1} \sum_{i,j,k} n_{ijk} (\overline{X}_{i...} - \overline{X}_{....})^2$. Here we use $MST_A = (a-1)^{-1} \sum_{i,j,k} (\widetilde{X}_{i...} - \widetilde{X}_{....})^2$.

Analogous versions of the above notation for means is also used when there are two or four factors in the model. Moreover, similar notation applies when X is replaced by e.

2.2. Testing of no main effects

As mentioned earlier, the data representation for a general theory of testing for no main effects of factor A can use a surrogate design with two factors, where factor B encompasses the level combinations of all other factors. Thus, let X_{ijk}, $k = 1, \ldots, n_{ij}$, be independent observations from some unknown distribution F_{ij}, $i = 1, \ldots, a$, $j = 1, \ldots, b$. Consider the decomposition $E(X_{ijk}) = \mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$, where $\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} (\alpha\beta)_{ij} = \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0$. We want to test the hypothesis $H_0(A) : \alpha_i = 0$, for all i. Note that the decomposition of the expected values is equivalent to generalized linear models using dummy variables for factors. The difference is that generalized linear models also need to specify the distribution and variance function beyond the model for the expected values. The variance and distribution of data are here left unspecified as long as certain moment conditions, which are specified in the theorems, are satisfied.

When a is small we only consider the case that b is large since, if b also small, the asymptotic theory requires a large number of replications in each cell and

belongs to the well-studied classical framework. In the small a and large b case, the test statistic is

$$Q_X(A) = N\mathbf{W}'\mathbf{C}_A'(\mathbf{C}_A\widehat{\mathbf{V}}\mathbf{C}_A')^{-1}\mathbf{C}_A\mathbf{W}, \qquad (2.1)$$

where $\mathbf{W} = (\widetilde{X}_{1..}, \ldots, \widetilde{X}_{a..})'$, $\mathbf{C}_A = (\mathbf{1}_{a-1}| - I_{a-1})$, $\widehat{\mathbf{V}} = diag(\hat\eta_1, \ldots, \hat\eta_a)$, and

$$\hat\eta_i = \frac{N}{b^2}\sum_{j=1}^{b}\frac{S_{ij,X}^2}{n_{ij}} \text{ with } S_{ij,X}^2 = (n_{ij}-1)^{-1}\sum_{k=1}^{n_{ij}}(X_{ijk} - \overline{X}_{ij.})^2.$$

When a is large, regardless of whether b is small or large, the statistic is $F_X(A)$, where

$$F_X(A) = \frac{MST_A}{MSE}, \ MST_A = \frac{1}{a-1}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\widetilde{X}_{i..} - \widetilde{X}_{...}\right)^2, \ MSE = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{S_{ij,X}^2}{n_{ij}}.$$
$$(2.2)$$

The next theorem is derived under a sequence of local alternatives of the form

$$F_{R,ij}(x) = F_{0,ij}(x) + R(a, b, N)A_i(x), \qquad (2.3)$$

where the means $\mu_{0,ij}$ of $F_{0,ij}(x)$ satisfy the null hypothesis and the $A_i(x)$ are the deviations, at the distribution level, from the null hypothesis. This formulation implies local alternatives of the form of $\mu_{R,ij} = \mu_{0,ij} + R(a, b, N)\int x dA_i(x)$, where integration is in the Lebesgue-Stieltjes sense. We assume that the $\int xd\, A_i(x)$ are uniformly bounded for all i. With this formulation, the distribution and variances under the alternatives can be different from those under the null, as is necessary in some cases. For example, when the data are from Gamma distribution with shape parameter α and scale parameter β, the mean of the distribution is $\alpha\beta$ and the variance is $\alpha\beta^2$. It is often impractical to consider location alternatives alone since the variance also changes with α and β. Similarly, under the Binomial, Poisson, or negative binomial distributions, the variance is a function of the mean and thus a shift in the mean entails a change in the variance. The alternatives described in (2.3) allow changes in the variance, or other aspects of the distribution, as the mean shifts from the null hypothesis, whereas the traditional location alternatives do not.

Theorem 2.1.
(a) *Let $Q_X(A)$ be given in (2.1). Consider the local alternatives (2.3) with $R(a, b, N) = N^{-1/2}$, and let $\sigma_{R,ij}^2$ be the variance of $F_{R,ij}(x)$. Assume that for all i,*

$$\left(\frac{1}{b}\sum_{j=1}^{b}\frac{\sigma_{R,ij}^2}{n_{ij}}\right)^{-2}\frac{1}{b}\sqrt{\frac{1}{b}\sum_{j=1}^{b}\frac{1}{n_{ij}^6}} \to 0,$$

and

$$\limsup_{b\to\infty} b^{-1} \sum_{j=1}^{b} E^2 (X_{ijm} - E(X_{ijm}))^4 < \infty. \tag{2.4}$$

If a remains bounded then, regardless of whether the $n_{ij} \geq 2$ stay bounded or tend to ∞,

$$Q_X(A) \xrightarrow{p} \chi^2_{a-1}(\delta_{11}) \ as \ b \to \infty,$$

where $\delta_{11} = \lim_{b\to\infty} \mathbf{g}' \mathbf{C}'_A (\mathbf{C}_A \widehat{\mathbf{V}} \mathbf{C}'_A)^{-1} \mathbf{C}_A \mathbf{g}$ with $\mathbf{g} = (\int x dA_1(x), \ldots, \int x dA_a(x))'$.

(b) *Let $F_X(A)$ be the statistic given in (2.2), and consider the local alternative in (2.3) with $R(a, b, N) = a^{-1/4} b^{-1/2}$. Further, let*

$$\tau_A = \frac{2}{ab^2} \sum_{i=1}^{a} \left(\sum_{j=1}^{b} \frac{\sigma^2_{R,ij}}{n_{ij}} \right)^2,$$

$$\sigma^2_A = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{\sigma^2_{R,ij}}{n_{ij}},$$

$$\phi_A = \frac{2}{ab^2} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{\sigma^4_{R,ij}}{n^2_{ij}(n_{ij} - 1)},$$

$$\delta_{12} = (a-1)^{-1} \sum_{i=1}^{a} \left[\int x dA_i(x) - \int x d\overline{A}.(x) \right]^2.$$

Assume that $\limsup(ab)^{-1} \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}^{-1} E[X_{ijm} - E(X_{ijm})]^4 < \infty$. Then as $a \to \infty$,

$$\sqrt{a} \left(F_X(A) - 1 \right) \xrightarrow{d} N \left(\lim_{a\to\infty} \frac{\delta_{12}}{\sigma^2_A}, \quad \lim_{a\to\infty} \frac{\tau_A + \phi_A}{\sigma^4_A} \right)$$

regardless of whether b and the $n_{ij} \geq 2$ stay bounded or tend to infinity, provided $\min_{ij}\{n_{ij}\}$ and $\max_{ij}\{n_{ij}\}$ are of the same order.

Remark 2.1. If the fourth moments of the X_{ijk} are uniformly bounded, the first part of condition (2.4) becomes

$$\left(\frac{1}{b} \sum_{j=1}^{b} \frac{\sigma^2_{R,ij}}{n_{ij}} \right)^{-2} \frac{1}{b^2} \sum_{j=1}^{b} \frac{1}{n^3_{ij}} \to 0.$$

2.3. Testing of no two-way interaction effects

Let A, B denote the two factors whose interactions are of interest. According to the discussion in Section 2.1, a general testing theory for this testing problem can be given in the context of a three-factor design where the third factor, C, encompasses the level combinations of all other factors. Let X_{ijkm}, $m = 1, \ldots, n_{ijk}$, be independent observations from F_{ijk}, $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, c$. Consider the decomposition

$$E(X_{ijkm}) = \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk},$$

where $\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{k=1}^{c} \gamma_k = \sum_{i=1}^{a}(\alpha\beta)_{ij} = \sum_{i=1}^{a}(\alpha\gamma)_{ik} = 0$, $\sum_{j=1}^{b}(\alpha\beta)_{ij} = \sum_{j=1}^{b}(\beta\gamma)_{jk} = \sum_{k=1}^{c}(\alpha\gamma)_{ik} = \sum_{k=1}^{c}(\beta\gamma)_{jk} = 0$, and $\sum_{i=1}^{a}(\alpha\beta\gamma)_{ijk} = \sum_{j=1}^{b}(\alpha\beta\gamma)_{ijk} = \sum_{k=1}^{c}(\alpha\beta\gamma)_{ijk} = 0$. We want to test the hypothesis $H_0(AB)$: $(\alpha\beta)_{ij} = 0$, for all i, j.

When a, b are small and c is large, the test statistic is

$$Q_X(AB) = N\mathbf{W}'\mathbf{C}'_{AB} \left(\mathbf{C}_{AB}\widehat{\mathbf{V}}\mathbf{C}'_{AB} \right)^{-1} \mathbf{C}_{AB}\mathbf{W}, \tag{2.5}$$

where $\mathbf{W} = (\widetilde{X}_{11..}, \ldots, \widetilde{X}_{1a..}, \ldots, \widetilde{X}_{a1..}, \ldots, \widetilde{X}_{ab..})'$, $\mathbf{C}_{AB} = \mathbf{M}_a \otimes \mathbf{M}_b$, where $\mathbf{M}_b = (\mathbf{1}_{b-1}| - I_{b-1})$, and $\widehat{\mathbf{V}} = \text{diag}\{\widehat{\eta}_{11}, \ldots, \widehat{\eta}_{1b}, \ldots, \widehat{\eta}_{b1}, \ldots, \widehat{\eta}_{ab}\}$ with $\widehat{\eta}_{ij} = (N/c^2)\sum_{k=1}^{c} S_{ijk,X}^2/n_{ijk}$.

When at least one of a, b are large, regardless of whether or not c is large, the test statistic is based on $F_X(AB)$, where

$$F_X(AB) = \frac{MST_{AB}}{MSE}, \; MST_{AB} = c\sum_{i,j} \frac{(\widetilde{X}_{ij..} - \widetilde{X}_{i...} - \widetilde{X}_{.j..} + \widetilde{X}_{....})^2}{(a-1)(b-1)},$$

$$MSE = \frac{1}{abc}\sum_{i,j,k} \frac{S_{ijk,X}^2}{n_{ijk}}. \tag{2.6}$$

We consider a sequence of local alternatives of the form

$$F_{R,ijk}(x) = F_{0,ijk}(x) + R(a, b, c, N)(AB)_{ij}(x), \tag{2.7}$$

where the means of the $F_{0,ijk}$ satisfy $H_0(AB)$ and $(AB)_{ij}$ are the deviations, at the distribution level, from the null hypothesis. Assume that the $\int x\, d\, (AB)_{ij}(x)$ are bounded uniformly for all i, j.

Theorem 2.2.

(a) *Let $Q_X(AB)$ be given at (2.5), and consider the alternatives given at (2.7) with $R(a,b,c,N) = N^{-1/2}$. Let $\sigma_{R,ijk}^2$ be the variance of $F_{R,ijk}$, and assume that $\forall i, j$,*

$$\left(\frac{1}{c}\sum_{k=1}^{c}\frac{\sigma_{R,ijk}^2}{n_{ijk}}\right)^{-2}\frac{1}{c}\sqrt{\frac{1}{c}\sum_{k=1}^{c}\frac{1}{n_{ijk}^6}}\to 0,\ \limsup_{c\to\infty} c^{-1}\sum_{k=1}^{c}E^2[(X_{ijkm}-E(X_{ijkm}))^4]<\infty.$$

(2.8)

Then as $c \to \infty$, if a, b stay bounded, regardless of whether the $n_{ijk} \geq 2$ are large or small,

$$Q_X(AB)\xrightarrow{d}\chi^2_{(a-1)\times(b-1)}(\delta_{21}),\ \text{where}\ \delta_{21}=\lim_{c\to\infty}\mathbf{G}'\mathbf{C}'_{AB}\Big(\mathbf{C}_{AB}\widehat{\mathbf{V}}\mathbf{C}'_{AB}\Big)^{-1}\mathbf{C}_{AB}\mathbf{G},$$

and $\mathbf{G} = \left(\int x d\,(AB)_{11}(x),\ldots,\int x d\,(AB)_{ab}(x)\right)'$.

(b) *Let $F_X(AB)$ be given at (2.6), and consider the local alternative given at (2.7) with $R(a,b,c,N) = (ab)^{-1/4}c^{-1/2}$. Assume $\limsup (abc)^{-1}\sum_{i,j,k} n_{ijk}^{-1}E[X_{ijkm} - E(X_{ijkm})]^4 < \infty$. Let $\tau_{AB} = (\tau_1^2 + \tau_2^2 + \tau_3^2)^{1/2}/\sigma_{AB}^2$, where*

$$\tau_1^2 = \frac{2}{b^2 c^2 a}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{\sigma_{R,ijk}^4}{n_{ijk}^2(n_{ijk}-1)},$$

$$\tau_2^2 = \frac{2(b-2)}{ab(b-1)^2c^2}\sum_{i=1}^{a}\sum_{j=1}^{b}\Big(\sum_{k=1}^{c}\frac{\sigma_{R,ijk}^2}{n_{ijk}}\Big)^2,$$

(2.9)

$$\tau_3^2 = \frac{2}{ab^2(b-1)^2c^2}\sum_{i=1}^{a}\Big(\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{\sigma_{R,ijk}^2}{n_{ijk}}\Big)^2,$$

$$\sigma_{AB}^2 = \frac{1}{abc}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{\sigma_{R,ijk}^2}{n_{ijk}}.$$

(2.10)

Then, regardless of whether $n_{ijk} \geq 2$ and c are large or small provided that $\min_{ijk}\{n_{ijk}\}$ and $\max_{ijk}\{n_{ijk}\}$ are of the same order,

$$\sqrt{ab}(F_X(AB)-1)\xrightarrow{d}N\Big(\lim_{\max\{a,b\}\to\infty}\frac{\delta_{22}}{\sigma_{AB}^2},\ \lim_{\max\{a,b\}\to\infty}b\tau_{AB}^2\Big),\ \text{as}\ \max\{a,b\}\to\infty,$$

where

$$\delta_{22} = \frac{1}{(a-1)(b-1)}\sum_{i,j}\Big[\int x d(AB)_{ij}(x) - \int x d\overline{(AB)}_{i\cdot}(x)$$

$$- \int x d\overline{(AB)}_{\cdot j}(x) + \int x d\overline{(AB)}_{\cdot\cdot}(x)\Big]^2.$$

Remark 2.4. Consistent estimators of τ_i^2, $i = 1, 2, 3$, and σ_{AB}^2 can be obtained by replacing $\sigma_{R,ijk}^2$ with the sample variances, and replacing $\sigma_{R,ijk}^4$ by an unbiased estimator $\widehat{\sigma}_{R,ijk}^4$. In cases that at least four replications are available, $\widehat{\sigma}_{R,ijk}^4 = [4P_{n_{ij}}^4]^{-1} \sum_{k_1 \neq k_2 \neq k_3 \neq k_4} (X_{ijk_1} - X_{ijk_2})^2 (X_{ijk_3} - X_{ijk_4})^2$ can be used, where $P_{n_{ij}}^4$ is the number of permutations of size 4 from n_{ij} distinct objects. The proof can be achieved by showing that the variance of the estimator of τ_i^2 converges to 0 and the expectation converges to τ_i^2. This is straightforward by noticing that the sample variances and $\widehat{\sigma}_{R,ijk}^4$ are unbiased estimators of $\sigma_{R,ijk}^2$ and $\sigma_{R,ijk}^4$, respectively (see the proof of Lemma A.4 for similar arguments). When some replications are less than four but at least two, unbiased estimation of $\sigma_{R,ijk}^4$ can be achieved by exploiting patterns of heteroscedasticity, e.g., $\sigma_{R,ijk} = \sigma_{R,jk}$, for all i. In absence of such patterns, we recommend a bias correction technique such as the bootstrap or jackknife. When replications are at least four, the performance with both unbiased estimators and bias correction procedures are evaluated in simulation studies that yield very close results. It should be noted that the computation using bias corrected estimators of the $\sigma_{R,ijk}^4$ is much faster than that using unbiased estimators.

2.4. Testing of no three-way interaction effects

Let A, B and C denote the three factors whose interactions are of interest. According to the discussion in Section 2.1, a general testing theory for this testing problem can be given in the context of a four-factor surrogate design where the fourth factor, D, encompasses the level combinations of all other factors. Let X_{ijklm}, $m = 1, \ldots, n_{ijkl}$, be independent observations from F_{ijkl}, $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, c$, $l = 1, \ldots, d$. Consider the decomposition

$$E(X_{ijklm}) = \mu_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\delta)_{il} + (\beta\gamma)_{jk} + (\beta\delta)_{jl}$$
$$+ (\gamma\delta)_{kl} + (\alpha\beta\gamma)_{ijk} + (\alpha\beta\delta)_{ijl} + (\beta\gamma\delta)_{jkl} + (\alpha\gamma\delta)_{ikl} + (\alpha\beta\gamma\delta)_{ijkl},$$

where $\sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{k=1}^c \gamma_k = \sum_{l=1}^d \delta_l = 0$; $\sum_{i=1}^a (\alpha\beta)_{ij} = \sum_{j=1}^b (\alpha\beta)_{ij} = 0$, and similar constraints for all the other two-way interaction effects; $\sum_{i=1}^a (\alpha\beta\gamma)_{ijk} = \sum_{j=1}^b (\alpha\beta\gamma)_{ijk} = \sum_{k=1}^c (\alpha\beta\gamma)_{ijk} = 0$, and similar constraints for all the other three-way interaction effects; and $\sum_{i=1}^a (\alpha\beta\gamma\delta)_{ijkl} = \sum_{j=1}^b (\alpha\beta\gamma\delta)_{ijkl} = \sum_{k=1}^c (\alpha\beta\gamma\delta)_{ijkl} = \sum_{l=1}^d (\alpha\beta\gamma\delta)_{ijkl} = 0$. We want to test the hypothesis $H_0(ABC) : (\alpha\beta\gamma)_{ijk} = 0$ for all i, j, k.

When a, b, c are small and d is large, the test statistic is

$$Q_X(ABC) = N\mathbf{W}'\mathbf{C}_{ABC}' \left(\mathbf{C}_{ABC} \widehat{\mathbf{V}} \mathbf{C}_{ABC}' \right)^{-1} \mathbf{C}_{ABC}\mathbf{W}, \qquad (2.11)$$

where $\mathbf{W} = (\widetilde{X}_{111..}, \ldots, \widetilde{X}_{11c..}, \widetilde{X}_{121..}, \ldots, \widetilde{X}_{12c..}, \ldots, \widetilde{X}_{ab1..}, \ldots, \widetilde{X}_{abc..})'$, $\mathbf{C}_{ABC} = \mathbf{M}_a \otimes \mathbf{M}_b \otimes \mathbf{M}_c$ with $\mathbf{M}_a = (\mathbf{1}_{a-1} | -I_{a-1})$, and $\widehat{\mathbf{V}} = diag\{\widehat{\eta}_{111}, \ldots, \widehat{\eta}_{11c}, \widehat{\eta}_{121}, \ldots, \widehat{\eta}_{12c}, \ldots, \widehat{\eta}_{ab1}, \ldots, \widehat{\eta}_{abc}\}$ with $\widehat{\eta}_{ijk} = (N/d^2) \sum_{l=1}^d (S_{ijkl,X}^2/n_{ijkl})$.

When at least one of a, b, c is large, and regardless of whether d is small (including zero, as is the case in a three-way design) or large, the test statistic is $F_X(ABC) = MST_{ABC}/MSE$, where

$$MST_{ABC} = \sum_{i,j,k,l} \frac{\left(\widetilde{X}_{ijk..} - \widetilde{X}_{ij...} - \widetilde{X}_{i.k..} - \widetilde{X}_{.jk..} + \widetilde{X}_{i....} + \widetilde{X}_{.j...} + \widetilde{X}_{..k..} - \widetilde{X}_{.....}\right)^2}{(a-1)(b-1)(c-1)},$$

$$MSE = \frac{1}{abcd} \sum_{i,j,k,l} \frac{S_{ijkl,X}^2}{n_{ijkl}}.$$

We consider a sequence of local alternatives of the form

$$F_{R,ijkl}(x) = F_{ijkl}(x) + R(a,b,c,d,N)(ABC)_{ijk}(x), \qquad (2.12)$$

where the means of the $F_{0,ijkl}$ satisfy $H_0(ABC)$ and $(ABC)_{ijk}$ are the deviations, at the distribution level, from the null hypothesis. Assume that the $\int x\, d\,(ABC)_{ijk}(x)$ are uniformly bounded for all i,j,k.

Theorem 2.3.

(a) *Let $Q_X(ABC)$ be the statistic given at (2.11), and consider the alternative given at (2.12) with $R(a,b,c,d,N) = N^{-1/2}$. Let $\sigma_{R,ijkl}^2$ be the variance of $F_{R,ijkl}$, and assume that $\forall i,j,k$,*

$$\left(\frac{1}{d} \sum_{l=1}^{d} \frac{\sigma_{R,ijkl}^2}{n_{ijkl}}\right)^{-2} \frac{1}{d} \sqrt{\frac{1}{d} \sum_{l=1}^{d} \frac{1}{n_{ijkl}^6}} \to 0,$$

$$\limsup_{d\to\infty} \frac{1}{d} \sum_{l=1}^{d} E^2[(X_{ijklm} - E(X_{ijklm}))^4] < \infty, \qquad (2.13)$$

Then as $d \to \infty$, if a, b, c stay bounded, regardless of whether the $n_{ijk} \geq 2$ are large or small,

$$Q_X(ABC) \xrightarrow{d} \chi_{(a-1)\times(b-1)\times(c-1)}^2(\delta_{31}),$$

$$\text{where } \delta_{31} = \lim_{d\to\infty} \mathbf{U}'\mathbf{C}_{ABC}'\left(\mathbf{C}_{ABC}\widehat{\mathbf{V}}\mathbf{C}_{ABC}'\right)^{-1}\mathbf{C}_{ABC}\mathbf{U},$$

with $\mathbf{U} = \left(\int x\, dC_{111}(x), \ldots, \int x\, dC_{abc}(x)\right)'$.

(b) *Consider the local alternative given at (2.12) with $R(a,b,c,d,N) = (abc)^{-1/4}d^{-1/2}$. Assume $\limsup(abcd)^{-1}\sum_{i,j,k,l} n_{ijkl}^{-1}E[X_{ijklm} - E(X_{ijklm})]^4 < \infty$. Let $\tau_{ABC1} = (\tau_4 + \tau_5 + \tau_6 + \tau_7 + \tau_8)^{1/2}/\sigma_{ABC}^2$, where*

$$\tau_4 = \frac{2}{ac^2b^2d^2} \sum_{i,j,k,l} \frac{\sigma_{R,ijkl}^4}{n_{ijkl}^2(n_{ijkl}-1)}, \qquad \tau_5 = \frac{2(b-2)}{ac^2d^2b(b-1)^2} \sum_{i,j,k} \left(\sum_{l=1}^{d} \frac{\sigma_{R,ijkl}^2}{n_{ijkl}}\right)^2,$$

$$\tau_6 = \frac{2}{ac^2 b^2 d^2 (b-1)^2} \sum_{i,k} \left(\sum_{j,l} \frac{\sigma^2_{R,ijkl}}{n_{ijkl}} \right)^2, \quad \sigma^2_{ABC} = \frac{1}{abcd} \sum_{i,j,k,l} \frac{\sigma^2_{R,ijkl}}{n_{ijkl}},$$

$$\tau_7 = \frac{2n^2(a,b,c,d)}{ac^2 d^2 (b-1)^2 (c-1)^2} \sum_{i,j} \sum_{k \neq k'}^{c} \sum_{l,l'} \frac{\sigma^2_{R,ijkl}}{n_{ijkl}} \frac{\sigma^2_{R,ijk'l'}}{n_{ijk'l'}},$$

$$\tau_8 = \frac{2n^2(a,b,c,d)}{ab^2 c^2 d^2 (b-1)^2 (c-1)^2} \sum_{i=1}^{a} \sum_{j,j'} \sum_{k \neq k'}^{c} \sum_{l,l'} \frac{\sigma^2_{R,ijkl}}{n_{ijkl}} \frac{\sigma^2_{R,ij'k'l'}}{n_{ij'k'l'}}.$$

Then as $\max\{a,b,c\} \to \infty$, *regardless of whether the* $n_{ijkl}(\geq 2)$, *and* d *are small or large provided that* $\min_{ijkl}\{n_{ijkl}\}$ *and* $\max_{ijkl}\{n_{ijkl}\}$ *are of the same order,*

$$\sqrt{abc}(F_X(ABC) - 1) \xrightarrow{d} N\left(\lim_{\max\{a,b,c\} \to \infty} \frac{\delta_{32}}{\sigma^2_{ABC}}, \quad \lim_{\max\{a,b,c\} \to \infty} bc\tau^2_{ABC1} \right),$$

where

$$\delta_{32} = \frac{1}{(a-1)(b-1)(c-1)} \sum_{i,j,k} \left[\int xd(ABC)_{ijk}(x) - \int xd\overline{(ABC)}_{ij.}(x) \right.$$

$$- \int xd\overline{(ABC)}_{.jk}(x) - \int xd\overline{(ABC)}_{i.k}(x) + \int xd\overline{(ABC)}_{i..}(x)$$

$$\left. + \int xd\overline{(ABC)}_{.j.}(x) + \int xd\overline{(ABC)}_{..k}(x) - \int xd\overline{(ABC)}_{...}(x) \right]^2.$$

2.5. Analytical comparison of the proposed $F_X(\cdot)$ statistics with the ANOVA F-test

In this subsection, we clarify the relationship of the proposed $F_X(\cdot)$ statistics with the ANOVA F-test for balanced homoscedastic normal data. In this case, all our $F_X(\cdot)$ type statistics are equivalent to the ANOVA F-statistics for corresponding hypotheses. The only difference in them lies in that the ANOVA F-test uses percentiles from the F distribution, while the proposed procedure uses percentiles from the normal distribution based on Theorems 2.1−2.3. To illustrate the connection between the two procedures, consider Gaussian data from a balanced design with constant variance σ^2. Let U_1 and U_2 be the sums of squares for the hypotheses and error, respectively, divided by σ^2 under the local alternatives. Then $U_1 \sim \chi_{df_h}(\delta)$, $U_2 \sim \chi_{df_e}$, where df_h and df_e are the degrees of freedom for the hypotheses and error respectively, and δ is the noncentrality parameter calculated from corresponding local alternatives. Then $F = \{U_1/df_h\}/\{U_2/df_e\} \sim F_{df_h,df_e}(\delta)$. The power of the F-test at level α is

$$P(F > C_\alpha) = P\left(\frac{\sqrt{df_h}(U_1/df_h - U_2/df_e)}{U_2/df_e} > \sqrt{df_h}(C_\alpha - 1) \right),$$

where C_α is the $1 - \alpha$ percentile from the central F_{df_h,df_e} distribution. Note that as $df_e \geq df_h \to \infty$, U_2/df_e converges in probability to 1 and $Y = \sqrt{df_h}(U_1/df_h - U_2/df_e)$ has the approximate distribution $N(\delta/\sqrt{df_h}, 2 + 4\delta/df_h + 2df_h/df_e)$. So

$$P(F > C_\alpha) \approx P(Y > \sqrt{df_h}(C_\alpha - 1)). \qquad (2.14)$$

When the data are generated under the null hypothesis, the approximate distribution is centered at 0 with variance $2 + 2df_h/df_e$.

For power calculations under the local alternatives, with the uniformly bounded condition for the elements in $\mathbf{g}$ in Theorem 2.1, $\mathbf{G}$ in Theorem 2.2, and $\mathbf{U}$ in Theorem 2.3, we have $\delta/df_h = o(1)$. The power of the proposed tests based on $F_X(\cdot)$ statistics is equivalent to that from $Z \sim N(\delta/\sqrt{df_h}, 2 + 2df_h/df_e)$. Thus the asymptotic level and power of the two procedures coincide. For any finite df_h, however, the term δ/df_h may not be negligible. As a consequence, the power for the proposed procedure calculated from the asymptotic distribution may be higher than the actual power of the ANOVA F-tests. The discussions in this subsection only apply to homoscedastic balanced Gaussian data.

2.6. Techniques for obtaining the asymptotic distributions

As a tool for developing asymptotics in non-classical situations, Akritas and Papadatos (2004) suggested a bold use of Hájek's projection method applied to the difference between the mean squares for treatment and the mean squares for error, $MST - MSE$. Wang and Akritas (2006) followed Akritas and Papadatos (2004) and extended the results to two-way designs. However application of this technique, which involves writing $MST - MSE$ in matrix form, becomes almost intractable in the present multi-way setting.

Our approach partly adopts the projection idea in the sense that we decompose only MST under the null hypothesis into an asymptotically negligible term and one whose form is more amenable to asymptotic derivations. MSE is never decomposed. As an illustration, consider MST_A given at (2.2). Under the null hypothesis, we write $MST_A = P_A(\mathbf{e}) + T_A^*(\mathbf{e})$, where

$$P_A(\mathbf{e}) = \frac{b}{a}\sum_{i=1}^{a} \widetilde{e}_{i..}^2, \quad T_A^*(\mathbf{e}) = -\frac{b}{a(a-1)}\sum_{i \neq i'}^{a} \widetilde{e}_{i..}\widetilde{e}_{i'..}, \text{ and } e_{ijk} = X_{ijk} - E(X_{ijk}).$$

The term $T_A^*(\mathbf{e})$ can be ignored for the purposes of obtaining the asymptotic distribution of $F_X(A) - 1 = (MST_A - MSE)/MSE$ in Theorem 2.1, since $\sqrt{a}\min_{1 \leq i \leq a, 1 \leq j \leq b}\{n_{ij}\}T_A^*(\mathbf{e}) \to 0$ in probability. Note that the need to scale also by $\min_{1 \leq i \leq a, 1 \leq j \leq b}\{n_{ij}\}$ stems from the fact that, if $\min_{1 \leq i \leq a, 1 \leq j \leq b}\{n_{ij}\} \to \infty$

then $MSE \to 0$ (see (2.2)). The results of these decompositions are given as propositions in Section 4.1.

3. Numerical Results

Throughout the simulation studies in this section we use NP to denote the proposed test procedures. Thus, NP denotes the $Q_X(\cdot)$ statistic when evaluating the performance of the proposed procedure for testing of no main factor B effect, and the $F_X(\cdot)$ statistics when evaluating the performance of the proposed procedures for the rest of the effects. In addition to the reported power simulations, we also report the theoretical power calculated based on the asymptotic distribution under local alternatives.

3.1. Simulation results for binomial and count data

In this subsection, we compare the proposed test with the deviance test and the rank tests in Brunner and Puri (2001) for binomial and count data. Upon the request of the reviewers, we also include ANOVA F-test in some of the simulation settings. We consider designs with three factors, A, B and C, having 20, 2, and 20 levels respectively. The group sizes are $n_{4,2,k} = n_{8,2,k} = n_{10,2,k} = n_{13,1,k} = n_{15,2,k} = n_{20,1,k} = 5$, $n_{6,2,k} = n_{9,1,k} = n_{10,1,k} = 6$, for all k, and are 4 for the rest of the groups. Since the deviance test from generalized linear models is computationally very intense in this data setting (see Section 3.3), we used 1,000 simulation runs when the deviance test was included in the comparisons. Otherwise the comparisons were based on 3,600 runs.

For the exponential family, the deviance of a model is equal to the likelihood ratio statistic (for comparing the model under consideration with the saturated model) multiplied by a dispersion parameter. The deviance test compares the deviance from a model with the effect under consideration to that from the model without the effect. The difference between the two deviances has approximately a chi-square distribution with degrees of freedom equal to the difference in the number of parameters estimated if the sample sizes are large. The deviance test in our simulation is conducted in R by first fitting a generalized linear model with command *glm* containing all main and interaction effects and then applying the *drop1* command with the specification of *test="Chisq"* to do single term deletion and obtain p-values based on chi-square distributions.

For count data, we generated observations from Poisson distributions with means μ_{ijk} equal to 1 with probability $1 - p/100$, and given by the following values with probability $p/100$:

(P1) $\mu_{ijk} = 1 + i\tau/a$, under the alternatives of no main factor A effect;

(P2) $\mu_{ijk} = 1 + j\tau/b$ under the alternatives of no main factor B effect.

Table 1. Rejection rates and power at level 0.05 for count data. The underlined numbers without parentheses are the empirical power under the alternative and those with parentheses are the asymptotic theoretical power. 'NP' refers to the $Q_X(\cdot)$ type statistic for the effect of factor B, and the $F_X(\cdot)$ type statistics for the rest of the hypotheses. 'BP' refers to the Brunner and Puri (2001) test based on quadratic forms.

X_{ijkm}	effect	GLM	BP	NP	GLM	BP	NP	GLM	BP	NP
			$\tau = 0$			$\tau = 0.2$			$\tau = 0.4$	
Poisson	A	0.099	0.046	0.070	0.069	0.372	0.469	0.096	0.965	0.990
							(0.517)			(0.999)
with	B	0.076	0.046	0.038	0.077	0.045	0.048	0.068	0.052	0.055
mean	AC	0.238	0.038	0.059	0.210	0.036	0.068	0.161	0.034	0.067
$1 + \frac{i\tau}{a}$	AB	0.098	0.049	0.072	0.096	0.049	0.068	0.082	0.051	0.078
	ABC	0.402	0.035	0.065	0.334	0.039	0.063	0.262	0.037	0.068
	A	0.099	0.046	0.070	0.076	0.051	0.046	0.077	0.046	0.067
Poisson	B	0.076	0.046	0.038	0.087	0.750	0.779	0.077	0.998	1.000
							(0.775)			(0.999)
with	AC	0.238	0.038	0.059	0.213	0.038	0.061	0.162	0.032	0.053
mean	AB	0.098	0.049	0.072	0.084	0.052	0.052	0.081	0.049	0.078
$1 + \frac{j\tau}{b}$	ABC	0.402	0.035	0.065	0.320	0.041	0.069	0.245	0.034	0.054

The results are reported for $p = 100$ with $\tau = 0, 0.2, 0.4$, and $p = 5$ with $\tau = 3, 6, 9, 12$. The data with $p = 5$ refers to the case that the majority of the samples are from a pure distribution and 5% of the samples are contaminated.

For binomial data, we considered two settings. In the first setting, we generated Bernoulli observations with success probabilities π_{ijk} given by

(B1) $\pi_{ijk} = 0.2 + 0.8i\tau/a$, for $\tau = 0, 0.2, 0.4, 0.6$;

(B2) $\pi_{ijk} = 0.2 + 0.8j\tau/b$, for $\tau = 0.1, 0.15, 0.2$.

In the second setting, we generated observations from a mixture of two binomial distributions with 5 trials. For $p\%$ of the samples, the success probability took value $\pi_{ijk} = 0.2 + 0.72i/a$ for the alternative hypothesis of no main factor A effect and $\pi_{ijk} = 0.2 + 0.72j/b$ for the alternative hypothesis of no main factor B effect. For the rest of the samples, the success probability was 0.2. We let p ranges from 5% to 10% so that only a small percentage of the samples was contaminated.

Tables 1 and 2 give the rejection rates at level $\alpha = 0.05$ for testing the hypotheses of no main effect of factor A, B, and no A-B, A-C, and A-B-C interaction effect when the data were Poisson under settings (P1) and (P2) with contamination proportion $p = 100$ and $p = 5$, respectively. Table 3 presents the rejection rates at level $\alpha = 0.05$ for Bernoulli distributions (B1), (B2). The underlined numbers in Tables 1 and 3 are power estimates under the alternatives,
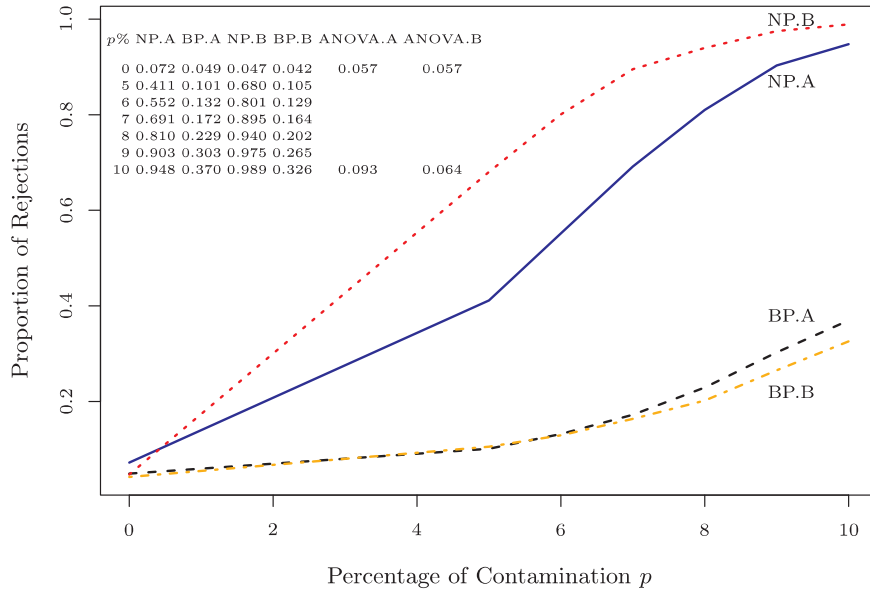
Estimated Power under Binomial Mixtures

Figure 1. Power Comparison for no main factor A and no main factor B effects at level 0.05 for the proposed tests (NP.A and NP.B, respectively) and Brunner and Puri (2001) tests (BP.A and BP.B, respectively) based on 3,600 runs. For the test of no main factor B effect, the data Y_{ijkm} were generated from Binomial distribution with 5 trials and success probability π_{ij}, where π_{ij} take value $0.2 + 0.72j/b$ with probability $p/100$, and 0.2 with probability $1 - p/100$. For the test of no main factor A effect, the data Y_{ijkm} were generated from Binomial distribution with 5 trials and success probability π_{ij}, where π_{ij} take value $0.2 + 0.72i/a$ with probability $p/100$, and 0.2 with probability $1 - p/100$. The value of p ranges from 0 to 10%. Value $p = 0$ corresponds to the null hypotheses. The listed values and plots present the rejection rates. The proposed tests demonstrate much better power than Brunner and Puri (2001) to detect the contaminations in the sample.

while the rest of the values in these two tables are type I error estimates. Figure 1 gives the power estimate under the mixture of two binomial distributions.

For the Poisson count data, the deviance tests tend to be very liberal for smaller values of τ, especially for hypotheses involving a high-dimensional parameter. Specifically, for τ values range from 0 to 0.4 in Table 1, the deviance test had type I errors at level 0.05 ranging from 0.245 to 0.402 for no A-B-C three-way interaction effect; from 0.161 to 0.238 for no A-C interaction effect; from 0.081 to 0.098 for no A-B interaction effect; from 0.077 to 0.099 for no main factor A effect; from 0.068 to 0.077 for no factor B effect. On the contrary,

Table 2. Rejection rates at level 0.05 for the proposed tests (NP), Brunner and Puri (2001) tests (BP), and ANOVA F-tests when data were generated under Poisson mixture based on 3,600 runs. The data Y_{ijkm} in the top panel labeled with m_A were generated from Poisson distribution with mean λ_{ij}, where λ_{ij} take value $1 + i\tau/a$ with probability 0.05, and 1 with probability 0.95. For the bottom panel labeled with m_B, the data Y_{ijkm} were generated from Poisson distribution with mean λ_{ij}, where λ_{ij} take value $1 + j\tau/b$ with probability 0.05, and 1 with probability 0.95. Hence, the data in the top and bottom panels are under the alternative hypotheses for main effect of A and B respectively. The row with $\tau = 0$ are type I error estimates. The underlined values are power estimates. The proposed tests have good power while the Brunner and Puri tests or ANOVA F-tests show no power to detect the alternatives.

	τ	NP					BP					ANOVA F-tests		
		A	B	AC	AB	ABC	A	B	AC	AB	ABC	A	B	AC
	0	0.070	0.038	0.059	0.072	0.065	0.046	0.046	0.038	0.049	0.035	0.060	0.054	0.058
	3	<u>0.237</u>	0.043	0.064	0.070	0.065	<u>0.077</u>	0.045	0.036	0.051	0.040	<u>0.084</u>	0.028	0.077
m_A	6	<u>0.614</u>	0.041	0.059	0.066	0.067	<u>0.088</u>	0.044	0.037	0.048	0.039	<u>0.119</u>	0.009	0.118
	9	<u>0.839</u>	0.056	0.048	0.071	0.069	<u>0.087</u>	0.058	0.036	0.049	0.036	<u>0.151</u>	0.002	0.138
	12	<u>0.910</u>	0.051	0.052	0.071	0.067	<u>0.075</u>	0.043	0.037	0.046	0.035	<u>0.164</u>	0.000	0.157
	3	0.068	<u>0.454</u>	0.066	0.069	0.066	0.046	<u>0.115</u>	0.041	0.054	0.040	0.025	<u>0.053</u>	0.001
m_B	6	0.067	<u>0.827</u>	0.058	0.068	0.062	0.049	<u>0.073</u>	0.038	0.049	0.039	0.012	<u>0.060</u>	0.000
	9	0.063	<u>0.935</u>	0.051	0.072	0.055	0.047	<u>0.058</u>	0.031	0.052	0.043	0.008	<u>0.071</u>	0.000
	12	0.063	<u>0.963</u>	0.051	0.070	0.052	0.046	<u>0.050</u>	0.038	0.047	0.036	0.004	<u>0.081</u>	0.000

the proposed test had type I errors ranging from 0.054 to 0.068 for no A-B-C interaction effect; from 0.052 to 0.078 for no A-B interaction effect; from 0.053 to 0.068 for no A-C interaction effect; 0.046 to 0.07 for no factor A effect; 0.038 to 0.055 for no factor B effect. The Brunner and Puri (2001) tests tended to be conservative, especially when testing for no A-C or A-B-C interaction effects where the achieved type I error rate ranged between 0.032 and 0.041 for Poisson count data and between 0.025 and 0.041 for binomial data. A possible explanation for this lies in the fact that these two hypotheses involve a high-dimensional parameter while the Brunner and Puri (2001) tests are developed for models with low-dimensional parameters.

Except for the test of no main factor B effect, the type I error estimates for the deviance test in binary data were much worse in that some error estimates were close to 1. Under the null hypotheses, the deviance test had type I error ranging from 0.925 to 0.998 for no A-B-C interaction effect; 0.885 to 0.941 for no A-C interaction; 0.127 to 0.263 for no A-B interaction; 0.111 to 0.140 for no factor A effect; and from 0.019 to 0.036 for no factor B effect. The proposed tests had type I error estimates close to the true level in nearly all data generation

Table 3. Rejection rates and power at level 0.05 for binary data. The data for the top panel labeled with 1 were generated from Bernoulli distribution with success probability $0.2+0.8i\tau/a$, and those for the bottom panel labeled with 2 were generated from Bernoulli distribution with success probability $0.2+0.8j\tau/b$. The underlined numbers without parentheses are the empirical power under the alternative and those with parentheses are the asymptotic theoretical power. The table legend is same as those in Table 1.

	effect	GLM	BP	NP	GLM	BP	NP	GLM	BP	NP	BP	NP	BP	NP
			$\tau=0$			$\tau=0.1$			$\tau=0.12$			$\tau=0.15$		$\tau=0.2$
	A	0.111	0.049	0.072	0.129	0.399	0.462	0.142	0.554	0.628	0.801	0.842	0.980	0.989
							(0.494)							(0.998)
	B	0.019	0.042	0.047	0.029	0.049	0.049	0.030	0.047	0.047	0.052	0.052	0.042	0.050
1	AC	0.911	0.025	0.072	0.923	0.025	0.054	0.927	0.031	0.060	0.032	0.058	0.041	0.052
	AB	0.127	0.051	0.065	0.190	0.047	0.066	0.210	0.054	0.074	0.052	0.070	0.049	0.079
	ABC	0.925	0.030	0.061	0.988	0.024	0.055	0.991	0.025	0.047	0.029	0.054	0.032	0.052
			$\tau=0$			$\tau=0.075$			$\tau=0.1$			$\tau=0.15$		$\tau=0.2$
	A	0.111	0.049	0.072	0.118	0.038	0.063	0.120	0.046	0.060	0.049	0.075	0.047	0.087
	B	0.019	0.042	0.047	0.038	0.538	0.538	0.039	0.767	0.757	0.977	0.967	0.999	0.999
							(0.529)			(0.757)		(0.971)		(0.999)
2	AC	0.911	0.025	0.072	0.928	0.028	0.052	0.941	0.033	0.068	0.035	0.070	0.039	0.057
	AB	0.127	0.051	0.065	0.198	0.035	0.058	0.227	0.052	0.065	0.047	0.080	0.048	0.080
	ABC	0.925	0.030	0.061	0.991	0.029	0.061	0.992	0.032	0.063	0.037	0.063	0.040	0.070

schemes.

Table 2, Figure 1 and the underlined values in Tables 1 and 3 are the power estimates under the alternatives for the test of no main factor A and factor B effects. The GLM deviance tests are not included in Table 2 or Figure 1 since they failed to maintain reasonable type I error, as discussed in previous two paragraphs, and had consistently low power with both the binomial and count data (see the underlined values in Tables 1 and 3). The proposed tests and Brunner and Puri (2001) tests have similar power in Tables 1 and 3. However, for the mixture alternatives of Table 2 and Figure 1, where only a small percentage of the observations are generated under non-null mean values, the Brunner and Puri (2001) rank tests have very low power. This is reasonable since rank procedures are relatively less sensitive to small percentage of large values. On the contrary, the proposed tests exhibited good power to detect these mixture-type alternatives.

Upon the request of the reviewers, we also included the ANOVA F-tests in power comparison for the same data generated from binomial or Poisson mixture models (Table 2 and Figure 1). The performance of the F-tests was similar for the binomial and Poisson models and we only summarize the patterns under the Poisson mixtures. When the data were under the alternatives with main factor A effects, all the F-tests on effects not related to factor A (such as the test of

no main B effect in Table 2) had very conservative type I error estimates, while those tests of effects related to factor A had rejection rates range between 0.077 and 0.164. Similarly, when the data were under the alternatives with main B effects, the F-tests of no main A or A-C interaction effects had conservative type I errors and almost no power for the test of no main B effects. The F-tests of no A-B and A-B-C interaction effects had similar rejection rates as the column for the A-C interaction effect in the m_A data setting and as the column for main B effect in the m_B data setting. We omitted them from the table. For the binomial data, the type I error for the F-tests of no main factor A or B effects and their power estimates when 10% of the data were generated from corresponding mixture alternatives are given in Figure 1. It is clearly seen that the F-tests had almost no power to detect these alternatives.

In summary, the proposed tests constitute a useful addition to the testing methodology, especially for detecting mixture-type alternatives having a small percentage of non-null signals. Such alternatives arise in applications such as monitoring disease outbreaks.

3.2. Simulation results for continuous data

We first compare the performance of the proposed tests and the ANOVA F-test in designs with three factors and continuous data. Factor A has $a = 20$, 30, or 50 levels, while factors B and C have $b = 2$ and $c = 20$ levels, respectively. The results are summarized in Tables 4 and 5. With the exception of the right panel of Table 5, the designs in both tables are unbalanced, though the unbalancedness in Table 5 is mild. The cell sample sizes are given in the captions. All simulation results in this section were based on 2,000 runs.

Table 4 gives the achieved type I error rates, at nominal level of significance $\alpha = 0.05$, with heteroscedastic data. The ANOVA F-test was surprisingly liberal for all the hypotheses and all three values of a considered. The average (over the three values of a) type I error rates for $H_0(AC)$ and $H_0(ABC)$ was over 0.4, while that for $H_0(B)$ was nearly 1. In contrast, the proposed nonparametric tests had reliable type I error estimates for all cases when $a = 30$ or 50. For the smaller value of $a = 20$, the proposed test had slightly elevated type I error for $H_0(A)$ and $H_0(AC)$. Both the unbalancedness and heteroscedastic variances contributed to the poor performance of the ANOVA F-test. Due to the failure of the ANOVA F-test to achieve a reasonably accurate type I error rate, no power comparisons were performed under the setting of Table 4.

Table 5 reports the rejection rates, both under the null and alternatives, for testing $H_0(A)$, $H_0(AC)$, and $H_0(B)$ for the proposed tests and the ANOVA F-tests when $a = 20$, $b = 2$ and $c = 20$. The cell sizes are given in the caption of the table. The theoretical power for the proposed test using the asymptotic

Table 4. Type I error estimate at level $\alpha = 0.05$, b=2, c=20, $X_{ijkm} \sim N(0, (4jk/bc)^2)$. The cell sizes used are as follows: When $a = 20$, $n_{i1k} = 12$ for $i = 1, \ldots, 10$, and all $k = 1, \ldots, 20$; $n_{i1k} = 10$ for $i = 11$, and $k = 1, \ldots, 20$; and $n_{i1k} = 5$ for $i = 12, \ldots, 20$; $n_{i2k} = 4$, for all i, k. When $a = 30$, $n_{i1k} = 12$ for $i = 1, \ldots, 10$, and all $k = 1, \ldots, 20$; $n_{i1k} = 10$ for $i = 11$, and $k = 1, \ldots, 20$; and $n_{i1k} = 5$ for $i = 12, \ldots, 30$; $n_{i2k} = 4$, for all i, k. When $a = 50$, $n_{i1k} = 12$ for $i = 1, \ldots, 10$, and all $k = 1, \ldots, 20$; $n_{i1k} = 10$ for $i = 11$, and $k = 1, \ldots, 20$; and $n_{i1k} = 5$ for $i = 12, \ldots, 50$; $n_{i2k} = 4$, for all i, k. 'NP' refers to the $Q_X(\cdot)$ type statistic for $H_0(B)$, and the $F_X(\cdot)$ type statistics for the rest of the hypotheses.

H_0	$a = 20$		$a = 30$		$a = 50$	
	ANOVA	NP	ANOVA	NP	ANOVA	NP
$H_0(A)$	0.097	0.070	0.099	0.066	0.107	0.066
$H_0(AB)$	0.130	0.050	0.122	0.059	0.093	0.049
$H_0(AC)$	0.491	0.079	0.469	0.061	0.475	0.067
$H_0((ABC)$	0.381	0.069	0.432	0.065	0.429	0.059
$H_0(B)$	0.999	0.066	0.999	0.063	0.999	0.058

Table 5. Empirical and asymptotic theoretical power at level $\alpha = 0.05$. a=20, b=2, c=20. For the left panel, $n_{4,2,k} = n_{8,2,k} = n_{10,2,k} = n_{13,1,k} = n_{15,2,k} = n_{20,1,k} = 5$, $n_{6,2,k} = n_{9,1,k} = n_{10,1,k} = 6$, for all k. The rest of the cell sizes are 4. NP.alt is theoretical power using asymptotic distribution under the alternatives. For the right panel, the cell sizes are 4. ANOVA.alt and NP.alt are the theoretical power of the ANOVA F-test and the proposed test, respectively, calculated using the true variance.

	$H_0(A)$			$H_0(B)$		
	$\text{Normal}\left(\frac{i\tau}{2a}, \frac{(4jk)^2}{(bc)^2}\right)$			$\text{Normal}\left(\frac{j\tau}{4b}, \frac{(4jk)^2}{(bc)^2}\right)$		
τ	ANOVA	$F_X(A)$	NP.alt	ANOVA	$Q_X(B)$	NP.alt
0	0.048	0.064	0.050	0.060	0.062	0.050
0.5	0.180	0.223	0.204	0.165	0.164	0.161
1	0.763	0.811	0.945	0.490	0.485	0.485
1.5	0.993	0.995	1.000	0.823	0.820	0.822

	$H_0(AC)$			
	$\text{Normal}\left(\frac{ik\tau}{ac}, 1\right)$			
τ	ANOVA	ANOVA.alt	$F_X(AC)$	NP.alt
0	0.045	0.05	0.056	0.050
0.5	0.076	0.074	0.096	0.074
1	0.176	0.188	0.206	0.195
1.5	0.500	0.499	0.548	0.548
2	0.860	0.871	0.891	0.932
3	1.000	1.000	1.000	1.000
4				

distribution under corresponding alternatives is given in columns with heading NP.alt. The corresponding theoretical powers of the F-test is reported only in the homoscedastic setting. The left panel of Table 5 corresponds to heteroscedastic case for hypothesis $H_0(A)$ with data generated from $N(i\tau/(2a), [4jk/(bc)]^2)$, and for $H_0(B)$ with data generated from $N(j\tau/(4b), [4jk/(bc)]^2)$. The proposed test had comparable power to the F-test for the considered alternatives to $H_0(B)$, and slightly better power than the F-test for the considered alternatives to $H_0(A)$.

In the homoscedastic setting of testing for $H_0(AC)$ (right panel of Table 5), the ANOVA F-test is the UMPI test for any given values of a and c. In this setting, however, the proposed $F_X()$ statistic is equivalent to the ANOVA F-test (see Section 2.5). The slight difference in the empirical powers of the two tests is due to the fact that the NP test uses the limiting normal distribution while the ANOVA F-test uses its exact F-distribution (see Section 2.5). Specifically, the ANOVA F-statistic has noncentrality parameter

$$\delta(\tau) = \frac{\tau^2 nb}{a^2 c^2} \sum_{i,k} \left[ik - \frac{i(c+1)}{2} - \frac{k(a+1)}{2} + \frac{(a+1)(c+1)}{4} \right]^2 = O(ac\tau^2). \quad (3.1)$$

Thus, the power of the ANOVA F-test is equivalent to that calculated from $N(\delta(\tau)/\sqrt{ac}, 4\delta(\tau)/(ac) + 2 + 1/(n-1))$. On the other hand, the asymptotic power of the proposed test is calculated from $N(\delta(\tau)/\sqrt{ac}, 2 + 1/(n-1))$. Because $\delta(\tau)/(ac) = o(1)$ under the uniformly bounded condition for large a and c, the asymptotic distributions, and thus the power functions, of the two tests coincide. For the values of a, c of Table 5, the differences in the variances are 0.061, 0.245, 0.551, 0.980, and 2.205, for $\tau = 0.5, 1.0, 1.5, 2.0, 3.0$ respectively. This explains the difference between the empirical and asymptotic power observed for the proposed test.

Next, we compare the test by Brunner, Dette, and Munk (1997) (BDM) with the proposed Chi-square test on data generated from Gamma distribution as follows (the τ values range from 0 to 1.7):

$$Y_{ijkm} = X_{ijkm} - (a+1)j + \tau j/b, \text{ where } X_{ijkm} \sim \text{Gamma}(ij, 2),$$
$$\text{for } i = 1, \ldots, a = 20; \ j = 1, 2, 3, 4; \ k = 1, \ldots, 20; \text{ and } n_{ijk} = 6 - j.$$

Value $\tau = 0$ corresponds to the null hypothesis of no main factor B effect. This is a highly heteroscedastic case as the cell variances $4ij$ range from 4 to 160. The estimated power at level 0.01 based on 2,000 runs is presented in Figure 2 with numerical values given inside of the plot. Both tests had accurate type I error estimates. As τ increases, the proposed test clearly had much better power than the BDM test. In fact, the power of the BDM test was only about half of that for the proposed test. The loss of power for BDM test was also noticed in some cases by Brunner, Dette, and Munk (1997) in their simulations. However, it was only marginal there because the number of cells was not large and they used Normal data. Since the derivation of the BDM test is based on the normal approximation of the cell means, the approximate $F-$distribution of the final test statistic could be far from the true distribution when the data are highly skewed and only very few observations per cell are available. This discrepancy is amplified as the number of cells increases.
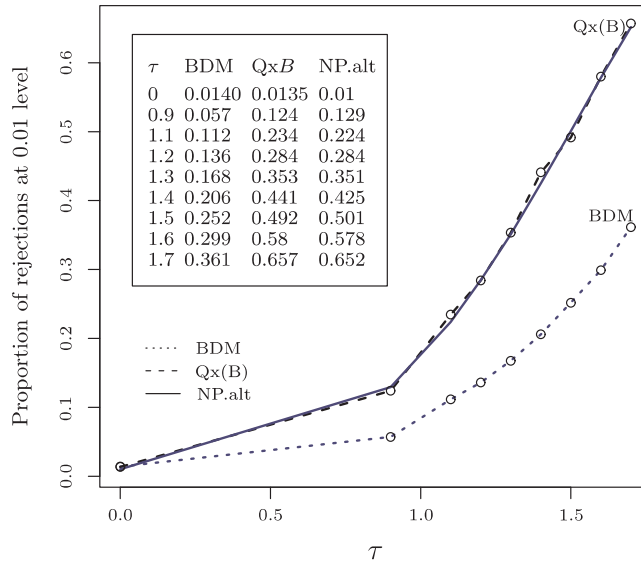
Power comparison for data from Gamma distribution

τ	BDM	QxB	NP.alt
0	0.0140	0.0135	0.01
0.9	0.057	0.124	0.129
1.1	0.112	0.234	0.224
1.2	0.136	0.284	0.284
1.3	0.168	0.353	0.351
1.4	0.206	0.441	0.425
1.5	0.252	0.492	0.501
1.6	0.299	0.58	0.578
1.7	0.361	0.657	0.652

Figure 2. Comparison of the estimated power at level 0.01 from 2,000 runs for the test of no main factor B effect when the data were generated from $Y_{ijkm} = X_{ijkm} - j(a+1) + \tau j/b$, where $X_{ijkm} \sim \text{Gamma}(ij, 2)$. The null hypothesis corresponds to $\tau = 0$. The 2nd and 3rd columns of the numerical values in the plot are the rejection rates for each τ value for the BDM test (Brunner, Dette and Munk 1997) and the proposed test (NP). The last column is the average power calculated based on the asymptotic distribution under the local alternatives in part (a) of Theorem 2.1. The power of BDM test is only about half of that for NP.

Remark 5.1. The simulation results reported for NP used the unbiased estimator for σ_{ijk}^4 as suggested in Remark 3.4 when $n_{ijk} \geq 4$, and the Jackknife bias-corrected estimator of σ_{ijk}^4 when $2 \leq n_{ijk} < 4$.

3.3. Comparison of computational time

All simulations in this manuscript were carried out in R 2.7.1. Commands *lm* and *glm* were used for fitting the ANOVA models and generalized linear models, respectively, and command *drop1* was used to obtain the deviance test from the fitted generalized linear models. The BDM test and proposed tests are programmed in R.

When applied to the traditional ANOVA designs (i.e., small number of factor levels) the needed computational time for the usual tests is almost negligible, even with large cell sample sizes. For example, in a three-way design with $a = 4$, $b = 2$, $c = 5$, and balanced sample size $n = 140$ (total 5,600 response values),

the average computing time (in seconds, with standard error in parenthesis) for one single test based on $1,000$ runs was 0.026 (2.7×10^{-4}) for the BDM test of no main factor B effect, 0.136 (4.7×10^{-4}) for the ANOVA F-test (for all main effects and their interactions), and 1.357 (2.4×10^{-3}) for the generalized linear models using the deviance test (for all effects, among which 0.180 second was for the glm model fit and 1.176 second was for the deviance test). This performance was observed using a PC with Intel Pentium(R) M processor 1.86 GHz, 1.00GB of RAM.

The computational time to carry out a test in HANOVA increased dramatically as the number of factor levels increased. Consider 5,600 response values as above, but with $a = 20$, $b = 4$, $c = 20$, and $n_{ijk} = 6 - j$, where $i = 1, \ldots, a$, $j = 1, 2, 3, 4$, $k = 1, \ldots, c$. Using exactly the same R-code and the same computer, the average computing time (in seconds) for one single test based on 100 runs was 18.74 (0.01) for the BDM test of no main factor B effect, 82.03 (0.181) for the ANOVA F-test, and $2{,}687.74$ (2.997) for the deviance test with generalized linear models (among which 411.17 seconds was used for fitting the generalized linear models and $2{,}276.57$ seconds for the deviance test). The average computational time for the proposed NP in this data setting was 0.085 (0.001) for no main factor B effect, and $9.555(0.045)$ for the tests of all main and interaction effects reported.

Thus, when applied to HANOVA designs, the tests developed under the traditional settings took much more computational time, in addition to their reported disadvantages in the form of inaccurate type I error rate and low power. In contrast, the proposed tests were computationally more efficient, in addition to achieving fairly accurate type I error rates and good power.

3.4. Analysis of stress response gene expression data

In this subsection, we apply our tests on gene expression data to study *Arabidopsis thaliana* responses to abiotic stresses. The expression data for shoot and root tissue under nine stresses (cold/freezing, osmotic, salt, drought, genotoxic, oxidative, UV-B, wounding, heat) and their corresponding controls were downloaded from the Bio-Array Resource (Toufighi et al. (2005)). The experiment under each stress was conducted at six time points (0.5h, 1 h, 3h, 6h, 12h, and 24h) with two replications per time point. Due to the small plant size, multiple plants were often used to produce enough material for the microarray chip of one sample. So the observations at different time points are independent. We focus our discussion on the analysis of gene PLDα3 (At5g25370), a gene that potentially catalyzes lipid reactions in response to stress.

The median expression levels over all stress categories at different time points are given in the left panel of Table 6. It can be seen that the expression level of

Table 6. Left panel: tissue time interaction; Right panel: interaction of tissue with stress vs control contrast.

Time	Root	Shoot
0.5 h	14.41	3.82
1.0 h	12.70	1.85
3.0 h	13.12	2.25
4.0 h	13.32	2.97
6.0 h	15.41	1.87
12.0 h	10.98	5.05
24.0 h	14.97	3.00

Stress	Root	Shoot
Cold	-2.73	1.63
Drought	-1.50	2.25
Genotoxic	-2.08	2.13
Heat	-0.82	0.76
Osmotic	-0.80	3.81
Oxidative	-1.99	1.27
Salt	-2.13	2.32
UV-B	-2.72	0.53
Wounding	-2.42	3.84

Interaction of Tissue and Control vs Stress Contrast
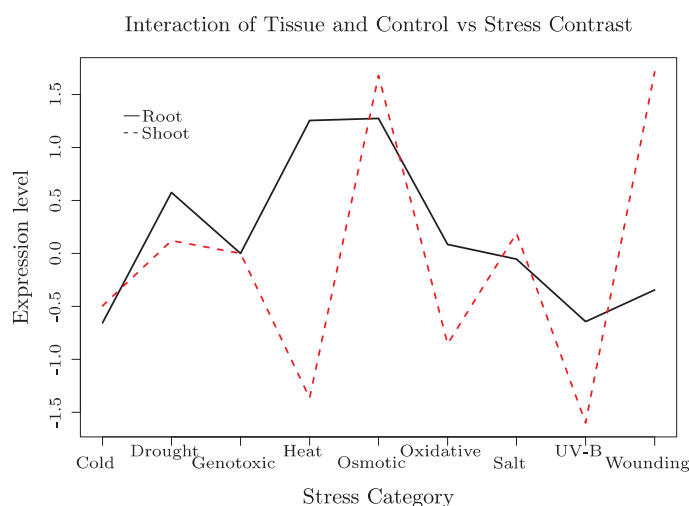
Figure 3. Stress and tissue interactions.

this gene in the root tissue continued to increase from one to six hours following the start of the experiment, then dropped to a low at twelve hours. On the other hand, it reached a high in the shoot tissue 12 hours after the experiment started. This suggests an interaction effect between tissue and time. The right panel of Table 6 gives the contrast of the median expression levels for each stress with its control for root and shoot tissues separately. It is clear that all the median expressions of the stresses are higher than their control in the root tissue, but, completely opposite for the shoot tissue. This suggests a strong tissue effect. The plot beside Table 6 is the median expression of the contrast between control and stress for each tissue type with the tissue effect removed. The plot indicate a stress-tissue interaction effect. We applied part (a) of Theorem 2.1 for the main effect of tissue and time, part (b) of Theorem 2.1 of Theorem 2.2 (b) of Theorem

Table 7. $p-$values for all effects for gene PLDα3.

Effect	ANOVA	NP
stress	0.552	0.888
tissue	0.000	0.000
time	0.011	1.47×10^{-9}
stress-tissue	0.869	0.095
stress-time	0.322	0.959
time-tissue	0.132	0.000
time-tissue-stress	0.988	0.971

2.2 for the stress-tissue and stress-time interaction, and Part (b) of Theorem 2.3 for the time-tissue-stress interaction effects. The $p-$values along with those from the ANOVA type III sum of squares are reported in Table 7.

Both ANOVA and NP detected a highly significant tissue effect. The NP found significant tissue-time interaction effect but the ANOVA failed to detect it. The NP also identified a highly significant time effect (the median expression levels at the six time points are 10.73, 7.98, 8.10, 8.06, 8.46, 7.48, and 9.43 respectively). The ANOVA test of no main time effect was not significant at the 0.01 level. Although the $p-$value of stress-tissue interaction effect from the NP is small, it is not significant. For all other effects, NP and ANOVA gave similar results.

3.5. Other potential applications

In the previous subsection, we gave an example of the analysis of stress response gene expression data for *Arabidopsis* under many different experimental conditions over multiple time points. This is a typical example of modern data collection technology. Beyond *Arabidopsis*, gene expression data from many other organisms such as drosophila melanogaster, yeast, Hessian fly from many genotypes under many experimental conditions are available to be compared for novel gene discovery or identification of important genes related to certain metabolism. Such data can be downloaded from public databases such as the Stanford Microarray Database (`http://smd.stanford.edu/`) and NCBI database (`http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed`). Due to the large number of experimental conditions and a small number of replications, traditional methods have inaccurate type I error rates and low power for these data. This paper offers a set of tools for effectively analyzing such data.

Another potential application of the proposed tests is in bio-surveillance for the early detection of disease outbreaks. For example, electronic syndromic surveillance units provide daily electronic syndromic information throughout the United States and many other countries (see Mnatsakanyan et al. (2008) and the

references therein) with the goal of early detection of public health events. These systems use data such as the count of positive lab tests from multiple categories of illnesses (for example, influenza, adenovirus, rhinovirus, etc.) to identify respiratory viral pathogen surveillance, or percentages of visits for severe influenza-like illness (ILI) reported by medical providers to detect upcoming epidemics. While they have been useful for the quick identification of potentially large outbreaks, all such systems continue to have high false-positive rates and most systems also face a sensitivity problem, in that an anomalous event may remain undetected if there are only a few cases (Reis, Pagano, and Mandl (2003), Marsden-Haug N (2005), Bravata et al. (2004)).

As illustrated in the simulation studies in Section 3.1, the proposed tests have best power when the data under the alternatives are from mixture distributions with only a small proportion of observations (5% to 10%) having expected values different from those under the null. This can be used to improve the sensitivity of bio-surveillance early event detection when applied in appropriate settings. The large number of factor levels might be different age groups, geographical areas, or number of weeks being monitored. Our results offer a potentially useful set of tools for these types of applications.

4. Summary and Discussion

In this paper we develop hypothesis testing theory for testing the main and interaction effects of up to the third order for general HANOVA with arbitrary number of factors in the presence of heteroscedastic variances and unbalanced sample sizes. Our models allow the response variable to be continuous or discrete, and leave the distribution of the data completely unspecified. The asymptotic results of the test statistics were obtained under nonclassical asymptotic settings in which the cell sample sizes can be very small. Two types of test statistics are constructed. One has a limiting χ^2 distribution and is applicable for testing hypotheses where the parameter of interest is of low dimension while the nuisance parameters are in a high-dimensional space. The other has a limiting normal distribution and is applicable for testing hypotheses where the parameter of interest is high dimensional. Extensive empirical studies were conducted. In addition to exhibiting reliable type I error rate and high power compared to competing test procedures, the proposed tests also showed significant computational advantage over the traditional methods in HANOVA settings.

Appendix: Proofs

A.1. Propositions and lemmas for projection and weak convergence

The following propositions give the decompositions and validate the projections of the the MST terms. They make it possible to simplify the test statistics before derivation of the asymptotic distribution. The proof of Proposition A.1 is given in supplementary material. The other propositions follow similar arguments and are thus omitted. Please see Wang (2004) for details of the omitted proofs.

Proposition A.1. *Let* $n(a,b,c,d) = \min\limits_{i,j,k,l}\{n_{ijkl}\}$, $e_{ijklm} = X_{ijklm} - E(X_{ijklm})$, *and set*

$$P_{1,ABC}(\mathbf{e}) = d(abc)^{-1} \sum_{i,k} \left[\sum_{j=1}^{b} \widetilde{e}_{ijk..}^2 - (b-1)^{-1} \sum_{j\neq j'} \widetilde{e}_{ijk..}\widetilde{e}_{ij'k..} \right],$$

$$P_{2,ABC}(\mathbf{e})$$

$$= P_{1,ABC}(\mathbf{e}) + \frac{bd}{(c-1)ac} \sum_{i=1}^{a} \sum_{k\neq k'}^{c} \widetilde{e}_{i.k..}\widetilde{e}_{i.k'..} - \frac{d}{(b-1)(c-1)ac} \sum_{i,j} \sum_{k\neq k'}^{c} \widetilde{e}_{ijk..}\widetilde{e}_{ijk'..},$$

$P_{3,ABC}(\mathbf{e}) = d \sum_{i,j,k} \widetilde{e}_{ijk..}^2 / (abc)$. *Then under the settings and assumptions of part* (b) *of Theorem* 2.3, *under* $H_0(ABC)$,

(1) *as* $a,c \to \infty$, *b remains bounded,*
$$T_1^*(\mathbf{e}) = n(a,b,c,d)\sqrt{ac}(MST_{ABC} - P_{1,ABC}(\mathbf{e})) \xrightarrow{p} 0;$$

(2) *as* $a \to \infty$, *b, c remain bounded,*
$$T_2^*(\mathbf{e}) = n(a,b,c,d)\sqrt{a}(MST_{ABC} - P_{2,ABC}(\mathbf{e})) \xrightarrow{p} 0 \;;$$

(3) *as* $a,b,c \to \infty$, $T_3^*(\mathbf{e}) = n(a,b,c,d)\sqrt{abc}(MST_{ABC} - P_{3,ABC}(\mathbf{e})) \xrightarrow{p} 0.$

Proof of Proposition A.1. Under $H_0(ABC)$, MST_{ABC} can be expressed in terms of the e_{ijklm}'s and we have

$$T_1^*(\mathbf{e}) = D_4(\mathbf{e}) + D_5(\mathbf{e}) + D_6(\mathbf{e}) - \frac{D_1(\mathbf{e})}{b-1} - \frac{D_2(\mathbf{e})}{b-1} - \frac{D_3(\mathbf{e})}{b-1},$$

$$T_2^*(\mathbf{e}) = \frac{D_5(\mathbf{e})}{\sqrt{c}} + \frac{D_6(\mathbf{e})}{\sqrt{c}} - \frac{D_2(\mathbf{e}) + D_3(\mathbf{e})}{(b-1)\sqrt{c}}, \quad T_3^*(\mathbf{e}) = \sqrt{b}T_1^*(\mathbf{e}) + D_7(\mathbf{e}),$$

where

$$D_1(\mathbf{e}) = -\frac{bd\,n(a,b,c,d)}{(c-1)\sqrt{ac}} \sum_{i=1}^{a} \sum_{k\neq k'}^{c} \widetilde{e}_{i.k..}\widetilde{e}_{i.k'..},$$

$$D_2(\mathbf{e}) = -\frac{bd\,n(a,b,c,d)}{(a-1)\sqrt{ac}} \sum_{i\neq i'}^{a} \sum_{k=1}^{c} \widetilde{e}_{i.k..}\widetilde{e}_{i'.k..},$$

$$D_3(\mathbf{e}) = \frac{bd\,n(a,b,c,d)}{(a-1)(c-1)\sqrt{ac}} \sum_{i\neq i'}^{a} \sum_{k\neq k'}^{c} \widetilde{e}_{i.k..}\widetilde{e}_{i'.k'..},$$

$$D_4(\mathbf{e}) = -\frac{d\,n(a,b,c,d)}{(b-1)(c-1)\sqrt{ac}} \sum_{i,j} \sum_{k\neq k'}^{c} \widetilde{e}_{ijk..}\widetilde{e}_{ijk'..},$$

$$D_5(\mathbf{e}) = \frac{d\,n(a,b,c,d)}{(a-1)(b-1)(c-1)\sqrt{ac}} \sum_{i\neq i'}^{a} \sum_{j=1}^{b} \sum_{k\neq k'}^{c} \widetilde{e}_{ijk..}\widetilde{e}_{i'jk'..},$$

$$D_6(\mathbf{e}) = -\frac{d\,n(a,b,c,d)}{(a-1)(b-1)\sqrt{ac}} \sum_{i\neq i'}^{a} \sum_{j,k} \widetilde{e}_{ijk..}\widetilde{e}_{i'jk..},$$

$$D_7(\mathbf{e}) = \frac{dn(a,b,c,d)}{(b-1)\sqrt{abc}} \sum_{i,k} \sum_{j\neq j'}^{b} \widetilde{e}_{ijk..}\widetilde{e}_{ij'k..}.$$

For all three cases, it suffices to show $D_t(\mathbf{e}) = o_p(1)$ for $t = 1,2,3,7$, and $\sqrt{b}D_s(\mathbf{e}) = o_p(1)$ for $s = 4,5,6$. The proofs of $D_7(\mathbf{e}) = o_p(1)$, $\sqrt{b}D_4(\mathbf{e}) = o_p(1)$, and $\sqrt{b}D_6(\mathbf{e}) = o_p(1)$ are similar; the proofs of $D_2(\mathbf{e}) = o_p(1)$ and $D_1(\mathbf{e}) = o_p(1)$ are similar; and the proofs of $D_3(\mathbf{e}) = o_p(1)$ and $\sqrt{b}D_5(\mathbf{e}) = o_p(1)$ are similar. So it is enough to prove $D_7(\mathbf{e}) = o_p(1)$, $D_2(\mathbf{e}) = o_p(1)$ and $D_3(\mathbf{e}) = o_p(1)$. Using (A.5), we have

$$E[D_7(\mathbf{e})]^2 = \frac{2n^2(a,b,c,d)}{(b-1)^2d^2abc} \sum_{i,k} \sum_{j\neq j'}^{b} \left(\sum_{l=1}^{d} \frac{\sigma_{ijkl}^2}{n_{ijkl}}\right)\left(\sum_{l'=1}^{d} \frac{\sigma_{ij'kl'}^2}{n_{ij'kl'}}\right)$$

$$\leq \frac{2n^2(a,b,c,d)}{(b-1)^2d^2abc} \sum_{i,k} \left(\sum_{j,l} \frac{\sigma_{ijkl}^2}{n_{ijkl}}\right)^2$$

$$\leq \frac{2n^2(a,b,c,d)}{(b-1)^2dac} \sum_{i,j,k,l} \frac{\sigma_{ijkl}^4}{n_{ijkl}^2} = O(b^{-1}),$$

$$E[D_2(\mathbf{e})]^2 = \frac{2n^2(a,b,c,d)}{(a-1)^2b^2d^2ac} \sum_{i\neq i'}^{a} \sum_{k=1}^{c} \left(\sum_{j,l} \frac{\sigma_{ijkl}^2}{n_{ijkl}}\right)\left(\sum_{j',l'} \frac{\sigma_{i'j'kl'}^2}{n_{i'j'kl'}}\right)$$

$$\leq \frac{2n^2(a,b,c,d)}{(a-1)^2b^2d^2ac} \sum_{k=1}^{c} \left(\sum_{i,j,l} \frac{\sigma_{ijkl}^2}{n_{ijkl}}\right)^2$$

$$\leq \frac{2n^2(a,b,c,d)}{(a-1)^2 bdc} \sum_{i,j,k,l} \frac{\sigma_{ijkl}^4}{n_{ijkl}^2} = O(a^{-1}),$$

$$E[D_3(\mathbf{e})]^2 = \frac{4n^2(a,b,c,d)}{(a-1)^2(c-1)^2 b^2 d^2 ac} \sum_{i \neq i'}^{a} \sum_{k \neq k'}^{c} \left(\sum_{j,l} \frac{\sigma_{ijkl}^2}{n_{ijkl}} \right) \left(\sum_{j',l'} \frac{\sigma_{i'j'kl'}^2}{n_{i'j'k'l'}} \right)$$

$$\leq \frac{4n^2(a,b,c,d)}{(a-1)^2(c-1)^2 b^2 d^2 ac} \left(\sum_{i,j,k,l} \frac{\sigma_{ijkl}^2}{n_{ijkl}} \right)^2$$

$$\leq \frac{2n^2(a,b,c,d)}{(a-1)^2 bd} \sum_{i,j,k,l} \frac{\sigma_{ijkl}^4}{n_{ijkl}^2} = O(a^{-1}).$$

So $D_7(\mathbf{e}) = o_p(1)$, $D_2(\mathbf{e}) \xrightarrow{p} 0$ and $D_3(\mathbf{e}) \xrightarrow{p} 0$, and the proof is complete.

Proposition A.2. *Set* $n(a,b,c) = \min_{i,j,k}\{n_{ijk}\}$, $P_{1,AB}(\mathbf{e}) = [c/(ab)] \sum_{i=1}^{a} \sum_{j=1}^{b} \widetilde{e}_{ij..}^2$, *where* $e_{ijkm} = X_{ijkm} - E(X_{ijkm})$, *and* $P_{2,AB}(\mathbf{e}) = P_{1,AB}(\mathbf{e}) - [c/[ab(b-1)]] \sum_{i=1}^{a} \sum_{j \neq j'}^{b} \widetilde{e}_{ij..}\widetilde{e}_{ij'...}$ *Then under* $H_0(AB)$ *and the settings and assumptions of part* (b) *of Theorem 2.2, we have*

(1) $T_{1AB}^*(\mathbf{e}) = n(a,b,c)\sqrt{ab}(MST_{AB} - P_{1,AB}(\mathbf{e})) \xrightarrow{p} 0$, *as* $a,b \to \infty$;

(2) $T_{2AB}^*(\mathbf{e}) = n(a,b,c)\sqrt{a}(MST_{AB} - P_{2,AB}(\mathbf{e})) \xrightarrow{p} 0$, *as* $a \to \infty$, b *remains fixed.*

Proposition A.3. *Let* $P_A(\mathbf{e}) = (b/a) \sum_{i=1}^{a} \widetilde{e}_{i..}^2$, *where* $e_{ijk} = X_{ijk} - E(X_{ijk})$. *Assume the settings and assumptions of part* (b) *of Theorem 2.1 are satisfied. Then under* $H_0(A)$, *we have* $T_A^*(\mathbf{e}) = n(a,b)\sqrt{a}(MST_A - P_A(\mathbf{e})) \xrightarrow{p} 0$, *as* $a,b \to \infty$, *regardless of whether the* n_{ij} *remain fixed or go to* ∞.

The following lemmas give weak consistency of the MSE terms. The proof of Lemma A.4 is given while the other two proofs are omitted.

Lemma A.4. *Under the settings and assumptions of part* (b) *of Theorem 2.1, we have* $n(a,b)(MSE - \sigma_A^2) \xrightarrow{p} 0$.

Proof of Lemma A.4. Note that

$$E(MSE) = E\left[\frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{S_{ij,X}^2}{n_{ij}} \right] = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{\sigma_{ij}^2}{n_{ij}} = \sigma_A^2,$$

and

$$(ab)^2 \mathrm{Var}\,(MSE)$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{1}{n_{ij}^2(n_{ij}-1)^2} \mathrm{Var}\left(\frac{n_{ij}-1}{n_{ij}} \sum_{m=1}^{n_{ij}} e_{ijm}^2 - \frac{1}{n_{ij}} \sum_{m \neq m'}^{n_{ij}} e_{ijm}e_{ijm'} \right)$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{1}{n_{ij}^4} \left[\sum_{m=1}^{n_{ij}} E(e_{ijm}^4) - \frac{n_{ij}(n_{ij}-3)}{n_{ij}-1} \sigma_{ij}^4 \right].$$

So $\mathrm{Var}\,(n(a,b)MSE) \leq 1/(ab)^2 \sum_{i=1}^{a} \sum_{j=1}^{b} n^2(a,b)/n_{ij}^4 \sum_{m=1}^{n_{ij}} E(e_{ijm}^4) \to 0$ as $a \to \infty$, $b \to \infty$. Therefore, $n(a,b)[MSE - \sigma_A^2] \xrightarrow{p} 0$.

Lemma A.5. *Under the settings and assumptions of part* (b) *of Theorem 2.2,* $n(a,b,c)(MSE - \sigma_{AB}^2) \xrightarrow{p} 0$.

Lemma A.6. *Under the settings and assumptions of part* (b) *of Theorem 2.3,* $n(a,b,c,d)(MSE - \sigma_{ABC}^2) \xrightarrow{p} 0$.

A.2. Proofs of the Theorems

Proof of Theorem 2.1. We first show the asymptotic results under the null hypothesis and then apply them in the proofs under the alternatives. Let $\sigma_{ij} = \mathrm{Var}\,(X_{ijk})$ under the null hypothesis.

(a). Under $H_0(A)$, $\mathbf{C}_A E(\mathbf{W}) = \mathbf{0}$, so we have $\mathbf{C}_A \mathbf{W} = \mathbf{C}_A(\mathbf{W} - E(\mathbf{W}))$. Thus the result follows by showing $\sqrt{N}(\mathbf{W} - E(\mathbf{W})) \xrightarrow{d} N_a(\mathbf{0}, \mathbf{V})$, where $\mathbf{V} = diag\{\eta_1, \ldots, \eta_a\}$, and using the Continuous Mapping and Slutsky's Theorems, since the $\widehat{\eta}_i$ are consistent estimators of η_i, $i = 1, \ldots, a$. By the independence of the $\widetilde{X}_{i..}$, and since a is finite, the asymptotic normality of $\mathbf{W}$ follows from that of each $\widetilde{X}_{i..}$. Write $\sqrt{N}(\widetilde{X}_{i..} - E(\widetilde{X}_{i..})) = \sqrt{N}\widetilde{e}_{i..}$, where $e_{ijk} = X_{ijk} - E(X_{ijk})$. By Lyapounov's Theorem, $\sqrt{N}\widetilde{e}_{i..}/\eta_{i,N} \to N(0,1)$, where $\eta_{i,N} = \mathrm{Var}\,(\sqrt{N}\widetilde{e}_{i..}) = N \sum_{j=1}^{b} \mathrm{Var}\,(\bar{e}_{ij.})/b^2 = N \sum_{j=1}^{b} \sigma_{ij}^2/(n_{ij}b^2)$, because

$$L(b) = \frac{1}{\eta_{i,N}^2} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} E \left| \frac{\sqrt{N}}{bn_{ij}} e_{ijk} \right|^4 = \left(\sum_{j=1}^{b} \frac{\sigma_{ij}^2}{n_{ij}} \right)^{-2} \sum_{j=1}^{b} \left(\frac{1}{n_{ij}^3} E|e_{ijk}^4| \right) \to 0,$$

as a result of assumption (2.4) and Hölder's inequality.

Under the alternative given by (2.3), $X_{ijk} \sim F_{N,ij}(x) = F_{ij}(x) + N^{-1/2} A_i(x)$, so $E(X_{ijk}) = \int x dF_{ij}(x) + N^{-1/2} \int x dA_i(x)$, and

$$\sqrt{N}\mathbf{C}_A \mathbf{W} = \sqrt{N}\mathbf{C}_A[\mathbf{W} - E(\mathbf{W})] + \sqrt{N}\mathbf{C}_A E(\mathbf{W}) = \sqrt{N}\mathbf{C}_A[\mathbf{W} - E(\mathbf{W})] + \mathbf{C}_A \mathbf{g}.$$

From above proof under the null, we have $\sqrt{N}(\mathbf{W} - E(\mathbf{W})) \xrightarrow{d} N_a(\mathbf{0}, \mathbf{V})$. So $\sqrt{N}\mathbf{C}_A \mathbf{W} \xrightarrow{d} N_a(\mathbf{C}_A \mathbf{g}, \quad \mathbf{C}_A \mathbf{V} \mathbf{C}_A')$. Thus premultiplication by the inverse of the symmetric square root of the covariance matrix yields $[\mathbf{C}_A \mathbf{V} \mathbf{C}_A']^{-1/2} \sqrt{N}\mathbf{C}_A \mathbf{W} \xrightarrow{d}$

$N_a\left([\mathbf{C}_A\mathbf{V}\mathbf{C}_A']^{-1/2}\mathbf{C}_A\,\mathbf{g},\quad \mathbf{I}_a\right)$. Then the asymptotic noncentral Chi-square distribution of $N\mathbf{W}'\mathbf{C}_A'(\mathbf{C}_A\widehat{\mathbf{V}}\mathbf{C}_A')^{-1}\mathbf{C}_A\mathbf{W}$ follows directly from Slutsky's Theorem and the definition of noncentral Chi-square distribution.

(b). We give only the proof for the case when both a and b go to ∞. The case that $a \to \infty$ while b stays bounded can be shown similarly. By Lemma A.4 and Proposition A.3, we only need to consider the asymptotic distribution of $n(a,b)\sqrt{a}(P_A(\mathbf{e})-MSE)$ under $H_0(A)$, where $n(a,b) = \min_{i,j}\{n_{ij}\}$, and $P_A(\mathbf{e}) = \dfrac{b}{a}\sum_{i=1}^{a}\widetilde{e}_{i..}^2$. Write

$$n(a,b)\sqrt{a}(P_A(\mathbf{e}) - MSE) = T_{1A}(\mathbf{e}) + T_{3A}(\mathbf{e}), \tag{A.1}$$

where

$$T_{1A}(\mathbf{e}) = \frac{n(a,b)}{b\sqrt{a}}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\bar{e}_{ij.}\bar{e}_{ij'.},\quad T_{3A}(\mathbf{e}) = \frac{n(a,b)}{b\sqrt{a}}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m\neq m'}^{n_{ij}}\frac{e_{ijm}e_{ijm'}}{n_{ij}(n_{ij}-1)}.$$

It is easy to see that $E(T_{1A}(\mathbf{e})) = E(T_{3A}(\mathbf{e})) = 0$ and, as $a,b \to \infty$,

$$\mathrm{Var}\left(T_{3A}(\mathbf{e})\right) = \frac{2n^2(a,b)}{ab^2}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{\sigma_{ij}^4}{n_{ij}(n_{ij}-1)} \le \frac{4}{ab^2}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\frac{n(a,b)}{n_{ij}}\sigma_{ij}^2\right)^2 \to 0, \tag{A.2}$$

$$\mathrm{Var}\left(T_{1A}(\mathbf{e})\right) = \frac{2n^2(a,b)}{b^2a}\sum_{i=1}^{a}\sum_{j\neq j'}^{b}\left(\frac{\sigma_{ij}^2}{n_{ij}}\right)\left(\frac{\sigma_{ij'}^2}{n_{ij'}}\right) = n^2(a,b)\tau_A + o(1). \tag{A.3}$$

By (A.1), (A.2), and (A.3) it suffices to find the asymptotic distribution of $T_{1A}(\mathbf{e})$. We use Lyapounov's Theorem. Since (A.3) is bounded, Lyapounov's condition is satisfied if $L(a,b) = \sum_{i=1}^{a} E\left|(ab^2)^{-1/2}\sum_{j\neq j'}^{b}n(a,b)\bar{e}_{ij.}\bar{e}_{ij'.}\right|^4 \to 0$. We have

$$L(a,b) = \sum_{i=1}^{a}\sum_{j\neq j'}^{b}\sum_{j_1\neq j_1'}^{b}\sum_{j_2\neq j_2'}^{b}\sum_{j_3\neq j_3'}^{b}\frac{n^4(a,b)}{a^2b^4}E(\bar{e}_{ij.}\bar{e}_{ij'.}\bar{e}_{ij_1.}\bar{e}_{ij_1'.}\bar{e}_{ij_2.}\bar{e}_{ij_2'.}\bar{e}_{ij_3.}\bar{e}_{ij_3'.})$$

$$= O\left(\sum_{i=1}^{a}\sum_{j\neq j'\neq j_1\neq j_1'}^{b}\frac{n^4(a,b)}{a^2b^4}E(\bar{e}_{ij.})^2 E(\bar{e}_{ij'.})^2 E(\bar{e}_{ij_1.})^2 E(\bar{e}_{ij_1'.})^2\right)$$

$$= O\left(\frac{1}{a^2b^4}\sum_{i=1}^{a}\left(\sum_{j=1}^{b}\frac{n(a,b)}{n_{ij}}\sigma_{ij}^2\right)^4\right) = O(a^{-1}), \tag{A.4}$$

where the second equality follows from the fact that, when the number of different elements among $\{j, j', j_1, j_1', j_2, j_2', j_3, j_3'\}$ is five or more, the expectation on the

right hand side of (A.4) is zero, and the fourth equality is due to the inequality

$$\left|\sum_{i=1}^{m} z_i\right|^p \leq m^{p-1} \sum_{i=1}^{m} |z_i|^p, \ m \geq 1, \ p \geq 1, \tag{A.5}$$

which for $p > 1$ follows from Hölder's inequality. This completes the proof under the null.

When the observations are from the alternative given by (2.3) with $R(a, b, N) = a^{-1/4}b^{-1/2}$, $X_{ijk} \sim F_{N,ij}(x) = F_{ij}(x) + a^{-1/4}b^{-1/2}A_i(x)$, and $E(X_{ijk}) = \int x dF_{ij}(x) + a^{-1/4}b^{-1/2}\int x dA_i(x)$. Denote the error as $e_{ijk} = X_{ijk} - E(X_{ijk})$. We can write MST_A as

$$MST_A = \sum_{i,j} \frac{(\widetilde{e}_{i\cdot\cdot} - \widetilde{e}_{\cdots})^2}{a-1} + \sum_{i=1}^{a} \frac{\left[\int x dA_i(x) - \int x d\overline{A}_\cdot(x)\right]^2}{\sqrt{a}(a-1)}$$
$$+ 2\sum_{i=1}^{a} \frac{(\widetilde{e}_{i\cdot\cdot} - \widetilde{e}_{\cdots})}{a^{1/4}(a-1)} \left[\int x dA_i(x) - \int x d\overline{A}_\cdot(x)\right]$$
$$= \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(\widetilde{e}_{i\cdot\cdot} - \widetilde{e}_{\cdots})^2}{a-1} + \frac{\delta_{12}}{\sqrt{a}} + O_p(a^{-3/4}).$$

Therefore

$$\sqrt{a}(F_X(A)-1) = \sqrt{a}\left(\frac{1}{a-1}\sum_{i=1}^{a}\sum_{j=1}^{b}(\widetilde{e}_{i\cdot\cdot}-\widetilde{e}_{\cdots})^2 - MSE\right)\frac{1}{MSE} + \frac{\delta_{12}}{MSE} + o_p(1).$$

Note that the first term resembles the test statistic under the null hypothesis though the variances may be different from that under the null. Apply the result under the null with the variance given under the alternative to complete the proof.

Proof of Theorem 2.2. Part (a) can be shown similarly as part (a) of Theorem 2.1, and is omitted.

(b) When the observations are from the alternative given by (2.7) with $R(a, b, c, N) = (ab)^{-1/4}c^{-1/2}$, $X_{ijkm} \sim F_{N,ijk}(x) = F_{ijk}(x) + (ab)^{-1/4}c^{-1/2}(AB)_{ij}(x)$, and $E(X_{ijkm}) = \int x dF_{ijk}(x) + (ab)^{-1/4}c^{-1/2}\int x d(AB)_{ij}(x)$. Then MST_{AB} can be written as

$$MST_{AB} = c\sum_{i,j} \frac{(\widetilde{\epsilon}_{ij\cdot\cdot} - \widetilde{\epsilon}_{i\cdots} - \widetilde{\epsilon}_{\cdot j\cdot} + \widetilde{\epsilon}_{\cdots})^2}{(a-1)(b-1)} + \frac{\delta_{22}}{\sqrt{ab}} + c\sum_{i,j} \frac{(\widetilde{\epsilon}_{ij\cdot\cdot} - \widetilde{\epsilon}_{i\cdots} - \widetilde{\epsilon}_{\cdot j\cdot} + \widetilde{\epsilon}_{\cdots})}{\sqrt{b}(a-1)(b-1)}$$
$$\times \left[\int x d(AB)_{ij}(x) - \int x d\overline{(AB)}_{i\cdot}(x) - \int x d\overline{(AB)}_{\cdot j}(x) + \int x d\overline{(AB)}_{\cdots}(x)\right]$$

$$= c \sum_{i,j} \frac{(\widetilde{\epsilon}_{ij..} - \widetilde{\epsilon}_{i...} - \widetilde{\epsilon}_{.j..} + \widetilde{\epsilon}_{....})^2}{(a-1)(b-1)} + \frac{\delta_{22}}{\sqrt{ab}} + O_p(a^{-1/2}b^{-1}).$$

Note that the first term is equivalent to MST_{AB} calculated under the null hypothesis, but with error terms having variance given under the alternatives, so if we can prove the result under the null hypothesis, the rest of the proof follows from Lemma A.5 with an application of Slutsky's Theorem.

In the rest of the proof, we consider the test statistics under the null hypothesis. Let $\sigma_{ijk} = \mathrm{Var}\,(X_{ijkm})$ under the null. By Proposition A.2 and Lemma A.5, we only need to consider the asymptotic distribution of $n(a,b,c)\sqrt{ab}(P_{2,AB}(\mathbf{e}) - MSE)$ under $H_0(AB)$, where $e_{ijk} = X_{ijkm} - E(X_{ijkm})$. Write $n(a,b,c)\sqrt{ab}$ $(P_{2,AB}(\mathbf{e}) - MSE) = a^{-1/2}\sum_{i=1}^{a} W_{i,AB}$, where the

$$W_{i,AB} = \frac{n(a,b,c)}{bc}\Bigg[-\frac{c^2}{b-1}\sum_{j \neq j'}^{b} \widetilde{e}_{ij..}\widetilde{e}_{ij'..} + \sum_{j=1}^{b}\sum_{k \neq k'}^{c} \widetilde{e}_{ijk.}\widetilde{e}_{ijk'.}$$

$$+ \sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m \neq m'}^{n_{ijk}} \frac{e_{ijkm}e_{ijkm'}}{n_{ijk}(n_{ijk}-1)}\Bigg]$$

are independent with zero mean. We use Lyapounov's Theorem. Some algebra shows that

$$\mathrm{Var}\,\Big(n(a,b,c)\sqrt{ab}(P_{2,AB}(\mathbf{e}) - MSE)\Big)$$

$$= \frac{2n^2(a,b,c)}{bc^2a}\sum_{i=1}^{a}\Bigg[\frac{1}{(b-1)^2}\sum_{j \neq j'}^{b}\Big(\sum_{k=1}^{c}\frac{\sigma_{ijk}^2}{n_{ijk}}\Big)\Big(\sum_{k'=1}^{c}\frac{\sigma_{ij'k'}^2}{n_{ij'k'}}\Big)$$

$$+ \sum_{j=1}^{b}\sum_{k \neq k'}^{c}\Big(\frac{\sigma_{ijk}^2}{n_{ijk}}\Big)\Big(\frac{\sigma_{ijk'}^2}{n_{ijk'}}\Big)\Bigg] + \frac{2n^2(a,c)}{bc^2a}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{\sigma_{ijk}^4}{n_{ijk}(n_{ijk}-1)}$$

$$= n^2(a,b,c)b[\tau_1^2 + \tau_2^2 + \tau_3^2],$$

where τ_1^2, τ_2^2, and τ_3^2 are defined in the statement of the theorem, stays bounded away from zero and ∞. Lyapounov's condition is satisfied if $\sum_{i=1}^{a} E|(1/\sqrt{a})$ $W_{i,AB}|^4 \to 0$. Using (A.5), we have

$$E(W_{i,AB}^4) \leq \frac{n^4(a,b,c)}{b^4c^4}3^3\Bigg[E\Big|\frac{c^2}{b-1}\sum_{j \neq j'}^{b}\widetilde{e}_{ij..}\widetilde{e}_{ij'..}\Big|^4 + E\Big|\sum_{j=1}^{b}\sum_{k \neq k'}^{c}\overline{e}_{ijk.}\overline{e}_{ijk'.}\Big|^4$$

$$+ E\Big|\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m \neq m'}^{n_{ijk}}\frac{e_{ijkm}e_{ijkm'}}{n_{ijk}(n_{ijk}-1)}\Big|^4\Bigg]. \tag{A.6}$$

For the rest of the proof, we need to separate the cases that b stays bounded or goes to infinity. The techniques are similar so we only give the case that b is bounded. When b is bounded, the second term above gives

$$\frac{n^4(a,b,c)}{a^2c^4} \sum_{i=1}^{a} E \Big| \sum_{j=1}^{b} \sum_{k \neq k'}^{c} \overline{e}_{ijk.} \overline{e}_{ijk'.} \Big|^4$$

$$= \sum_{i=1}^{a} \sum_{j,j_1,j_2,j_3} \sum_{k \neq k'}^{c} \sum_{k_1 \neq k_1'} \sum_{k_2 \neq k_2'}^{c} \sum_{k_3 \neq k_3'}^{c}$$

$$\frac{n^4(a,b,c)}{a^2c^4} E(\overline{e}_{ijk.} \overline{e}_{ijk'.} \overline{e}_{ij_1k_1.} \overline{e}_{ij_1k_1'.} \overline{e}_{ij_2k_2.} \overline{e}_{ij_2k_2'.} \overline{e}_{ij_3k_3.} \overline{e}_{ij_3k_3'.})$$

$$= O\Big(\sum_{i=1}^{a} \sum_{j,j_1}^{b} \sum_{k \neq k' \neq k_1 \neq k_1'}^{c} \frac{n^4(a,b,c)}{a^2c^4} \frac{\sigma_{ijk}^2 \sigma_{ijk'}^2 \sigma_{ij_1k_1}^2 \sigma_{ij_1k_1'}^2}{n_{ijk} n_{ijk'} n_{ij_1k_1} n_{ij_1k_1'}} \Big)$$

$$= O\Big(\frac{1}{a^2c^4} \sum_{i=1}^{a} \Big(\sum_{j=1}^{b} \sum_{k=1}^{c} \frac{n(a,b,c)}{n_{ijk}} \sigma_{ijk}^2 \Big)^4 \Big)$$

$$= O\Big(\frac{b^3c^3}{a^2c^4} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sigma_{ijk}^8 \Big) = O(a^{-1}), \tag{A.7}$$

where the second equality is because the expectation under the summation is zero when the number of different elements in $\{j, j_1, j_2, j_3\}$ is more than two, or the number of different elements in $\{k, k', k_1, k_1', k_2, k_2', k_3, k_3'\}$ is more than four, and the 5th equality is due to (A.5). Similarly, the first term in (A.6) gives $(n^4(a,c)/a^2c^4) \sum_{i=1}^{a} E \Big| \sum_{j \neq j'}^{b} \widetilde{e}_{ij..} \widetilde{e}_{ij'..} \Big|^4 = O(a^{-1})$. Finally, the last term in (A.6) gives

$$\frac{n^4(a,b,c)}{a^2c^4} \sum_{i=1}^{a} E \Big| \sum_{j,k} \sum_{m \neq m'}^{n_{ijk}} \frac{e_{ijkm} e_{ijkm'}}{n_{ijk}(n_{ijk}-1)} \Big|^4$$

$$= \frac{n^4(a,b,c)}{a^2c^4} \sum_{i=1}^{a} \sum_{j,j_1,j_2,j_3}^{b} \sum_{k,k_1,k_2,k_3}^{c} \sum_{m \neq m'}^{n_{ijk}} \sum_{m_1 \neq m_1'}^{n_{ij_1k_1}} \sum_{m_2 \neq m_2'}^{n_{ij_2k_2}} \sum_{m_3 \neq m_3'}^{n_{ij_3k_3}}$$

$$\frac{E(e_{ijkm} e_{ijkm'} e_{ij_1k_1m_1} e_{ij_1k_1m_1'} e_{ij_2k_2m_2} e_{ij_2k_2m_2'} e_{ij_3k_3m_3} e_{ij_3k_3m_3'})}{n_{ijk}(n_{ijk}-1) n_{ij_1k_1}(n_{ij_1k_1}-1) n_{ij_2k_2}(n_{ij_2k_2}-1) n_{ij_3k_3}(n_{ij_3k_3}-1)}$$

$$= O\Big(\frac{n^4(a,b,c)}{a^2c^4} \sum_{i=1}^{a} \sum_{j,j_1}^{b} \sum_{k,k_1}^{c} \sum_{m \neq m'}^{n_{ijk}} \sum_{m_1 \neq m_1'}^{n_{ij_1k_1}} \frac{E(e_{ijkm}^2) E(e_{ijkm'}^2) E(e_{ij_1k_1m_1}^2) E(e_{ij_1k_1m_1'}^2)}{n_{ijk}^2(n_{ijk}-1)^2 n_{ij_1k_1}^2(n_{ij_1k_1}-1)^2} \Big)$$

$$= O\Big(\frac{n^4(a,b,c)}{a^2c^4} \sum_{i=1}^{a} \Big(\sum_{j,k} \frac{\sigma_{ijk}^4}{n_{ijk}(n_{ijk}-1)} \Big)^2 \Big) = O\left(a^{-1}c^{-2} \right). \tag{A.8}$$

Thus Lyapounov's condition is satisfied and this completes the proof.

Proof of Theorem 2.3. Part (a) of the theorem can be similarly proved as part (a) of Theorem 2.1. The proof for part (b) includes three cases regarding whether b and c stay bounded, or go to infinity. These proofs follow similar arguments so we give only that for part (b). Please refer to Wang (2004) for detailed proof of the other parts.

(b). We only give the proof when both a, c are large and b is small. For other cases, see Wang (2004). Similar to the proof of part (b) of Theorem 2.2, we only need to consider the asymptotic distribution under the null hypothesis. The proof below assumes the null hypothesis is satisfied. By Slutsky's Theorem, Proposition A.1, and Lemma A.6, it suffices to consider the asymptotic distribution of $n(a,b,c,d)\sqrt{ac}(P_{1,ABC}(\mathbf{e}) - MSE)$ under $H_0(ABC)$, where $e_{ijklm} = X_{ijklm} - E(X_{ijklm})$. To apply Lyapounov's Theorem, write $n(a,b,c,d)\sqrt{ac}(P_{1,ABC}(\mathbf{e}) - MSE) = (\sqrt{ac})^{-1}\sum_{i=1}^{a}\sum_{k=1}^{c} W_{ik}$, where the

$$W_{ik} = \frac{n(a,b,c,d)}{bd}\Bigg[\sum_{j,l}\sum_{m\neq m'}\frac{e_{ijklm}e_{ijklm'}}{n_{ijkl}(n_{ijkl}-1)} + \sum_{j=1}^{b}\sum_{l\neq l'}^{d}\overline{e}_{ijkl.}\overline{e}_{ijkl'.}$$
$$+\frac{d^2}{(b-1)}\sum_{j\neq j'}^{b}\widetilde{e}_{ijk..}\widetilde{e}_{ij'k..}\Bigg]$$

are independent with zero mean. Thus,

$$acb^2 d^2 \text{Var}\left(\sqrt{ac}(P_{1,ABC}(\mathbf{e}) - MSE)\right)$$

$$= 2\sum_{i,j,k,l}\frac{\sigma_{ijkl}^4}{n_{ijkl}(n_{ijkl}-1)} + 2\sum_{i,j,k}\sum_{l\neq l'}^{d}\frac{\sigma_{ijkl}^2}{n_{ijkl}}\frac{\sigma_{ijkl'}^2}{n_{ijkl'}}$$

$$+\frac{2}{(b-1)^2}\sum_{i,k}\sum_{j\neq j'}^{b}\Big(\sum_{l=1}^{d}\frac{\sigma_{ijkl}^2}{n_{ijkl}}\Big)\Big(\sum_{l'=1}^{d}\frac{\sigma_{ij'kl'}^2}{n_{ij'kl'}}\Big)$$

$$= 2\sum_{i,k}\Bigg[\sum_{j,l}\frac{\sigma_{ijkl}^4}{n_{ijkl}^2(n_{ijkl}-1)} + \frac{b^2-2b}{(b-1)^2}\sum_{j}\Big(\sum_{l=1}^{d}\frac{\sigma_{ijkl}^2}{n_{ijkl}}\Big)^2 + \frac{1}{(b-1)^2}\Big(\sum_{j,l}\frac{\sigma_{ijkl}^2}{n_{ijkl}}\Big)^2\Bigg]$$

$$= ac^2 b^2 d^2(\tau_4 + \tau_5 + \tau_6),$$

which is bounded away from zero and ∞. So, Lyapounov's condition holds if

$$\sum_{i=1}^{a}\sum_{k=1}^{c}E\left|\frac{1}{\sqrt{ac}}W_{ik}\right|^4 = \frac{1}{a^2 c^2}\sum_{i=1}^{a}\sum_{k=1}^{c}E(W_{ik}^4) \to 0. \qquad (A.9)$$

By inequality (A.5),

$$E|W_{ik}^4| \leq 3^3\frac{[n(a,b,c,d)]^4}{(bd)^4}\Bigg[E\Big|\sum_{j,l}\sum_{m\neq m'}\frac{e_{ijklm}e_{ijklm'}}{n_{ijkl}(n_{ijkl}-1)}\Big|^4 + E\Big|\sum_{j=1}^{b}\sum_{l\neq l'}^{d}\overline{e}_{ijkl.}\overline{e}_{ijkl'.}\Big|^4$$

$$+\Big[\frac{1}{(b-1)}\Big]^4 E\Big|\sum_{j\neq j'}^{b}\sum_{l,l'}\overline{e}_{ijkl.}\overline{e}_{ij'kl'.}\Big|^4\Big].$$

Similar to the proof of (A.7) and (A.8), we obtain that (A.9) is bounded by

$$3^3\frac{n^4(a,b,c,d)}{a^2c^2b^4d^4}\sum_{i,k}\Big\{E\Big|\sum_{j,l}\sum_{m\neq m'}\frac{e_{ijklm}e_{ijklm'}}{n_{ijkl}(n_{ijkl}-1)}\Big|^4+E\Big|\sum_{j=1}^{b}\sum_{l\neq l'}^{d}\overline{e}_{ijkl.}\overline{e}_{ijkl'.}\Big|^4$$

$$+\frac{1}{(b-1)^4}E\Big|\sum_{j\neq j'}^{b}\sum_{l,l'}\overline{e}_{ijkl.}\overline{e}_{ij'kl'.}\Big|^4\Big\}$$

$$= O\left(a^{-1}c^{-1}b^{-2}d^{-2}\right)+O\left(a^{-1}c^{-1}\right)+O\left(a^{-1}c^{-1}(b-1)^{-4}\right).$$

Thus Lyapounov's condition is satisfied and the proof is complete.

References

Akritas, M. and Arnold, S. (2000). Asymptotics for ANOVA when the number of levels is large. *J. Amer. Statist. Assoc.* **95**, 212-226.

Akritas, M. and Papadatos, N. (2004). Heteroscedastic one-way ANOVA and lack-of-fit test. *J. Amer. Statist. Assoc.* **99**, 368-382.

Bathke, A. (2002). ANOVA for a large number of treatments. *Math. Methods Statist.* **11**, 118-132.

Bathke, A. (2004). The ANOVA F test can still be used in some balanced designs with unequal variances and nonnormal data. *J. Statist. Plann. Inference* **126**, 413-422.

Bathke, A. and Harrar, S. (2008). Nonparametric methods in multivariate factorial designs for large number of factor levels. *J. Statist. Plann. Inference* **138**, 588-610.

Boos, D. and Brownie, C. (1995). ANOVA and rank tests when the number of treatments is large. *Statist. Probab. Lett.* **23**, 183-191.

Bravata, D. M., McDonald, K. M., Smith, W. M., Rydzak, C., Szeto, H., Buckeridge, D. L., Haberland, C. and Owens, D. K. (2004). Systematic review: Surveillance systems for early detection of bioterrorism-related diseases. *Ann. Intern. Med.* **140**, 910-922.

Brunner, E., Dette, H. and Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *J. Amer. Statist. Assoc.* **92**, 1494-1502.

Brunner, E. and Puri, M. (2001). Nonparametric methods in factorial designs. *Statistical Papers* **42**, 1-52.

Choi, S., Hall, W. and Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.* **24**, 841-861.

Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91**, 674-688.

Fan, J. and Lin, S. (1998). Test of significance when data are curves. *J. Amer. Statist. Assoc.* **93**, 1007-1021.

Haberman, S. (1977). Log-linear models and frequency tables with small expected cell counts. *Ann. Statist.* **5**, 1148-1169.

Inglota, T. and Ledwina, T. (2005). Intermediate efficiency of some max-type statistics. *J. Statist. Plann. Inference* **136**, 2918-2935.

Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**, 295-327.

Li, H., Lindsay, B. and Waterman, R. (2003). Efficiency of projected score methods in rectangular array asymptotics. *J. Roy. Statist. Soc. Ser. B* **65**, 191-208.

Marsden-Haug N, Foster V, Hakre, S., Anderson, A. and Pavlin, J. (2005). Evaluation of joint services installation pilot project and BioNet Syndromic Surveillance Systems - United States, 2004. *MMWR* **54**, 194.

Mnatsakanyan, Z., D. Mollura, J. Ticehurst, M. Hashemian and L. Hung (2008). Electronic medical record (EMR) utilization for public health surveillance. *AMIA Annu. Symp. Proc.* **2008**, 480-484.

Neymann, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1-32.

Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16**, 356-366.

Reis, B., Pagano, M. and Mandl, K. (2003). Using temporal context to improve biosurveillance. *Proc. Natl. Acad. Sci. USA* **100**, 1961-1965.

Sahai, H. and Ageel, M. (2000). *The Analysis of Variance : Fixed, Random, and Mixed Models.* Springer Verlag, New York.

Sarandasa, H. and Altan, S. (1998). The analysis of small-sample multivariate data. *J. Biopharm. Statist.* **8**, 163-186.

Toufighi, K., S. M. Brady, R. Austin, E. Ly and N. J. Provart (2005). The botany array resource: e-Northerns, expression angling, and promoter analyses. *The Plant J.* **43**, 153-163.

Wang, H. (2004). Testing in multifactor heteroscedastic ANOVA and repeated measures designs with large number of levels. Ph.D. thesis, Pennsulvania State University.

Wang, H. and M. Akritas (2004). Rank tests for ANOVA with large number of factor levels. *J. Nonparametr. Stat.* **16**, 563-589.

Wang, L. and M. Akritas (2006). Two-way heteroscedastic ANOVA when the number of levels is large. *Statist. Sinica* **16**, 1387-1408.

Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the different classes. *J. Amer. Statist. Assoc.* **29**, 51-56.

Department of Statistics, Kansas State University, Manhattan KS 66506-0802, USA.

E-mail: hwang@ksu.edu

Department of Statistics, Pennsylvania State University, 326 Thomas Bldg., University Park, PA 16802, USA.

E-mail: mga@stat.psu.edu