Statistica Sinica **24** (2014), 43-61
doi:http://dx.doi.org/10.5705/ss.2011.133

TWO-FOLD NESTED DESIGNS: THEIR ANALYSIS AND CONNECTION WITH NONPARAMETRIC ANCOVA

Shu-Min Liao and Michael G. Akritas

Amherst College and The Pennsylvania State University

Abstract: In the context of a nonparametric model for the unbalanced heteroscedastic two-fold nested design, we consider the problem of testing for the sub-class effect. The asymptotic theory pertains to cases with a large number of sub-classes, and small number of classes. It is shown that the classical F-statistic is very sensitive to departures from homoscedasticity, even in balanced designs. We propose new test statistics when heteroscedasticity is of the between-classes type, as well as for the general heteroscedastic design. Their asymptotic distributions, both under the null and local alternative hypotheses, are established. The ramifications of these results to the hypothesis of no covariate effect in the nonparametric analysis of covariance are discussed. Simulation studies compare the finite sample performance of the proposed statistics with those of the classical F-test and the GEE test. Two data sets are analyzed.

Key words and phrases: Asymptotic distributions, heteroscedasticity, nonparametric testing, non-normality, sub-class effect, unbalanced designs.

1. Introduction

The classical ANOVA model assumes that the error terms are i.i.d. normal, in which case F-statistics have certain optimality properties (cf., Arnold (1981, Chap. 7)). Arnold (1980) showed that the classical F-test is robust to the normality assumption if the sample sizes are large while the number of factor levels or groups is small. The past decade has witnessed the generation of large data sets, involving a multitude of factor levels, in several areas of scientific investigation. For example, in agricultural trials it is not uncommon to see a large number of treatments with a small number of replications per treatment. Another application arises in certain type of microarray data in which the nested factor corresponds to a large number of genes. As a consequence, testing in designs with a large number of factor levels has attracted considerable attention. The seminal paper by Neyman and Scott (1948) highlights some interesting features of high dimensional inference. See also Simons and Yao (1999), Li, Lindsay, and Waterman (2003), and Hall, Marron, and Neeman (2005) for some representative publications.

The motivating application for the present work comes from the Mussel Watch Project of the National Oceanic and Atmospheric Administration (NOAA), which monitors chemical and biological contaminant trends in sediment and bivalve tissue collected from hundreds of EDAs (Estuarine Drainage Areas) on the West Coast, the East Coast (North, Middle, and South Atlantic), the Gulf of Mexico, and the Great Lakes. Since each coastal region has its own EDAs, results of crossed designs are not appropriate for studying differences among the different EDAs. In this data set, the number of EDAs within each coastal region is relatively large, ranging from 30-60, while the cell sizes within each sub-class is small. While normality and homoscedasticity are difficult to ascertain with small sample sizes, Figure 1 suggests that these assumptions are violated. Thus, there is need for a test procedure for the subclass effects in a two-fold nested design which accommodates a large number of subclasses, small and unequal sample sizes, and non-normal and heteroscedastic errors.

We consider the model

$$Y_{ijk} = \mu_{ij} + \sigma_{ij} \cdot e_{ijk} = \mu + \alpha_i + \delta_{ij} + \sigma_{ij} \cdot e_{ijk}, \tag{1.1}$$

$i = 1, \ldots, r$, $j = 1, \ldots, c_i$, $k = 1, \ldots, n_{ij}$, where the class effects α_i and subclass effects δ_{ij} satisfy

$$\sum_{i=1}^{r} n_{i\cdot}\alpha_i = 0\,, \text{ where } n_{i\cdot} = \sum_{j=1}^{c_i} n_{ij}\,, \text{ and } \sum_{j=1}^{c_i} n_{ij}\delta_{ij} = 0,\, \forall i,$$

and the e_{ijk} are independent with

$$E(e_{ijk}) = 0, \quad Var(e_{ijk}) = 1. \tag{1.2}$$

In this context, we develop a procedure for testing $H_0 : \delta_{ij} = 0$ for all i, j. This hypothesis is also relevant in the seemingly different context of nonparametric ANCOVA. To explain the connection, consider observations $(V_{i\ell}, X_{i\ell})$, $i = 1, \ldots, r$, $\ell = 1, \ldots, n_i$, such that, given $X_{i\ell} = x$,

$$V_{i\ell} = m_i(x) + \sigma_{ix} e_{ix} = m + a_i + d_i(x) + \sigma_i(x) e_{ix}, \tag{1.3}$$

where $m = r^{-1} \sum_{i=1}^{r} E\left[m_i(X_{i\ell})\right]$, $a_i = E\left[m_i(X_{i\ell})\right] - m$, and $d_i(x) = m_i(x) - m - a_i$. Thus, the covariate effect in the ANCOVA model (1.3) is nonlinear and can depend on the class i. Next, for each i, order the covariate values within the ith class, and define c_i subclasses with the jth subclass corresponding to n_{ij} consecutive covariate values. Let Y_{ijk}, $k = 1, \ldots, n_{ij}$, denote the $V_{i\ell}$'s in the jth subclass. Then the Y_{ijk} behave as if they came from a two-fold nested design (1.1) and the test procedure for the hypothesis $H_0 : \delta_{ij} = 0$ can be used for testing the hypothesis of no covariate effects in (1.3).

The purpose of the present article is to provide valid test procedures that can perform well in unbalanced and/or heteroscedastic designs when the number of sub-classes is large. The proposed test statistics are of the general form $MST - MSE$, but the MSE is chosen so that $E(MSE) = E(MST)$ under the null hypothesis. Note that this last relation does not hold under heteroscedasticity for the classical definition MSE. The basic asymptotic technique we apply is based on finding the joint limiting distribution of (MST, MSE) through a suitable representation by a simpler, asymptotically equivalent, random vector.

The use of test statistics of the form $MST-MSE$ is justified by the fact that, when the number of levels goes to infinity, both the numerator and denominator degrees of freedom of the F statistic tend to infinity. Hence it makes sense to study the asymptotic distribution of $F - 1$ or, equivalently, of $MST - MSE$. Three distinct general approaches for obtaining the asymptotic distribution of such test statistics are developed in Akritas and Arnold (2000), Bathke (2002), and Akritas and Papadatos (2004). Wang and Akritas (2006) applied the Akritas and Papadatos (2004) approach to two-way designs, Gupta, Harrar, and Fujikoshi (2006) considered designs with multivariate data, while Wang and Akritas (2004) and Bathke and Harrar (2008) considered methods based on ranks. Wang and Akritas (2009, 2010) further extended their results to heteroscedastic functional data and high dimensional ANOVA in their most recent works. In all cases, the asymptotic distribution of $MST-MSE$ (with some scaling that depends on the number of factor levels) is normal with zero mean, and the test procedure rejects the null hypothesis for large values of the test statistic. For the proposed test statistics, this practice is justified by the fact that the asymptotic distribution of the test statistic under alternatives has positive mean.

It is known that the classical, normality-based F-test is sensitive to departures from the homoscedasticity assumption, especially when the design is unbalanced. This is confirmed by the simulation results shown in Table 2 where 10 of the 20 type I error rates the F achieved are 80% or higher at nominal level $\alpha = 0.05$. Moreover, for the nested design we consider, even under homoscedasticity, the classical F-test may not be asymptotically valid in the unbalanced design if the cell sizes are small, unless the model is normal. The theoretical explanation for this phenomenon, given in Section 3, is confirmed by the simulations reported in Table 1, where 10 of the 16 type I error rates the F test achieved with non-normal errors are 10% or higher at $\alpha = 0.05$.

The rest of this manuscript is organized as follows. The next section describes the statistical model for the unbalanced heteroscedastic two-fold nested design, and reviews the classical F-test procedure for the hypothesis of no sub-class effect. In Section 3 we present the asymptotic theory for $F - 1$, where F is the classical F-statistic in the homoscedastic case, and conclude that without normality the classical F statistic may not be asymptotically valid in the unbalanced case even under homoscedasticity. In Section 4 we propose a new test statistic for the between-classes heteroscedastic model, and present its asymptotic distribution. In Section 5 we propose a test statistic for the model with general heteroscedasticity, and present its asymptotic theory. All statistics developed can be adapted for testing the hypothesis $H_0 : d_i(x) = 0$, for all i and x, of no covariate effect in the nonparametric ANCOVA model (1.3). The adaptation of the test under general heteroscedasticity is given in Section 5. All simulation results are shown in Section 6. Two data sets are analyzed in Section 7, the second of which illustrates the application to an ANCOVA setting. Finally, Section 8 states conclusions. Proofs of the results presented in Sections 3−5 are given in the online appendix.

2. The Statistical Model and Test Statistic

Consider the two-fold nested model (1.1), (1.2) with the μ_{ij} and σ_{ij} bounded, and note that it does not assume that the errors e_{ijk} are normally, or even identically, distributed. Thus, ordinal discrete data are included in this formulation. Let

$$C = \sum_{i=1}^{r} c_i, \quad n_{i\cdot} = \sum_{j=1}^{c_i} n_{ij}, \quad N_C = \sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij} = \sum_{i=1}^{r} n_{i\cdot\cdot}.$$

We are mainly interested in testing H_0: $\delta_{ij} = 0$ (no sub-class effect). Let

$$MS\delta = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot})^2}{C - r}, \tag{2.1}$$

$$MSE = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}}(Y_{ijk} - \bar{Y}_{ij\cdot})^2}{N_C - C}, \tag{2.2}$$

where $\bar{Y}_{ij\cdot}$ and $\bar{Y}_{i\cdot\cdot}$ are the corresponding unweighted means of Y_{ijk} within each sub-class and within each class. Then, the usual F-test statistic for testing H_0: $\delta_{ij} = 0$ is

$$F_C^{\delta} = \frac{MS\delta}{MSE}. \tag{2.3}$$

Under the normal homoscedastic model where the e_{ijk} are assumed to be iid $N(0, 1)$ and all $\sigma_{ij} = \sigma$, we have that

$$F_C^{\delta} \sim F_{C-r, N_C-C}, \text{ under } H_0 : \delta_{ij} = 0. \tag{2.4}$$

In what follows we examine the robustness of this procedure to departures from the assumptions of normality and homoscedasticity as the number of sub-classes gets large. We use the notations

$$\bar{n}_{ic_i} = \frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij} = \frac{n_{i\cdot}}{c_i}, \quad \underline{n}_{ic_i} = \frac{1}{c_i}\sum_{j=1}^{c_i}\frac{1}{n_{ij}}, \quad \bar{n}_C = \frac{1}{C}\sum_{i=1}^{r} n_{i\cdot} = \frac{N_C}{C}. \tag{2.5}$$

All results, except those of Section 5, are derived under conditions on the sample sizes: here exist numbers $\lambda_i \in (0,1)$, $\bar{n}_i > 1$, and $\underline{n}_i \in (0,\infty)$ such that, as $\min(c_i) \to \infty$,

$$\sqrt{C}\left(\frac{c_i}{C} - \lambda_i\right) \to 0, \quad \sqrt{c_i}\,(\bar{n}_{ic_i} - \bar{n}_i) \to 0, \quad \underline{n}_{ic_i} \longrightarrow \underline{n}_i. \tag{2.6}$$

Finally, we set $\bar{n} = \sum_{i=1}^{r} \lambda_i \bar{n}_i$.

3. Homoscedastic Designs

In this section we consider the unbalanced two-fold nested design with homoscedastic errors and derive the asymptotic distribution of F_C^δ, defined in (2.3). As a corollary of Theorem 1, we find that the usual, normal-based, F-test procedure is not robust to departures from the normality assumption even under homoscedasticity.

Theorem 1. *Under* (1.1) *with* $\sigma_{ij} = \sigma$, (1.2), (2.6), *and the decomposition of the means given in* (1.1), *assume that*

$$E(e_{ijk}^3) = 0, \quad E(e_{ijk}^4) = \kappa_i, \text{ and } E|e_{ijk}|^{4+2\epsilon} < \infty \text{ for some } \epsilon > 0.$$

Then, under alternatives δ_{ij} *that satisfy*

$$\sqrt{c_i}\left(\frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\frac{\delta_{ij}^2}{\sigma^2} - \theta_i\right) \to 0, \quad \text{as } \min(c_i) \to \infty \text{ while } r,\ n_{ij} \text{ stay fixed,}$$

where $\theta_i \geq 0$, $i = 1, \ldots, r$, *with some being strictly positive,*

$$\sqrt{C}\left(F_C^\delta - (1+\theta)\right) \xrightarrow{d} N\left(0,\, \Sigma_s\right), \tag{3.1}$$

where, with λ_i, $\bar{n}_i$, $\underline{n}$, *and* $\bar{n}$ *given in* (2.6), $\theta = \sum_{i=1}^{r} \lambda_i \theta_i$, *and*

$$\Sigma_s = 2 + 4\theta + \frac{2(1+\theta)^2}{\bar{n}-1} + \sum_{i=1}^{r}\left[(\kappa_i - 3)\lambda_i \frac{(2\theta+\bar{n})(\bar{n}\underline{n}_i - 1) + (2\theta+1)(\bar{n}_i - \bar{n}) + \theta^2(\bar{n}_i + \underline{n}_i - 2)}{(\bar{n}-1)^2}\right].$$

Under the null hypothesis $H_0 : \delta_{ij} = 0$, *which results in* $\theta = 0$, *we then have*

$$\sqrt{C}\left(F_C^\delta - 1\right) \xrightarrow{d} N\left(0,\, 2 + \frac{2}{\bar{n}-1} + \sum_{i=1}^{r}\left[\frac{(\kappa_i - 3)\lambda_i(\bar{n}^2\underline{n}_i - 2\bar{n} + \bar{n}_i)}{(\bar{n}-1)^2}\right]\right). \tag{3.2}$$

Corollary 1. *Under the model and assumptions of Theorem* 1, *the classical, normality-based, F-test procedure for the hypothesis* $H_0 : \delta_{ij} = 0$, *shown in* (2.4), *is* not *asymptotically valid when the model is not normal, unless* $n_{ij} = n$, $\forall i,\ j$, *or* $\kappa_i = 3$, $\forall i$.

It can be shown that if normality holds, the test procedure implied by Theorem 1 is asymptotically equivalent to the classical F-test procedure under $H_0 : \delta_{ij} = 0$.

4. Between-classes Heteroscedastic Designs

In this section we consider the heteroscedastic unbalanced two-fold nested design, but assume we have *between-classes heteroscedasticity*, $\sigma_{ij} = \sigma_i$, in (1.1). It can be shown that if the design is unbalanced then, under heteroscedasticity, it is no longer true that $E(MSE) = E(MS\delta)$ under the null hypothesis $H_0 : \delta_{ij} = 0$. Thus it is clear that the usual F-test procedure is not valid even under normality.

The idea of our proposed test statistic is to replace MSE by a different linear combination of the cell sample variances in order to match the expected value of $MS\delta$ under the null hypothesis. This achieved by replacing MSE by MSE^*, defined as

$$MSE^* = \frac{1}{C-r} \sum_{i=1}^{r} \frac{c_i - 1}{n_{i\cdot} - c_i} \sum_{j=1}^{c_i} \sum_{k=1}^{n_{ij}} \left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)^2 .$$

It is easily seen that MSE^* satisfies $E(MSE^*) = E(MS\delta)$ under the null hypothesis. Then we define the corresponding test statistic as

$$F_C^* - 1 = \frac{MS\delta}{MSE^*} - 1. \tag{4.1}$$

It is easy to verify that, in the balanced case, $F_C^* = F_C^\delta$, where F_C^δ is the classical F-statistic given in (2.3).

The asymptotic distribution of this test statistic, $F_C^* - 1$, is given in Theorem 2. In Corollary 1 we obtain that, under heteroscedasticity, the classical F-test procedure is not valid in the balanced case (where $F_C^* = F_C^\delta$) even under normality.

Theorem 2. *Under* (1.1)*, but with* $\sigma_{ij} = \sigma_i$*, consider alternatives* δ_{ij} *that satisfy*

$$\sqrt{c_i}\left(\frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\frac{\delta_{ij}^2}{\sigma_i^2} - \theta_i\right) \to 0, \quad \textit{as } \min(c_i) \to \infty \textit{ while } r,\ n_{ij} \textit{ stay fixed,}$$

where $\theta_i \geq 0$*,* $i = 1, \ldots, r$*, with some being strictly positive. Then*

$$\sqrt{C}\left(F_C^* - (1+\theta^*)\right) \xrightarrow{d} N\left(0, \Sigma_s^*\right), \tag{4.2}$$

where, with λ_i*,* $\bar{n}_i$*,* $\underline{n}_i$ *and* $\bar{n}$ *given in* (2.6)*,*

$$\theta^* = \frac{\theta^\sigma}{\beta}, \quad \textit{where} \quad \theta^\sigma = \sum_{i=1}^{r} \sigma_i^2 \lambda_i \theta_i \quad \textit{and} \quad \beta = \sum_{i=1}^{r} \sigma_i^2 \lambda_i, \quad \textit{and}$$

$$\Sigma_s^* = \sum_{i=1}^r \frac{\lambda_i \sigma_i^4}{\beta^2} \left\{ 2 + 4\theta_i + \frac{2(1+\theta^*)^2}{\bar{n}_i - 1} \right.$$
$$\left. + \frac{\kappa_i - 3}{(\bar{n}_i - 1)^2} \left[(\bar{n}_i + 2\theta^*)(\bar{n}_i \underline{n}_i - 1) + \theta^{*2}(\bar{n}_i + \underline{n}_i - 2) \right] \right\}.$$

Under the null hypothesis $H_0 : \delta_{ij} = 0$, which results in $\theta^* = 0$, we have

$$\sqrt{C}\,(F_C^* - 1) \xrightarrow{d} N\left(0, \sum_{i=1}^r \frac{\lambda_i \sigma_i^4}{\beta^2} \left[2 + \frac{2}{\bar{n}_i - 1} + \frac{(\kappa_i - 3)\bar{n}_i(\bar{n}_i \underline{n}_i - 1)}{(\bar{n}_i - 1)^2} \right] \right). \quad (4.3)$$

Corollary 2. *Under the assumptions of Theorem* 2*, if the design is balanced, $c_i = c$ and $n_{ij} = n$, then the test statistic F_C^* is equal to the classical F-test statistic F_C^δ and, as $c \to \infty$,*

$$\sqrt{C}\,(F_C^* - 1) \xrightarrow{d} N\left(0, \sum_{i=1}^r \frac{\lambda_i \sigma_i^4}{\beta^2} \left[2 + \frac{2}{n - 1} \right] \right) \quad (4.4)$$

under the null hypothesis $H_0 : \delta_{ij} = 0$.

5. General Heteroscedastic Designs

In this section we consider the general unbalanced heteroscedastic two-fold nested model at (1.1). Here $E(MSE) = E(MS\delta)$ is no longer true if the design is unbalanced. We thus replace MSE by

$$MSE^{**} = \frac{1}{C - r} \sum_{i=1}^r \sum_{j=1}^{c_i} \left(1 - \frac{n_{ij}}{n_{i\cdot}} \right) S_{ij}^2, \quad (5.1)$$

where $S_{ij}^2 = \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij\cdot})^2 / (n_{ij} - 1)$. The statistic for the general heteroscedastic case is then defined as

$$F_C^{**} - 1 = \frac{MS\delta}{MSE^{**}} - 1. \quad (5.2)$$

It is easy to verify that, in the balanced case, $F_C^{**} = F_C^* = F_C^\delta$, where F_C^δ is the classical F-statistic given in (2.3), and F_C^* is the test statistic under between-classes heteroscedastic designs given in (4.1). The asymptotic distribution of the test statistic $F_C^{**} - 1$ is given next.

Theorem 3. *Under* (1.1), (1.2), *and the decomposition of the means given in* (1.1), *assume that there exist κ_{ij}, λ_i, a_{1i}, a_{2i}, b_{1i}, b_{2i}, and b_{3i} such that, as $\min(c_i) \to \infty$, $E(e_{ijk}^3) = 0$, $E(e_{ijk}^4) = \kappa_{ij}$, and $E|e_{ijk}|^{4+2\epsilon} < \infty$ for some $\epsilon > 0$;*

$$\sqrt{C}\left(\frac{c_i}{C} - \lambda_i \right) \to 0,\ \sqrt{c_i}\left(\frac{1}{c_i} \sum_{j=1}^{c_i} \sigma_{ij}^2 - a_{1i} \right) \to 0, \quad \frac{1}{n_{i\cdot}} \sum_{j=1}^{c_i} n_{ij} \sigma_{ij}^2 \longrightarrow a_{2i}, \quad (5.3)$$

$$\frac{1}{c_i}\sum_{j=1}^{c_i}\sigma_{ij}^4 \longrightarrow b_{1i}, \quad \frac{1}{c_i}\sum_{j=1}^{c_i}\frac{\sigma_{ij}^4}{n_{ij}-1} \longrightarrow b_{2i}, \qquad \frac{1}{c_i}\sum_{j=1}^{c_i}\frac{\sigma_{ij}^4(\kappa_{ij}-3)}{n_{ij}} \longrightarrow b_{3i}.$$

Then, under alternatives δ_{ij} that satisfy, as $\min(c_i) \to \infty$ *while* $r,\ n_{ij}$ *stay fixed,*

$$\sqrt{c_i}\left(\frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\delta_{ij}^2 - \theta_{1i}\right) \to 0, \quad \frac{1}{c_i}\sum_{j=1}^{c_i} n_{ij}\delta_{ij}^2\sigma_{ij}^2 \longrightarrow \theta_{2i},$$

where $\theta_{1i} \geq 0$, $\theta_{2i} \geq 0$, $i = 1, \ldots, r$, with some being strictly positive, we have

$$\sqrt{C}\left(F_C^{**} - (1+\theta^{**})\right) \xrightarrow{d} N\left(0,\ \frac{1}{a_1^2}\left[2(b_1+b_2) + 4(\theta_2 + b_2\theta^{**}) + (2b_2+b_3)\theta^{**2}\right]\right),$$

where

$$\theta^{**} = \frac{\theta_1}{a_1}, \qquad \theta_1 = \sum_{i=1}^{r}\lambda_i\theta_{1i}, \quad a_1 = \sum_{i=1}^{r}\lambda_i a_{1i}, \tag{5.4}$$
$$b_1 = \sum_{i=1}^{r}\lambda_i b_{1i}, \quad b_2 = \sum_{i=1}^{r}\lambda_i b_{2i}, \quad b_3 = \sum_{i=1}^{r}\lambda_i b_{3i}, \quad \theta_2 = \sum_{i=1}^{r}\lambda_i\theta_{2i}.$$

Under the null hypothesis $H_0 : \delta_{ij} = 0$, which results in $\theta^{**} = 0$, we have

$$\sqrt{C}\left(F_C^{**} - 1\right) \xrightarrow{d} N\left(0,\ \frac{2b_1 + 2b_2}{a_1^2}\right). \tag{5.5}$$

To extend the above theory to the context of nonparametric ANCOVA, let $(V_{i\ell}, X_{i\ell})$, $i = 1, \ldots, r$, $\ell = 1, \ldots, n_i$, be observations according to the ANCOVA model (1.3), arranged within each i according to increasing covariate values. Assume that the variance function $\sigma_i^2(x)$ is continuous in x. Let n be a fixed number and, for each i, define $c_i = [n_i/n]$ subclasses corresponding to n consecutive X_{ij}. (The tail $n_i - nc_i$ observations can form their own group or be ignored.) Let Y_{ijk}, $k = 1, \ldots, n$, denote the $V_{i\ell}$s in the jth subclass, and let X_{ijk}, e_{ijk} denote the corresponding covariates and error terms. Set

$$\sigma_{ijk}^2 = \sigma_i^2(X_{ijk}), \quad \sigma_{ij}^2 = \frac{1}{n}\sum_{k=1}^{n}\sigma_{ijk}^2$$

and define F_C^{**} as in (5.2) using the present Y_{ijk}s.

Corollary 3. *Under* (1.3), *the assumptions of Theorem* 3, *and the null hypothesis $H_0 : d_i(x) = 0$, for all i and x,*

$$\sqrt{C}(F_C^{**} - 1) \xrightarrow{d} N\left(0, \frac{2b_1 + 2b_2}{a_1^2}\right),$$

where b_1, b_2, and a_1 are as defined in Theorem 3.

6. Simulations

Simulations were used to compare the performances of various test procedures. Let *CF* denote the classical F-test procedure, shown in (2.4), and *HOM*, *UW*, *HET* denote the test procedures implied by the asymptotic results of (3.2), (4.3), and (5.5), respectively. The procedure *CF* is compared with *HOM* for homoscedastic nested designs (Section 6.1), and with *UW* for between-classes heteroscedastic nested designs (Section 6.2). In Section 6.3 the *HET* procedure is compared to *UW* and *HOM* for both homoscedastic and heteroscedastic nested designs. These four procedures under nested designs are then further compared to the classical F-test based ANCOVA (denoted as *FtACV*) and the rank-based Drop test of McKean and Sheather (1991) (denoted as *DROP*) in the context of analysis of covariance in Section 6.4.

For all simulations under nested designs (Sections 6.1−6.3, the number of classes used was five ($r = 5$). The different combinations of numbers of sub-classes, with the average $\bar{c}$ in each case, were: (I) $\bar{c} = 15 \Leftrightarrow (c_1,\ c_2,\ c_3,\ c_4,\ c_5) = (7,\ 11,\ 15,\ 19,\ 23)$; (II) $\bar{c} = 30 \Leftrightarrow (c_1,\ c_2,\ c_3,\ c_4,\ c_5) = (15,\ 23,\ 30,\ 37,\ 45)$; (III) $\bar{c} = 100 \Leftrightarrow (c_1,\ c_2,\ c_3,\ c_4,\ c_5) = (50,\ 75,\ 100,\ 125,\ 150)$; (IV) $\bar{c} = 500 \Leftrightarrow (c_1,\ c_2,\ c_3,\ c_4,\ c_5) = (250,\ 375,\ 500,\ 625,\ 750)$. The number of observations in each sub-class (n_{ij}) was generated by $n_{ij} = Z_{ij} + \upsilon_i \times I(Z_{ij} = 0)$, where $I(\cdot)$ is an indicator function and $Z_{ij} \sim Poisson(\upsilon_i),\ i = 1, \ldots, 5;\ j = 1, \ldots, c_i$. The values of υ_i used in our simulations were $(\upsilon_1,\ \upsilon_2,\ \upsilon_3,\ \upsilon_4, \upsilon_5)' = (2,\ 2,\ 2,\ 12,\ 2)$ for homoscedastic cases, and $(\upsilon_1,\ \upsilon_2,\ \upsilon_3,\ \upsilon_4, \upsilon_5)' = (5,\ 5,\ 5,\ 12,\ 5)$ for heteroscedastic cases. The values of the other parameters in the decomposition (1.1) were as follows: $\mu = 0$, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)' = (-3, -2, -1, 2)'$ and α_5 was chosen so that $\sum_i n_{i\cdot}\alpha_i = 0$. After generating the n_{ij} and fixing all parameters, we randomly generated errors e_{ijk} from one of five distributions: (*i*) **Normal**: the standard normal; (*ii*) **Exponen**: the exponential distribution with $\lambda = 1$; (*iii*) **LogNorm**: the log-normal distribution whose logarithm has mean 0 and standard deviation 1; (*iv*) **Mixture**: the mixture distribution defined as $U_1 \cdot X_1 + (1 - U_1) \cdot Y_1$, where $U_1 \sim$ Bernoulli$(p = 0.9)$, $X_1 \sim N(-1, 1)$ and $Y_1 \sim N(9, 1)$; and (*v*) **Multi-d**: when $r = 1,\ 2,\ 3,\ 4$, generate e_{ijk} from *Normal*, *Exponen*, *LogNorm*, and *Mixture* as described above, respectively. When $r = 5$, the e_{ijk} were generated from the mixture distribution $U_2 \cdot X_2 + (1 - U_2) \cdot Y_2$, where $U_2 \sim$ Bernoulli$(p = 0.5)$, $X_2 \sim N(-3, .5)$ and $Y_2 \sim N(3, .5)$. All e_{ijk} were standardized to have mean 0 and standard deviation 1. As for the variances, we used $\sigma_{ij} = \sigma = 1,\ \forall i,\ j$, for homoscedastic designs, $(\sigma_{ij}) = (\sigma_i) = (\sigma_1,\ \sigma_2,\ \sigma_3,\ \sigma_4,\ \sigma_5) = (1,\ 1,\ 5,\ 1,\ 1),\ \forall j$, for between-classes heteroscedastic designs, and $\sigma_{ij} = 4 \cdot I(i = 3) + 5 \cdot I(j < 0.3\ c_i) + (j/c_i),\ \forall i,\ j$, for general heteroscedastic designs. All simulations were done over $10,000$ runs at nominal $\alpha = 0.05$.

In Section 6.4, we consider the ANCOVA model (1.3) with $r = 2$ and assume $m_1(x) = 0$ and $m_2(x) = 2$ under the null hypothesis of no covariate effects. We generated the covariate values X_{il} using $X_{1l} \sim U(2,\ 3)$ and $X_{2l} \sim U(1.5,\ 2.5)$. Let $V_{il} = m_i(x) + \sigma_{ix} e_{ix}$, with heteroscedasticity function $\sigma_i(x)$ and the error terms e_{ix} specified as (1) Case 1: $\sigma_1(X_{1l}) = 5$, $\sigma_2(X_{2l}) = 1$, and e_{ix} generated from the standard normal distribution; (2) Case 2: $\sigma_1(X_{1l}) = 5 + 10 \times I(X_{1l} < 2.2)$, $\sigma_2(X_{2l}) = 1 + 6 \times I(X_{2l} < 1.7)$, and e_{ix} generated from the standard normal distribution; (3) Case 3: $\sigma_1(X_{1l})$ and $\sigma_2(X_{2l})$ are the same as those in Case 2, but e_{ix} generated from the log-normal distribution whose logarithm has mean 0 and standard deviation 1. Three sample sizes were considered: $(n_1,\ n_2) = (150,\ 250), (400,\ 600)$, and $(800,\ 1,200)$.

6.1. Simulations under homoscedastic nested designs

We first compare the achieved sizes of two procedures, *CF* and *HOM*, under homoscedastic designs. The first procedure, *CF*, based on the classical normality-based F-test theorem, rejects at level α if

$$F_C^{\delta} > F^{\alpha}_{C-r,N_C-C}, \tag{6.1}$$

where F_C^{δ} is defined in (2.3) and F^{α}_{C-r,N_C-C} is the $(1-\alpha)100$th percentile of the F_{C-r,N_C-C} distribution. The second procedure, *HOM*, using the asymptotic null distribution shown in (3.2), rejects at level α if

$$\sqrt{C}(F_C^{\delta} - 1) > \sqrt{2 + \frac{2}{\bar{n}_C - 1} + \sum_{i=1}^{r}\left[\frac{(\hat{\kappa}_i - 3)\hat{\lambda}_i(\bar{n}_C^2 \underline{n}_{ic_i} - 2\bar{n}_C + \bar{n}_{ic_i})}{(\bar{n}_C - 1)^2}\right]}\, Z_{\alpha}, \tag{6.2}$$

where F_C^{δ} is as before and Z_{α} is the $(1-\alpha)100$th percentile of the standard normal distribution. In addition, $\hat{\lambda}_i$ and $\hat{\kappa}_i$ are the empirical versions of λ_i and κ_i, namely

$$\hat{\lambda}_i = \frac{c_i}{C}, \quad \hat{\kappa}_i = \frac{\hat{\nu}_i}{(MSE)^2}, \quad \text{where } \hat{\nu}_i = \frac{1}{n_{i\cdot}} \sum_j \sum_k (Y_{ijk} - \bar{Y}_{i\cdot\cdot})^4, \tag{6.3}$$

while $\bar{n}_{ic_i}$, $\underline{n}_{ic_i}$, and $\bar{n}_C$ are as defined in (2.5). It can be easily verified that, under the null hypothesis, $\hat{\kappa}_i \xrightarrow{P} \kappa_i$, as $\min(c_i) \to \infty$. The simulated sizes are shown in Table 1.

The results in Table 1 confirm the conclusions stated in Corollary 1. Thus, the classical *CF* is liberal in this unbalanced design for all non-normal distributions, with the achieved α-level increasing with the number of sub-classes. On the other hand, the proposed procedure *HOM* performed well for all distributions.

Table 1. Achieved α-levels under homoscedastic designs.

	$\bar{c}=15$		$\bar{c}=30$		$\bar{c}=100$		$\bar{c}=500$	
	CF	HOM	CF	HOM	CF	HOM	CF	HOM
Normal	0.0523	0.0680	0.0516	0.0604	0.0525	0.0581	0.0502	0.0518
Exponen	0.1004	0.0617	0.1044	0.0554	0.1110	0.0488	0.1151	0.0465
LogNorm	0.1725	0.0604	0.1917	0.0534	0.2314	0.0507	0.2679	0.0470
Mixture	0.1011	0.0711	0.0919	0.0576	0.0980	0.0561	0.1021	0.0535
Multi-d	0.0806	0.0595	0.0806	0.0573	0.0742	0.0526	0.0839	0.0521

Table 2. Achieved α-levels under between-classes heteroscedastic designs.

	$\bar{c}=15$		$\bar{c}=30$		$\bar{c}=100$		$\bar{c}=500$	
	CF	UW	CF	UW	CF	UW	CF	UW
Normal	0.3392	0.0843	0.5450	0.0723	0.9132	0.0630	1.0000	0.0554
Exponen	0.3135	0.0754	0.5217	0.0720	0.9066	0.0684	0.9998	0.0563
LogNorm	0.2937	0.0702	0.4594	0.0604	0.8464	0.0698	0.9930	0.0571
Mixture	0.2971	0.0784	0.5214	0.0734	0.9030	0.0653	0.9997	0.0590
Multi-d	0.3172	0.0783	0.5156	0.0723	0.9031	0.0685	0.9996	0.0588

6.2. Simulations under between-classes heteroscedastic nested designs

Here we compare the achieved sizes of *CF* and *UW*. The procedure, *CF*, is as shown in (6.1), while the other procedure, *UW*, using the asymptotic null distribution shown in (4.3), rejects at level α if

$$\sqrt{C}(F_C^* - 1) > \sqrt{\sum_{i=1}^{r} \frac{\hat{\lambda}_i \hat{\sigma}_i^4}{\hat{\beta}^2} \left[2 + \frac{2}{\bar{n}_{ic_i} - 1} + \frac{(\tilde{\kappa}_i - 3)\bar{n}_{ic_i}(\bar{n}_{ic_i}\underline{n}_{ic_i} - 1)}{(\bar{n}_{ic_i} - 1)^2} \right]} Z_\alpha, \quad (6.4)$$

where F_C^* is defined in (4.1). The empirical quantities $\hat{\lambda}_i$, $\bar{n}_{ic_i}$, $\underline{n}_{ic_i}$, and $\bar{n}_C$ are as defined in (6.3). Moreover, $\hat{\beta}$, $\hat{\sigma}_i^4$, and $\tilde{\kappa}_i$ above, are as follows:

$$\hat{\beta} = \sum_{i=1}^{r} \hat{\lambda}_i \hat{\sigma}_i^2, \quad \text{where } \hat{\sigma}_i^2 = S_i^2; \quad \hat{\sigma}_i^4 = (\hat{\sigma}_i^2)^2; \quad \text{and } \tilde{\kappa}_i = \frac{\hat{\nu}_i}{\hat{\sigma}_i^4}, \qquad (6.5)$$

where $\hat{\nu}_i$ is also defined in (6.3). Again, it can be easily verified that, as $\min(c_i) \to \infty$, $\tilde{\kappa}_i$ converges in probability to κ_i under the null hypothesis. The corresponding simulated sizes under heteroscedastic designs, based on 10,000 runs, are shown in Table 2.

Table 2 makes it clear that the traditional *CF* procedure is quite inappropriate for between-classes heteroscedastic designs. More specifically, when $\bar{c}$ is large enough, regardless of the underlying distribution, the *CF* procedure rejects the null hypothesis almost all the times under the null hypothesis. On the other

Table 3. Achieved α-levels under general heteroscedastic designs.

	$\bar{c} = 15$		$\bar{c} = 30$		$\bar{c} = 100$		$\bar{c} = 500$	
	UW	HET	UW	HET	UW	HET	UW	HET
Normal	0.0882	0.0854	0.0985	0.0837	0.1479	0.0701	0.1560	0.0575
Exponen	0.0865	0.0748	0.0896	0.0717	0.1333	0.0597	0.1545	0.0530
LogNorm	0.0726	0.0631	0.0738	0.0568	0.1092	0.0491	0.1225	0.0461
Mixture	0.0856	0.0655	0.0852	0.0567	0.1378	0.0554	0.1507	0.0508
Multi-d	0.0897	0.0778	0.0910	0.0718	0.1366	0.0635	0.1576	0.0550

hand, the proposed procedure UW performed well for all distributions, though liberally in the case of small number of sub-classes.

6.3. Simulations under general heteroscedastic nested designs

The simulations in the previous subsection demonstrate that the classical CF procedure is very liberal under between-classes heteroscedasticity. Simulations under general heteroscedasticity, not shown here, revealed similar behavior. Thus, the tables in this subsection exclude the CF procedure.

In Table 3 we compare the achieved α-levels of UW and HET, under general heteroscedasticity. The former procedure is described in (6.4), while the latter uses the statistic F_C^{**} given in (5.2) and its asymptotic null distribution shown in (5.5). Thus, the HET procedure rejects at level α if

$$\sqrt{C}(F_C^{**} - 1) > \sqrt{\frac{2\hat{b}_1 + 2\hat{b}_2}{\hat{a}_1^2}} Z_\alpha, \tag{6.6}$$

where $\hat{a}_1$, $\hat{b}_1$ and $\hat{b}_2$ are consistent estimators of a_1, b_1 and b_2. Note that consistent estimation of b_1 and b_2 needs unbiased estimation of each σ_{ij}^4. For such unbiased estimation we used the U-statistics with the kernel $(Y_{ij1} - Y_{ij2})^2/2 \times (Y_{ij3} - Y_{ij4})^2/2$. As a consequence, the application of procedure HET requires $n_{ij} \geq 4$, although Theorem 3 requires only $n_{ij} \geq 2$. From Table 3 we see that both procedures are liberal when the average number of sub-classes is 5, but HET becomes less so as $\bar{c}$ increases. On the other hand, UW becomes more liberal as $\bar{c}$ increases, a behavior which is expected in view of the fact that it is not designed to allow the present type of heteroscedasticity.

Table 4 performs a more detailed comparison of the procedures UW and HET under the setting of between-classes heteroscedasticity when both are asymptotically valid. It suggests that the achieved α-levels of the two procedures are comparably close to the nominal level, with HET slightly less liberal for non-normal distributions. It also shows the achieved powers of these two procedures, only for the case of $\bar{c} = 100$, under the alternatives $\delta_{ij} = t \times (2j/c_i - 1)$, for

Table 4. Achieved α-levels and Powers under between-classes heteroscedastic designs.

	$\bar{c} = 15$		$\bar{c} = 30$		$\bar{c} = 100$		$\bar{c} = 500$	
α-levels	UW	HET	UW	HET	UW	HET	UW	HET
Normal	0.0799	0.0899	0.0701	0.0771	0.0606	0.0612	0.0513	0.0539
Exponen	0.0779	0.0817	0.0720	0.0698	0.0628	0.0560	0.0566	0.0519
LogNorm	0.0722	0.0780	0.0655	0.0529	0.0651	0.0525	0.0624	0.0440
Mixture	0.0785	0.0752	0.0750	0.0633	0.0650	0.0536	0.0584	0.0494
Multi-d	0.0768	0.0797	0.0711	0.0655	0.0626	0.0567	0.0560	0.0505
($\bar{c} = 100$)	t=0.6		t=0.8		t=1.0		t=1.2	
Powers	UW	HET	UW	HET	UW	HET	UW	HET
Normal	0.3353	0.3410	0.6249	0.6214	0.9204	0.9173	1.0000	0.9998
Exponen	0.3308	0.3361	0.6210	0.6436	0.9083	0.9259	0.9997	0.9997
LogNorm	0.9750	0.9986	0.9941	1.0000	0.9986	1.0000	0.9996	1.0000
Mixture	0.3289	0.3218	0.6115	0.6417	0.9135	0.9350	0.9998	1.0000
Multi-d	0.3325	0.3347	0.6243	0.6549	0.9101	0.9304	0.9994	0.9997

$t = 0.6,\ 0.8,\ 1.0,\ 1.2$ and $i = 1, \ldots, 5$, $j = 1, \ldots, c_i - 1$. For each i, δ_{ic_i} is chosen so that $\sum_j n_{ij}\delta_{ij} = 0$. Again the procedures have comparable power with *HET* being slightly more powerful for non-normal distributions.

The next table performs a more detailed comparison of the procedures *HOM*, *UW*, and *HET* under homoscedasticity when all three are asymptotically valid. The results reported in Table 5 suggest that the achieved α-levels of the three procedures are comparably close to the nominal level (the results for $\bar{c} = 500$ are very close to those for $\bar{c} = 100$, so they are omitted). It also compares the achieved powers of these three procedures, for the case of $\bar{c} = 100$, under alternatives $\delta_{ij} = t \times (2j/c_i - 1)$, for $t = 0.20,\ 0.25,\ 0.35$ and $i = 1, \ldots, 5$, $j = 1, \ldots, c_i - 1$. For each i, δ_{ic_i} is chosen so that $\sum_j n_{ij}\delta_{ij} = 0$. Note that the cell sizes used here are larger than those used in Table 1, as required for the applicability of *HET*. The results suggest that, even though procedure *HET* estimates more parameters, this does not compromise its power.

Table 6 shows simulation results comparing the achieved α-levels of *CF*, *HET*, and *GEE*. Because the latest GEE package in R, *geepack*, appears to work only when the numbers of subclasses are all the same, we generated the data using the same setting as in Table 3, but let $c_i = 30$, $\forall i$ (while allowing the n_{ij} to be different), for the heteroscedastic unbalanced case. For comparison purposes, Table 6 also shows the corresponding results under homoscedasticity (with $\sigma_{ij} = 1$) and balancedness (with $c_i = 30$ and $n_{ij} = 100$) when all procedures are asymptotically valid. Clearly, *HET* is the only one which works in the context of unbalanced two-fold nested designs under general heteroscedasticity.

Table 5. Achieved α-levels and Powers under homoscedastic designs.

	$\overline{c}=15$			$\overline{c}=30$			$\overline{c}=100$		
α-levels	HOM	UW	HET	HOM	UW	HET	HOM	UW	HET
Normal	0.072	0.070	0.075	0.058	0.057	0.058	0.053	0.053	0.053
Exponen	0.072	0.066	0.065	0.061	0.058	0.056	0.056	0.053	0.050
LogNorm	0.065	0.055	0.053	0.062	0.053	0.044	0.057	0.053	0.041
Mixture	0.069	0.063	0.057	0.064	0.064	0.058	0.058	0.059	0.052
Multi-d	0.067	0.067	0.066	0.066	0.062	0.062	0.059	0.058	0.056
($\overline{c}=100$)	$t=0.20$			$t=0.25$			$t=0.35$		
Powers	HOM	UW	HET	HOM	UW	HET	HOM	UW	HET
Normal	0.436	0.429	0.427	0.735	0.725	0.724	0.993	0.992	0.992
Exponen	0.411	0.418	0.426	0.706	0.716	0.732	0.992	0.993	0.994
LogNorm	0.977	0.989	1.000	0.995	0.998	1.000	0.998	1.000	1.000
Mixture	0.417	0.421	0.428	0.711	0.714	0.727	0.992	0.992	0.995
Multi-d	0.419	0.420	0.422	0.716	0.720	0.725	0.992	0.992	0.993

Table 6. Achieved α-levels under general heteroscedastic and homoscedastic designs.

($c_i=30$)	Heteroscedastic Unbalanced				Homoscedastic Balanced			
	CF	UW	HET	GEE	CF	UW	HET	GEE
Normal	0.258	0.095	0.073	0.996	0.058	0.046	0.046	0.074
Exponen	0.226	0.077	0.062	1.000	0.052	0.054	0.050	0.160
LogNorm	0.230	0.076	0.053	1.000	0.048	0.048	0.042	0.320
Mixture	0.226	0.069	0.042	1.000	0.060	0.044	0.045	0.224
Multi-d	0.238	0.102	0.074	1.000	0.048	0.056	0.056	0.130

6.4. Simulations in the context of analysis of covariance

In Table 7 we compare the achieved α-levels of the aformentioned procedures to those of $FtACV$ and $DROP$ in the context of analysis of covariance. One can easily see that all proposed procedures perform similarly as they do in the two-fold nested designs. In addition, the procedure HET significantly outperforms the classical FtACV and the rank-based Drop test in all simulations.

7. Data Analyses: Two Empirical Studies

7.1. Mussel watch project data

One application for our methodology can be found through the National Oceanic and Atmospheric Administration's National Status and Trends Program. In 1986, this division undertook a very large scale project to monitor the levels of numerous chemical contaminants and organic chemical constituents in marine sediment and bivalve (mollusk) tissue samples. This project, dubbed the Mussel Watch Project, is still on-going and there are no apparent plans to discontinue it in the near future. There are currently over 300 coastal sites at which sediment and bivalve samples are collected and analyzed for the project. Each site is

Table 7. Achieved α-levels in the context of ANCOVA with $r = 2$.

Case 1:	Between-Groups Heteroscedasticity with Normal Errors						
	(n_1, n_2)	FtACV	DROP	CF	HOM	UW	HET
	(150, 250)	0.1224	0.2312	0.1262	0.1459	0.0731	0.0792
	(400, 600)	0.1158	0.2221	0.1304	0.1436	0.0666	0.0695
	(800, 1200)	0.1095	0.2195	0.1320	0.1419	0.0630	0.0632
Case 2:	General Heteroscedasticity with Normal Errors						
	(n_1, n_2)	FtACV	DROP	CF	HOM	UW	HET
	(150, 250)	0.2139	0.3067	0.1826	0.2016	0.1585	0.0974
	(400, 600)	0.1999	0.2893	0.1951	0.2076	0.1625	0.0857
	(800, 1200)	0.1970	0.2968	0.2020	0.2108	0.1577	0.0715
Case 3:	General Heteroscedasticity with Log-normal Errors						
	(n_1, n_2)	FtACV	DROP	CF	HOM	UW	HET
	(150, 250)	0.2240	0.9133	0.1443	0.1630	0.1264	0.0813
	(400, 600)	0.2107	0.9990	0.1640	0.1762	0.1310	0.0650
	(800, 1200)	0.2018	1.000	0.1669	0.1748	0.1285	0.0521

categorized as being within a certain Estuarine Drainage Area (EDA). See NOAA (1998) for more details. For our data analysis, we chose to analyze the Lead concentrations from years 1998 to 2005. We chose to analyze concentrations of Lead in tissue samples, specifically in the *Crassostrea virginica*, or American Oysters, from two different regions: Middle and South Atlantic, and the Gulf of Mexico. Due to the fact that nested in each region there are many EDAs, it is natural to consider regions as classes, and EDAs as sub-classes in our analysis. The main interest of our study is the sub-class effect. The boxplots of the lead concentration levels at each EDA, shown in Figure 1, suggest heteroscedasticity among different EDAs in the same region (general heteroscedasticity). Thus, the procedure *HET* seems to be an appropriate one for analyzing this data set. However, the results of application of the other procedures mentioned in this paper (i.e. *CF*, *HOM*, and *UW*) are also included for comparison purposes. Because the *HET* procedure requires at least four observations within each sub-class, we remove four EDAs with less than four observations from our data, resulting in 58 EDAs in total. (Another approach would be to impute values, but this is pursued elsewhere.)

Application of the procedures *CF*, *HOM*, *UW*, and *HET* on this data set yields p-values of 0.1076, 0.3136, 0.2008, and 0.0246, respectively, for the hypothesis of no EDA effect. Note that only procedure *HET* detects the effect of EDA at $\alpha = 0.05$. A closer examination of the data reveals that the largest sample variance estimate from EDA 'G120x' in the Gulf of Mexico region is 69.21, while the second largest one is only 2.45. This high variance of the data in EDA 'G120x' in fact results from a few outliers in a site named 'CBPP', as shown in Figure 1. After four data points from 'CBPP' are removed, the sample variance

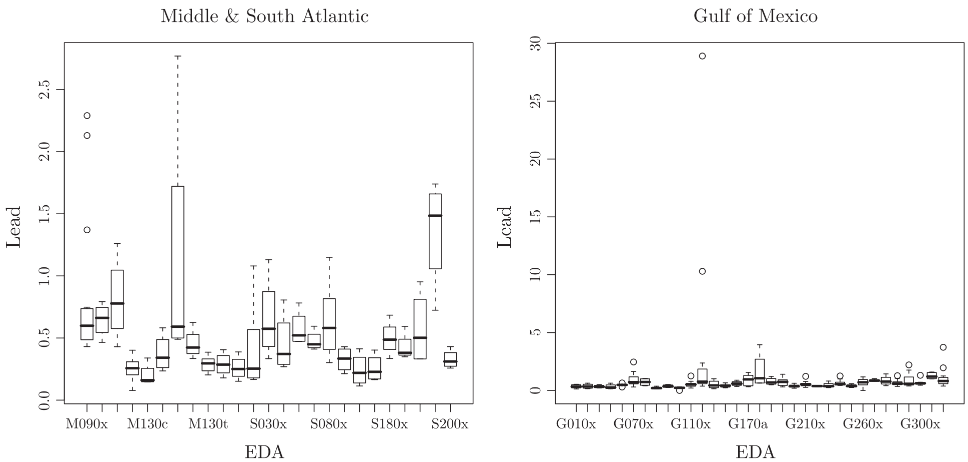

Figure 1. Mussel Watch Project. The boxplots of the Lead concentration levels at EDAs nested in Middle and South Atlantic (left) and in the Gulf of Mexico (right).

estimate of the EDA 'G120x' becomes 0.0921 and heteroscedasticity is not so pronounced. Thus, the p-values of four procedures are all very close to zero (less than 10^{-12}). This dramatic change confirms the instability of procedures *CF*, *HOM* and *UW* under general heteroscedasticity.

It is worth noting that a simple logarithmic transformation cannot be used here to resolve the heteroscedasticity issue as some data points have zero values. Even with those zero values being removed, model fitting with the log-transformed data is highly questionable because the corresponding diagnostic plots reveal severe violations of normality and homoscedastic assumptions.

7.2. NADP data: ramification for nonparametric ANCOVA

Another application for our methodology can be found through the National Atmospheric Deposition Program (2009), which monitors geographical and temporal long-term trends on the chemistry of precipitation. Starting from only 22 stations in 1978, NAPD has grown as a nationwide network of over 250 sites for which precipitation samples are collected and analyzed in the Central Analytical Laboratory (CAL) weekly. For our data analysis, we chose to analyze the pH level (reported as the negative log of hydrogen ion concentration) of precipitation samples as measured in the CAL from the first week of January 2003 to the last week of December 2007. We compared the data in two North Carolina towns, Lewiston and Coweeta, interested in the effect of time. Among the total 233 weeks in this period, there are several weeks in which data were missing at one or both locations. After removing those missing values, there are 180 weeks of data for each of these two towns, $n_1 = n_2 = 180$, although this balancedness is simply a coincidence; the missing data in fact are at different time points for the two locations.

This data set can be analyzed as a simple one-way ANCOVA model with locations as groups and time as the covariate. We applied the classical F-test based ANCOVA and the rank-based Drop test of McKean and Sheather (1991) to the data for the covariate effect of time and got p-values of .5737 and .7527, respectively, both indicating that the time effect is not significant. (Log-transforming the data did not change those p-values much in this example.) However, this insignificance may result from their lack of power when the effect of interest is not linear.

Here, we utilize the asymptotic results for the two-fold nested model to the one-way analysis of covariance. More specifically, we think of two locations as two *classes* and form artificial *sub-classes* by dividing the observations in the same class into non-overlapping 'windows' of a fixed size 5: the first time sub-class consists of observations from weeks 1–5, the second time sub-class consists of observations from weeks 6–10, and so on. Since there are 180 observations in each of two locations, this division results in $180/5 = 36$ sub-classes each class, $c_1 = c_2 = 36$. The boxplots of the pH levels at each of these 36 times for the two locations are shown in Figure 2. Note that this study does not present a typical repeated measures structure as there is only one observation per group/class at each time point. In addition, a simple time series analysis does not indicate meaningful correlation over time, so it appears reasonable to implement our methodology in this study.

For the sub-class effect of time, the four aforementioned procedures, *CF*, *HOM*, *UW*, and *HET*, give p-values of 0.0929, 0.0757, 0.0760, and 0.0508, respectively. (The p-values on the log-transformed data for *CF*, *HOM*, *UW*, and *HET* are 0.0673, 0.0495, 0.0499, and 0.0292, respectively.) Since the assumption of homoscedasticity is clearly violated as shown in Figure 2, we learn from theoretical derivations and numerical simulations that procedure *HET* is the only valid procedure under this type of heteroscedasticity. It is indeed the procedure that gives the lowest p-value among the five procedures.

8. Concluding Comments

We have established via theoretical derivations and numerical evidence that, when the number of sub-classes is large, the classical F-test procedure is very sensitive to departures from homoscedasticity regardless of whether the model is balanced or unbalanced. Even under homoscedasticity, it is still not asymptotically valid in unbalanced designs with non-normal errors. For this reason, we develop procedures that are asymptotically valid under heteroscedasticity. We

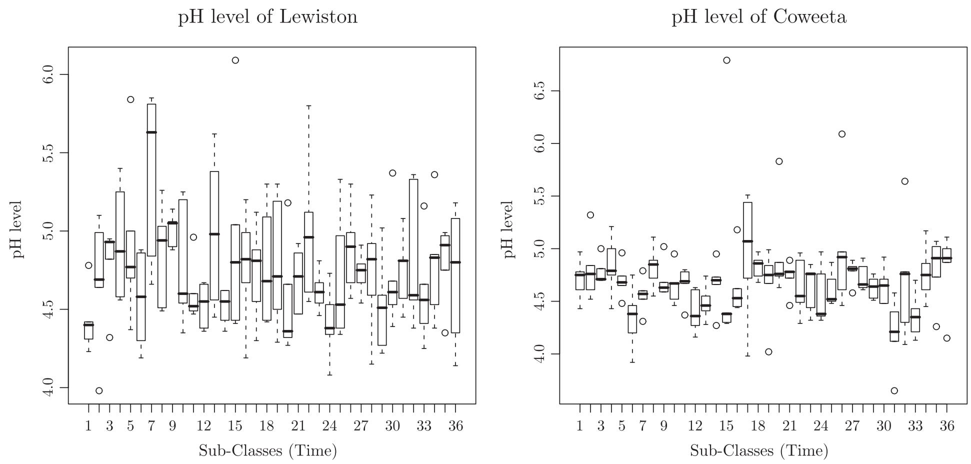

Figure 2. NADP Rain Data. The boxplots of the pH levels of precipitation at different Times from January 2003 to December 2007 in Lewiston (left) and Coweeta (right).

distinguish between what we call general heteroscedasticity and between-classes heteroscedasticity, and develop corresponding test procedures, *HET* and *UW*, for each case. Simulations indicate that the *HET* procedure is very competitive against the *CF* and the *UW* procedures in cases where the last two are valid. Thus, we recommend the procedure *HET* for general applicability provided $n_{ij} \geq 4$ for all $i,\ j$. The procedure *UW* is preferable to *HET* when the between-classes heteroscedasticity assumption appears tenable. The procedure *CF* is preferable to *HET* when the assumptions of normality and homoscedasticity appear tenable.

We have also demonstrated, via data analysis and simulations, that a simple ramification of the present methodology for the nested model can work reasonably well for analysis of covariance. Preliminary work suggests that using overlapping windows results in improved power but, due to the dependence caused, this approach is the subject of a different paper.

References

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

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

Arnold, S. F. (1980). Asymptotic validity of F Tests for the ordinary linear model and the multiple correlation model. *J. Amer. Statist. Assoc.* **75**, 890-894.

Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis.* Wiley, New York.

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

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

Gupta, A., Harrar, S. and Fujikoshi, Y. (2006). Asymptotics for testing hypothesis in some multivariate variance components model under non-normality. *J. Multivariate Anal.* **97**, 148-178.

Hall, P., Marron, J. S. and Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *J. Roy. Statist. Soc. Ser. B* **67**, 427-444.

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

McKean, J. W. and Sheather, S. J. (1991). Small sample properties of robust analyses of linear models based on R-estimates: A survey. *Inst. Math. Appl.* **34**, 1-19.

National Atmospheric Deposition Program (NRSP-3) (2009). Illinois State Water Survey. NADP Program Office, 2204 Griffith Dr., Champaign, IL 61820.

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

Nationl Oceanic and Atmospheric Administration (NOAA) (1988). Chemical Contaminants in Oysters and Mussels, by Tom O'Connor. NOAA's State of the Coast Report. Silver Spring, MD: NOAA.

Simons, G. and Yao, Y. (1999). Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons. *Ann. Statist.* **27**, 1041-1060.

Wang, H. and Akritas, M. G. (2004). Rank tests for ANOVA with large number of factor levels. *J. Nonparametr. Stat.* **16**, 563-590.

Wang, H. and Akritas, M. G. (2009). Rank tests in heteroscedastic multi-way HANOVA. *J. Nonparametr. Stat.* **21**, 663-681.

Wang, H. and Akritas, M. G. (2010). Rank test for heteroscedastic functional data. *J. Nonparametr. Stat.* **101**, 1791-1805.

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

Department of Mathematics, Amherst College Box 2239, P.O. 5000, Amherst, MA 01002-5000, USA.

E-mail: sliao@amherst.edu

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

E-mail: mga@stat.psu.edu

(Received June 2011; accepted November 2012)