Heteroscedastic One-Way ANOVA and Lack-of-Fit Tests

M. G. AKRITAS and N. PAPADATOS

Recent articles have considered the asymptotic behavior of the one-way analysis of variance (ANOVA) F statistic when the number of levels or groups is large. In these articles, the results were obtained under the assumption of homoscedasticity and for the case when the sample or group sizes n_i remain fixed as the number of groups, a, tends to infinity. In this article, we study both weighted and unweighted test statistics in the heteroscedastic case. The unweighted statistic is new and can be used even with small group sizes. We demonstrate that an asymptotic approximation to the distribution of the weighted statistic is possible only if the group sizes tend to infinity suitably fast in relation to a. Our investigation of local alternatives reveals a similarity between lack-of-fit tests for constant regression in the present case of replicated observations and the case of no replications, which uses smoothing techniques. The asymptotic theory uses a novel application of the projection principle to obtain the asymptotic distribution of quadratic forms.

KEY WORDS: Large number of factor levels; Local alternatives; Projection method; Quadratic forms; Regression; Unbalanced models.

1. INTRODUCTION

It is well known that the F test is robust to the normality assumption if the number of factor levels or groups is small (fixed) and the sample or group sizes are large (tend to infinity); see Arnold (1980). In this setting the theory of weighted least squares statistics is also well understood; see Arnold (1981, chap. 13). However, the case where the number of factor levels is large (tends to infinity) is still not adequately developed.

When the number of levels, a, goes to infinity, the asymptotic distribution of the F statistic $F = \mathrm{MST}/\mathrm{MSE}$, where MST is the mean square for treatment and MST is the mean square for error, is found by obtaining the asymptotic distribution of $a^{1/2}(F - 1)$. The MSE typically converges in probability to a constant and thus, by Slutsky's theorem, the foregoing expression reduces to finding the asymptotic distribution of $a^{1/2}(\mathrm{MST} - \mathrm{MSE})$. Boos and Brownie (1995) presented some results in this direction, but used a specialized technique applicable only to few models. Because $\mathrm{MST} - \mathrm{MSE}$ is a quadratic form, the most direct way to find its asymptotic distribution is to apply results for the asymptotic normality of quadratic forms; see de Jong (1987) and Jiang (1996) and references therein. However, it is not straightforward to apply these results. Akritas and Arnold (2000) developed an approach that is based on finding the joint limiting distribution of (MST, MSE). With this technique they covered a very general class of models and also obtained the asymptotic distribution of the statistics under fixed alternatives. Independently and using different asymptotic techniques, Bathke (2002) also generalized the results of Boos and Brownie (1995) to fixed effects balanced multifactor designs.

The aforementioned results all pertain to the homoscedastic case with small (fixed) group sizes. However, the assumption that a large number of populations are homoscedastic is difficult to ascertain when the group size from each population is small. As demonstrated by Scheffé (1959, chap. 10), the F test is sensitive to departures from the homoscedasticity assumption, particularly in the unbalanced case. The homoscedastic procedure based on the asymptotic theory for large a is equally sensitive. For example, 1,000 simulation replications with $a = 30$ levels, group sizes 14, 15, 7, 4, 4, 4, 4, 6, 5, 4, 4, 4, 6, 5, 5, 4, 6, 6, 8, 4, 5, 4, 5, 6, 4, 5, 4, 6, 4, 5, and heteroscedastic normal errors with corresponding variances $(1 + .133i)^2$, $i = 1, \ldots, 30$, at $\alpha = .05$, yielded achieved α levels of .141 and .178 for the classical F test and that based on Theorem 2.2(a), respectively. [For the same setting but homoscedastic errors, the procedure of Theorem 2.2(a) achieved an α level of .070.] For the same setting, the unweighted heteroscedastic test procedure for large a (see Theorem 2.5), achieved an α level of .073. [Also see Remark 2.3(ii) for a similar simulation in the balanced case.]

Even under homoscedasticity, the usual F test is not asymptotically valid in the unbalanced case if the group sizes are small; see Section 2.1. Although an asymptotically valid procedure using the F-test statistic is provided in Section 2.1, it requires estimation of the fourth moment.

The purpose of the present article is to provide test procedures that are valid and perform well in unbalanced and/or heteroscedastic situations when a tends to infinity. We consider both the classical weighted statistic and an unweighted statistic that appears to be new. Using exact calculations under normality, we demonstrate that the classical weighted statistic is very unstable if the group sizes are small, which explains Krutchkoff's (1989) observation. Asymptotic approximation to the distribution of the weighted statistic requires the average group size to tend to infinity faster than $a^{1/2}$. The procedure that uses the new unweighted statistic is applicable also with small group sizes. Its asymptotic and small sample properties are preferable to those of the procedure based on the F-test statistic, even in the homoscedastic case. Indeed, it does not require estimation of the fourth moment and its asymptotic theory uses weaker conditions.

The technique we apply is based on an application of the projection principle. It allows us to study, directly and elegantly, the asymptotic null distribution of the quadratic form $\mathrm{MST} - \mathrm{MSE}$. The novelty of the technique rests on the choice of the class of

M. G. Akritas is Professor, Department of Statistics, Pennsylvania State University, University Park, PA 16802 (E-mail: *mga@stat.psu.edu*). N. Papadatos is Assistant Professor, Department of Mathematics, Section of Statistics and O.R., University of Athens, 15785 Athens, Greece (E-mail: *npapadat@cc.uoa.gr*). The authors thank the Editor for a most efficient handling of the manuscript, and the Associate Editor and two referees for a number of interesting suggestions which led to a significant improvement of the material. This research was supported in part by National Science Foundation grant SES-99-86592. N. Papadatos also acknowledges partial support from the University of Athens (research grant 70/4/5637).

Journal of the American Statistical Association
June 2004, Vol. 99, No. 466, Theory and Methods
DOI 10.1198/016214504000000412

random variables onto which to project. Although for simplicity we focus on the one-way model, it is rather obvious that the projection principle and the idea of choosing the class of variables onto which to project, applies to general multifactor models; see Wang (2003) and Wang (2004). The basic technique is demonstrated for the homoscedastic case, where the transparency of the method permits derivation of the asymptotic theory under weaker assumptions on the moments and group sizes in the unbalanced case than those of Akritas and Arnold (2000). In doing so, we also consider the case where the group sizes are allowed to tend to infinity together with the number of levels. This setting was also considered by Portnoy (1984), but from the M-estimation point of view.

The one-way layout F statistic coincides with the lack-of-fit statistic for testing the hypothesis of constant regression against a general alternative with replicated observations, and thus the present results have direct bearing on this problem as well. The literature of lack-of-fit testing in regression is quite extensive: see Eubank and Hart (1992), Müller (1992), Härdle and Mammen (1993), Hart (1997), and Dette and Munk (1998), to mention a few. It is quite interesting that the asymptotic validity of the common lack-of-fit test in the case of replicated observations has never been considered. Our study of local alternatives reveals that the classical lack-of-fit test with replicated observations cannot detect alternatives that converge to the null hypothesis at rate $a^{-1/2}$, but rather at rates that resemble those in the nonparametric literature. Given the calculation in the case of normal variables with known variance presented in Fan (1996), this is not surprising.

Section 2 gives the test statistics and their limiting null distribution with some comments on their performance. In Section 3 we present the projection method in the context of quadratic forms. Section 4 gives asymptotic results under local alternatives. Some simulation results are discussed in Section 5. Proofs of the results presented in Sections 2 and 4 are given in the Appendix.

In all that follows, X_{ij}, $i = 1, \dots, a$, $j = 1, \dots, n_i$, denotes a double sequence of independent random variables, $S_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_{i\cdot})^2$ and

$$\mathrm{MST} = \frac{1}{a-1} \sum_{i=1}^{a} n_i (\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2,$$

$$\mathrm{MSE} = \frac{1}{N-a} \sum_{i=1}^{a} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_{i\cdot})^2, \qquad (1.1)$$

$$F_a = \frac{\mathrm{MST}}{\mathrm{MSE}},$$

where $\overline{X}_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$ and $\overline{X}_{\cdot\cdot} = N^{-1} \sum_{i=1}^{a} \sum_{j=1}^{n_i} X_{ij}$ with $N = n_1 + \cdots + n_a$. In case the group sizes $n_i = n_i(a) \to \infty$ as $a \to \infty$, we also write $S_i^2(a)$ and $N(a)$.

2. MAIN RESULTS

In this section we present and discuss the asymptotic null distribution of the proposed test statistics. Corresponding results under local alternatives are stated in Section 4.

2.1 Homoscedastic Models

Let U_a have an $F_{a-1,N-a}$ distribution, which is the distribution of F_a under homoscedasticity and normality. Let also $N/a \to b$, $a^{-1} \sum_{i=1}^{a} n_i^{-1} \to b_1$; thus, in the balanced case, $b = n$ and $bb_1 = 1$. It is easily verified that if $b < \infty$, $a^{1/2}(U_a - 1) \xrightarrow{d} \mathrm{N}(0, 2b/(b-1))$ as $a \to \infty$, and if N/a also tends to infinity with a, then $a^{1/2}(U_a - 1) \xrightarrow{d} \mathrm{N}(0, 2)$. Thus, Theorem 2.1 asserts that the usual F test is asymptotically, as $a \to \infty$, correct in the balanced homoscedastic case even without the normality assumption. Theorem 2.2, however, shows that the usual F procedure for unbalanced models is robust to departures from the normality assumption only if $bb_1 = 1$ or the group sizes are also large.

Theorem 2.1 (Balanced case). Let X_{ij}, $i = 1, \dots, a$, $j = 1, \dots, n$, be an iid (independent, identically distributed) sequence of random variables with $\mathbb{E}X_{ij} = \mu$ and $0 < \operatorname{Var} X_{ij} = \sigma^2 < \infty$.

(a) If $n \ge 2$ remains fixed, then

$$a^{1/2}(F_a - 1) \xrightarrow{d} \mathrm{N}\left(0, \frac{2n}{n-1}\right) \qquad \text{as } a \to \infty.$$

(b) If $n = n(a) \to \infty$, as $a \to \infty$, then

$$a^{1/2}(F_a - 1) \xrightarrow{d} \mathrm{N}(0, 2) \qquad \text{as } a \to \infty.$$

Note that the result of part (b) of Theorem 2.1 is easily guessed from part (a). Its proof, however, involves a rather interesting application of the Lindeberg condition.

Theorem 2.2 (Unbalanced case). Let X_{ij}, $i = 1, \dots, a$, $j = 1, \dots, n_i$, be an iid sequence of random variables with $\mathbb{E}X_{ij} = \mu$, $0 < \operatorname{Var} X_{ij} = \sigma^2 < \infty$.

(a) If, for some $\delta > 0$, $\mathbb{E}|X_{ij}|^{4+\delta} < \infty$, $\sup_{a \ge 1} a^{-1} \times \sum_{i=1}^{a} n_i^{4+\delta} < \infty$,

$$\bar{n} = \bar{n}(a) = \frac{1}{a} \sum_{i=1}^{a} n_i \to b \in (1, \infty),$$

and

$$\frac{1}{a} \sum_{i=1}^{a} \frac{1}{n_i} \to b_1 \qquad \text{as } a \to \infty,$$

then

$$a^{1/2}(F_a - 1) \xrightarrow{d} \mathrm{N}(0, \tau^2) \qquad \text{as } a \to \infty,$$

where, letting $\mu_4 = \mathbb{E}[(X_{ij} - \mu)^4/\sigma^4]$,

$$\tau^2 = \frac{2b}{b-1} + (\mu_4 - 3)\frac{b(bb_1 - 1)}{(b-1)^2}.$$

(b) Let $n_i = n_i(a)$, and set $n(a) = \min\{n_i(a); i = 1, \dots, a\}$ and $\kappa(a) = \max\{n_i(a); i = 1, \dots, a\}$. Assume that

$$n(a) \to \infty \qquad \text{as } a \to \infty,$$

and

$$\kappa(a)/n(a) \le C < \infty \qquad \text{for all } a.$$

If $\mathbb{E}X_{ij}^4 < \infty$, then

$$a^{1/2}(F_a - 1) \xrightarrow{d} \mathrm{N}(0, 2) \qquad \text{as } a \to \infty.$$

The reason we need higher moments in the unbalanced case is because the square terms (i.e., X_{ij}^2) do not cancel [see (A.6)]. The preceding results with fixed n_i's overlap with those of Boos and Brownie (1995) and Akritas and Arnold (2000), although the present assumptions are slightly weaker. The new results under homoscedasticity pertain to the case given in Section 4, where the group sizes also tend to infinity and the limiting distribution is under local alternatives.

2.2 Heteroscedastic Models

2.2.1 The Possible Statistics. Under heteroscedasticity it is possible to have both weighted and unweighted statistics. In this section we first introduce the two statistics and then present their asymptotic theory.

The unweighted statistic, whose version for the unbalanced case is new, is based on the observation that in the balanced case, $\mathbb{E}(\text{MST}) = \mathbb{E}(\text{MSE})$ under the null hypothesis, so that the statistic MST − MSE is still centered. In the unbalanced case this is not true, but centering can be achieved by replacing MSE with $\text{MSE}^* = (a-1)^{-1}\sum_{i=1}^a (1 - n_i/N)S_i^2$. This leads to the statistic

$$T_a = a^{-1/2}\sum_{i=1}^a \left[n_i(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2 - \left(1 - \frac{n_i}{N}\right)S_i^2 \right], \quad (2.1)$$

which, for the balanced case only, is closely connected to F_a via the relationship

$$T_a = \left(1 - \frac{1}{a}\right)\left(a^{1/2}(F_a - 1)\right)\left(\frac{1}{a}\sum_{i=1}^a S_i^2\right)$$
$$= \left(1 - \frac{1}{a}\right)\left(a^{1/2}(\text{MST} - \text{MSE})\right). \quad (2.2)$$

The Wald-type weighted statistic is

$$T_{\mathrm{W}} = \overline{\mathbf{X}}_{\cdot}'\mathbf{C}'(\mathbf{C}\mathbf{V}\mathbf{C}')^{-1}\mathbf{C}\overline{\mathbf{X}}_{\cdot}, \quad (2.3)$$

where $\overline{\mathbf{X}}_{\cdot} = (\overline{X}_{1\cdot}, \dots, \overline{X}_{a\cdot})'$ is the vector of sample means with dispersion matrix $\mathbf{V} = \operatorname{Var}\overline{\mathbf{X}}_{\cdot} = \operatorname{diag}(\sigma_1^2/n_1, \dots, \sigma_a^2/n_a)$, and $\mathbf{C} = (\mathbf{1}_{a-1}|-\mathbf{I}_{a-1})$, where $\mathbf{1}_k = (1, \dots, 1)' \in \mathbb{R}^k$ and $\mathbf{I}_k$ denotes the k-dimensional identity matrix. Note that this is similar to the weighted scheme that would be used if $n_i \to \infty$ and a is fixed. Also note that, writing $\mathbf{J} = \mathbf{1}_a\mathbf{1}_a'$,

$$T_{\mathrm{W}} = \overline{\mathbf{X}}_{\cdot}'\left(\mathbf{V}^{-1} - \frac{1}{\operatorname{tr}(\mathbf{V}^{-1})}\mathbf{V}^{-1}\mathbf{J}\mathbf{V}^{-1}\right)\overline{\mathbf{X}}_{\cdot}$$
$$= \sum_{i=1}^a \frac{n_i}{\sigma_i^2}\overline{X}_{i\cdot}^2 - \frac{1}{\sum_{i=1}^a n_i/\sigma_i^2}\left(\sum_{i=1}^a \frac{n_i}{\sigma_i^2}\overline{X}_{i\cdot}\right)^2, \quad (2.4)$$

so that it is easily calculated without inverting a matrix. Note that in the homoscedastic case and with the usual estimation of the common variance, T_{W} becomes the usual one-way analysis of variance (ANOVA) statistic.

Remark 2.1. An alternative derivation of T_{W} is first to standardize $\overline{\mathbf{X}}_{\cdot}$ and then center it before forming the quadratic form. In particular, $\mathbf{V}^{-1/2}\overline{\mathbf{X}}_{\cdot}$ has mean value $\mathbf{V}^{-1/2}\boldsymbol{\mu}$ and covariance matrix $\mathbf{I}_a$. Thus, the appropriate quadratic form is $(\overline{\mathbf{X}}_{\cdot} - \widehat{\boldsymbol{\mu}})'\mathbf{V}^{-1}(\overline{\mathbf{X}}_{\cdot} - \widehat{\boldsymbol{\mu}})$. The minimum variance unbiased estimator of $\boldsymbol{\mu}$ under the hypothesis of a common mean is $\widehat{\boldsymbol{\mu}} = \widehat{\mu}\mathbf{1}_a$, where $\widehat{\mu} = \sum_{i=1}^a w_i\overline{X}_{i\cdot}$, where the w_i are proportional to n_i/σ_i^2 and sum to 1. With this choice of $\widehat{\boldsymbol{\mu}}$ it is easy to verify that the preceding quadratic form equals T_{W}. However, T_{W} involves unknown variances that must be estimated. The resulting generalized or weighted least squares statistic is

$$\widehat{T}_{\mathrm{W}} = \sum_{i=1}^a \frac{n_i}{S_i^2}\overline{X}_{i\cdot}^2 - \frac{1}{\sum_{i=1}^a n_i/S_i^2}\left(\sum_{i=1}^a \frac{n_i}{S_i^2}\overline{X}_{i\cdot}\right)^2. \quad (2.5)$$

It will be shown that $\widehat{T}_{\mathrm{W}}$ is asymptotically equivalent to T_{W} only if the group sizes tend to infinity as $a \to \infty$ suitably fast. The difficulties with fixed group sizes are highlighted by the following proposition, which indicates that $\widehat{T}_{\mathrm{W}}$, applied to the simplest case (i.e., to the balanced homoscedastic normal case), is quite unstable. The assertion is stated here without proof.

Proposition 2.3. Let X_{ij} be iid $\mathrm{N}(\mu, \sigma^2)$, $i = 1, \dots, a$, $j = 1, \dots, n$, $\sigma^2 > 0$, n fixed. Then, provided $n \ge 6$,

$$a^{-1/2}\left(\widehat{T}_{\mathrm{W}} - \frac{n-1}{n-3}a\right) \xrightarrow{d} \mathrm{N}\left(0, 2\frac{(n-1)^2(n-2)}{(n-3)^2(n-5)}\right) \qquad \text{as } a \to \infty.$$

Note that the requirement $n \ge 6$ arises from the fact that $\operatorname{Var}[(\overline{X}_{i\cdot} - \mu)^2/S_i^2] = 2(n-1)^2(n-2)n^{-2}(n-3)^{-2}(n-5)^{-1}$. Thus, for small group sizes, one can expect unstable behavior of the Wald-type weighted statistic.

Remark 2.2. Although weighted statistics are known to have better power properties, the results of Proposition 2.3 and Section 2.2.3 suggest that their use requires the group sizes to be large also. For example, in the last simulation setting in Section 5 ($a = 30, n = 10$), the achieved α level of the weighted test statistic is .295, but if n is increased to 70 and 80, the achieved α levels become .096 and .076, respectively. Thus, we do not recommend using the weighted test statistic if the group sizes are less than 80. On the other hand, the unweighted test statistic proposed here can be used also with small n.

2.2.2 The Unweighted Statistic. The asymptotic distribution of T_a is given separately for the balanced and unbalanced cases, and small or large group sizes.

Theorem 2.4 (Balanced case, small n). For $n \ge 2$ fixed, let X_{ij}, $i = 1, \dots, a$, $j = 1, \dots, n$, be a double sequence of independent random variables with $\mathbb{E}X_{ij} = \mu$ and $0 < \operatorname{Var}X_{ij} = \sigma_i^2 < \infty$, and suppose that in each row i, the random variables X_{ij}, $j = 1, \dots, n$, have the same distribution. Moreover, assume that

$$\frac{1}{a}\sum_{i=1}^a \sigma_i^4 \to s^4 \in (0, \infty) \qquad \text{as } a \to \infty \quad (2.6)$$

and that for some $\delta > 0$, $\sup_{a \ge 1} a^{-1}\sum_{i=1}^a (\mathbb{E}|X_{i1}|^{2+\delta})^2 < \infty$. Then, with T_a given by (2.1),

$$T_a \xrightarrow{d} \mathrm{N}\left(0, \frac{2ns^4}{n-1}\right) \qquad \text{as } a \to \infty.$$

Remark 2.3. (i) Assuming that the MSE is consistent for $\sigma^2 = \lim a^{-1}\sum_{i=1}^a \sigma_i^2$, (2.2) implies that the assertion of Theorem 2.4 is equivalent to

$$a^{1/2}(F_a - 1) \xrightarrow{d} \mathrm{N}\left(0, \frac{2ns^4}{(n-1)\sigma^4}\right) \qquad \text{as } a \to \infty. \tag{2.7}$$

Similar observations are valid for the results of Theorems 2.6, 4.3, and 4.5(a).

(ii) Because $s^4 \geq \sigma^4$ in (2.7), it follows that the usual F test, or that based on Theorem 2.1, will be liberal under heteroscedasticity. To illustrate this effect, we conducted small simulation with 1,000 runs and $X_{ij} \sim \mathrm{N}(0, \sigma_i^2)$, $i = 1, \dots, 30$, $j = 1, \dots 10$, where $\sigma_i = 1$ if $i \leq 15$ and $\sigma_i = 5$ if $i \geq 16$, at nominal $\alpha = .05$. The achieved α levels of the classical F test and the procedure based on Theorem 2.1 are .109 and .129, respectively, whereas that based on Theorem 2.4 achieved .074.

(iii) If $\sigma_i^2 = \sigma^2$ for all i, then $s^4 = \sigma^4$ and (2.7) coincides with the result of Theorem 2.1(a). However, the results use different conditions. Theorem 2.4 does not require X_{ij} to be iid, but uses extra moment assumptions to control the variation of the summands.

Theorem 2.5 (Unbalanced case, small n_i). Let X_{ij}, $i = 1, \dots, a$, $j = 1, \dots, n_i$, be a double sequence of independent random variables with $\mathbb{E}X_{ij} = \mu$ and $0 < \operatorname{Var} X_{ij} = \sigma_i^2 < \infty$, where $n_i \geq 2$ for each i. Suppose that in each row i, the random variables X_{ij}, $j = 1, \dots, n_i$, have the same distribution. Moreover, assume that

$$\frac{1}{a}\sum_{i=1}^a \sigma_i^4 \to s^4 \in (0, \infty),$$

$$\frac{1}{a}\sum_{i=1}^a \frac{\sigma_i^4}{n_i - 1} \to \gamma^4 \in (0, \infty) \qquad \text{as } a \to \infty,$$

and that for some $\delta > 0$, $\sup_{a \geq 1} \frac{1}{a}\sum_{i=1}^a n_i^{2+\delta} < \infty$ and $\sup_{i \geq 1} \mathbb{E}|X_{i1}|^{2+\delta} < \infty$. Then, with T_a given by (2.1),

$$T_a \xrightarrow{d} \mathrm{N}\big(0, 2(s^4 + \gamma^4)\big) \qquad \text{as } a \to \infty.$$

Remark 2.4. Compared to its homoscedastic counterpart, Theorem 2.5 does not require the X_{ij} to be iid, uses weaker moment conditions, and the limiting result does not involve μ_4. This highlights the differences between T_a and F_a in the unbalanced case.

Next we present versions of Theorems 2.4 and 2.5 when the group sizes tend to infinity with the number of factor levels. Here, the balanced and unbalanced cases are combined in one theorem, where the asymptotic result is shown under two different sets of conditions, one being weaker in moment assumptions and more restrictive in conditions on group sizes than the other.

Theorem 2.6 (Large n_i). Let X_{ij}, $i = 1, \dots, a$, $j = 1, \dots, n_i(a)$, be independent with $\mathbb{E}X_{ij} = \mu$ and $0 < \operatorname{Var} X_{ij} = \sigma_i^2 < \infty$, and suppose that for each i, X_{ij}, $j = 1, \dots, n_i(a)$, have the same distribution. Set $n(a) = \min\{n_i(a); i = 1, \dots, a\}$, $\kappa(a) = \max\{n_i(a); i = 1, \dots, a\}$, and assume that $2 \leq n(a) \to \infty$, $\kappa(a)/n(a) \leq C < \infty$ for all a, and

$$\frac{1}{a}\sum_{i=1}^a \sigma_i^4 \to s^4 \in (0, \infty) \qquad \text{as } a \to \infty.$$

If for some $\delta > 0$, either

$$\sup_{a \geq 1} \frac{1}{a}\sum_{i=1}^a (\mathbb{E}|X_{i1}|^{2+\delta})^2 < \infty \qquad \text{and} \qquad n(a) = o\big(a^{\delta/(4+2\delta)}\big) \tag{2.8}$$

or

$$\sup_{a \geq 1} \frac{1}{a}\sum_{i=1}^a \mathbb{E}|X_{i1}|^{4+\delta} < \infty, \tag{2.9}$$

then, with T_a given by (2.1),

$$T_a \xrightarrow{d} \mathrm{N}(0, 2s^4) \qquad \text{as } a \to \infty.$$

2.2.3 The Weighted Statistic. Given the difficulties with bounded group sizes when the unknown variances are estimated (see Proposition 2.3), this section considers the case where the group sizes tend to ∞.

Let T_{W} be given in (2.3) or (2.4). A straightforward calculation reveals that

$$\mathbb{E}T_{\mathrm{W}} = a - 1 + \sum_{i=1}^a \frac{n_i}{\sigma_i^2}\mu_i^2 - \frac{1}{\sum_{i=1}^a n_i/\sigma_i^2}\left(\sum_{i=1}^a \frac{n_i}{\sigma_i^2}\mu_i\right)^2$$
$$\geq a - 1$$

with equality if and only if the null hypothesis $H_0: \mu_1 = \cdots = \mu_a$ is true. Thus we are led to consideration of the asymptotic distribution, under the null hypothesis, of

$$T_{\mathrm{W}} - (a-1) = \sum_{i=1}^a \left(\frac{n_i}{\sigma_i^2}\overline{X}_{i\cdot}^2 - 1\right) - \frac{1}{t(a)}\sum_{i=1}^a \frac{n_i}{\sigma_i^2}\left(\frac{n_i}{\sigma_i^2}\overline{X}_{i\cdot}^2 - 1\right) - \frac{1}{t(a)}\sum_{i \neq j} \frac{n_i}{\sigma_i^2}\frac{n_j}{\sigma_j^2}\overline{X}_{i\cdot}\overline{X}_{j\cdot},$$

where $t(a) = \sum_{i=1}^a n_i/\sigma_i^2$.

Proposition 2.7. Assume that X_{ij} are independent random variables such that $\mathbb{E}X_{ij} = \mu$ and $\operatorname{Var} X_{ij} = \sigma_i^2 > 0$, $i = 1, \dots, a$, $j = 1, \dots, n_i(a)$, and suppose that in each row i, the random variables X_{ij}, $j = 1, \dots, n_i(a)$, have the same distribution. Moreover, assume that for some $\delta > 0$, $\sup_{i \geq 1} \mathbb{E}|(X_{i1} - \mu)/\sigma_i|^{4+\delta} < \infty$. If $n(a) = \min\{n_i(a);\ i = 1, \dots, a\} \to \infty$ as $a \to \infty$, then, for T_{W} given in (2.3) or (2.4),

$$a^{-1/2}\big(T_{\mathrm{W}} - (a-1)\big) \xrightarrow{d} \mathrm{N}(0, 2) \qquad \text{as } a \to \infty.$$

Theorem 2.8. Consider the setting of Proposition 2.7 and, moreover, assume that the variances σ_i^2, $i \geq 1$, are bounded away from zero and that the random variables X_{ij} are uniformly bounded by some constant M. Let $\widehat{T}_{\mathrm{W}}$ be the statistic given in (2.5), let $n_i = n_i(a)$ and set $n(a) = \min\{n_i(a);\ i = 1, \dots, a\}$, $\bar{n} = \bar{n}(a) = a^{-1}\sum_{i=1}^a n_i$. If condition (C.1) of Lemma C.2 holds and, in addition,

$$\begin{aligned} &\text{(a)} \quad a^{-1/2}\sum_{i=1}^a \frac{(\log n_i)^4}{n_i} = o(1), \\ &\text{(b)} \quad \frac{a^{1/2}}{\bar{n}} = o(1), \qquad \text{and} \\ &\text{(c)} \quad n(a) \to \infty, \end{aligned} \tag{2.10}$$

then

$$a^{-1/2}\big(\widehat{T}_W - (a-1)\big) \xrightarrow{d} \mathrm{N}(0,2) \qquad \text{as } a \to \infty.$$

We remark that the condition of uniform boundedness of the random variables, which will be used to apply Bernstein's inequality in the proof of Lemmas C.1 and C.2, can be replaced by a condition that imposes a bound on the rate of growth of the absolute moments (cf. Shorack and Wellner, 1986, p. 855).

3. THE PROJECTION METHOD FOR QUADRATIC FORMS

A common method for obtaining the limit distribution of a sequence of statistics S_a is to show that it is asymptotically equivalent to a sequence $\widetilde{S}_a$ of which the limit behavior is easy to derive. For certain classes of statistics (e.g., U statistics, R statistics) the suitable sequence $\widetilde{S}_a$ is found by the method of projection. For a nice presentation of the projection method, see van der Vaart (1998, chap. 11). If S_a is based on $U_1, \dots, U_a$ and has finite second moment, its projection onto the class of random variables of the form $\sum_{i=1}^a g_i(U_i)$, where g_i are measurable with $\sum_{i=1}^a \mathbb{E}[g_i^2(U_i)] < \infty$, is given by

$$\widetilde{S}_a = \sum_{i=1}^a \mathbb{E}(S_a|U_i) - (a-1)\mathbb{E}S_a.$$

This is known as the Hájek projection of S_a. In this section we show that a slight modification of the Hájek projection method yields a sequence of statistics that is asymptotically equivalent to the sequence of quadratic forms MST − MSE. For simplicity only, the formulation of this section is limited to homoscedastic models, but includes both cases where the possibly unequal group sizes are either fixed or tend to infinity. In fact (see Appendix A), the proofs given in the present article are all based on the same projection method.

Let $\mathbf{X}_i = (X_{i1}, \dots, X_{in_i})'$, $i = 1, \dots, a$, be independent random vectors with independent components, and without loss of generality assume that $\mathbb{E}X_{ij} = 0$ and $\operatorname{Var} X_{ij} = 1$ for all i, j. We are interested in the limiting distribution of $F_a - 1$, or equivalently, whereas $\mathrm{MSE} \xrightarrow{p} 1$, in the limiting distribution of MST − MSE, as $a \to \infty$. To do so, we use the projection method, projecting onto the class of random variables

$$\mathcal{C}_{\mathrm{H}} = \left\{ \sum_{i=1}^a g_i(\mathbf{X}_i) : g_i \text{ measurable with} \sum_{i=1}^a \mathbb{E}[g_i^2(\mathbf{X}_i)] < \infty \right\}. \tag{3.1}$$

It is easily seen that the projection onto the class $\mathcal{C}_{\mathrm{H}}$ of any statistic S_a (satisfying $\mathbb{E}S_a^2 < \infty$), based on $\mathbf{X}_i$, $i = 1, \dots, a$, is given by

$$\widetilde{S}_a = \sum_{i=1}^a \mathbb{E}(S_a|\mathbf{X}_i) - (a-1)\mathbb{E}S_a. \tag{3.2}$$

This is a slight variation of Hájek's projection in the sense that, to define the class of random variables onto which we project, the original set of independent variables is grouped into independent vectors.

Remark 3.1. We stipulate that such variations of Hájek's projection produce asymptotically equivalent statistics for most quadratic forms arising in statistics. Indeed, the interesting part in most quadratic forms is due to blocks that lie along the diagonal. In the present case, if we let $\mathbf{X} = (X_{11}, \dots, X_{1n_1}, \dots, X_{a1}, \dots, X_{an_a})'$, then MST − MSE can be written as a quadratic form in $\mathbf{X}$, $\mathrm{MST} - \mathrm{MSE} = \mathbf{X}'\mathbf{A}\mathbf{X}$, with $\mathbf{A}$ given by

$$\mathbf{A} = \begin{pmatrix} \mathbf{B}_1 & -c_3\mathbf{1}_{n_1}\mathbf{1}'_{n_2} & \cdots & -c_3\mathbf{1}_{n_1}\mathbf{1}'_{n_a} \\ -c_3\mathbf{1}_{n_2}\mathbf{1}'_{n_1} & \mathbf{B}_2 & \cdots & -c_3\mathbf{1}_{n_2}\mathbf{1}'_{n_a} \\ \vdots & \vdots & \vdots & \vdots \\ -c_3\mathbf{1}_{n_a}\mathbf{1}'_{n_1} & -c_3\mathbf{1}_{n_a}\mathbf{1}'_{n_2} & \cdots & \mathbf{B}_a \end{pmatrix},$$

where $\mathbf{1}_k = (1, \dots, 1)' \in \mathbb{R}^k$ and $\mathbf{B}_i$ are $n_i \times n_i$ matrices with elements $b_{i,sk}$ given by

$$b_{i,sk} = \begin{cases} \dfrac{c_1}{n_i} - c_2 - c_3 & \text{if } s = k \\ \dfrac{c_1}{n_i} - c_3 & \text{if } s \neq k, \end{cases} \tag{3.3}$$

where

$$c_1 = \frac{N-1}{(N-a)(a-1)}, \quad c_2 = \frac{1}{N-a}, \quad c_3 = \frac{1}{N(a-1)}. \tag{3.4}$$

Thus, all elements of $\mathbf{A}$ outside the block diagonal equal c_3, which is an order of magnitude smaller than the other elements.

Using the quadratic form representation of MST − MSE, the notation introduced in Remark 3.1, and under $\mathbb{E}X_{ij} = 0$ and $\operatorname{Var} X_{ij} = 1$, we have

$$\mathbf{X}'\mathbf{A}\mathbf{X} = \sum_{i=1}^a \mathbf{X}_i'\mathbf{B}_i\mathbf{X}_i - c_3 \sum_{i_1 \neq i_2} \sum_{j_1}^{n_{i_1}} \sum_{j_2}^{n_{i_2}} X_{i_1 j_1} X_{i_2 j_2} \tag{3.5}$$

and thus

$$\begin{aligned} \mathbb{E}(\mathbf{X}'\mathbf{A}\mathbf{X}|\mathbf{X}_r) &= \mathbf{X}_r'\mathbf{B}_r\mathbf{X}_r + \sum_{i \neq r, i=1}^a \sum_{s=1}^{n_i} b_{i,ss} \\ &= \mathbf{X}_r'\mathbf{B}_r\mathbf{X}_r - \sum_{s=1}^{n_r} b_{r,ss}, \end{aligned}$$

where the second equality follows from the fact that the diagonal elements of $\mathbf{A}$ sum to zero. According to (3.2), the projection of $\mathbf{X}'\mathbf{A}\mathbf{X}$ onto the class $\mathcal{C}_{\mathrm{H}}$ given in (3.1) is

$$\begin{aligned} \sum_{r=1}^a \mathbb{E}(\mathbf{X}'\mathbf{A}\mathbf{X}|\mathbf{X}_r) &= \sum_{r=1}^a \left[\mathbf{X}_r'\mathbf{B}_r\mathbf{X}_r - \sum_{s=1}^{n_r} b_{r,ss} \right] \\ &= \sum_{r=1}^a \mathbf{X}_r'\mathbf{B}_r\mathbf{X}_r \\ &= \mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X}, \end{aligned} \tag{3.6}$$

where $\mathbf{A}_{\mathrm{D}} = \operatorname{diag}\{\mathbf{B}_1, \dots, \mathbf{B}_a\}$. Using the facts that the X_{ij} are independent with zero mean, and the diagonal elements of $\mathbf{A}$

(and hence also of $\mathbf{A}_D$) sum to zero, it is easy to see that $\mathbb{E}(\mathbf{X}'\mathbf{A}\mathbf{X}) = \mathbb{E}(\mathbf{X}'\mathbf{A}_D\mathbf{X}) = 0$. Using the notation in (3.4) and a straightforward computation,

$$\mathbf{X}'\mathbf{A}_D\mathbf{X} = \sum_{r=1}^{a} \mathbf{X}_r'\mathbf{B}_r\mathbf{X}_r = \sum_{i=1}^{a}\left[\left(\frac{c_1}{n_i} - c_3\right)\left(\sum_{j=1}^{n_i} X_{ij}\right)^2 - c_2\sum_{j=1}^{n_i} X_{ij}^2\right]. \quad (3.7)$$

Because this is a sum of a independent random variables with zero mean, it is reasonable to expect that

$$\left(\mathrm{Var}(\mathbf{X}'\mathbf{A}_D\mathbf{X})\right)^{-1/2}\mathbf{X}'\mathbf{A}_D\mathbf{X} \xrightarrow{d} \mathrm{N}(0, 1). \quad (3.8)$$

Direct computations given in Appendix A show that, with no restriction on the n_i,

$$\mathrm{Var}(\mathbf{X}'\mathbf{A}_D\mathbf{X}) = O(a^{-1}). \quad (3.9)$$

This suggests that the proper scaling of $\mathbf{X}'\mathbf{A}_D\mathbf{X}$ is $a^{-1/2}$, even if the group sizes tend to infinity. Relationships (3.5), (3.6), (3.8), and (3.9) imply that for the projection method to be successful in finding the asymptotic distribution of $\mathbf{X}'\mathbf{A}\mathbf{X}$, we must have $\mathbf{X}'\mathbf{A}\mathbf{X} - \mathbf{X}'\mathbf{A}_D\mathbf{X} = o_p(a^{-1/2})$. This fact, in the present particular case of homoscedastic models, is given in the following proposition.

Proposition 3.1. Under $\mathbb{E}X_{ij} = 0$ and $\mathrm{Var}\, X_{ij} = 1$ (note there is no assumption regarding a common distribution, no requirement of existence of higher moments, and no restriction on the n_i; the only requirement is the independence of X_{ij}), we have

$$a^{1/2}(\mathbf{X}'\mathbf{A}\mathbf{X} - \mathbf{X}'\mathbf{A}_D\mathbf{X}) \xrightarrow{p} 0.$$

Remark 3.2. Since MST/MSE is invariant to a common location–scale transformation of X_{ij}, the statement of Proposition 3.1 includes the general homoscedastic case and pertains to the null hypothesis of equality of means.

Proof of Proposition 3.1. From (3.5) and (3.6), it follows that

$$a\mathbb{E}(\mathbf{X}'\mathbf{A}\mathbf{X} - \mathbf{X}'\mathbf{A}_D\mathbf{X})^2 = ac_3^2 \sum_{i_1\neq i_2}\sum_{i_3\neq i_4}\sum_{j_1}^{n_{i_1}}\sum_{j_2}^{n_{i_2}}\sum_{j_3}^{n_{i_3}}\sum_{j_4}^{n_{i_4}} \mathbb{E}\left(X_{i_1j_1}X_{i_2j_2}X_{i_3j_3}X_{i_4j_4}\right).$$

Each of these expectations is zero unless either $i_1 = i_3$ and $i_2 = i_4$ and $j_1 = j_3$ and $j_2 = j_4$, or $i_1 = i_4$ and $i_2 = i_3$ and $j_1 = j_4$ and $j_2 = j_3$, in which case the expectation equals 1. Thus,

$$a\mathbb{E}(\mathbf{X}'\mathbf{A}\mathbf{X} - \mathbf{X}'\mathbf{A}_D\mathbf{X})^2 = \frac{2a}{(a-1)^2}\left(1 - \frac{1}{N^2}\sum_{i=1}^{a} n_i^2\right) \le \frac{2a}{(a-1)^2} \to 0 \qquad \text{as } a \to \infty.$$

In Section 2.1, as well as in the next section, this result is used to find the asymptotic distribution of $a^{1/2}\mathbf{X}'\mathbf{A}\mathbf{X}$.

4. LOCAL ALTERNATIVES

To obtain the limiting distribution of the statistics under local alternatives, it is convenient to assume that the observed random variables are simple translations of the original ones described in the corresponding theorems, that is,

$$Y_{ij} = X_{ij} + \mu_i(a), \qquad i = 1, \ldots, a,\ j = 1, \ldots, n_i, \quad (4.1)$$

where the constants $\mu_i(a) \to 0$ as $a \to \infty$ in a suitable rate, so that the limiting distribution exists. It turns out that we can find convenient alternatives of the form (4.1) with

$$\mu_i(a) = a^{3/4} n_i^{-1/2} \int_{(i-1)/a}^{i/a} g(t)\, dt, \quad (4.2)$$

where g is a continuous function on [0, 1] such that

$$\int_0^1 g(t)\, dt = 0. \quad (4.3)$$

It is also useful to define the "departure parameter,"

$$\theta^2 = \int_0^1 g^2(t)\, dt, \quad (4.4)$$

which vanishes if and only if the null hypothesis is correct. The form of (4.2) reveals that the statistics cannot detect alternatives that converge to the null hypothesis faster than $a^{-1/4}$ for bounded (fixed) group sizes and faster than $a^{-1/4}(n_i(a))^{-1/2}$ for group sizes going to infinity.

4.1 The Homoscedastic Case

The two results of this section generalize Theorems 2.1 and 2.2 by presenting the asymptotic distribution of F_a under local alternatives. Here and in the proofs we use the notation $F_a(\mathbf{X})$ and $F_a(\mathbf{Y})$ for the F_a statistic in (1.1) evaluated on the X_{ij}'s and Y_{ij}'s, respectively.

Theorem 4.1 (Balanced case). Under the conditions of Theorem 2.1 imposed on the X_{ij}, and if $\mu_i(a)$ are given by (4.2) with $n_i \equiv n \ge 2$, we have the following results:

(a) If $n \ge 2$ remains fixed and θ^2 denotes the departure parameter given by (4.4), then

$$a^{1/2}\left(F_a(\mathbf{Y}) - 1\right) \xrightarrow{d} \mathrm{N}\left(\frac{\theta^2}{\sigma^2}, \frac{2n}{n-1}\right) \qquad \text{as } a \to \infty.$$

(b) If $n(a) \to \infty$, then, with θ^2 as in (a),

$$a^{1/2}\left(F_a(\mathbf{Y}) - 1\right) \xrightarrow{d} \mathrm{N}\left(\frac{\theta^2}{\sigma^2}, 2\right) \qquad \text{as } a \to \infty.$$

Theorem 4.2 (Unbalanced case). Define the function $s_a(t) = \sum_{i=1}^{a} n_i^{1/2} I(i - 1 \le at < i)$, $0 \le t < 1$, and let $\mu_i(a)$ be given by (4.2) and Y_{ij} be given by (4.1). Then we have the following results:

(a) In the case where the n_i remain fixed, if the conditions of Theorem 2.2(a) are satisfied and also

$$\lim_{a\to\infty} \int_0^1 s_a(t) g(t)\, dt = \rho_1$$

exists, then with θ^2 denoting the departure parameter given by (4.4), we have

$$a^{1/2}\big(F_a(\mathbf{Y})-1\big) \xrightarrow{d} \mathrm{N}\left(\frac{\theta^2-\rho_1^2/b}{\sigma^2}, \tau^2\right) \qquad \text{as } a\to\infty.$$

(b) If the $n_i(a)\to\infty$ as $a\to\infty$, consider the assumptions of Theorem 2.2(b) and also assume that the limits

$$\lim_{a\to\infty}\int_0^1 \frac{s_a(t)}{\kappa(a)^{1/2}} g(t)\,dt = \rho_2, \qquad \lim_{a\to\infty}\frac{\kappa(a)}{\bar n(a)} = \beta$$

exist, where $\bar n(a) = a^{-1}\sum_{i=1}^a n_i(a)$. Then, with θ^2 is as in (a),

$$a^{1/2}\big(F_a(\mathbf{Y})-1\big) \xrightarrow{d} \mathrm{N}\left(\frac{\theta^2-\beta\rho_2^2}{\sigma^2}, 2\right) \qquad \text{as } a\to\infty.$$

Remark 4.1. (1) As mentioned in Section 1 and demonstrated in the present section, the rates that local alternatives must converge to the null hypothesis to allow detection resemble those found in the literature for lack-of-fit testing in nonparametric regression. Proposition 3.1, however, shows that MST − MSE is essentially a sum of independent random variables. Thus, it should be possible to define and study a corresponding partial sum process that would allow ideas such as the cumulative sum and the adaptive Neyman test to be applied to improve the performance of the local alternatives. This research, however, is beyond the scope of the present article.

(2) Requiring the existence of $\lim_{a\to\infty}\int_0^1 s_a(t)g(t)\,dt = \rho_1$ (which is necessarily finite, as the proof of Theorem 4.2 shows) does not seem to be very restrictive. For instance, let the n_i be uniformly bounded, assuming values in $\{1,\dots,m\}$ for some m, and define $A_{k,a} = \{i\in\{1,\dots,a\}: n_i = k\}$ and $I_{k,a} = \bigcup_{i\in A_{k,a}}[(i-1)/a, i/a)$ for $k=1,\dots,m$. Then, by the dominated convergence theorem, we have

$$\begin{aligned}\int_0^1 s_a(t)g(t)\,dt &= \sum_{k=1}^m k^{1/2}\int_{I_{k,a}} g(t)\,dt \\ &\to \sum_{k=1}^m k^{1/2}\int_{I_k} g(t)\,dt \\ &= \rho_1,\end{aligned}$$

provided $\lim_{a\to\infty} I_{k,a} = I_k$ exists for all k. Similar observations hold for the existence of the limit ρ_2 in Theorem 4.2(b).

4.2 The Heteroscedastic Case

Here and in the proofs, we write $T_a(\mathbf{X})$ and $T_a(\mathbf{Y})$ for the T_a statistic in (2.1) evaluated on the X_{ij}'s and Y_{ij}'s, respectively. We again note the close connection between T_a and F_a for the balanced case, described in (2.2). The first two results deal, respectively, with the balanced and unbalanced heteroscedastic cases with small group sizes.

Theorem 4.3. Assume that the conditions of Theorem 2.4 imposed on the X_{ij} hold, and let Y_{ij} and $\mu_i(a)$ be as in (4.1) and (4.2) with $n_i\equiv n\ge 2$. Then, with θ^2 denoting the departure parameter given in (4.4) and $T_a(\mathbf{Y})$ denoting the statistic (2.1) calculated in $\mathbf{Y}$,

$$T_a(\mathbf{Y}) \xrightarrow{d} \mathrm{N}\left(\theta^2, \frac{2ns^4}{n-1}\right) \qquad \text{as } a\to\infty.$$

Theorem 4.4. Suppose that the conditions of Theorem 2.5 imposed on the X_{ij} and n_i are satisfied, and let Y_{ij} and $\mu_i(a)$ be as in (4.1) and (4.2). Furthermore, let ρ_1 be as in Theorem 4.2(a) and set $b = \lim_{a\to\infty} a^{-1}\sum_{i=1}^a n_i$ (assumed to exist). Then

$$T_a(\mathbf{Y}) \xrightarrow{d} \mathrm{N}\big(\theta^2-\rho_1^2/b, 2(s^4+\gamma^4)\big) \qquad \text{as } a\to\infty,$$

where $T_a(\mathbf{Y})$ is the statistic given by (2.1) calculated in $\mathbf{Y}$ and θ^2 is given by (4.4).

When the group sizes also tend to infinity, we have the following result.

Theorem 4.5. (a) In the balanced case, assume that either one of the two sets of conditions of Theorem 2.6 imposed on the X_{ij} hold, and let Y_{ij} and $\mu_i(a)$ be as in (4.1) and (4.2) with $2\le n_i\equiv n(a)\to\infty$. Then

$$T_a(\mathbf{Y}) \xrightarrow{d} \mathrm{N}(\theta^2, 2s^4) \qquad \text{as } a\to\infty,$$

where $T_a(\mathbf{Y})$ is the statistic given by (2.1) calculated in $\mathbf{Y}$ and θ^2 is the departure parameter given in (4.4).

(b) In the unbalanced case, assume that either one of the two sets of conditions of Theorem 2.6 imposed on the X_{ij} and $n_i(a)$ are satisfied, and let Y_{ij} and $\mu_i(a)$ be as in (4.1) and (4.2) [with $n_i = n_i(a)\to\infty$]. Furthermore, let ρ_2 and β be as in Theorem 4.2(b) (assumed to exist). Then

$$T_a(\mathbf{Y}) \xrightarrow{d} \mathrm{N}(\theta^2-\beta\rho_2^2, 2s^4) \qquad \text{as } a\to\infty,$$

with $T_a(\mathbf{Y})$ and θ^2 as in (a).

5. SIMULATION RESULTS

The simulations reported in Section 1 and Remark 2.3(ii) suggest that the heteroscedastic procedures should be used whenever homoscedasticity cannot be ascertained. The purpose of this section is to investigate possible disadvantages of using heteroscedastic procedures on homoscedastic data. The simulations focus mainly on the unweighted statistics that can be used also with small group sizes. Indeed, the simulation reported in Remark 2.2 indicates that for the Type I error rate of the weighted statistic to be acceptable, there must be at least 80 observations per group. Thus it is not included in the simulations in this section. However, as an indication of the power advantage of the weighted statistic, our last simulation uses heteroscedastic data and includes the statistic that standardizes by the true variances.

All simulations are based on 1,000 replications and use normal and lognormal distributions. In this section, CF denotes the classical F test, and BHOM and BHET denote the test procedures of Theorems 2.1(a) and 2.4, respectively.

It should be pointed out that consistent estimation of s^4 in Theorem 2.4 requires unbiased estimation of each σ_i^4. For this we used U statistics, each with kernel $(X_{i1}-X_{i2})^2\times(X_{i3}-X_{i4})^2$. Thus, even though the asymptotic result in Theorem 2.4 requires only that $n\ge 2$, actual application of BHET requires $n\ge 4$.

To investigate the achieved α levels, we considered balanced groups of size $n=5$ and number of groups $a=10,15,20,25$, with nominal $\alpha=.05$. In the case of Lognormal(0, 1) samples, the achieved α levels were satisfactory even with $a=10$ (.040, .074, and .060 for CF, BHOM, and BHET), with BHOM always somewhat more liberal than BHET. Under normality, BHET is more liberal than BHOM, and both are more liberal than in the lognormal case. For example, with $a = 25$ the achieved α levels of CF, BHOM, and BHET were .050, .073, and .077.

Power simulations in the homoscedastic case used $a = 20$ and balanced groups of size $n = 5$, at nominal $\alpha = .05$. Data values were generated as $X_{ij} \sim i\tau/a + \text{error}$, where the error was again either Normal(0, 1) or Lognormal(0, 1). In the lognormal case, BHET seems to outperform both CF and BHOM. For example, the achieved α levels (i.e., for $\tau = 0$) were .048, .075, and .054 for CF, BHOM, and BHET, and their achieved powers at $\tau = 3$ were, respectively, .670, .732, and .810. Due to the already mentioned liberal tendency of BHOM and BHET in the normal case, their achieved powers in the simulation with normal errors turned out to be higher than that of CF.

For comparison purposes with the weighted statistic, our last simulation uses heteroscedastic data with $a = 30$, $n = 10$, and nominal $\alpha = .05$. The heteroscedasticity in the generated data, $X_{ij} \sim \mathrm{N}(i\tau/a, 1 + 25i/a)$, was chosen so that the achieved α level of CF is in an acceptable range (although that of BHOM is not). As expected, the weighted statistic (with weights based on the true variances) outperforms the other procedures. For example, the achieved α levels were .071, .095, .071, and .056 for CF, BHOM, BHET, and the weighted statistic, whereas their powers at $\tau = 2$ were .249, .299, .252, and .360, respectively.

6. CONCLUSIONS

Recent work by Boos and Brownie (1995), Akritas and Arnold (2000), and Bathke (2002) considered the behavior of the classical F statistic, F_a, in homoscedastic ANOVA models with a large number of levels and small group sizes. In this article we consider the heteroscedastic one-way ANOVA model with a large number of levels and small or large group sizes.

Our theoretical results, backed by numerical evidence, show that the F procedure (using F_a with F distribution critical points) is sensitive to departures from homoscedasticity in both the balanced and the unbalanced case. Even under homoscedasticity, it is not asymptotically valid in the unbalanced case unless the group sizes are also large. An asymptotically valid procedure using F_a in the unbalanced homoscedastic case with small group sizes is suggested in Theorem 2.2, but it requires estimation of the fourth moment and assumes the existence of $4 + \delta$ moments.

Because of this sensitivity of the F procedure to departures from homoscedasticity, our investigation also considered two statistics designed to accommodate heteroscedasticity. One is the classical weighted statistic, $\widehat{T}_{\mathrm{W}}$, and the other is an unweighted statistic, T_a, which appears to be new. Statistic T_a differs from F_a only in the unbalanced case, and performs well even with small group sizes. It is recommended in any situation where homoscedasticity cannot be ascertained, although if the group sizes are also large (we recommend at least 80), $\widehat{T}_{\mathrm{W}}$ can also be used for improved power.

Although they were designed for the heteroscedastic case, the asymptotic properties of T_a are preferable to those of F_a, even in the homoscedastic case. For example, the asymptotic null results for T_a assume only that the variables in the same group have the same distribution, whereas those for F_a require all observations to be iid. Moreover, in the unbalanced case, the asymptotic distribution of T_a uses fewer moment conditions and its limiting distribution does not involve the fourth moment. On the basis of this, we recommend T_a over F_a in all unbalanced situations (homoscedastic or not).

The theoretical results do not clearly identify a preferable procedure in the balanced homoscedastic case. (Although the asymptotic result for F_a requires all observations to be iid, that for T_a uses slightly stronger moment assumptions.) However, the simulations in Section 5 suggest that T_a behaves at least comparably to F_a, even in the balanced homoscedastic case.

The combined overall recommendation is that T_a be used in all situations with small group sizes. For heteroscedastic models with large group sizes, we also recommend $\widehat{T}_{\mathrm{W}}$.

APPENDIX A: PROOFS

Proof of Theorems 2.1 and 4.1. We firstly study MSE. If n if fixed, it is immediate (using an obvious notation) that $\mathrm{MSE}(\mathbf{Y}) = \mathrm{MSE}(\mathbf{X}) \xrightarrow{p} \sigma^2$ as $a \to \infty$. In the nontrivial case where $n = n(a) \to \infty$, we evidently have $S_i^2 = S_i^2(a) \xrightarrow{p} \sigma^2$ as $a \to \infty$, and because the $S_i^2(a)$, $a \geq 1$, are nonnegative random variables with $\mathbb{E}(S_i^2(a)) = \sigma^2$, it follows by Scheffé's lemma (see the subsequent proof of Lemma D.1) that $\mathbb{E}|S_i^2(a) - \sigma^2| \to 0$. Thus, the assertion $\mathrm{MSE} = \mathrm{MSE}(\mathbf{X}) = \mathrm{MSE}(\mathbf{Y}) \xrightarrow{p} \sigma^2$ follows from $\mathbb{E}|\mathrm{MSE} - \sigma^2| \leq (1/a)\sum_{i=1}^a \mathbb{E}|S_i^2(a) - \sigma^2| = \mathbb{E}|S_1^2(a) - \sigma^2| \to 0$. Therefore, writing

$$a^{1/2}\big(F_a(\mathbf{Y}) - 1\big) = a^{1/2}\big(F_a(\mathbf{X}) - 1\big) + \frac{1 - a^{-1}}{\mathrm{MSE}(\mathbf{X})}\big(h(a) + H_a\big), \tag{A.1}$$

where

$$h(a) = a^{-1/2} n \sum_{i=1}^a \mu_i^2(a), \qquad H_a = 2na^{-1/2}\sum_{i=1}^a \mu_i(a)\overline{X}_i. \tag{A.2}$$

[note that $n^{1/2}(\mu_1(a) + \cdots + \mu_a(a)) = \int_0^1 g(t)\,dt = 0$], the desired results follow if we show parts (a) and (b) of Theorem 2.1 together with $h(a) \to \theta^2$ and $H_a \xrightarrow{p} 0$ as $a \to \infty$.

Observe that $h(a) = a\sum_{i=1}^a (\int_{(i-1)/a}^{i/a} g(t)\,dt)^2 = a^{-1} \times \sum_{i=1}^a g^2(t_{i,a}) \to \theta^2$, where $t_{i,a} \in [(i-1)/a, i/a]$. Also, $\mathbb{E}H_a = 0$ and $\operatorname{Var} H_a = 4\sigma^2 a^{-1/2} h(a)$; thus $H_a = o_p(1)$. Hence, it remains to establish the null distribution results [parts (a) and (b) of Theorem 2.1], and without loss of generality, we may assume that $\mathbb{E}X_{ij} = 0$ and $\operatorname{Var} X_{ij} = 1$. Proposition 3.1 implies that in both cases it suffices to consider $\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X}$. Using (3.7) for the present balanced case, it follows that

$$\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X} = \frac{1}{a}\sum_{i=1}^a U_i, \tag{A.3}$$

where

$$U_i = \frac{1}{n-1}\left[\left(\sum_{j=1}^n X_{ij}\right)^2 - \sum_{j=1}^n X_{ij}^2\right]$$

are iid random variables. Clearly, $\mathbb{E}U_i = 0$, whereas

$$\begin{aligned}\mathbb{E}\big((n-1)^2U_i^2\big) &= \mathbb{E}\left(\sum_{j_1=1}^n \sum_{j_2=1, j_2\neq j_1}^n 2X_{ij_1}^2 X_{ij_2}^2\right) \\ &= 2n(n-1).\end{aligned} \tag{A.4}$$

[Note that the computation in (A.4) requires the finiteness of only the second moment.] The classical central limit theorem and the preceding variance computation yield $a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X} \xrightarrow{d} \mathrm{N}(0, 2n/(n-1))$ as $a \to \infty$, showing part (a) of Theorem 2.1. For part (b), write (A.3) as

$$\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X} = \frac{1}{a}\sum_{i=1}^{a} U_{a,i},$$

where

$$U_{a,i} = \frac{n(a)}{n(a)-1}\left[\left(\frac{1}{n(a)^{1/2}}\sum_{j=1}^{n(a)} X_{ij}\right)^2 - \frac{1}{n(a)}\sum_{j=1}^{n(a)} X_{ij}^2\right],$$

$i = 1, \dots, a$, are iid random variables. As previously remarked, $\mathbb{E}U_{a,i} = 0$ and [see (A.4)] $\operatorname{Var} U_{a,i} = 2n(a)/(n(a)-1)$. Note the similarity between $U_{a,1}$ and U_a of Lemma D.1(a). Using the result of this lemma,

$$\begin{aligned}
&\frac{a(n(a)-1)}{2n(a)}\sum_{i=1}^{a}\mathbb{E}\left[\frac{1}{a^2}U_{a,i}^2 I\left(|U_{a,i}| \ge a\left(\frac{2n(a)}{a(n(a)-1)}\right)^{1/2}\epsilon\right)\right] \\
&= \frac{n(a)-1}{2n(a)}\mathbb{E}\left[U_{a,1}^2 I\left(|U_{a,1}| \ge \left(\frac{2an(a)}{n(a)-1}\right)^{1/2}\epsilon\right)\right] \\
&\to 0
\end{aligned}$$

as $a \to \infty$. It follows that Lindeberg's condition is satisfied for $\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X}$ and part (b) of Theorem 2.1 is proved.

Proof of Theorems 2.2 and 4.2. It is easy to see that, in both cases [with n_i fixed or $n_i = n_i(a) \to \infty$], $\mathrm{MSE} = \mathrm{MSE}(\mathbf{X}) = \mathrm{MSE}(\mathbf{Y})$ is a consistent estimator of σ^2 as $a \to \infty$, because it is unbiased and $a(\bar{n}-1)\operatorname{Var}(\mathrm{MSE})/\sigma^4 = \mu_4 - 1 - (\mu_4 - 3)(1 - (1/a)\sum_{i=1}^{a} 1/n_i)/(\bar{n}-1)$, where $\mu_4 = \mathbb{E}[(X_{ij}-\mu)^4/\sigma^4]$. Again split $a^{1/2}(F_a(\mathbf{Y}) - 1)$ as in (A.1), where now

$$\begin{aligned}
h(a) &= a^{-1/2}\sum_{i=1}^{a} n_i\big(\mu_i(a) - \overline{\mu}(a)\big)^2, \\
H_a &= 2a^{-1/2}\sum_{i=1}^{a} n_i\big(\mu_i(a) - \overline{\mu}(a)\big)\overline{X}_{i\cdot},
\end{aligned} \tag{A.5}$$

with $\overline{\mu}(a) = (1/N)\sum_{i=1}^{a} n_i\mu_i(a)$. We observe that to establish both parts (a) and (b) of Theorem 4.2, it is sufficient to verify the null-hypothesis results [parts (a) and (b) of Theorem 2.2] together with the fact that $h(a)$ tends to an appropriate constant as $a \to \infty$. Regarding the last limit, write $h(a) = a^{-1/2}\sum_{i=1}^{a} n_i\mu_i^2(a) - Na^{-1/2}\overline{\mu}^2(a)$ and observe that the first term tends to θ^2, as in the previous proof, whereas the second term is $Na^{-1/2}\overline{\mu}^2(a) = aN^{-1}(\sum_{i=1}^{a} n_i^{1/2}\int_{(i-1)/a}^{i/a} g(t)\,dt)^2 = \bar{n}^{-1}(\int_0^1 s_a(t)g(t)\,dt)^2$, which, when the n_i are fixed, converges to $\rho_1^2/b \le \theta^2$; in the case where $n_i = n_i(a) \to \infty$, we have $N(a)a^{-1/2}\overline{\mu}^2(a) = \kappa(a)\bar{n}(a)^{-1}(\int_0^1 s_a(t) \times \kappa(a)^{-1/2}g(t)\,dt)^2 \to \beta\rho_2^2 \le \theta^2$. Therefore, it remains to verify the null distribution results [parts (a) and (b) of Theorem 2.2], and without loss of generality, we may assume that $\mathbb{E}X_{ij} = 0$ and $\operatorname{Var} X_{ij} = 1$, so that $\mu_4 = \mathbb{E}X_{ij}^4$.

(a) By Proposition 3.1 and the fact that $\mathrm{MSE} \xrightarrow{p} 1$, it suffices to show that $a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X} \xrightarrow{d} \mathrm{N}(0, \tau^2)$. By (3.7), it follows that

$$a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X} = \sum_{i=1}^{a} U_{a,i},$$

where

$$U_{a,i} = a^{1/2}\left(\frac{c_1}{n_i} - c_3\right)\left(\sum_{j=1}^{n_i} X_{ij}\right)^2 - c_2 a^{1/2}\sum_{j=1}^{n_i} X_{ij}^2,$$

$i = 1, \dots, a$, are independent random variables. It is easily seen that $\mathbb{E}U_{a,i} = a^{1/2}[c_1 - n_i(c_2 + c_3)]$, so that $\sum_{i=1}^{a}\mathbb{E}U_{a,i} = 0$. Thus $a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X}$ is centered and part (a) of Theorem 2.2 follows by an application of Lyapounov's theorem. To do this, we make use of the variance expression

$$\begin{aligned}
\operatorname{Var}(a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X}) = a\sum_{i=1}^{a}\Bigg\{&\left(\frac{c_1}{n_i} - c_2 - c_3\right)^2(\mu_4 - 1)n_i \\
&+ 2n_i(n_i - 1)\left(\frac{c_1}{n_i} - c_3\right)^2\Bigg\},
\end{aligned} \tag{A.6}$$

which can be easily verified after some algebra. (Note that in the balance case ($n_i = n$, for all i), (A.6) reduces to $\operatorname{Var}(a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X}) = 2n/(n-1)$, [see the proof of Theorem 2.1(a)], because $c_1/n - c_2 - c_3 = 0$, $c_1/n - c_3 = c_2 = [a(n-1)]^{-1}$. Of course, this derivation assumes a finite fourth moment, which is not assumed in Theorem 2.1(a), and the reason is that the square terms, in contrast to the balanced case, do not cancel here.) Using $\lim_{a\to\infty} a^2c_1^2 = (b/(b-1))^2$, $\lim_{a\to\infty} a^2(c_2 + c_3)^2 = (b-1)^{-2}$, $\lim_{a\to\infty} a^2c_3^2 = \lim_{a\to\infty} a^2c_1c_3 = 0$, and $\lim_{a\to\infty} a^2c_1c_2 = b/(b-1)^2$, it can be easily verified that

$$\begin{aligned}
\operatorname{Var}(a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X}) &\to (\mu_4 - 1)\frac{b(bb_1 - 1)}{(b-1)^2} + 2\frac{b^2(1-b_1)}{(b-1)^2} \\
&= \tau^2,
\end{aligned} \tag{A.7}$$

where the last equality can be seen by adding and subtracting $2b/(b-1)$. Because of (A.7), Lyapounov's condition is satisfied if, for some $\delta > 0$,

$$\sum_{i=1}^{a}\mathbb{E}|U_{a,i} - \mathbb{E}U_{a,i}|^{2+\delta} \to 0 \qquad \text{as } a \to \infty. \tag{A.8}$$

This is shown in Appendix B.

(b) Using (3.7), write

$$a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X} = a^{-1/2}\sum_{i=1}^{a} U_{a,i}, \tag{A.9}$$

where now

$$\begin{aligned}
U_{a,i} = {}& ac_1(a)\left[\left(n_i(a)^{-1/2}\sum_{j=1}^{n_i(a)} X_{ij}\right)^2 - 1\right] \\
&+ N(a)c_2(a)\frac{an_i(a)}{N(a)}\left[1 - \frac{1}{n_i(a)}\sum_{j=1}^{n_i(a)} X_{ij}^2\right] \\
&+ aN(a)c_3(a)\frac{n_i(a)}{N(a)}\left[1 - \left(n_i(a)^{-1/2}\sum_{j=1}^{n_i(a)} X_{ij}\right)^2\right].
\end{aligned} \tag{A.10}$$

From the assumptions, we have

$$\begin{aligned}
ac_1(a) &\to 1, \\
N(a)c_2(a) &\to 1, \\
aN(a)c_3(a) &\to 1, \\
\frac{n_i(a)}{N(a)} &\to 0, \\
\frac{an_i(a)}{N(a)} &\le C.
\end{aligned} \tag{A.11}$$

Using (A.9), (A.10), (A.11), and Chebyshev's inequality, it is easy to see that

$$\begin{aligned} a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X} &= (ac_1(a))a^{-1/2}\sum_{i=1}^{a}\left[\left(n_i(a)^{-1/2}\sum_{j=1}^{n_i(a)}X_{ij}\right)^2-1\right]+o_P(1) \\ &= (ac_1(a))a^{-1/2}\sum_{i=1}^{a}V_{a,i}+o_P(1), \end{aligned} \tag{A.12}$$

where $V_{a,i}$ are defined in the preceding equation. Also

$$v_a^2=\mathrm{Var}\left[a^{-1/2}\sum_{i=1}^{a}V_{a,i}\right]=2+(\mu_4-3)\frac{1}{a}\sum_{i=1}^{a}\frac{1}{n_i(a)}\to 2.$$

It remains to verify Lindeberg's condition, that is, to show that for any $\epsilon>0$,

$$\frac{1}{av_a^2}\sum_{i=1}^{a}\mathbb{E}\left[V_{a,i}^2 I(|V_{a,i}|\ge \epsilon v_a a^{1/2})\right]\to 0 \qquad \text{as } a\to\infty. \tag{A.13}$$

This, however, follows immediately from Lemma D.2.

Proof of Theorem 2.4. Without loss of generality, we may assume that $\mathbb{E}X_{ij}=0$. The limiting distribution of T_a coincides with that of $a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X}=a^{1/2}(\mathrm{MST}-\mathrm{MSE})$, because $a\mathbb{E}(\mathbf{X}'\mathbf{A}\mathbf{X}-\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X})^2\le 2a^{-1}(a-1)^{-2}(\sum_{i=1}^{a}\sigma_i^2)^2\to 0$. From

$$a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X}=\sum_{i=1}^{a}U_{a,i},$$

where

$$U_{a,i}=\frac{1}{a^{1/2}(n-1)}\left[\left(\sum_{j=1}^{n}X_{ij}\right)^2-\sum_{j=1}^{n}X_{ij}^2\right],$$

it follows that $\mathbb{E}U_{a,i}=0$, $\mathrm{Var}\,U_{a,i}=2n\sigma_i^4/((n-1)a)$ and, therefore,

$$\mathrm{Var}(a^{1/2}\mathbf{X}'\mathbf{A}_{\mathrm{D}}\mathbf{X})=\frac{2n}{(n-1)a}\sum_{i=1}^{a}\sigma_i^4\to\frac{2ns^4}{n-1}. \tag{A.14}$$

As in the proof of Theorem 2.2(a), we proceed to verify Lyapounov's condition (A.8); this can be easily done using (B.4) and the assumptions as follows:

$$\begin{aligned} &\sum_{i=1}^{a}\mathbb{E}|U_{a,i}|^{2+\delta} \\ &\quad=\left(a(n-1)^2\right)^{-1-\delta/2}\sum_{i=1}^{a}\mathbb{E}\left|\left(\sum_{j=1}^{n}X_{ij}\right)^2-\sum_{j=1}^{n}X_{ij}^2\right|^{2+\delta} \\ &\quad\le n^{2+\delta}a^{-1-\delta/2}\sum_{i=1}^{a}\left(\mathbb{E}|X_{i1}|^{2+\delta}\right)^2\to 0. \end{aligned}$$

Proof of Theorem 2.5. The existence of $\lim_{a\to\infty}a^{-1}\sum_{i=1}^{a}\sigma_i^4=s^4<\infty$ implies that $\sigma_a^4=o(a)$ as $a\to\infty$. Thus, for any $\epsilon>0$, $\sigma_a^4/a<\epsilon$ for all $a\ge a_0$, showing that for large enough a, $a^{-1}\times\max_{1\le i\le a}\{\sigma_i^4\}\le a^{-1}\max\{\sigma_1^4,\dots,\sigma_{a_0-1}^4\}+a^{-1}\max\{\sigma_{a_0}^4,\dots,\sigma_a^4\}\le C_0/a+\epsilon$, which yields

$$\lim_{a\to\infty}\frac{1}{a}\max_{1\le i\le a}\{\sigma_i^4\}=0. \tag{A.15}$$

We find the asymptotic distribution of T_a by projecting it onto the class $\mathcal{C}_{\mathrm{H}}$ given in (3.1), and without loss of generality we may assume that $\mathbb{E}X_{ij}=0$. From (3.2) and $\mathbb{E}T_a=0$, the projected statistic is $\widetilde{T}_a=\sum_{i=1}^{a}\mathbb{E}(T_a|\mathbf{X}_i)$, where easy calculations yield

$$\begin{aligned} \mathbb{E}(T_a|\mathbf{X}_i)&=\frac{1}{\sqrt{a}}\left(1-\frac{n_i}{N}\right)(n_i\overline{X}_{i\cdot}^2-S_i^2) \\ &=\frac{1}{\sqrt{a}}\left(1-\frac{n_i}{N}\right)\frac{1}{n_i-1}\sum_{j_1,j_2=1,\,j_2\ne j_1}^{n_i}X_{ij_1}X_{ij_2}. \end{aligned}$$

After some algebra, we get $\widetilde{T}_a-T_a=a^{-1/2}N^{-1}\sum_{i_1\ne i_2,i_1,i_2=1}^{a}\times\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}X_{i_1j_1}X_{i_2j_2}$ and, therefore,

$$\begin{aligned} \mathbb{E}(\widetilde{T}_a-T_a)^2&=\frac{2}{aN^2}\left(\left(\sum_{i=1}^{a}n_i\sigma_i^2\right)^2-\sum_{i=1}^{a}n_i\sigma_i^4\right) \\ &\le\frac{2}{a}\max_{1\le i\le a}\{\sigma_i^4\}\to 0 \end{aligned}$$

by (A.15). Moreover,

$$\begin{aligned} \mathrm{Var}\,\widetilde{T}_a&=2\left(1-\frac{1}{N}\right)^2\left(\frac{1}{a}\sum_{i=1}^{a}\sigma_i^4+\frac{1}{a}\sum_{i=1}^{a}\frac{\sigma_i^4}{n_i-1}\right) \\ &\qquad-\frac{2}{N}\left(2-\frac{1}{N}\right)\frac{1}{a}\sum_{i=1}^{a}n_i\sigma_i^4+\frac{2}{N^2}\frac{1}{a}\sum_{i=1}^{a}n_i^2\sigma_i^4, \end{aligned}$$

and by using the assumptions, (A.15), and the fact that $N\to\infty$ as $a\to\infty$, we conclude that $\mathrm{Var}\,\widetilde{T}_a\to 2(s^4+\gamma^4)\in(0,\infty)$ as $a\to\infty$. Therefore, it remains to verify Lyapounov's condition for $\widetilde{T}_a$, that is, it suffices to show that

$$L(a)=a^{-1-\delta/2}\sum_{i=1}^{a}\left(\left(1-\frac{n_i}{N}\right)\frac{1}{n_i-1}\right)^{2+\delta}\mathbb{E}|U_i|^{2+\delta}\to 0 \tag{A.16}$$

for some $\delta>0$, as $a\to\infty$, where $U_i=\sum_{j_1\ne j_2,j_1,j_2=1}^{n_i}X_{ij_1}X_{ij_2}$. Using (B.4) again, it follows that $\mathbb{E}|U_i|^{2+\delta}\le(n_i(n_i-1))^{2+\delta}\times(\mathbb{E}|X_{i1}|^{2+\delta})^2$, and using the assumptions that $a^{-1}\sum_{i=1}^{a}n_i^{2+\delta}\le C_1<\infty$ and $\max_{1\le i\le a}\mathbb{E}|X_{i1}|^{2+\delta}\le C_2<\infty$ for all a and the fact that $(1-n_i/N)<1$, we get $L(a)\le a^{-\delta/2}C_1C_2^2$, from which (A.16) follows.

Proof of Theorem 2.6. Without loss of generality we may assume that $\mathbb{E}X_{ij}=0$. According to (3.2), the projection of T_a onto the class $\mathcal{C}_{\mathrm{H}}$, in (3.1), is $\widetilde{T}_a=\sum_{i=1}^{a}\mathbb{E}(T_a|\mathbf{X}_i)$, where

$$\mathbb{E}(T_a|\mathbf{X}_i)=\frac{1}{\sqrt{a}}\left(1-\frac{n_i(a)}{N(a)}\right)\frac{1}{n_i(a)-1}\sum_{j_1,j_2=1,\,j_2\ne j_1}^{n_i(a)}X_{ij_1}X_{ij_2}.$$

As in the proof of Theorem 2.5, we have $\mathbb{E}(\widetilde{T}_a-T_a)^2\le 2a^{-1}N(a)^{-1}\times\sum_{i=1}^{a}n_i(a)\sigma_i^4\le 2\kappa(a)a^{-2}n(a)^{-1}\sum_{i=1}^{a}\sigma_i^4\to 0$ by the assumption that $\kappa(a)/n(a)\le C$. Moreover,

$$\mathrm{Var}\,\widetilde{T}_a\le\frac{2}{a}\left(1-\frac{n(a)}{a\kappa(a)}\right)^2\frac{n(a)}{n(a)-1}\sum_{i=1}^{a}\sigma_i^4\to 2s^4$$

and, similarly, $\mathrm{Var}\,\widetilde{T}_a\ge 2a^{-1}(1-a^{-1}\kappa(a)/n(a))^2\sum_{i=1}^{a}\sigma_i^4\to 2s^4$, showing that $\mathrm{Var}\,\widetilde{T}_a\to 2s^4$ as $a\to\infty$. Therefore, it remains to verify Lyapounov's condition for $\widetilde{T}_a$, that is, it suffices to show that

$$L(a)=a^{-1-\delta/2}\sum_{i=1}^{a}\left(\left(1-\frac{n_i(a)}{N(a)}\right)\frac{1}{n_i(a)-1}\right)^{2+\delta}\mathbb{E}|U_{a,i}|^{2+\delta}\to 0$$

for some $\delta > 0$, as $a \to \infty$, where $U_{a,i} = \sum_{j_1 \neq j_2, j_1, j_2 = 1}^{n_i(a)} X_{ij_1} X_{ij_2}$. However, because

$$L(a) \le R(a) = \frac{a^{-1-\delta/2}}{(n(a)-1)^{2+\delta}} \sum_{i=1}^{a} \mathbb{E}|U_{a,i}|^{2+\delta},$$

the desired result is proved if we show that $R(a) \to 0$ as $a \to \infty$. Assuming first (2.8) and using (B.4), it follows that $\mathbb{E}|U_{a,i}|^{2+\delta} \le (n_i(a)(n_i(a)-1))^{2+\delta}(\mathbb{E}|X_{i1}|^{2+\delta})^2$ and, thus,

$$R(a) \le a^{-\delta/2}\left(\frac{\kappa(a)(\kappa(a)-1)}{n(a)-1}\right)^{2+\delta} a^{-1}\sum_{i=1}^{a}(\mathbb{E}|X_{i1}|^{2+\delta})^2 \to 0,$$

because $\kappa(a)/n(a) \le C$, $n(a) \to \infty$, $a^{-1}\sum_{i=1}^{a}(\mathbb{E}|X_{i1}|^{2+\delta})^2 \le C_2 < \infty$ for all a, and $\kappa(a)^{2+\delta} \le Cn(a)^{2+\delta} = o(a^{\delta/2})$. Assume next (2.9). It is known that for any $p \ge 2$, there exists a finite positive constant A_p (depending only on p) such that for any iid random variables $V_1, \dots, V_n$ with $\mathbb{E}V_i = 0$,

$$\mathbb{E}|V_1 + \cdots + V_n|^p \le A_p n^{p/2}\mathbb{E}|V_1|^p; \tag{A.17}$$

this fact follows if we first use the Marcinkiewich–Zygmund inequality, $\mathbb{E}|V_1 + \cdots + V_n|^p \le A_p \mathbb{E}(V_1^2 + \cdots + V_n^2)^{p/2}$, $p \ge 1$, and then apply (B.1) to the last sum (see Chow and Teicher 1997, pp. 386–387). Using (A.17) and (B.1), it can be shown that

$$\mathbb{E}\left|\left(\sum_{j=1}^{n_i(a)} X_{ij}\right)^2 - \sum_{j=1}^{n_i(a)} X_{ij}^2\right|^{2+\delta} \le D_\delta n_i(a)^{2+\delta}\mathbb{E}|X_{i1}|^{4+2\delta}, \tag{A.18}$$

where D_δ is a finite positive constant that depends only on δ. Use of (A.18) yields

$$\begin{aligned} R(a) &\le D_\delta \frac{a^{-1-\delta/2}}{(n(a)-1)^{2+\delta}} \sum_{i=1}^{a} n_i(a)^{2+\delta}\mathbb{E}|X_{i1}|^{4+2\delta} \\ &\le D_\delta\, a^{-\delta/2}\left(\frac{\kappa(a)}{n(a)-1}\right)^{2+\delta} a^{-1}\sum_{i=1}^{a}\mathbb{E}|X_{i1}|^{4+2\delta} \to 0. \end{aligned}$$

Proof of Proposition 2.7. Without loss of generality, we may assume that $\mathbb{E}X_{ij} = 0$ for all i, j. Let T_W be given as in (2.3) or (2.4) and set $S_a = T_W - (a-1)$. A straightforward calculation yields that the projection of $T_W - (a-1)$ onto $\mathcal{C}_H$, defined in (3.1), is

$$\begin{aligned} \widetilde{S}_a &= \sum_{i=1}^{a} \mathbb{E}(T_W - (a-1)|\mathbf{X}_i) \\ &= \sum_{i=1}^{a}\left(1 - \frac{1}{t(a)}\frac{n_i}{\sigma_i^2}\right)\left(\frac{n_i}{\sigma_i^2}\overline{X}_{i\cdot}^2 - 1\right). \end{aligned} \tag{A.19}$$

It follows easily using Chebyshev's inequality that $a^{-1/2}(S_a - \widetilde{S}_a) \xrightarrow{p} 0$ as $a \to \infty$ holds without any restrictions on the group sizes $n_1, \dots, n_a$ and with no higher moment assumptions. Thus, the asymptotic distribution of S_a is the same as that of its projection $\widetilde{S}_a$, given in (A.19). Now, with $t(a) = \sum_{i=1}^{a} n_i/\sigma_i^2$,

$$\begin{aligned} \mathrm{Var}(a^{-1/2}\widetilde{S}_a) &= 2\frac{1}{a}\sum_{i=1}^{a}\left(1 - \frac{1}{t(a)}\frac{n_i}{\sigma_i^2}\right)^2 \\ &\quad + \frac{1}{a}\sum_{i=1}^{a}\left(1 - \frac{1}{t(a)}\frac{n_i}{\sigma_i^2}\right)^2 \frac{\mathbb{E}X_{i1}^4 - 3\sigma_i^4}{n_i\sigma_i^4} \to 2, \end{aligned}$$

because the first term on the right of the equality converges to 2 as is seen by $\sum_{i=1}^{a}(1 - (1/t(a))(n_i/\sigma_i^2))^2 = a - (2/t(a))\sum_{i=1}^{a}(n_i/\sigma_i^2) + (1/t(a)^2)\sum_{i=1}^{a}(n_i^2/\sigma_i^4) = a - 2 + c_a$, $c_a \in [0, 1]$, while the second is easily seen to converge to 0 by the assumptions made. Hence, it suffices to show Lyapounov's condition, that is,

$$\begin{aligned} L(a) &= a^{-1-\delta/2}\sum_{i=1}^{a}\left(1 - \frac{1}{t(a)}\frac{n_i}{\sigma_i^2}\right)^{2+\delta} \\ &\qquad \times \frac{1}{n_i^{2+\delta}\sigma_i^{4+2\delta}}\mathbb{E}\left|\left(\sum_j X_{ij}\right)^2 - n_i\sigma_i^2\right|^{2+\delta} \to 0 \end{aligned}$$

as $a \to \infty$, for some $\delta > 0$. If follows that

$$\begin{aligned} &\mathbb{E}\left|\left(\sum_j X_{ij}\right)^2 - n_i\sigma_i^2\right|^{2+\delta} \\ &\quad \le 2^{1+\delta}\left\{\mathbb{E}\left|\sum_j (X_{ij}^2 - \sigma_i^2)\right|^{2+\delta} + \mathbb{E}\left|\sum_{j_1 \neq j_2} X_{ij_1}X_{ij_2}\right|^{2+\delta}\right\} \\ &\quad \le 2^{1+\delta}\left(A_{2+\delta}n_i^{1+\delta/2}\mathbb{E}|X_{ij}^2 - \sigma_i^2|^{2+\delta} + D_\delta n_i^{2+\delta}\mathbb{E}|X_{ij}|^{4+2\delta}\right) \\ &\quad \le K_\delta n_i^{2+\delta}\mathbb{E}|X_{ij}|^{4+2\delta}, \end{aligned}$$

where the first inequality follows by (B.1), the second by (A.17) and (A.18), the definitions of A_p and D_δ are also given there, and $K_\delta = 2^{3+2\delta}A_{2+\delta} + 2^{1+\delta}D_\delta$. Thus,

$$L(a) \le a^{-1-\delta/2}\sum_{i=1}^{a}\left(1 - \frac{1}{t(a)}\frac{n_i}{\sigma_i^2}\right)^2 K_\delta \frac{\mathbb{E}|X_{ij}|^{4+2\delta}}{\sigma_i^{4+2\delta}} \to 0$$

by the assumptions made.

Proof of Theorem 2.8. We first note that assumptions (2.10) imply

$$\begin{aligned} &\text{(a)} \quad \frac{1}{\bar{n}^{1/2}a}\sum_{i=1}^{a}(\log n_i)^2 \to 0, \\ &\text{(b)} \quad \frac{1}{\bar{n}^2 a^{1/2}}\sum_{i=1}^{a}(\log n_i)^2 \to 0, \\ &\text{(c)} \quad \frac{1}{\bar{n}a^{3/2}}\sum_{i=1}^{a} n_i^{1/2}\log n_i \to 0. \end{aligned} \tag{A.20}$$

In view of Proposition 2.7, it suffices to show that $a^{-1/2}(\widehat{T}_W - T_W) \xrightarrow{p} 0$. Use the notation in Remark 2.1 to write $T_W = \sum_{i=1}^{a}(n_i/\sigma_i^2)(\overline{X}_{i\cdot} - \widehat{\mu})^2$ and $\widehat{T}_W = \sum_{i=1}^{a}(n_i/S_i^2)(\overline{X}_{i\cdot} - \widetilde{\mu})^2$, with $\widetilde{\mu} = \sum_{i=1}^{a}\widehat{w}_i\overline{X}_{i\cdot}$, where $\widehat{w}_i$ are proportional to n_i/S_i^2 and sum to 1. Relationship $a^{-1/2}(\widehat{T}_W - T_W) \xrightarrow{p} 0$ is implied by showing

$$V_1(a) = a^{-1/2}\sum_{i=1}^{a}\left(\frac{1}{\sigma_i^2} - \frac{1}{S_i^2}\right)n_i(\overline{X}_{i\cdot} - \widetilde{\mu})^2 = o_p(1), \tag{A.21}$$

$$V_2(a) = a^{-1/2}\sum_{i=1}^{a}\frac{n_i}{\sigma_i^2}\left[(\overline{X}_{i\cdot} - \widehat{\mu})^2 - (\overline{X}_{i\cdot} - \widetilde{\mu})^2\right] = o_p(1). \tag{A.22}$$

Write

$$\begin{aligned} V_1(a) &= a^{-1/2}\sum_{i=1}^{a}\frac{S_i^2 - \sigma_i^2}{S_i^2\sigma_i^2}n_i\left[(\overline{X}_{i\cdot} - \widehat{\mu})^2 + (\widetilde{\mu} - \widehat{\mu})^2\right. \\ &\qquad\qquad \left. - 2(\widetilde{\mu} - \widehat{\mu})(\overline{X}_{i\cdot} - \widehat{\mu})\right] \\ &= V_{1,1}(a) + V_{1,2}(a) + V_{1,3}(a). \end{aligned}$$

Using Lemma C.2, we have that $\overline{X}_{i\cdot} - \widehat{\mu} = \overline{X}_{i\cdot} - \mu - \sum_{i_1=1}^{a} w_{i_1} \times (\overline{X}_{i_1\cdot} - \mu) = O_p(\log n_i/n_i^{1/2}) + O_p((\bar{n}a)^{-1/2}) = O_p(\log n_i/n_i^{1/2})$

uniformly in i. Assuming, without loss of generality, that $\mu = 0$ and using, in addition, Lemma C.1, it is seen that

$$a^{-1/2}\sum_{i=1}^{a}\frac{(S_i^2-\sigma_i^2)^2}{\sigma_i^4 S_i^2}n_i(\overline{X}_{i\cdot}-\widehat{\mu})^2 = O_p\left(a^{-1/2}\sum_{i=1}^{a}\frac{(\log n_i)^4}{n_i}\right) = o_p(1)$$

by assumption (2.10)(a). Thus, $V_{1,1}(a) = a^{-1/2}\sum_{i=1}^{a}[(S_i^2-\sigma_i^2)/\sigma_i^4]n_i(\overline{X}_{i\cdot}-\widehat{\mu})^2 + o_p(1)$. Using now

$$a^{-1/2}\sum_{i=1}^{a}\frac{S_i^2-\sigma_i^2}{\sigma_i^4}n_i\widehat{\mu}^2 = O_p\left(\frac{1}{\bar{n}a^{3/2}}\sum_{i=1}^{a}n_i^{1/2}\log n_i\right) = o_p(1) \qquad \text{by (A.20)(c)}$$

and

$$a^{-1/2}\sum_{i=1}^{a}\frac{S_i^2-\sigma_i^2}{\sigma_i^4}n_i\widehat{\mu}\overline{X}_{i\cdot} = O_p\left(\frac{1}{\bar{n}^{1/2}a}\sum_{i=1}^{a}(\log n_i)^2\right) = o_p(1) \qquad \text{by (A.20)(a),}$$

we have

$$\begin{aligned}
V_{1,1}(a) &= a^{-1/2}\sum_{i=1}^{a}\frac{1}{\sigma_i^4}\left[\frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij}^2-\sigma_i^2) - \frac{n_i}{n_i-1}\left(\overline{X}_{i\cdot}^2-\frac{\sigma_i^2}{n_i}\right)\right]n_i\overline{X}_{i\cdot}^2 + o_p(1)\\
&= a^{-1/2}\sum_{i=1}^{a}\frac{1}{\sigma_i^4}\frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij}^2-\sigma_i^2)\frac{1}{n_i}\sum_{j_1=1}^{n_i}\sum_{j_2=1}^{n_i}X_{ij_1}X_{ij_2}\\
&\quad - a^{-1/2}\sum_{i=1}^{a}\frac{1}{\sigma_i^4}\frac{n_i}{n_i-1}\frac{1}{n_i^2}\sum_{j=1}^{n_i}(X_{ij}^2-\sigma_i^2)\times\frac{1}{n_i}\sum_{j_1=1}^{n_i}\sum_{j_2=1}^{n_i}X_{ij_1}X_{ij_2}\\
&\quad - a^{-1/2}\sum_{i=1}^{a}\frac{1}{\sigma_i^4}\frac{n_i}{n_i-1}\frac{1}{n_i^2}\sum_{r_1\neq r_2}X_{ir_1}X_{ir_2}\times\frac{1}{n_i}\sum_{j_1=1}^{n_i}\sum_{j_2=1}^{n_i}X_{ij_1}X_{ij_2} + o_p(1)\\
&= (A)+(B)+(C)+o_p(1). \qquad \text{(A.23)}
\end{aligned}$$

Direct calculations, using the fact that the σ_i stay bounded away from zero, yield that $\mathbb{E}[(A)^2]$, $\mathbb{E}[(C)^2]$, and, of course, $\mathbb{E}[(B)^2]$ are all $O((a^{-1/2}\sum_{i=1}^{a}n_i^{-1})^2)$. Thus, by assumption (2.10)(a), $V_{1,1}(a) = o_p(1)$. To show that $V_{1,2}(a) = o_p(1)$ and $V_{1,3}(a) = o_p(1)$ we first consider the rate with which $\widetilde{\mu}-\widehat{\mu}$ goes to zero. Adding and subtracting $(\sum_{i=1}^{a}n_i/\sigma_i^2)^{-1}(n_r/S_r^2)$ inside $(\widehat{w}_r - w_r)$ gives

$$\begin{aligned}
\widetilde{\mu}-\widehat{\mu} &= \sum_{r=1}^{a}(\widehat{w}_r-w_r)(\overline{X}_{r\cdot}-\mu)\\
&= \frac{1/(\bar{n}a)\sum_{i=1}^{a}n_i/\sigma_i^2 - 1/(\bar{n}a)\sum_{i=1}^{a}n_i/S_i^2}{1/(\bar{n}a)^2(\sum_{i=1}^{a}n_i/S_i^2)(\sum_{i=1}^{a}n_i/\sigma_i^2)}\frac{1}{\bar{n}a}\sum_{r=1}^{a}\frac{n_r}{S_r^2}(\overline{X}_{r\cdot}-\mu)\\
&\quad + \frac{1}{1/(\bar{n}a)\sum_{i=1}^{a}n_i/\sigma_i^2}\frac{1}{\bar{n}a}\sum_{r=1}^{a}\frac{n_r}{S_r^2\sigma_r^2}(\overline{X}_{r\cdot}-\mu)(S_r^2-\sigma_r^2)\\
&= O_p(1)\times\left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}n_i\frac{S_i^2-\sigma_i^2}{\sigma_i^2S_i^2}\right)\left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}\frac{n_i}{S_i^2}(\overline{X}_{i\cdot}-\mu)\right)\\
&\quad + O(1)\times\left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}\frac{n_i}{S_i^2\sigma_i^2}(\overline{X}_{i\cdot}-\mu)(S_i^2-\sigma_i^2)\right). \qquad \text{(A.24)}
\end{aligned}$$

The expression in the second term on the right side of (A.24) equals (assuming, without loss of generality, that $\mu = 0$)

$$\begin{aligned}
&\frac{1}{\bar{n}a}\sum_{i=1}^{a}\frac{n_i}{\sigma_i^4}\overline{X}_{i\cdot}(S_i^2-\sigma_i^2) + \frac{1}{\bar{n}a}\sum_{i=1}^{a}\frac{n_i}{S_i^2\sigma_i^4}(S_i^2-\sigma_i^2)^2\overline{X}_{i\cdot}\\
&= O_p\left(\frac{1}{\bar{n}}\right) + O_p\left(\frac{\sum_{i=1}^{a}n_i^{-1/2}(\log n_i)^3}{\sum_{i=1}^{a}n_i}\right)\\
&= O_p\left(\frac{1}{\bar{n}}\right) \qquad \text{(A.25)}
\end{aligned}$$

by Lemmas C.1 and C.2 and the fact that $\mathbb{E}[\{a^{-1}\sum_{i=1}^{a}(n_i/\sigma_i^4)\times(S_i^2-\sigma_i^2)\overline{X}_{i\cdot}\}^2] = O(1)$ as follows by direct calculations and assumption that the σ_i stay bounded away from zero. The expression in the first term on the right side of (A.24) equals (assuming, without loss of generality, that $\mu = 0$)

$$\begin{aligned}
&\left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}n_i\frac{S_i^2-\sigma_i^2}{\sigma_i^4}\right)\left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}\frac{n_i}{\sigma_i^2}\overline{X}_{i\cdot}\right)\\
&\quad - \left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}n_i\frac{S_i^2-\sigma_i^2}{\sigma_i^4}\right)\left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}n_i\overline{X}_{i\cdot}\frac{S_i^2-\sigma_i^2}{S_i^2\sigma_i^2}\right)\\
&\quad + \left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}n_i\frac{(S_i^2-\sigma_i^2)^2}{\sigma_i^4S_i^2}\right)\left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}\frac{n_i}{\sigma_i^2}\overline{X}_{i\cdot}\right)\\
&\quad + \left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}n_i\frac{(S_i^2-\sigma_i^2)^2}{\sigma_i^4S_i^2}\right)\left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}n_i\overline{X}_{i\cdot}\frac{S_i^2-\sigma_i^2}{S_i^2\sigma_i^2}\right)\\
&= O_p\left(\frac{1}{(\bar{n}a)^{1/2}}\right)\times O_p\left(\frac{1}{(\bar{n}a)^{1/2}}\right)\\
&\quad + O_p\left(\frac{1}{(\bar{n}a)^{1/2}}\right)\times O_p\left(\frac{1}{\bar{n}}\right)\\
&\quad + O_p\left(\frac{a^{-1}\sum_i(\log n_i)^2}{\bar{n}}\right)\times O_p\left(\frac{1}{(\bar{n}a)^{1/2}}\right)\\
&\quad + O_p\left(\frac{a^{-1}\sum_i(\log n_i)^2}{\bar{n}}\right)\times O_p\left(\frac{1}{\bar{n}}\right), \qquad \text{(A.26)}
\end{aligned}$$

as is easily seen by decomposing $S_i^2-\sigma_i^2$ as was done for (A.23), direct moment calculations, use of Lemmas C.1 and C.2 and (A.25). Relationships (A.24), (A.25), and (A.26) imply that

$$\widetilde{\mu}-\widehat{\mu} = \sum_{r=1}^{a}(\widehat{w}_r-w_r)(\overline{X}_{r\cdot}-\mu) = O_p\left(\frac{1}{\bar{n}}\right). \qquad \text{(A.27)}$$

Thus, by (A.20)(b),

$$\begin{aligned}
V_{1,2}(a) &= (\widetilde{\mu}-\widehat{\mu})^2a^{-1/2}\sum_{i=1}^{a}n_i\left[\frac{S_i^2-\sigma_i^2}{\sigma_i^4}-\frac{(S_i^2-\sigma_i^2)^2}{\sigma_i^4S_i^2}\right]\\
&= O_p\left(\frac{1}{\bar{n}^2}\right)\times O_p(\bar{n}^{1/2})\\
&\quad + O_p\left(\frac{1}{\bar{n}^2}\right)\times O_p\left(\frac{1}{a^{1/2}}\sum_{i=1}^{a}(\log n_i)^2\right)\\
&= o_p(1), \qquad \text{(A.28)}
\end{aligned}$$

and, using also (A.25),

$$V_{1,3}(a) = 2(\widetilde{\mu} - \widehat{\mu})a^{-1/2}\sum_{i=1}^{a}\frac{S_i^2 - \sigma_i^2}{S_i^2\sigma_i^2}n_i(\overline{X}_{i\cdot} - \mu)$$
$$= O_p\left(\frac{1}{\bar{n}}\right) \times O_p(a^{1/2})$$
$$= o_p(1)$$

by (2.10)(b). Next, we show (A.22). With some algebra we have

$$V_2(a) = a^{-1/2}\sum_{i=1}^{a}\frac{n_i}{\sigma_i^2}\Bigg[\sum_{r=1}^{a}(\widehat{w}_r - w_r)\overline{X}_{r\cdot} \times \left\{2\overline{X}_{i\cdot} - \sum_{m=1}^{a}(\widehat{w}_m + w_m)\overline{X}_{m\cdot}\right\}\Bigg]$$
$$= a^{-1/2}\sum_{i=1}^{a}\frac{n_i}{\sigma_i^2}\Bigg[\sum_{r=1}^{a}(\widehat{w}_r - w_r)\overline{X}_{r\cdot}\{2\overline{X}_{i\cdot} - 2\mu\} - \sum_{r=1}^{a}(\widehat{w}_r - w_r)\overline{X}_{r\cdot}\left\{\sum_{m=1}^{a}(\widehat{w}_m + w_m)\overline{X}_{m\cdot} - 2\mu\right\}\Bigg].$$
$$= V_{2,1}(a) - V_{2,2}(a). \tag{A.29}$$

Using the assumption that the σ_i^2 are bounded away from zero, we have

$$V_{2,1}(a) = \bar{n}^{1/2}\sum_{r=1}^{a}(\widehat{w}_r - w_r)\overline{X}_{r\cdot} \times \left\{\left(\sum_{i=1}^{a}\frac{n_i}{\sigma_i^2}\right)^{-1/2}\sum_{i=1}^{a}\frac{n_i}{\sigma_i^2}(2\overline{X}_{i\cdot} - 2\mu)\right\}\left(\frac{1}{\bar{n}a}\sum_{i=1}^{a}\frac{n_i}{\sigma_i^2}\right)^{1/2}$$
$$= \bar{n}^{1/2}\sum_{r=1}^{a}(\widehat{w}_r - w_r)(\overline{X}_{r\cdot} - \mu) \times O_p(1) \times O(1)$$
$$= o_p(1) \tag{A.30}$$

by (A.27). Next, using (A.27) again and a direct variance calculation,

$$\sum_{m=1}^{a}(\widehat{w}_m + w_m)\overline{X}_{m\cdot} - 2\mu$$
$$= \sum_{m=1}^{a}(\widehat{w}_m + w_m)(\overline{X}_{m\cdot} - \mu)$$
$$= \sum_{m=1}^{a}(\widehat{w}_m - w_m)(\overline{X}_{m\cdot} - \mu) + 2\sum_{m=1}^{a}w_m(\overline{X}_{m\cdot} - \mu)$$
$$= O_p\big((\bar{n}a)^{-1/2}\big).$$

It follows that

$$V_{2,2}(a) = a^{-1/2} \times O(\bar{n}a) \times O_p\left(\frac{1}{\bar{n}}\right) \times O_p\big((\bar{n}a)^{-1/2}\big)$$
$$= O_p\big((\bar{n})^{-1/2}\big). \tag{A.31}$$

Relationships (A.29), (A.30), and (A.31) show (A.22), completing the proof.

Proof of Theorem 4.3. It is easy to see that, with $h(a)$ and H_a as in (A.2),

$$T_a(\mathbf{Y}) = T_a(\mathbf{X}) + h(a) + H_a. \tag{A.32}$$

Therefore, we may repeat the same arguments as in the proof of Theorem 4.1(a); the only difference arises from the fact that $\operatorname{Var} H_a = 4na^{-1}\sum_{i=1}^{a}\mu_i^2(a)\sigma_i^2$ depends on the different σ_i^2. This can be easily treated, however, because from (A.15), which follows from (2.6), we have $\operatorname{Var} H_a \le 4\max_{1\le i\le a}\{\sigma_i^2\}a^{-1/2}h(a) \to 0$.

Proof of Theorem 4.4. Again split $T_a(\mathbf{Y})$ as in (A.32), where now $h(a)$ and H_a are given by (A.5). Therefore, $h(a) \to \theta^2 - \rho_1^2/b \ge 0$, as in the proof Theorem 4.2(a) (with fixed n_i). On the other hand, the variance of H_a goes to zero by (A.15), as in the proof of Theorem 4.3.

Proof of Theorem 4.5. (a) The conclusion follows by using the same arguments as in the proof of Theorem 4.3 in combination with the proof of Theorem 4.1(b), because $h(a)$ depends only on the function g [and not on $n(a)$], and similarly for the expectation and the variance of H_a.

(b) Use (A.32) as in the proof of Theorem 4.4, with $h(a)$ and H_a as in (A.5). Therefore, $h(a) \to \theta^2 - \beta\rho_2^2 \ge 0$, as in the proof of Theorem 4.2(b), and also the variance of H_a goes to zero by (A.15).

APPENDIX B: PROOF OF (A.8)

Let $0 < \delta < 1$ and write $\sum_{i=1}^{a}\mathbb{E}|U_{a,i} - \mathbb{E}U_{a,i}|^{2+\delta} = [(ac_1)^{2+\delta}/a^{\delta/2}]a^{-1}\sum_{i=1}^{a}R_{a,i}$, where

$$R_{a,i} = \mathbb{E}\left|\left(\frac{1}{n_i} - \frac{c_2 + c_3}{c_1}\right)\left(\sum_{j=1}^{n_i}X_{ij}\right)^2 + \frac{c_2}{c_1}\sum_{j_1,j_2=1,j_2\ne j_1}^{n_i}X_{ij_1}X_{ij_2} - \left[1 - n_i\frac{c_2 + c_3}{c_1}\right]\right|^{2+\delta}.$$

Because $ac_1 \to b/(b-1)$ and $a^{\delta/2} \to \infty$, (A.8) is satisfied if we show that $a^{-1}\sum_{i=1}^{a}R_{a,i}$ remains bounded. In the following discussion, we make use of the inequality

$$\left|\sum_{i=1}^{m}z_i\right|^p \le m^{p-1}\sum_{i=1}^{m}|z_i|^p, \qquad m \ge 1,\ p \ge 1. \tag{B.1}$$

(For $p > 1$, the inequality follows from Hölder's inequality.) From (B.1) we have

$$R_{a,i} \le 3^{1+\delta}\left\{\left|\frac{1}{n_i} - \frac{c_2 + c_3}{c_1}\right|^{2+\delta}\mathbb{E}\left|\sum_{j=1}^{n_i}X_{ij}\right|^{4+2\delta} + \left|\frac{c_2}{c_1}\right|^{2+\delta}\mathbb{E}\left|\sum_{j_1,j_2=1,j_2\ne j_1}^{n_i}X_{ij_1}X_{ij_2}\right|^{2+\delta} + \left|1 - n_i\frac{c_2 + c_3}{c_1}\right|^{2+\delta}\right\}, \tag{B.2}$$

$$\mathbb{E}\left|\sum_{j=1}^{n_i}X_{ij}\right|^{4+2\delta} \le n_i^{3+2\delta}\sum_{j=1}^{n_i}\mathbb{E}|X_{ij}|^{4+2\delta}$$
$$= n_i^{4+2\delta}\mathbb{E}|X_{ij}|^{4+2\delta}, \tag{B.3}$$

and

$$\mathbb{E}\left|\sum_{j_1,j_2=1,j_2\ne j_1}^{n_i}X_{ij_1}X_{ij_2}\right|^{2+\delta}$$
$$= 2^{2+\delta}\mathbb{E}\left|\sum_{1\le j_1<j_2\le n_i}X_{ij_1}X_{ij_2}\right|^{2+\delta}$$

$$\le 2^{2+\delta}\left(\frac{n_i(n_i-1)}{2}\right)^{1+\delta}\sum_{1\le j_1<j_2\le n_i}\mathbb{E}\left|X_{ij_1}X_{ij_2}\right|^{2+\delta}$$
$$=\left(n_i(n_i-1)\right)^{2+\delta}\left(\mathbb{E}|X_{ij}|^{2+\delta}\right)^2. \quad \text{(B.4)}$$

Using (B.3) and (B.4), (B.2) becomes [set $M(\delta)=\mathbb{E}|X_{ij}|^{4+2\delta}\ge(\mathbb{E}|X_{ij}|^{2+\delta})^2$]

$$R_{a,i}\le 3^{1+\delta}\left\{\left|\frac{1}{n_i}-\frac{c_2+c_3}{c_1}\right|^{2+\delta}n_i^{4+2\delta}M(\delta)+\left|\frac{c_2}{c_1}\right|^{2+\delta}n_i^{4+2\delta}M(\delta)+\left|1-n_i\frac{c_2+c_3}{c_1}\right|^{2+\delta}\right\}$$
$$\le 3^{1+\delta}n_i^{4+2\delta}[2M(\delta)+1]$$

by the facts $|1/n_i-(c_2+c_3)/c_1|\le 1$ [which follows from $0<(c_2+c_3)/c_1\le 1$], $c_2/c_1\le 1$, and $|1-n_i(c_2+c_3)/c_1|\le n_i$. Thus $a^{-1}\sum_{i=1}^a R_{a,i}$ remains bounded, showing (A.8).

APPENDIX C: LEMMAS USED IN THE PROOF OF THEOREM 2.8

Lemma C.1. Consider the setting of Theorem 2.8. If the $n_i\to\infty$ so that $n(a)>c\log(a)$, where the constant c is given in the proof, then, with probability as high as desired, the S_i^2 stay bounded away from zero for all a large enough.

Proof. Pick $\delta\in(0,\inf\{\sigma_i^2;\ i\ge 1\})$. Then

$$\mathbb{P}[S_i^2>\delta]$$
$$=1-\mathbb{P}[S_i^2-\sigma_i^2<\delta-\sigma_i^2]$$
$$\ge 1-\mathbb{P}[|S_i^2-\sigma_i^2|>|\delta-\sigma_i^2|]$$
$$\ge 1-\mathbb{P}\left[\left|\frac{1}{n_i-1}\sum_{j=1}^{n_i}[(X_{ij}-\mu_i)^2-\sigma_i^2]\right|>\frac{1}{2}|\delta-\sigma_i^2|\right]$$
$$-\mathbb{P}\left[\frac{n_i}{n_i-1}\left|(\overline{X}_{i\cdot}-\mu_i)^2-\frac{\sigma_i^2}{n_i}\right|>\frac{1}{2}|\delta-\sigma_i^2|\right]$$
$$=1-\mathbb{P}\left[\left|\sum_{j=1}^{n_i}[(X_{ij}-\mu_i)^2-\sigma_i^2]\right|>(n_i-1)\frac{1}{2}|\delta-\sigma_i^2|\right]$$
$$-\mathbb{P}\left[\left|\sum_{j=1}^{n_i}(X_{ij}-\mu_i)\right|>n_i\left(\frac{1}{2}|\delta-\sigma_i^2|\frac{n_i-1}{n_i}+\frac{\sigma_i^2}{n_i}\right)^{1/2}\right],$$

where the last equality holds for all $n(a)$ larger than a constant that does not depend on i. Using the preceding relationship, Bernstein's inequality, and the notation $t_i=(1-n_i^{-1})|\delta-\sigma_i^2|/2$, $\tilde{t}_i=(t_i+\sigma_i^2/n_i)^{1/2}$, we have

$$\log\mathbb{P}\left[\bigcap_{i=1}^a\{S_i^2>\delta\}\right]$$
$$=\log\left(\prod_{i=1}^a\mathbb{P}[S_i^2>\delta]\right)$$
$$\ge\sum_{i=1}^a\log\left(1-\exp\left[-\frac{n_i^2t_i^2}{2n_i\operatorname{Var}[(X_{ij}-\mu_i)^2]+(2/3)Mn_it_i}\right]-\exp\left[-\frac{n_i^2\tilde{t}_i^2}{2n_i\operatorname{Var}X_{ij}+(2/3)Mn_i\tilde{t}_i}\right]\right)\to 0,$$

provided $n(a)>c\log(a)$, some

$$c>\sup_{i\ge1}\max\left\{\frac{2\operatorname{Var}[(X_{ij}-\mu_i)^2]+(2/3)M\tilde{t}_i}{t_i^2},\frac{2\operatorname{Var}X_{ij}+(2/3)M\tilde{t}_i}{\tilde{t}_i^2}\right\}.$$

Lemma C.2. Consider the setting of Theorem 2.8. If

$$\sum_{i=1}^a\exp\{-(\log n_i)^2\}\to 0, \quad \text{(C.1)}$$

then

$$\frac{n_i^{1/2}}{\log n_i}(S_i^2-\sigma_i^2)=O_p(1),$$
$$\frac{n_i^{1/2}}{\log n_i}(\overline{X}_{i\cdot}-\mu_i)=O_p(1)$$

uniformly in $i=1,\dots,a$, as $a\to\infty$.

Remark. In the balanced case, condition (C.1) reduces to $(\log n)^2-\log a\to\infty$, which is implied by (2.10)(b).

Proof of Lemma C.2. Let δ be any positive number. Working as in the proof of Lemma C.1, we have

$$\mathbb{P}[|S_i^2-\sigma_i^2|\ge\delta]$$
$$\le\mathbb{P}\left[\left|\sum_{j=1}^{n_i}[(X_{ij}-\mu_i)^2-\sigma_i^2]\right|\ge(n_i-1)\frac{\delta}{2}\right]$$
$$+\mathbb{P}\left[\left|\sum_{j=1}^{n_i}(X_{ij}-\mu_i)\right|\ge n_i\left(\frac{\delta}{2}\frac{n_i-1}{n_i}+\frac{\sigma_i^2}{n_i}\right)^{1/2}\right]$$
$$\le\exp\left[-\frac{(n_i-1)^2\delta^2/4}{2n_i\operatorname{Var}[(X_{ij}-\mu_i)^2]+(2/3)M(n_i-1)\delta/2}\right]$$
$$+\exp\left[-n_i^2[\delta(n_i-1)/(2n_i)+\sigma_i^2/n_i]\big/\left(2n_i\operatorname{Var}X_{ij}+(2/3)Mn_i[\delta(n_i-1)/(2n_i)+\sigma_i^2/n_i]^{1/2}\right)\right].$$

Thus, with $K=9M^4$,

$$\mathbb{P}\left[\bigcup_{i=1}^a\{|S_i^2-\sigma_i^2|\ge Kn_i^{-1/2}\log n_i\}\right]$$
$$\le\sum_{i=1}^a\mathbb{P}\left[|S_i^2-\sigma_i^2|\ge Kn_i^{-1/2}\log n_i\right]\to 0$$

as follows by a straightforward calculation using (C.1). This shows the first of the two statements of the lemma. The second statement follows similarly.

APPENDIX D: TWO USEFUL LEMMAS

Lemma D.1. (a) Let X_i, $i\ge 1$, be a sequence of iid random variables with $\mathbb{E}X_i=0$ and $\mathbb{E}X_i^2=1$. Let $n(a)$ be a sequence of positive integers tending to ∞ as $a\to\infty$ and set

$$U_a=\frac{2}{n(a)}\sum_{j<k,j,k=1}^{n(a)}X_jX_k$$
$$=\left(\frac{1}{n(a)^{1/2}}\sum_{j=1}^{n(a)}X_j\right)^2-\frac{1}{n(a)}\sum_{j=1}^{n(a)}X_j^2,\qquad a\ge 1.$$

Then the sequence of random variables U_a^2, $a \geq 1$, is uniformly integrable.

(b) If, furthermore, $\mathbb{E}X_i^4 < \infty$, then the same is true for the sequence V_κ^2, $\kappa \geq 1$, with

$$V_\kappa = \left(\frac{1}{n(\kappa)^{1/2}} \sum_{j=1}^{n(\kappa)} X_j\right)^2 - 1, \qquad \kappa \geq 1.$$

Proof. (a) Using the central limit theorem, the law of large numbers, and Slutsky's theorem it follows that

$$U_a^2 = \left[\left(\frac{1}{n(a)^{1/2}} \sum_{j=1}^{n(a)} X_j\right)^2 - \frac{1}{n(a)} \sum_{j=1}^{n(a)} X_j^2\right]^2 \xrightarrow{d} (Z^2 - 1)^2 \quad \text{as } a \to \infty,$$

where Z is a random variable that has standard normal distribution. Thus,

$$\lim_{a\to\infty} \mathbb{E}U_a^2 = \lim_{a\to\infty} \frac{2(n(a)-1)}{n(a)} = 2 = \mathbb{E}\left(\lim_{a\to\infty} U_a^2\right),$$

which, together with $U_a^2 \geq 0$, implies that $U_a^2, a \geq 1$, are uniformly integrable by Scheffé's lemma; see theorem 5.4 of Billingsley (1968). Part (b) can be proved similarly.

Lemma D.2. Let X_{ij}, $i, j \geq 1$, be a double sequence of iid random variables with $\mathbb{E}X_{ij} = 0$, $\mathbb{E}X_{ij}^2 = 1$, and $\mathbb{E}X_{ij}^4 < \infty$. Assume that $n_i(a)$, $i = 1, \ldots, a$, $a \geq 1$, are sequences of positive integers (group sizes) that satisfy $\min\{n_i(a); i = 1, \ldots, a\} \to \infty$ as $a \to \infty$ and define the double sequence of random variables

$$V_{a,i} = \left(\frac{1}{n_i(a)^{1/2}} \sum_{j=1}^{n_i(a)} X_{ij}\right)^2 - 1, \qquad 1 \leq i \leq a, a \geq 1.$$

Then, for any sequence $c(a)$ tending to ∞, as $a \to \infty$,

$$\frac{1}{a} \sum_{i=1}^{a} \mathbb{E}\left[V_{a,i}^2 I\left(|V_{a,i}| \geq c(a)\right)\right] \to 0 \qquad \text{as } a \to \infty. \tag{D.1}$$

Proof. Define the sequence of "pooled group sizes" $m(\kappa)$, $\kappa \geq 1$, as

$$\begin{array}{lll} m(1) = n_1(1), & & \\ m(2) = n_1(2), & m(3) = n_2(2), & \\ m(4) = n_1(3), & m(5) = n_2(3), & m(6) = n_3(3), \\ \vdots & \vdots & \vdots \end{array}$$

and define the sequence of random variables $V_\kappa = (m(\kappa)^{-1/2} \times \sum_{j=1}^{m(\kappa)} X_j)^2 - 1$, $\kappa \geq 1$, where $X_1, X_2, \ldots$ is an iid sequence that has the same distribution as the X_{ij}'s. It is clear from the preceding construction that $m(\kappa) \to \infty$ as $\kappa \to \infty$, and from Lemma D.1(b), we conclude that the sequence V_κ^2, $\kappa \geq 1$, is uniformly integrable. It is also clear that the left side of (D.1) is less than or equal to

$$\sup_{\kappa \geq 1} \mathbb{E}\left[V_\kappa^2 I\left(|V_\kappa| \geq c(a)\right)\right] \to 0 \qquad \text{as } a \to \infty.$$

[Received May 2002. Revised October 2003.]

REFERENCES

Akritas, M., and Arnold, S. (2000), "Asymptotics for ANOVA When the Number of Levels Is Large," *Journal of the American Statistical Association*, 95, 212–226.

Arnold, S. F. (1980), "Asymptotic Validity of F-Tests for the Ordinary Linear Model and the Multiple Correlation Model," *Journal of the American Statistical Association*, 75, 890–895.

——— (1981), *The Theory of Linear Models and Multivariate Analysis*, New York: Wiley.

Bathke, A. (2002), "ANOVA for a Large Number of Treatments," *Mathematical Methods of Statistics*, 11, 118–132.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: Wiley.

Boos, D. D., and Brownie, C. (1995), "ANOVA and Rank Tests When the Number of Treatments Is Large," *Statistics and Probability Letters*, 23, 183–191.

Chow, Y. S., and Teicher, H. (1997), *Probability Theory: Independence, Interchangeability, Martingales* (3rd ed.), New York: Springer-Verlag.

de Jong, P. (1987), "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75, 261–277.

Dette, H., and Munk, A. (1998), "Validation of Linear Regression Models," *The Annals of Statistics*, 26, 778–800.

Eubank, R. L., and Hart, J. D. (1992), "Testing Goodness-of-Fit in Regression via Order Selection Criteria," *The Annals of Statistics*, 20, 1412–1425.

Fan, J. (1996), "Test of Significance Based on Wavelet Thresholding and Neyman's Truncation," *Journal of the American Statistical Association*, 91, 674–688.

Härdle, W., and Mammen, E. (1993), "Comparing Nonparametric Versus Parametric Regression Fits," *The Annals of Statistics*, 21, 1926–1947.

Hart, J. D. (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, New York: Springer-Verlag.

Jiang, J. (1996), "REML Estimation: Asymptotic Behavior and Related Topics," *The Annals of Statistics*, 24, 255–286.

Krutchkoff, R. G. (1989), "Two-Way Fixed Effects Analysis of Variance When the Error Variances May Be Unequal," *Journal of Statistical Computation and Simulation*, 32, 177–183.

Müller, H. G. (1992), "Goodness of Fit Diagnostics for Regression Models," *Scandinavian Journal of Statistics*, 19, 157–172.

Portnoy, S. (1984), "Asymptotic Behavior of M-Estimators of p Regression Parameters When p^2/n Is Large. I. Consistency," *The Annals of Statistics*, 12, 1298–1309.

Scheffé, H. (1959), *The Analysis of Variance*, New York: Wiley.

Shorack, G. R., and Wellner, J. A. (1986), *Empirical Processes With Applications to Statistics*, New York: Wiley.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, New York; Cambridge, UK: Cambridge University Press.

Wang, H. (2004), "Rank Tests in Multifactor Heteroscedastic ANOVA and Repeated Measures Designs With Large Number of Levels," unpublished Ph.D. dissertation, The Pennsylvania State University.

Wang, L. (2003), "Testing in Heteroscedastic ANOVA With Large Number of Levels and in Nonparametric ANCOVA," unpublished Ph.D. dissertation, The Pennsylvania State University.