Simple t-distribution Based Tests for Meta-Analysis

J. Hartung[1]; **K.H. Makambi**[2]

[1,2]Department of Statistics, University of Dortmund,

D-44221 Dortmund, Germany

[2]Department of Mathematics and Statistics,

Jomo Kenyatta University of Agriculture and Technology

P.O. Box 62000, Nairobi, Kenya.

Key words: Fixed and random effects models; type I error rate; estimated degrees of freedom; balanced and unbalanced samples; homoscedastic and heteroscedastic error variances; replication

---

Summary: The variance function of the optimal estimator of the overall mean in a heteroscedastic one-way ANOVA model is dominated by positive semi-definite quadratic functions. This makes it possible to develop closely related tests on the nullity of the overall mean parameter, in one-way fixed and random effects ANOVA models, which make use of the quantiles of the t-distribution. These tests are founded on the convexity arguments similar to Hartung (1976). Simulation results indicate that the proposed tests attain type I error rates which are far more acceptable than those of the commonly used tests.

---

[1]**Correspondence Address:** Prof. Dr. Joachim Hartung, Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany.

Research partly supported by the Germany Academic Exchange Programme (DAAD) and the Germany Research Community (DFG).

1. Introduction

Combining results from different experiments (or studies) has become common in many fields of scientific inquiry. One has, for example, balanced or unbalanced, homoscedastic or heteroscedastic samples to assess the overall treatment effect. With treatment-by-centre interaction in such samples, we get a random effects model, otherwise we have a fixed effects model.

The possibility of many false positives in meta-analysis due to the underestimate of the variance of the estimate of the overall treatment effect cannot be overemphasized as indicated by Li et al. (1994) and Boeckenhoff/Hartung (1998). Suggested corrections for the fixed effects model with the resulting test statistics being normally distributed do not extend naturally to the random effects model.

By noting that the estimate of the overall treatment effect is dominated by a positive semi-definite quadratic form and estimating its distribution by a χ^2-distribution by equating its first two moments, we obtain tests of significance for the overall effect which are based on the t-distribution. Two related tests, cf. section 2, for the fixed effects model are suggested and one test, cf. section 3, for the random effects model. Accompanying simulation results, cf. Tables I and II, indicate that our suggested test statistics improve greatly the attained type I error rates.

2. Fixed Effects Model

For $K \geq 2$ independent experiments, let y_{ij} be the observation on the j-th subject of the i-th experiment, $i = 1, \ldots, K$ and $j = 1, \ldots, n_i$, such that

$$y_{ij} \quad = \quad \mu + e_{ij}, \; i = 1, \ldots, K, \; j = 1, \ldots, n_i, \qquad (1)$$

where μ is the common mean for all the K homogeneous experiments, e_{ij} are error terms which are assumed to be mutually stochastically independent and normally distributed, that is, $e_{ij} \sim N(0, \sigma_i^2);\ i = 1, \ldots, K,\ j = 1, \ldots, n_i$. The best estimate for μ in each study (experiment) is the individual sample mean $\hat{\mu}_i = \sum_{j=1}^{n_i} y_{ij}/n_i = \bar{y}_{i.}$ with variance $\sigma_i^2/n_i,\ i = 1, \ldots, K$. This means that we have a fixed effects combinations model such that $\hat{\mu}_i \sim N(\mu, \sigma_i^2/n_i),\ i = 1, \ldots, K$.

Our interest is in testing the hypothesis $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ at some type I error rate, α.

Now, the best unbiased estimator of μ which traces back to Cochran (1954) (see also Whitehead and Whitehead, 1991) is:

$$\tilde{\mu} = \frac{\sum_{i=1}^{K} \frac{n_i}{\sigma_i^2} \cdot \hat{\mu}_i}{\sum_{i=1}^{K} n_i/\sigma_i^2} \tag{2}$$

with variance $\sigma_{\tilde{\mu}}^2 = \left(\sum_{i=1}^{K} n_i/\sigma_i^2\right)^{-1}$. Under H_0 the statistic

$$T = \frac{\tilde{\mu}}{\sqrt{\sigma_{\tilde{\mu}}^2}} \sim N(0,1). \tag{3}$$

In most practical situations, however, the individual error variances are unknown and on estimating them by $\hat{\sigma}_i^2$, we obtain the estimate of the overall mean to be

$$\hat{\mu} = \frac{\sum_{i=1}^{K} \frac{n_i}{\hat{\sigma}_i^2} \cdot \hat{\mu}_i}{\sum_{i=1}^{K} n_i/\hat{\sigma}_i^2} \tag{4}$$

so that when $\mu = 0$, the test statistic

$$T_1 = \frac{\hat{\mu}}{\sqrt{\hat{\sigma}_{\tilde{\mu}}^2}} \overset{approx}{\sim} N(0,1) \tag{5}$$

In our experience (cf: also Li et al., 1994 and Boeckenhoff/Hartung, 1998) this test attains type I error rates which are much greater than the nominal level, α.

Consider now a positive discrete random variable d taking on realizations

$d_i = 1/x_i$ with probabilities ω_i, for $i = 1, \ldots, K$, and the convex function $g(d) = 1/d$, then Jensen's inequality

$$g(E(d)) = \frac{1}{\sum_{i=1}^K \omega_i \cdot d_i} \leq E(g(d)) = \sum_{i=1}^K \omega_i \cdot \frac{1}{d_i}$$

provides us with the well known inequality between the harmonic and arithmetic means.

Lemma 1:

For $x_i > 0$, $\omega_i \geq 0$, $i = 1, \ldots, K$, $\sum_{i=1}^K \omega_i = 1$, there holds

$$\bar{x}_{\omega,h} = \frac{1}{\sum_{i=1}^K \omega_i \cdot \frac{1}{x_i}} \leq \sum_{i=1}^K \omega_i \cdot x_i = \bar{x}_{\omega,a}.$$

Next, let

$$f_{\hat{\mu},h}(s^2) = \hat{\sigma}^2_{\tilde{\mu}} = \frac{1}{\sum_{i=1}^K n_i/s_i^2} = \frac{1}{N} \cdot \frac{1}{\sum_{i=1}^K \frac{n_i/N}{s_i^2}}, \tag{6}$$

where $s_i^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2/(n_i - 1)$ is an unbiased estimate of σ_i^2 from the i-th experiment. Using Lemma 1 above and setting $\omega_i = n_i/N$ we get

$$f_{\hat{\mu},h}(s^2) = \frac{1}{N} \cdot \frac{1}{\sum_{i=1}^K \omega_i/s_i^2} \leq \frac{1}{N} \cdot \sum_{i=1}^K \omega_i \, s_i^2 =: f_{\hat{\mu},a}(s^2) \tag{7}$$

with $x_i = s_i^2$. Clearly $f_{\hat{\mu},a}(s^2)$ is a positive semi-definite quadratic form in the random variables, which dominates the function $f_{\hat{\mu},h}(s^2)$. Thus, the approximate distribution of $f_{\hat{\mu},h}(s^2)$ can be obtained as follows:

Let

$$Q(f_{\hat{\mu},h}) = \nu \cdot \frac{1}{E f_{\hat{\mu},h}(s^2)} \cdot f_{\hat{\mu},h}(s^2),$$

then $Q(f_{\hat{\mu},h}) \overset{approx}{\sim} \chi^2_\nu$, where according to Patnaik (1949)

$$\nu = 2 \cdot \frac{(E f_{\hat{\mu},h}(s^2))^2}{Var f_{\hat{\mu},h}(s^2)}$$

By convexity arguments of Hartung (1976, sec. 1), cf: also Boeckenhoff/Hartung (1998), we have

$$E(f_{\hat{\mu},h}(s^2)) \leq \sigma^2_{\tilde{\mu}},$$

and for the variance $Varf_{\hat{\mu},h}(s^2)$, we have the following upper estimates:

$$Varf_{\hat{\mu},h}(s^2) \leq E\left(\left(\sum_{i=1}^{K}\frac{n_i}{s_i^2}\right)^{-2} - \left(\sum_{i=1}^{K}\frac{\sqrt{n_i^2-1}}{n_i-3}\cdot\frac{n_i}{s_i^2}\right)^{-2}\right) = E(\hat{V}_1) \quad (8)$$

$$Varf_{\hat{\mu},h}(s^2) \leq \left(\sum_{i=1}^{K}\sqrt{\frac{n_i-1}{n_i+1}}\cdot\frac{n_i}{\sigma_i^2}\right)^{-2} - \left(\sum_{i=1}^{K}\frac{n_i-1}{n_i-3}\cdot\frac{n_i}{\sigma_i^2}\right)^{-2} = V_2 \quad (9)$$

For the estimated degrees of freedom, ν, we will make use of $\hat{V}_j,\ j=1,2$, as given in (8) and (9) above with the parameters $\sigma_i^2,\ i=1,\ldots,K$, in V_2 replaced by their estimators to obtain $\hat{V}_2$. That is,

$$\hat{\nu}_j = 2\cdot\frac{\left(f_{\hat{\mu},h}(s^2)\right)^2}{\hat{V}_j},\ j=1,2.$$

In the following, however, we propose to introduce a "compensation factor" to the numerator of $\nu_j,\ j=1,2$, to avoid adverse underestimation. Let this factor be given by $\delta_j = \kappa\cdot\sqrt{\hat{V}_j},\ j=1,2,\ \kappa>0$. Thus we have the modified operational $\nu_j,\ j=1,2$, given by

$$\hat{\nu}_j(\kappa) = 2\cdot\frac{\left(f_{\hat{\mu},h}(s^2)+\delta_j\right)^2}{\hat{V}_j},\ j=1,2.$$

So, we can summarise the above considerations to formulate the following theorem.

Theorem 1: The test statistics $T_1^{(t)}$, $t=1,2$, under $H_0:\mu=0$, are such that:
a)

$$T_1^{(1)} = \frac{\hat{\mu}}{\sqrt{f_{\hat{\mu},h}(s^2)}} \overset{approx}{\sim} t_{\hat{\nu}_1(\kappa)}$$

b)

$$T_1^{(2)} = \frac{\hat{\mu}}{\sqrt{f_{\hat{\mu},h}(s^2)}} \overset{approx}{\sim} t_{\hat{\nu}_2(\kappa)}$$

Note that $T_1^{(1)}$ and $T_1^{(2)}$ differ only in the associated estimated degrees of freedom.

Using $T_1^{(1)}$ and $T_1^{(2)}$ with $\kappa=0.5$ we now demonstrate through a simulation study that the two proposed tests attains type I error rates which are closer to the nominal level than the commonly used test T_1 which attains levels well above the ideal level, α, especially for small sample sizes. For comparison, we have also considered in our simulations $T_1^* = \hat{\mu}/(\sum_{i=1}^{K} n_i/\sigma_i^2)^{-1/2}$ with the true σ_i^2 in the variance term of T_1, and the critical values are taken from the standard normal distribution, as for T_1.

Table I: Actual type I error rates (10 000 runs) for K=3 and K=6 at significance level $\alpha = 5\%$ using test statistics T_1^*, T_1, $T_1^{(1)}$ and $T_1^{(2)}$ for the fixed effects model.

nominal level,α=5%		Attained type I error rates, $\hat{\alpha}$%							
Sample sizes	Error variances	K=3				K=6 (1 Replication of K=3)			
(n_1, n_2, n_3)	$(\sigma_1^2, \sigma_2^2, \sigma_3^2)$	T_1^*	T_1	$T_1^{(1)}$	$T_1^{(2)}$	T_1^*	T_1	$T_1^{(1)}$	$T_1^{(2)}$
(5,5,5)	(1,3,5)	9.2	18.2	8.0	10.1	11.7	23.4	10.8	13.6
	(4,4,4)	8.3	18.6	8.2	10.5	11.4	23.6	10.9	13.7
(10,10,10)	(1,3,5)	6.6	10.0	4.9	5.4	7.0	11.0	5.4	6.0
	(4,4,4)	6.9	10.8	5.4	6.0	7.3	11.7	5.9	6.5
(20,20,20)	(1,3,5)	5.7	7.0	4.4	4.5	6.0	7.5	4.7	4.9
	(4,4,4)	5.9	7.2	4.5	4.8	6.0	7.5	4.6	4.8
(5,10,15)	(1,3,5)	7.3	13.3	5.9	6.9	9.5	16.8	7.6	9.0
	(4,4,4)	8.0	13.1	6.4	7.2	8.8	13.4	6.8	7.6
	(5,3,1)	7.2	10.1	5.6	6.0	8.4	12.3	6.3	6.8
(10,20,30)	(1,3,5)	6.5	9.3	4.8	5.2	6.5	9.4	5.0	5.4
	(4,4,4)	6.2	7.6	4.8	5.0	6.2	8.1	4.8	5.0
	(5,3,1)	5.9	6.9	4.7	4.8	6.0	7.2	4.9	5.0

We consider first K=3 with various constellations of sample sizes and error variances (see Table I below). In order to see the effect of increasing the number of experiments with all the other factors held constant, we make one independent replication of all the constellations of K=3 to obtain K=6. The results given are for testing $H_0 : \mu = 0$ against a two-sided alternative $H_1 : \mu \neq 0$.

We notice that the attained type I error rates in column 4 and 8 of Table I are far much greater than the nominal level of 5.0 percent . For small sample sizes, this liberality of T_1 is relatively higher for balanced samples and increases with the number of experiments (studies), that is, the attained levels are greater for K=6 than for K=3. The proposed tests, $T_1^{(1)}$ and $T_1^{(2)}$, improve the attained levels appreciably, despite showing some increase in the levels attained with increase in the number of studies.

For balanced samples greater than 10, the proposed tests attain reasonable stability with respect to increase in the number of experiments. This is also conspicuous for unbalanced samples in cases where the smallest sample size is equal to 10.

3 Random Effects Model

For the one-way random effects model we add a random effect $a_i \sim N(0, \sigma_a^2)$, $i = 1, \ldots, K$, to model (1), see section 2 above, to obtain

$$y_{ij} = \mu + a_i + e_{ij},\ i = 1, \ldots, K,\ j = 1, \ldots, n_i,$$

with $a_1, \ldots, a_K, e_{11}, \ldots, e_{Kn_K}$ being mutually stochastically independent, so that $\hat{\mu}_i \sim N(\mu, \sigma_a^2 + \sigma_i^2/n_i)$. Then the estimator of μ equivalent to (4) is given by

$$\hat{\mu} = \frac{\sum_{i=1}^{K} \frac{1}{v_i} \cdot \hat{\mu}_i}{\sum_{i=1}^{K} 1/v_i}, \tag{10}$$

where $v_i = \hat{\sigma}_a^2 + \hat{\sigma}_i^2/n_i = \hat{\sigma}_a^2 + \xi_i,\ i = 1, \ldots, K$. Therefore, we have the commonly used test statistic

$$T_{1(r)} = \frac{\hat{\mu}}{(\sum_{i=1}^{K} 1/v_i)^{-1/2}} \overset{approx}{\sim} N(0,1) \tag{11}$$

This test suffers from the same weaknesses as its fixed effects counterpart, with the situation here being compounded by the estimation of the variance of the random effect, σ_a^2.

Let $\tau_i^2 = \sigma_a^2 + \sigma_i^2/n_i$, and define the quadratic form $Q = \sum_{i=1}^K h_i(\hat{\mu}_i - \sum_{j=1}^K b_j\,\hat{\mu}_j)^2$, where $h_i > 0$ and $b_i > 0$ with $\sum_{i=1}^K b_i = 1,\ i = 1, \ldots, K$. By a somewhat lengthy derivation, it can be shown that, Hartung (1999), (cf: also, e.g., Hartung, 1981, Mathai/Provost, 1992):

$$E(Q) = \sum_{i=1}^K h_i(1-2b_i)\tau_i^2 + (\sum_{i=1}^K h_i)(\sum_{i=1}^K b_i^2\tau_i^2), \tag{12}$$

$$Var(Q) = 2 \cdot \left(\sum_{i=1}^K h_i^2 D_i^2 + \sum_{i=1}^K \sum_{i \neq j=1}^K h_i h_j C_{ij}^2 \right), \tag{13}$$

where

$$D_i = (1-2b_i)\tau_i^2 + \sum_{k=1}^K b_k^2\tau_k^2, \tag{14}$$

$$C_{ij} = \sum_{k=1}^K b_k^2\tau_k^2 - b_i\tau_i^2 - b_j\tau_j^2,\ i,j = 1, \ldots, K, \tag{15}$$

which are also estimated by replacing parameters by their estimates, yielding with special choices of

$$b_i = \frac{n_i/\sigma_i^2}{\sum_{i=1}^K n_i/\sigma_i^2}, \quad h_i = \frac{b_i}{1 - \sum_{i=1}^K b_i^2}$$

the Cochran (1954) estimator (cf: also DerSimonian/Laird, 1986; Whitehead/Whitehead, 1991)

$$\tilde{\sigma}_a^2 = \sum_{i=1}^K h_i(\hat{\mu}_i - \sum_{j=1}^K b_j\,\hat{\mu}_j)^2 - \sum_{i=1}^K r_i\,\xi_i, \tag{16}$$

with $r_i = (b_i - b_i^2)/(1 - \sum_{i=1}^K b_i^2),\ i = 1, \ldots, K$, which is an unbiased estimator of σ_a^2, and we get for its variance

$$Var(\tilde{\sigma}_a^2) = Var(Q) + \sum_{i=1}^K r_i^2 \cdot Var(\xi_i). \tag{17}$$

Also $Var(\xi_i) = 2 \cdot \sigma_i^4 / n_i^2(n_i - 1)$ and its best invariant unbiased estimator is given by $\widehat{Var}(\xi_i) = 2 \cdot \xi_i^2/(n_i + 1)$, Hartung/Voet (1986). Note that $\tilde{\sigma}_a^2$ has a positive probability of taking negative values. For a realization the parameter σ_i^2/n_i in b_i is replaced by ξ_i so that $\tilde{\sigma}_a^2$ becomes the estimator $\hat{\sigma}_a^2$.
Making use now Lemma 1 again, we have

$$\frac{1}{\sum_{i=1}^K 1/v_i} \leq \frac{1}{K} \cdot \sum_{i=1}^{K} \frac{1}{K} \cdot v_i = \frac{1}{K^2} \sum_{i=1}^{K} (\hat{\sigma}_a^2 + \xi_i), \tag{18}$$

and therefore,

$$\frac{1}{\sum_{i=1}^K 1/v_i} = \Delta \cdot (\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^{K} \xi_i),$$

where Δ is a positive random variable. Next,

$$\begin{aligned} \nu_r \cdot \left(E \left(\frac{1}{\sum_{i=1}^K 1/v_i} \right) \right)^{-1} \cdot \frac{1}{\sum_{i=1}^K 1/v_i} &= \nu_r \cdot \frac{\Delta \cdot (\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^K \xi_i)}{E\left(\Delta \cdot (\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^K \xi_i)\right)} \\ &\approx \nu_r \cdot \frac{(\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^K \xi_i)}{E(\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^K \xi_i))} \overset{approx}{\sim} \chi^2_{\nu_r}, \end{aligned}$$

where, if $\hat{\sigma}_a^2 > 0$ and by the independence of Q and $\xi_i,\ i = 1, \ldots, K$, ν_r is given by

$$\begin{aligned} \nu_r &= 2 \cdot \frac{\left(E\left(\Delta \cdot (\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^K \xi_i)\right)\right)^2}{Var\left(\Delta \cdot (\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^K \xi_i)\right)} \\ &\approx 2 \cdot \frac{\left(E(\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^K \xi_i)\right)^2}{Var\left(\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^K \xi_i\right)} \\ &= 2 \cdot \frac{\left(E(\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^K \xi_i)\right)^2}{Var(Q) + \frac{2}{K^2} \sum_{i=1}^K (r_i K - 1)^2 \cdot \sigma_i^4 / n_i^2}, \end{aligned}$$

where ν_r is estimated by

$$\hat{\nu}_r = 2 \cdot \frac{\left(\hat{\sigma}_a^2 + \frac{1}{K} \sum_{i=1}^K \xi_i)\right)^2}{\widehat{Var}(Q) + \frac{2}{K^2} \sum_{i=1}^K (\hat{r}_i K - 1)^2 \cdot \frac{\xi_i^2}{n_i + 1}}. \tag{19}$$

If $\hat{\sigma}_a^2 \leq 0$, then

$$\hat{\nu}_r = \frac{\left(\sum_{i=1}^K \xi_i\right)^2}{\sum_{i=1}^K \frac{\xi_i^2}{n_i+1}}. \tag{20}$$

So, for testing the hypothesis $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$, we can summarise the considerations above in the following theorem:

Theorem 2: Under H_0 there is

$$T_{1(r)}(\hat{\nu}_r) = \frac{\hat{\mu}}{(\sum_{i=1}^{K} 1/v_i)^{-1/2}} \tag{21}$$

distributed approximately as a central t-variable with $\hat{\nu}_r$ degrees of freedom, where $\hat{\nu}_r$ is given in (19) for $\hat{\sigma}_a^2 > 0$ (cf: equation (16)) and by (20) in the case when $\hat{\sigma}_a^2 \leq 0$.

Now the various test statistics are compared in a simulation study, cf: Table II. The values reported there under T_r^*, for K=3 and 6, are obtained by using the test statistic $T_r^* = \hat{\mu}/(\sum_{i=1}^{K} 1/\tau_i^2)^{-1/2}$ with the true values τ_i^2 in the variance term of $T_{1(r)}$ and the critical values are obtained from the standard normal distribution, as for T_1.

To obtain K=6 we independently replicated K=3 once , for $\sigma_a^2 = 0, 0.5, 5,$ 25. For $\sigma_a^2 = 0.0$, (see Table II), the proposed test $T_{1(r)}(\hat{\nu}_r)$ attains acceptable type I error rates, despite being a bit more liberal for K=6 and small sample sizes of 5 per experiment. Also for unbalanced samples, when relatively large individual error variances are paired with relatively small sample sizes, the test is conservative for K=3.

For values of σ_a^2 between 0.5 and 5, the proposed test attains levels far more acceptable than those of the commonly used statistic $T_{1(r)}$, save for some small traces of liberality especially for small sample size constellations.

For large values of σ_a^2, the attained type I error rates stabilize for all sample size and individual error variance combinations considered.

Table II: Actual type I error rates (10 000 runs) for K=3 and 6 at significance level $\alpha = 5\%$ using test statistics T_r^*, $T_{1(r)}$ and $T_{1(r)}(\hat{\nu}_r)$ for the random effects model.

Nominal level, α=5%			Attained type I error rates, $\hat{\alpha}$%					
	Sample sizes	Error variances		K=3		K=6 (1 Replication of K=3)		
σ_a^2	(n_1, n_2, n_3)	$(\sigma_1^2, \sigma_2^2, \sigma_3^2)$	T_r^*	$T_{1(r)}$	$T_{1(r)}(\hat{\nu}_r)$	T_r^*	$T_{1(r)}$	$T_{1(r)}(\hat{\nu}_r)$
0.0	(5,5,5)	(1,3,5)	8.1	8.8	5.8	9.4	9.7	7.5
		(4,4,4)	6.2	10.0	6.2	7.2	10.5	8.0
	(20,20,20)	(1,3,5)	7.5	5.0	3.8	6.7	5.0	4.2
		(4,4,4)	5.4	4.9	3.6	5.5	4.9	3.7
	(5,10,15)	(1,3,5)	5.9	7.8	5.6	6.7	8.0	6.4
		(4,4,4)	6.7	6.7	4.1	7.3	7.2	5.1
		(5,3,1)	11.5	5.2	2.9	10.8	5.3	3.7
	(10,20,30)	(1,3,5)	5.6	5.2	4.1	5.7	5.5	4.3
		(4,4,4)	6.7	5.2	3.7	6.0	4.9	3.7
		(5,3,1)	10.1	4.0	2.9	8.5	4.5	3.5
0.5	(5,5,5)	(1,3,5)	6.9	16.9	10.6	6.1	12.4	9.0
		(4,4,4)	6.2	13.5	8.1	6.2	11.1	7.9
	(20,20,20)	(1,3,5)	5.8	18.4	10.0	5.1	11.5	6.6
		(4,4,4)	5.2	14.2	7.7	4.9	10.1	5.3
	(5,10,15)	(1,3,5)	5.3	14.2	8.1	5.2	10.7	6.5
		(4,4,4)	5.6	13.5	8.1	5.2	10.4	6.8
		(5,3,1)	6.4	20.0	13.3	5.8	13.3	9.4
	(10,20,30)	(1,3,5)	5.4	16.2	7.9	5.5	10.9	5.5
		(4,4,4)	4.9	14.6	8.3	5.1	10.8	6.4
		(5,3,1)	5.7	20.7	13.7	5.3	13.8	9.2
1.0	(5,5,5)	(1,3,5)	5.7	18.4	11.1	5.5	12.6	8.9
		(4,4,4)	5.5	15.2	8.5	5.4	11.6	7.6
	(20,20,20)	(1,3,5)	5.0	18.5	8.4	5.3	11.6	5.7
		(4,4,4)	5.2	16.3	7.1	5.4	11.1	5.7

Table II: Cont.

Nominal level, α=5%			Attained type I error rates, $\hat{\alpha}$%					
	Sample sizes	Error variances	K=3			K=6 (1 Replication of K=3)		
σ_a^2	(n_1, n_2, n_3)	$(\sigma_1^2, \sigma_2^2, \sigma_3^2)$	T_r^*	$T_{1(r)}$	$T_{1(r)}(\hat{\nu}_r)$	T_r^*	$T_{1(r)}$	$T_{1(r)}(\hat{\nu}_r)$
1.0	(5,10,15)	(1,3,5)	5.3	16.7	8.4	5.0	10.9	6.0
		(4,4,4)	5.7	15.8	8.4	5.5	10.9	6.7
		(5,3,1)	5.4	21.0	13.4	5.6	13.5	9.2
	(10,20,30)	(1,3,5)	5.1	17.3	7.0	4.6	11.1	5.3
		(4,4,4)	4.9	16.9	8.3	5.0	10.9	6.0
		(5,3,1)	4.9	21.2	12.3	5.2	13.3	7.5
5.0	(5,5,5)	(1,3,5)	5.1	20.6	9.3	5.1	12.2	5.5
		(4,4,4)	5.4	18.2	7.7	5.1	11.6	5.7
	(20,20,20)	(1,3,5)	5.3	20.3	5.8	5.1	12.2	4.2
		(4,4,4)	5.0	19.2	6.1	4.6	11.0	4.9
	(5,10,15)	(1,3,5)	5.0	19.4	6.4	4.8	11.4	4.8
		(4,4,4)	4.9	19.5	7.1	4.8	10.9	4.9
		(5,3,1)	5.0	20.6	9.1	5.2	13.0	5.4
	(10,20,30)	(1,3,5)	5.3	19.1	5.7	4.8	11.3	4.8
		(4,4,4)	5.2	19.2	6.3	4.4	10.6	4.6
		(5,3,1)	4.8	20.4	7.2	5.0	13.5	4.7
25	(5,5,5)	(1,3,5)	4.7	19.3	5.4	5.1	13.2	4.2
		(4,4,4)	4.9	20.0	5.9	5.0	11.4	4.2
	(20,20,20)	(1,3,5)	4.8	19.5	4.5	4.7	12.0	4.0
		(4,4,4)	4.7	19.5	4.8	5.3	10.9	5.0
	(5,10,15)	(1,3,5)	4.7	18.9	4.4	5.0	11.8	4.4
		(4,4,4)	4.8	19.4	5.3	5.0	11.7	4.3
		(5,3,1)	5.1	20.8	5.7	4.8	13.6	4.1
	(10,20,30)	(1,3,5)	4.8	18.6	4.8	4.7	11.2	4.4
		(4,4,4)	5.1	20.0	4.7	4.9	11.6	5.0
		(5,3,1)	5.0	20.2	4.7	4.7	13.4	4.4

4. Conclusion

The problem of frequent liberal decisions is very common in meta-analysis. With our proposed tests, we see a great improvement in the attained type I error rates for both the fixed and random effects ANOVA models. We would recommend the use of these tests in place of the commonly used method to minimise the danger of registering too many significant results.

References

Boeckenhoff, A.; Hartung, J. (1998): Some corrections of the significance level in meta-analysis. *Biometrical Journal* 40, vol. 8, 937-947.

Cochran, W.G. (1954): The combination of estimates from different experiments. *Biometrics,* 10, 101-129.

DerSimonian, R.; Laird, N.M. (1986): Meta-analysis in clinical trials. *Controlled Clinical Trials* 7, 177-188.

Hartung, J. (1976): Some inequalities for a random function. *Teorija Verojatnostei i ee Primenenja* 21, 661-665 (also in *Theory of Probability and its applications,* SIAM 21,1977).

Hartung, J. (1981): Nonnegative Minimum Biased Invariant Estimation in Variance Component Models. *The Annals of Statistics,* 9, 278-292.

Hartung, J. (1999): *Variance Components Analysis.* Lectures Notes, Dortmund University, Dortmund (Unpublished).

Hartung, J.; Voet, B. (1986): Best invariant unbiased estimators for the mean squared error of variance component estimators. *Journal of the American Statistical Association* 81, 689-691.

Li, Y; Shi, L.; Roth, H.D. (1994): The bias of the commonly used estimate of variance in meta-analysis. *Communications in Statistics - Theory and Meth-*

ods 23, 1063-1085.

Mathai, A.M.; Provost, S.B. (1992): *Quadratic forms in random variables.* Marcel Decker, New York.

Patnaik, P. B. (1949): The non-central χ^2- and F-distributions and their applications. *Biometrika* 36, 202-232.

Whitehead, A.; Whitehead, J. (1991): A general parametric approach to meta-analysis of randomised clinical trials. *Statist. Med.* 10, 1665-1677.