Center for Advanced Studies in Measurement and Assessment

CASMA Research Report

Number 5

Bootstrap Procedures for Estimating Standard Errors of Estimated Variance Components for Two-Facet Designs*

Ye Tong and Robert L. Brennan†

February 2004

*A revised version of a paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, April, 2003.

†Ye Tong is Research Assistant, Iowa Testing Programs (ITP), College of Education, University of Iowa, Iowa City, IA 52242 (email:ye-tong@uiowa.edu). Robert L. Brennan is E. F. Lindquist Chair in Measurement and Testing and Director, Center for Advanced Studies in Measurement and Assessment (CASMA), 297 Lindquist North, College of Education, University of Iowa, Iowa City, IA 52242 (email: robert-brennan@uiowa.edu).

Center for Advanced Studies in
Measurement and Assessment (CASMA)
College of Education
University of Iowa
Iowa City, IA 52242
Tel: 319-335-5439
Fax: 319-384-0505
Web: www.uiowa.edu/~casma

Contents

Abstract

The estimation of standard errors of estimated variance components has long been a challenging task in generalizability theory. For two decades researchers have speculated about the potential applicability of the bootstrap for obtaining such estimates, but the same researchers have identified problems in using the bootstrap. Some of these problems are associated with the fact that bootstrap procedures generally produce biased estimates of variance components and their standard errors. Other problems are associated with the fact that, with ANOVA-type designs, there are many possibilities for obtaining bootstrap samples, and it has not been demonstrated that there is a single optimal bootstrap procedure for obtaining the "best" estimates for all variance components and standard errors.

Using normal-data simulations, this paper suggests solutions to both of these problems for the $p \times i \times h$ and $p \times (i : h)$ designs. Specifically, this paper proposes explicit bias-correction formulas and suggests rules for choosing a particular bootstrap procedure for particular variance components. The recommendations proposed are tentative, but they are largely confirmed by the simulations. Further investigations need to be conducted to extend these recommendations to dichotomous and polytomous data, and to more complicated designs.

1 Introduction

Estimated variance components in generalizability theory, like other statistics, are subject to sampling errors. Many procedures for estimating the standard errors involved in variance component estimation have been proposed over the years. The nonparametric bootstrap procedure is one such strategy. However, G studies involving two or more facets complicate the application of the bootstrap, since resampling can occur for any single facet or for any combination of facets. It has been shown in some empirical studies (mainly for the $p \times i$ design) that different bootstrap procedures tend to yield different results, both for estimated variance components and for the related standard errors (Brennan, Harris, & Hanson, 1987; Wiley, 2000; Brennan 2001). No specific bootstrap strategy has demonstrated superiority in accurately estimating the sampling variability of *all* estimated variance components.

Previous research has shown that different bootstrap sampling strategies employed in the $p \times i$ context provide widely divergent point and interval estimates for variance components and their standard errors (Brennan et al., 1987; Luecht & Smith, 1988; Othman, 1995; Wiley, 2000). Furthermore, these papers allude to sample size and disproportionality between the numbers of persons and items as possible factors contributing to the differences across bootstrap strategies. Wiley (2000) provided algorithms explaining why the discrepancies exist and how to correct them. He concluded that the variance component estimates provided by various bootstrap strategies differ in a systematic manner, and he proposed bias correction factors for each variance component estimate under each of the possible bootstrap sampling procedures for the $p \times i$ design. Refer to Appendix A for the complete set of bias correction factors for the $p \times i$ design and some bias correction factors for the $p \times i \times h$ design that Wiley (2000) proposed. But, so far, only the $p \times i$ design has been investigated rather intensively.

Even with the simple $p \times i$ design, however, the bias correction factors do not work equally well with all bootstrap procedures. In particular, previous research suggests that resampling along the dimension(s) represented by the component of interest [e.g., resampling persons to estimate the standard error of $\hat{\sigma}^2(p)$] tends to give the most accurate estimates of standard errors.

The current study suggests and examines bias correction factors and general rules for applying bootstrap procedures with the $p \times i \times h$ design (e.g., persons crossed with items crossed with occasions) and the $p \times (i\!:\!h)$design (e.g., persons crossed with items nested within occasions). In the terminology of generalizability theory, these are called two-facet designs, because the object of measurement (usually persons) is not called a "facet" per se. Here, however, to keep terminology simple, we will sometimes refer to the objects of measurement as a facet.

Specifically, for the $p \times i \times h$ and $p \times (i\!:\!h)$ designs, this paper proposes:

- complete sets of bias-correction factors for the estimated variance components, which extend results reported by Wiley (2000); and

- rules for determining which bootstrap sampling procedure to use—rules that "formalize" some speculative claims in Brennan (2001) and Wiley (2000).

The proposed rules for applying bootstrap sampling procedures, in conjunction with the bias-correction factors are:

Rule 1 : For main effects, to estimate variance components and their standard errors, use the corresponding bootstrap procedure. For example, to estimate $\sigma^2(p)$, use boot-p.

Rule 2 : For interaction effects, to estimate variance components and their standard errors, use a one-dimensional bootstrap procedure that relates to the interaction term and that has the largest sample size. For example, to estimate $\sigma^2(ih)$, select boot-i if $n_i > n_h$ or select boot-h if $n_h > n_i$.

Rule 3 : All other things being equal, resampling along a dimension involving relatively large sample sizes tends to provide more accurate results. For example, if n_p is very large, boot-p may work about as well, or sometimes better, than applying Rules 1 and 2.

It should be noted that the proposed bias-correction factors and rules are *not* formally proved in a strict sense, but they are demonstrated and supported by normal-data simulations.

2 Methodology

The designs being investigated in this paper are the random effects $p \times i \times h$ and $p \times (i\!:\!h)$designs. By definition, for these random effects designs, p, i, and h are all regarded as random facets. In this section, the simulation procedure used to obtain data is first described, followed by a discussion of how bootstrap strategies are employed for these designs. Then bias-correction factors for both designs are proposed.

2.1 Data

The sample sizes and parameters used for the simulations are listed in Table 1. The parameter values were chosen to be reasonably representative of real-world applications.

The score effects are assumed to be normally distributed here. Therefore, for the $p \times i \times h$ design, the following formula can be used to generate observed scores:

$$\begin{aligned} X_{pih} &= \mu + \sigma(p)z_p + \sigma(i)z_i + \sigma(h)z_h \\ &\quad + \sigma(pi)z_{pi} + \sigma(ph)z_{ph} + \sigma(ih)z_{ih} + \sigma(pih)z_{pih}, \end{aligned} \tag{1}$$

where z_p, z_i, z_h, z_{pi}, z_{ph}, z_{ih} and z_{pih} are randomly and independently sampled from the standard normal distribution. The parameter μ is not of concern here, since our focus is on variance components and their standard errors. Similarly, to generate observed scores for the $p \times (i:h)$ design, the following formula can be used:

$$X_{pi:h} = \mu + \sigma(p)z_p + \sigma(h)z_h + \sigma(ph)z_{ph} + \sigma(i{:}h)z_{i:h} + \sigma(pi{:}h)z_{pi:h}, \quad (2)$$

where the z scores are similarly defined as in Equation 1.

Based on the random-effects assumption, the actual standard errors of the estimated variance components $[\hat{\sigma}^2(\alpha)]$ can be obtained using the following formula (see Searle, 1971) referenced in Brennan (2001):

$$\sigma(\hat{\sigma}^2(\alpha)) = \frac{1}{\pi(\dot{\alpha})}\sqrt{\sum_{\beta}\frac{2[\mathbf{E}MS(\beta)]^2}{df(\beta)}}, \quad (3)$$

where $\pi(\dot{\alpha})$ is the product of the sample sizes for indices not in α, and β indexes the expected mean squares that enter $\sigma^2(\alpha)$. For example, under normality, the standard error of $\hat{\sigma}^2(i:h)$ for the nested $p \times (i:h)$ design is:

$$\begin{aligned}\sigma(\hat{\sigma}^2(i{:}h)) &= \frac{1}{n_p}\sqrt{\frac{2[\mathbf{E}MS(i{:}h)]^2}{n_h(n_i-1)} + \frac{2[\mathbf{E}MS(pi{:}h)]^2}{n_h(n_p-1)(n_i-1)}} \\ &= \frac{1}{100}\sqrt{\frac{2(908)^2}{2(20-1)} + \frac{2(208)^2}{2(100-1)(20-1)}} \\ &= 2.0309.\end{aligned}$$

It follows that all the parameters for variance components and their standard errors for the two designs being studied in the paper are known a priori.

2.2 Bootstrap Procedures

With respect to the $p \times i \times h$ design, the bootstrap can be applied using any one of the following seven procedures: boot-p, boot-i, boot-h, boot-p,i, boot-p,h, boot-i,h and boot-p,i,h. For example, for the boot-p procedure, $n_p = 100$ persons are sampled with replacement, and the variance components based on the bootstrap sample are estimated. For the boot-p,i procedure, an independent sample of $n_p = 100$ persons and an independent sample of $n_i = 20$ items (both sampled with replacement) are drawn and the sampled persons are matched with their responses to the sampled items.

For the $p \times (i:h)$ design, five possible bootstrap procedures are treated in this paper: boot-p, boot-h, boot-i,h, boot-p,h and boot-p,i,h. Boot-i,h and boot-p,i,h for the nested design are defined differently from those for the crossed design. Because the i facet is nested within the h facet, resampling of the i facet can only occur after resampling of the h facet. To be more specific, to conduct the boot-i,h procedure for the nested design, an independent sample of size $n_h = 2$ is drawn (with replacement) from the h facet; then, within each sampled level of h, a sample of size $n_i = 20$ is drawn (with replacement) from the i facet.

2.3 Bias Correction Factors

Brennan et al. (1987) demonstrated and Wiley (2000) proved that for the $p \times i$ design, the estimated variance components for bootstrap sampling procedures are usually biased. Wiley went on to suggest bias adjustments for each bootstrap procedure that might be employed with the $p \times i$ design. Tentative extensions for the $p \times i \times h$ and $p \times (i\!:\!h)$ designs are proposed in the current paper. Let

- α be the set of indices for the effect under consideration;
- K be the set of indices in the bootstrap procedure, with any one of them designated k_j; and
- c_i be a main effect index in α that is also in K.[1]

Then we speculate that the following formulas for bias correction can be used for the $p \times i \times h$ and the $p \times (i\!:\!h)$ designs:

- If K and α contain common facet(s):

$$\begin{aligned}
\hat{\sigma}^2(\alpha) &= \frac{\prod_{c_i} n_{c_i}}{\prod_{c_i} df(c_i)} \hat{\sigma}^2(\alpha|\text{boot}-K) \\
&\quad - \sum_{k_j \notin \alpha} \frac{\prod_{c_i} n_{c_i}}{df(k_j) \prod_{c_i} df(c_i)} \hat{\sigma}^2(\alpha k_j|\text{boot}-K) \\
&\quad + \sum_{\substack{k_j \notin \alpha, k_m \notin \alpha \\ k_j \neq k_m}} \frac{n_{c_i}}{df(k_m) df(k_j) df(c_i)} \hat{\sigma}^2(\alpha k_j k_m|\text{boot}-K). \qquad (4)
\end{aligned}$$

- If K and α do not contain any common facet:

$$\begin{aligned}
\hat{\sigma}^2(\alpha) &= \hat{\sigma}^2(\alpha|\text{boot}-K) \\
&\quad - \sum_{k_j} \frac{1}{df(k_j)} \hat{\sigma}^2(\alpha k_j|\text{boot}-K) \\
&\quad + \sum_{k_j \neq k_m} \frac{1}{df(k_m) df(k_j)} \hat{\sigma}^2(\alpha k_j k_m|\text{boot}-K). \qquad (5)
\end{aligned}$$

Note that, in Equation 4, if there are no c_i indices, then terms involving them are set to 1.

[1] In the terminology of Brennan, 2001, a main effect index is an index that appears before any colon in the designation of the effect. For example, for the effect $i\!:\!h$, i is a main effect index.

For example, to estimate $\hat{\sigma}^2(p)$ in the crossed design using the boot-i, h procedure, α is p, and K consists of i and h. Since α and K do not contain any common facet, Equation 5 can be used to obtain:

$$\begin{aligned}\hat{\sigma}^2(p) &= \hat{\sigma}^2(p|\text{boot}{-}i,h) - \frac{1}{n_i-1}\hat{\sigma}^2(pi|\text{boot}{-}i,h) \\ &\quad -\frac{1}{n_h-1}\hat{\sigma}^2(ph)|\text{boot}{-}i,h) + \frac{1}{(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}i,h).\end{aligned}$$

To estimate $\hat{\sigma}^2(i{:}\,h)$ for the nested design using the boot-p, h procedure, α contains i and h, K contains p and h, and there are no c_i indices. Since there is a common facet shared by α and K, namely h, Equation 4 can be used to obtain the bias-corrected estimate:

$$\hat{\sigma}^2(i{:}h) = \hat{\sigma}^2(i{:}h|\text{boot}{-}p,h) - \frac{1}{n_p-1}\hat{\sigma}^2(pi{:}h|\text{boot}{-}p,h).$$

The complete set of bias correction factors for the estimates of all the variance components for different bootstrap procedures is listed in Appendix B for the crossed design and Appendix C for the nested design.

3 Simulation Results

The simulations conducted in the paper involved 1000 trials. That is, 1000 arrays of size $n_p \times n_i \times n_h$ ($100 \times 20 \times 2$) were generated. For each bootstrap procedure, 100 bootstrap samples (replications) were drawn within *each* of the 1000 trials.[2] Within each trial, for each effect, the mean and variance, over the 100 replications, of the estimated variance components (both unadjusted and adjusted) were obtained. For each effect, the *averaged over trials* of the mean estimated variance components (both unadjusted and adjusted) was taken as the estimated variance component for the simulation; similarly, the square root of the variances of the estimated variance components *average over trials* (both unadjusted and adjusted) was taken as the estimated standard error of the estimated variance component for the simulation. The C programming language was used to conduct the simulation and the analysis.

3.1 The Crossed $p \times i \times h$ Design

Table 2 lists the variance component estimates for the $p \times i \times h$ design. The first row provides the parameter values for each of the seven variance components in the $p \times i \times h$ design. The second row provides the empirical values obtained (without bootstrapping) for the estimated variance components averaged across the 1000 trials. Rows 3–9 list the raw (i.e., unadjusted) estimates for the variance components based on each of the seven possible bootstrap procedures. Each of

[2] That is, the results for each of the bootstrap procedures for both designs were based on 100,000 bootstrap analyses.

the estimates in the table were based on 100,000 bootstrap calculations. But, as can be seen from Table 2, some of the estimates deviated quite a bit from the parameter values. For example, $\hat{\sigma}^2(pih)$ obtained from boot-p, i, h with a parameter value of 144 has an estimate of 67.15. The results in Table 2 empirically demonstrate that the raw results from bootstrap procedures are generally biased.

Table 3 provides the standard deviations of the unadjusted bootstrap estimates for the various bootstrap procedures.—i.e., estimated standard errors for the unadjusted estimated variance components. The values in the first row, are the parameters obtained using Equation 3. The empirical values in the second row are the standard deviations of the 1000 estimates from the 1000 trials, without bootstrapping. For each of the bootstrap procedures, Rows 3–9 provide the estimated standard errors of the estimated variance components.[3] Some of them are fairly close to the corresponding parameter values, such as $\hat{\sigma}(\sigma^2(p))$ for boot-p and $\hat{\sigma}(\sigma^2(i))$ for boot-i. However, many of them are not close to the parameters, especially $\hat{\sigma}(\sigma^2(pih))$ for the boot-p, i, h procedure.

Tables 4 and 5 provide the bias-corrected versions of the results reported in Tables 2 and 3, respectively. That is, for every effect and bootstrap procedure, the bias correction factors (Equations 4 or 5) were applied for each replication of each trial. Table 4 suggests that the bias-corrected estimates of the variance appear reasonably close to their corresponding parameters, even for values that were fairly deviant prior to bias correction, such as $\hat{\sigma}^2(pih)$ for boot-i, h. However, with respect to the standard error estimates, the bias correction factors produced mixed results. Some of the estimates are closer to their parameter values after the bias adjustment, such as $\hat{\sigma}(\sigma^2(pi))$ for boot-i; some of the estimates became even further away from their parameter values, such as $\hat{\sigma}(\sigma^2(pih))$ for boot-h.

Using Rules 1 and 2 proposed in the first section, in conjunction with the bias-correction Equations 4 and 5, gave the results in Table 6. From this table, it can be seen that all the estimates, for both variance components and their related standard errors, are fairly close to the parameter values. It appears that the proposed strategy for choosing a bootstrap procedure and correcting for bias worked reasonably well for this simulation for the $p \times i \times h$ design.

3.2 The Nested $p \times (i\!:\!h)$ Design

Tables 7 and 8 provide the raw (i.e., unadjusted) variance component estimates and estimated standard errors, respectively, for the nested design simulation. The layout of the tables is similar to that for the crossed design. From Tables 7 and 8, it can be seen that some of the raw results are fairly different from those of the parameter values, especially for the variance component estimates. One difference with respect to the crossed design is that the variance component estimate for $\sigma^2(ph)$ was negative for the boot-h and boot-p, h procedures. The

[3]More correctly, for each effect and each bootstrap procedure, the reported value is the square root of the average over trials of the variance of the estimated variance components.

small sample size for the h facet, in conjunction with the fact that it has another facet nested within itself, likely affected the estimates.

Tables 9 and 10 provide the bias-corrected versions of Tables 7 and 8, using the bias correction formulas proposed in this paper. From Table 9, it can be seen that, in terms of estimating variance components, the bias corrections did not work as well for the nested design as they did for the crossed design. The negative values for the variance component estimates remained negative and were even further away from the parameters after bias adjustments. Also, the variance component estimates for $\sigma^2(p)$ based on boot-h and boot-p,h appear to deviate quite a bit from the parameter values, given the fact that each estimate was based on 100,000 bootstrap calculations. A similar statement applies to the estimation of $\sigma^2(ph)$ through boot-h and boot-p,h. The rest of the estimated variance components were more similar to the parameter values after bias adjustment. As can be seen from Table 10, after bias corrections, many of the estimates for standard errors were still quite far away from the parameter values.

Using Rule 1 and Rule 2, Table 11 was obtained for the nested design and the original simulation, using only the bootstrap procedures speculated to work best for estimating a certain variance component and its standard error. From this table, it can be seen that the strategy worked fairly well for this simulation. The largest discrepancies occurred for $\hat{\sigma}^2(h)$, $\hat{\sigma}(\hat{\sigma}^2(h))$ and $\hat{\sigma}(\hat{\sigma}^2(i\!:\!h))$. It is speculated that the small sample size for the h facet, as well as the fact that h has another facet nested within it, might have made the estimate unstable.

Since the small sample size for the h facet might have contributed to the less-than-optimal results noted above, two more simulations were conducted, with the same parameter values for the variance components, the same number of trials and number of replications, but slightly different sample sizes for the h facet. The first simulation involved $n_h = 4$ and the second simulation $n_h = 6$. Refer to Table 12 to Table 15 for details. In these tables, estimated variance components and standard errors are reported after bias correction. The results tend to show that, after increasing the sample size for the h facet, the results became more similar to the parameter values.

3.3 Boot-p Procedure

Both of the two designs investigated in the current study both have a relatively large sample size for the p facet. Recall that the boot-p procedure only resamples persons. It seems, based on the results obtained from these simulations, that this larger sample size tends to produce more stable results, both in terms of the variance components, and the standard errors. Table 16 lists the estimates for variance components and standard errors using the boot-p procedure for both the crossed and nested designs. From this table, it seems that, for this simulation study, boot-p works fairly well for estimating the parameters, except when estimating the "main" effects for i and h. This is probably caused by the relatively large sample size for the p facet ($n_p = 100$) compared to the i and h facets ($n_i = 20$ and $n_h = 2$). This result is consistent with the conclusions drawn by previous researchers with respect to $p \times i$ design (see, for example, Brennan, 2001).

4 Discussion

For multi-facet designs, there are many possibilities for obtaining bootstrap samples, and it has not been demonstrated that there is a single optimal bootstrap procedure for obtaining the "best" estimates for all variance components and standard errors. Using normal-data simulations, this paper investigated bootstrap procedures for the $p \times i \times h$ design and the $p \times (i : h)$ designs. For these circumstances, the results from these simulation studies seem to warrant the following tentative conclusions:

- Bootstrap procedures for estimating variance components and standard errors generally yield biased results, but bias-correction formulas given by Equations 4 and 5 work reasonably well when used in conjunction with the rules discussed in this paper.
- Specifically, for main effects, the corresponding bootstrap procedure generally yields good results. For interaction effects, the one-dimensional bootstrap corresponding to the interaction term and with larger sample size generally produces adequate results.
- When the sample size is large for a particular facet, the bootstrap procedure corresponding to that facet tends to produce stable estimates. When the sample size is small for a facet, the estimates based on the corresponding bootstrap procedure tend to deviate from the parameter values. When a facet has another facet nested within it, some of its related estimates may deviate from the parameters.

The estimation of standard errors of estimated variance components has long been a challenging task in generalizability theory. For two decades researchers have speculated about the potential applicability of the bootstrap for obtaining such estimates, but the same researchers have identified problems in using the bootstrap. This paper proposes procedures that may ameliorate these problems and thus make the bootstrap a more accessible procedure for estimating standard errors of estimated variance components in generalizability theory. The recommendations in the current paper are limited, however, in that the simulations are for only two two-facet designs, based on normality assumptions. Further investigations need to be conducted to extend these recommendations to dichotomous and polytomous data, and to more complicated designs.

5 References

Brennan, R. L., Harris, D.J., & Hanson, B.A. (1987). *The bootstrap and other procedures for examining the variability of estimated variance components in testing contexts.* ACT Research Report Series 87-7. Iowa City, IA: American College Testing Program.

Brennan, R. L. (2001). *Generalizability Theory.* New York:Springer-Verlag.

Leucht, R.M., & Smith, P.L.(1989). *The effects of bootstrapping strategies on the estimation of variance components.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Searle, S. R. (1971). *Linear Models.* New York: Wiley.

Othman, A.R. (1995). *Examining task sampling variablity in science performance assessments.* Unpublished doctoral dissertation, University of California, Santa Barbara.

Wiley, E. W. (2000). *Bootstrap strategies for variance component estimation: theoretical and empirical results.* Unpublished doctoral dissertation, Stanford.

6 Tables

Table 1: Sample sizes and parameters used for both $p \times i \times h$ design and $p \times (i:h)$ design.

Sample size	$p \times i \times h$ design	$p \times (i:h)$ design
$n_p = 100$	$\sigma^2(p) = 16$	$\sigma^2(p) = 16$
$n_i = 20$	$\sigma^2(i) = 4$	
$n_h = 2$	$\sigma^2(h) = 1$	$\sigma^2(h) = 1$
	$\sigma^2(pi) = 64$	
	$\sigma^2(ph) = 2$	$\sigma^2(ph) = 2$
	$\sigma^2(ih) = 3$	$\sigma^2(i{:}h) = 7$
	$\sigma^2(pih) = 144$	$\sigma^2(pi{:}h) = 208$

Table 2: Raw (Unadjusted) Estimated Variance Components for the $p \times i \times h$ Design with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	$\hat{\sigma}^2(p)$	$\hat{\sigma}^2(i)$	$\hat{\sigma}^2(h)$	$\hat{\sigma}^2(pi)$	$\hat{\sigma}^2(ph)$	$\hat{\sigma}^2(ih)$	$\hat{\sigma}^2(pih)$
parameter	16.0000	4.0000	1.0000	64.0000	2.0000	3.0000	144.0000
empirical	16.0249	4.0722	0.9448	64.2565	2.0186	2.9566	143.9334
boot-p	15.8754	4.7047	0.9678	63.6142	2.0007	4.3933	142.5068
boot-i	19.2233	3.8700	1.0926	61.0402	9.2051	2.8156	136.7320
boot-h	17.0371	5.5484	0.4736	136.2177	1.0081	1.4792	71.9755
boot-p, i	19.0250	4.4917	1.1863	60.4233	9.1193	4.1763	135.3943
boot-p, h	16.8654	6.9290	0.4829	135.0929	0.9967	2.1863	71.0151
boot-i, h	23.8356	5.2745	0.5427	129.3172	4.6105	1.4039	68.4561
boot-p, i, h	23.6767	6.5808	0.5897	128.6116	4.5336	2.0691	67.1517

Table 3: Estimated Standard Errors of Estimated Variance Components Based on Unadjusted Results for the $p \times i \times h$ design with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	$\hat{\sigma}(\hat{\sigma}^2(p))$	$\hat{\sigma}(\hat{\sigma}^2(i))$	$\hat{\sigma}(\hat{\sigma}^2(h))$	$\hat{\sigma}(\hat{\sigma}^2(pi))$	$\hat{\sigma}(\hat{\sigma}^2(ph))$	$\hat{\sigma}(\hat{\sigma}^2(ih))$	$\hat{\sigma}(\hat{\sigma}^2(pih))$
parameter	3.4545	2.3399	1.7580	5.0178	1.3285	1.4412	4.6955
empirical	3.5775	2.4329	1.6142	5.1865	1.3120	1.4153	4.8151
boot-p	3.3992	1.5978	0.6725	4.9511	1.3140	1.2572	4.6280
boot-i	3.0017	2.1973	1.0348	4.7657	2.9621	1.3170	4.9127
boot-h	1.9042	2.0594	0.9345	72.0770	1.2030	1.6387	72.0000
boot-p, i	5.3159	3.0679	1.4190	8.0608	3.7049	2.1199	7.8443
boot-p, h	4.1468	3.2291	1.0572	71.6118	1.5085	2.4717	71.3805
boot-i, h	6.3720	3.1604	1.2205	68.7300	5.0988	1.8133	68.5025
boot-p, i, h	8.0113	4.3141	1.4148	68.5415	5.3098	2.6485	68.0118

Table 4: Bias-corrected Estimated Variance Components for the $p \times i \times h$ Design with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	$\hat{\sigma}^2(p)$	$\hat{\sigma}^2(i)$	$\hat{\sigma}^2(h)$	$\hat{\sigma}^2(pi)$	$\hat{\sigma}^2(ph)$	$\hat{\sigma}^2(ih)$	$\hat{\sigma}^2(pih)$
parameter	16.0000	4.0000	1.0000	64.0000	2.0000	3.0000	144.0000
empirical	16.0249	4.0722	0.9448	64.2565	2.0186	2.9566	143.9334
boot-p	16.0357	4.0621	0.9476	64.2568	2.0209	2.9539	143.9463
boot-i	16.0107	4.0737	0.9444	64.2528	2.0087	2.9638	143.9284
boot-h	16.0290	4.0692	0.9472	64.2422	2.0163	2.9584	143.9511
boot-p, i	16.0048	4.0856	0.9464	64.2460	2.0135	2.9565	143.9599
boot-p, h	16.0289	4.0954	0.9456	64.7250	2.0136	2.9379	143.4647
boot-i, h	16.0218	4.0743	0.9375	64.0642	2.0152	2.9555	144.1182
boot-p, i, h	16.0690	4.0956	0.9413	65.3481	2.0188	2.9280	142.7999

Table 5: Bias-corrected Estimated Standard Errors of Estimated Variance Components for the $p \times i \times h$ design with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	$\hat{\sigma}(\hat{\sigma}^2(p))$	$\hat{\sigma}(\hat{\sigma}^2(i))$	$\hat{\sigma}(\hat{\sigma}^2(h))$	$\hat{\sigma}(\hat{\sigma}^2(pi))$	$\hat{\sigma}(\hat{\sigma}^2(ph))$	$\hat{\sigma}(\hat{\sigma}^2(ih))$	$\hat{\sigma}(\hat{\sigma}^2(pih))$
parameter	3.4545	2.3399	1.7580	5.0178	1.3285	1.4412	4.6955
empirical	3.5775	2.4329	1.6142	5.1865	1.3120	1.4153	4.8151
boot-p	3.4336	1.6004	0.6727	5.0011	1.3272	1.2598	4.6748
boot-i	3.0475	2.3129	1.0407	5.0165	3.0800	1.3863	5.1712
boot-h	2.8225	3.5056	1.8690	144.0396	2.4061	3.2773	144.0001
boot-p, i	5.4185	3.2320	1.4286	8.5708	3.8657	2.2349	8.3406
boot-p, h	4.8605	4.1688	2.1055	144.2768	3.0475	3.7240	144.2031
boot-i, h	5.4463	4.5811	2.3857	144.2592	4.9791	3.8174	144.2159
boot-p, i, h	7.7691	5.5599	2.7501	144.7578	5.9670	4.5293	144.6290

Table 6: Estimated Variance Components and Standard Errors Obtained via Proposed Rules for the $p \times i \times h$ design with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	Mean		Standard Error	
	Parameter	Estimate	Parameter	Estimate
$\hat{\sigma}^2(p)$	16.0000	16.0357	3.4545	3.4336
$\hat{\sigma}^2(i)$	4.0000	4.0737	2.3399	2.3129
$\hat{\sigma}^2(h)$	1.0000	0.9472	1.7580	1.8690
$\hat{\sigma}^2(pi)$	64.0000	64.2568	5.0178	5.0011
$\hat{\sigma}^2(ph)$	2.0000	2.0209	1.3285	1.3272
$\hat{\sigma}^2(ih)$	3.0000	2.9638	1.4413	1.3863
$\hat{\sigma}^2(pih)$	144.0000	143.9463	4.6955	4.6748

Table 7: Raw Estimated Variance Components for the $p \times (i:h)$ Design with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	$\hat{\sigma}^2(p)$	$\hat{\sigma}^2(h)$	$\hat{\sigma}^2(i{:}h)$	$\hat{\sigma}^2(ph)$	$\hat{\sigma}^2(pi{:}h)$
parameter	16.0000	1.0000	7.0000	2.0000	208.0000
empirical	16.0949	1.0437	6.9130	2.0374	208.0525
boot-p	15.9183	1.0654	9.0043	2.0208	205.9771
boot-h	22.3243	0.3476	6.9127	-4.1789	208.0683
boot-i,h	22.3287	0.6890	6.5632	6.1978	197.6483
boot-p,h	22.0712	0.3119	8.9985	-4.1220	205.9803
boot-p,i,h	22.1245	0.7619	8.5558	6.1419	195.6856

Table 8: Estimated Standard Errors of Estimated Variance Components for the $p \times (i:h)$ Design Based on Unadjusted Results with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	$\hat{\sigma}(\hat{\sigma}^2(p))$	$\hat{\sigma}(\hat{\sigma}^2(h))$	$\hat{\sigma}(\hat{\sigma}^2(i{:}h))$	$\hat{\sigma}(\hat{\sigma}^2(ph))$	$\hat{\sigma}(\hat{\sigma}^2(pi{:}h))$
parameter	3.2761	2.0872	2.0836	1.7787	4.7959
empirical	3.3080	2.0493	2.0985	1.6724	4.8006
boot-p	3.5863	1.1662	1.5511	2.3351	4.8311
boot-h	4.1726	0.9777	1.4983	6.2818	3.3623
boot-i,h	4.1120	1.2920	2.3924	7.3199	5.9559
boot-p,h	4.0817	1.1421	2.3673	6.5084	6.7305
boot-p,i,h	4.0239	1.4362	3.2777	7.6514	9.2055

Table 9: Bias-corrected Estimated Variance Components for the $p\times(i:h)$ Design with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	$\hat\sigma^2(p)$	$\hat\sigma^2(h)$	$\hat\sigma^2(i{:}h)$	$\hat\sigma^2(ph)$	$\hat\sigma^2(pi{:}h)$
parameter	16.0000	1.0000	7.0000	2.0000	208.0000
empirical	16.0949	1.0437	6.9130	2.0374	208.0525
boot-p	16.0791	1.0449	6.9237	2.0412	208.0577
boot-h	26.5032	0.6952	6.9127	-8.3579	208.0683
boot-i,h	17.1274	1.0325	6.9086	1.9931	208.0509
boot-p,h	26.4578	0.7070	6.9179	-8.3273	208.0609
boot-p,i,h	16.1440	1.0534	6.9254	2.0046	208.0655

Table 10: Bias-corrected Estimated Standard Errors of Estimated Variance Components for the $p \times (i:h)$ Design with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	$\hat\sigma(\hat\sigma^2(p))$	$\hat\sigma(\hat\sigma^2(h))$	$\hat\sigma(\hat\sigma^2(i{:}h))$	$\hat\sigma(\hat\sigma^2(ph))$	$\hat\sigma(\hat\sigma^2(pi{:}h))$
parameter	3.2761	2.0872	2.0836	1.7787	4.7959
empirical	3.2947	2.1290	2.0710	1.7746	4.8287
boot-p	3.3003	0.8955	1.5225	1.7897	4.8020
boot-h	12.6916	2.4767	1.4981	12.5742	3.3446
boot-i,h	7.1942	3.5724	2.5248	14.4903	6.2879
boot-p,h	13.3314	2.7831	2.3677	12.8315	6.8022
boot-p,i,h	15.0108	4.1675	3.4589	15.3151	9.8046

Table 11: Estimated Variance Components and Standard Errors Obtained via Proposed Rules for the Nested Design with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	Mean		Standard Error	
	Parameter	Estimate	Parameter	Estimate
$\hat\sigma^2(p)$	16.0000	16.0791	3.2761	3.3003
$\hat\sigma^2(h)$	1.0000	0.6952	2.0872	2.4767
$\hat\sigma^2(i{:}h)$	7.0000	6.9086	2.0836	2.5248
$\hat\sigma^2(ph)$	2.0000	2.0412	1.7787	1.7897
$\hat\sigma^2(pih)$	208.0000	208.0577	4.8020	4.7523

Table 12: Bias-corrected Estimated Variance Components for the $p \times (i:h)$ Design with $n_p = 100$, $n_i = 20$, and $n_h = 4$.

	$\hat\sigma^2(p)$	$\hat\sigma^2(h)$	$\hat\sigma^2(i{:}h)$	$\hat\sigma^2(ph)$	$\hat\sigma^2(pi{:}h)$
parameter	16.0000	1.0000	7.0000	2.0000	208.0000
empirical	16.0086	1.0031	6.9757	2.0014	207.9615
boot-p	16.0158	1.0044	6.9780	2.0014	207.9485
boot-h	19.4788	0.8860	6.9784	-1.4657	207.9535
boot-i,h	16.5097	1.0060	6.9755	1.9951	207.9581
boot-p,h	19.4636	0.8886	6.9712	-1.4637	207.9793
boot-p,i,h	16.0055	1.0004	6.9749	1.9761	208.0008

Table 13: Bias-corrected Estimated Standard Errors for Estimated Variance Components for the $p \times (i:h)$ Design with $n_p = 100$, $n_i = 20$, and $n_h = 4$.

	$\hat\sigma(\hat\sigma^2(p))$	$\hat\sigma(\hat\sigma^2(h))$	$\hat\sigma(\hat\sigma^2(i{:}h))$	$\hat\sigma(\hat\sigma^2(ph))$	$\hat\sigma(\hat\sigma^2(pi{:}h))$
parameter	2.7266	1.2058	1.4734	1.0316	3.3912
empirical	2.7087	1.1880	1.5052	0.9937	3.4108
boot-p	2.7581	0.5101	1.0982	1.0295	3.4026
boot-h	3.2914	1.1466	1.2611	3.0799	2.8857
boot-i,h	3.0754	1.6417	1.9007	4.3195	4.7293
boot-p,h	4.5181	1.3310	1.9026	3.3841	5.3499
boot-p,i,h	5.1779	1.9182	2.5836	4.9160	7.3108

Table 14: Bias-corrected Estimated Variance Components for the $p \times (i:h)$ Design with $n_p = 100$, $n_i = 20$, and $n_h = 6$.

	$\hat\sigma^2(p)$	$\hat\sigma^2(h)$	$\hat\sigma^2(i{:}h)$	$\hat\sigma^2(ph)$	$\hat\sigma^2(pi{:}h)$
parameter	16.0000	1.0000	7.0000	2.0000	208.0000
empirical	15.8868	0.9701	7.0191	2.0161	207.9650
boot-p	15.8919	0.9700	7.0265	2.0142	207.9741
boot-h	17.9660	0.8971	7.0151	-0.0630	207.9659
boot-i,h	16.2144	0.9692	7.0267	2.0188	207.9684
boot-p,h	17.9645	0.8973	7.0217	-0.0620	207.9793
boot-p,i,h	15.8914	0.9721	7.0176	2.0083	207.9736

Table 15: Bias-corrected Estimated Standard Errors for Estimated Variance Components for the $p \times (i\!:\!h)$ Design with $n_p = 100$, $n_i = 20$, and $n_h = 6$.

	$\hat\sigma(\hat\sigma^2(p))$	$\hat\sigma(\hat\sigma^2(h))$	$\hat\sigma(\hat\sigma^2(i{:}h))$	$\hat\sigma(\hat\sigma^2(ph))$	$\hat\sigma(\hat\sigma^2(pi{:}h))$
parameter	2.5712	0.9342	1.2030	0.8003	2.7689
empirical	2.5485	0.9077	1.2081	0.8184	2.7450
boot-p	2.5742	0.3964	0.9125	0.8025	2.7768
boot-h	1.9093	0.8723	1.1014	1.6175	2.5265
boot-i, h	2.1052	1.2166	1.6135	2.7273	4.0029
boot-p, h	3.3932	1.0221	1.6344	1.9539	4.5261
boot-p, i, h	3.8340	1.4344	2.1749	3.2454	6.0722

Table 16: Estimated Variance Components and Standard Errors Obtained using boot-p for both Designs with $n_p = 100$, $n_i = 20$, and $n_h = 2$.

	Mean		Standard Error	
	Parameter	Estimate	Parameter	Estimate
$p \times i \times h$ design:				
$\hat\sigma^2(p)$	16.0000	16.0357	3.4545	3.4336
$\hat\sigma^2(i)$	4.0000	4.0621	2.3399	1.6004
$\hat\sigma^2(h)$	1.0000	0.9476	1.7580	0.6727
$\hat\sigma^2(pi)$	64.0000	64.2568	5.0178	5.0011
$\hat\sigma^2(ph)$	2.0000	2.0209	1.3285	1.3272
$\hat\sigma^2(ih)$	3.0000	2.9539	1.4413	1.2598
$\hat\sigma^2(pih)$	144.0000	143.9463	4.6955	4.6748
$p \times (i\!:\!h)$ design:				
$\hat\sigma^2(p)$	16.0000	16.0791	3.2761	3.3003
$\hat\sigma^2(h)$	1.0000	1.0449	2.0872	0.8955
$\hat\sigma^2(i{:}h)$	7.0000	6.9237	2.0836	1.5225
$\hat\sigma^2(ph)$	2.0000	2.0412	1.7787	1.7897
$\hat\sigma^2(pih)$	208.0000	208.0577	4.8020	4.7523

7 Appendix A: Wiley's (2000) Bias-Correction Factors

Wiley (2000) proposed the complete set of bias correction factors for the two-facet crossed design $p \times i$. These factors are listed below.
Adjusted Estimates Based on boot-p

$$\begin{aligned}
\hat{\sigma}^2(p) &= \frac{n_p}{n_p-1}\hat{\sigma}^2(p|\text{boot}{-}p) \\
\hat{\sigma}^2(i) &= \hat{\sigma}^2(i|\text{boot}{-}p) - \frac{1}{n_p-1}\hat{\sigma}^2(pi|\text{boot}{-}p) \\
\hat{\sigma}^2(pi) &= \frac{n_p}{n_p-1}\hat{\sigma}^2(pi|\text{boot}{-}p)
\end{aligned}$$

Adjusted Estimates Based on boot-i

$$\begin{aligned}
\hat{\sigma}^2(p) &= \hat{\sigma}^2(p|\text{boot}{-}i) - \frac{1}{n_i-1}\hat{\sigma}^2(pi|\text{boot}{-}i) \\
\hat{\sigma}^2(i) &= \frac{n_i}{n_i-1}\hat{\sigma}^2(i|\text{boot}{-}i) \\
\hat{\sigma}^2(pi) &= \frac{n_i}{n_i-1}\hat{\sigma}^2(pi|\text{boot}{-}i)
\end{aligned}$$

Adjusted Estimates Based on boot-p, i

$$\begin{aligned}
\hat{\sigma}^2(p) &= \frac{n_p}{n_p-1}\hat{\sigma}^2(p|\text{boot}{-}p,i) - \frac{n_p}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi|\text{boot}{-}p,i) \\
\hat{\sigma}^2(i) &= \frac{n_i}{n_i-1}\hat{\sigma}^2(i|\text{boot}{-}p,i) - \frac{n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi|\text{boot}{-}p,i) \\
\hat{\sigma}^2(pi) &= \frac{n_p n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi|\text{boot}{-}pi)
\end{aligned}$$

Wiley (2000) also proposed some of the bias correction factors for the three-facet crossed design, which are listed below:

$$\begin{aligned}
\hat{\sigma}^2(p) &= \frac{n_p}{n_p-1}\hat{\sigma}^2(p|\text{boot}{-}p) \\
\hat{\sigma}^2(i) &= \frac{n_i}{n_i-1}\hat{\sigma}^2(i|\text{boot}{-}i) \\
\hat{\sigma}^2(h) &= \frac{n_h}{n_h-1}\hat{\sigma}^2(h|\text{boot}{-}h) \\
\hat{\sigma}^2(pi) &= \frac{n_p n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi|\text{boot}{-}p,i) \\
\hat{\sigma}^2(ph) &= \frac{n_p n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}{-}p,h)
\end{aligned}$$

$$\hat{\sigma}^2(ih) = \frac{n_i n_h}{(n_i-1)(n_h-1)}\hat{\sigma}^2(ih|\text{boot}{-}i,h)$$
$$\hat{\sigma}^2(pih) = \frac{n_p n_i n_h}{(n_p-1)(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}p,i,h)$$

8 Appendix B: Bias-Correction Factors for the Crossed Design

Adjusted Estimates Based on boot-p

$$\hat{\sigma}^2(p) = \frac{n_p}{n_p-1}\hat{\sigma}^2(p|\text{boot}{-}p)$$
$$\hat{\sigma}^2(i) = \hat{\sigma}^2(i|\text{boot}{-}p) - \frac{1}{n_p-1}\hat{\sigma}^2(pi|\text{boot}{-}p)$$
$$\hat{\sigma}^2(h) = \hat{\sigma}^2(h|\text{boot}{-}p) - \frac{1}{n_p-1}\hat{\sigma}^2(po|\text{boot}{-}p)$$
$$\hat{\sigma}^2(pi) = \frac{n_p}{n_p-1}\hat{\sigma}^2(pi|\text{boot}{-}p)$$
$$\hat{\sigma}^2(ph) = \frac{n_p}{n_p-1}\hat{\sigma}^2(ph|\text{boot}{-}p)$$
$$\hat{\sigma}^2(ih) = \hat{\sigma}^2(ih|\text{boot}{-}p) - \frac{1}{n_p-1}\hat{\sigma}^2(pih|\text{boot}{-}p)$$
$$\hat{\sigma}^2(pih) = \frac{n_p}{n_p-1}\hat{\sigma}^2(pih|\text{boot}{-}p)$$

Adjusted Estimates Based on boot-i

$$\hat{\sigma}^2(p) = \hat{\sigma}^2(p|\text{boot}{-}i) - \frac{1}{n_i-1}\hat{\sigma}^2(pi|\text{boot}{-}i)$$
$$\hat{\sigma}^2(i) = \frac{n_i}{n_i-1}\hat{\sigma}^2(i|\text{boot}{-}i)$$
$$\hat{\sigma}^2(h) = \hat{\sigma}^2(h|\text{boot}{-}i) - \frac{1}{n_i-1}\hat{\sigma}^2(ih|\text{boot}{-}i)$$
$$\hat{\sigma}^2(pi) = \frac{n_i}{n_i-1}\hat{\sigma}^2(pi|\text{boot}{-}i)$$
$$\hat{\sigma}^2(ph) = \hat{\sigma}^2(ph|\text{boot}{-}i) - \frac{1}{n_i-1}\hat{\sigma}^2(pih|\text{boot}{-}i)$$
$$\hat{\sigma}^2(ih) = \frac{n_i}{n_i-1}\hat{\sigma}^2(ih|\text{boot}{-}i)$$
$$\hat{\sigma}^2(pih) = \frac{n_i}{n_i-1}\hat{\sigma}^2(pih|\text{boot}{-}i)$$

Adjusted Estimates Based on boot-h

$$\hat{\sigma}^2(p) = \hat{\sigma}^2(p|\text{boot}{-}h) - \frac{1}{n_h-1}\hat{\sigma}^2(ph|\text{boot}{-}h)$$

$$
\begin{aligned}
\hat{\sigma}^2(i) &= \hat{\sigma}^2(i|\text{boot}-h) - \frac{1}{n_h - 1}\hat{\sigma}^2(ih|\text{boot}-h) \\
\hat{\sigma}^2(h) &= \frac{n_h}{n_h - 1}\hat{\sigma}^2(h|\text{boot}-h) \\
\hat{\sigma}^2(pi) &= \hat{\sigma}^2(pi|\text{boot}-h) - \frac{1}{n_h - 1}\hat{\sigma}^2(pih|\text{boot}-h) \\
\hat{\sigma}^2(ph) &= \frac{n_h}{n_h - 1}\hat{\sigma}^2(ph|\text{boot}-h) \\
\hat{\sigma}^2(ih) &= \frac{n_h}{n_h - 1}\hat{\sigma}^2(ih|\text{boot}-h) \\
\hat{\sigma}^2(pih) &= \frac{n_h}{n_h - 1}\hat{\sigma}^2(pih|\text{boot}-h)
\end{aligned}
$$

Adjusted Estimates Based on boot-p, i

$$
\begin{aligned}
\hat{\sigma}^2(p) &= \frac{n_p}{n_p - 1}\hat{\sigma}^2(p|\text{boot}-p,i) - \frac{n_p}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi|\text{boot}-p,i) \\
\hat{\sigma}^2(i) &= \frac{n_i}{n_i - 1}\hat{\sigma}^2(i|\text{boot}-p,i) - \frac{n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi|\text{boot}-p,i) \\
\hat{\sigma}^2(h) &= \hat{\sigma}^2(h|\text{boot}-p,i) - \frac{1}{n_p - 1}\hat{\sigma}^2(ph|\text{boot}-p,i) \\
&\quad - \frac{1}{n_i - 1}\hat{\sigma}^2(ih)|\text{boot}-p,i) + \frac{1}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pih|\text{boot}-p,i) \\
\hat{\sigma}^2(pi) &= \frac{n_p n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi|\text{boot}-pi) \\
\hat{\sigma}^2(ph) &= \frac{n_p}{n_p - 1}\hat{\sigma}^2(ph|\text{boot}-p,i) - \frac{n_p}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pih|\text{boot}-p,i) \\
\hat{\sigma}^2(ih) &= \frac{n_i}{n_i - 1}\hat{\sigma}^2(ih|\text{boot}-p,i) - \frac{n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pih|\text{boot}-p,i) \\
\hat{\sigma}^2(pih) &= \frac{n_p n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pih|\text{boot}-p,i)
\end{aligned}
$$

Adjusted Estimates Based on boot-p, h

$$
\begin{aligned}
\hat{\sigma}^2(p) &= \frac{n_p}{n_p - 1}\hat{\sigma}^2(p|\text{boot}-p,h) - \frac{n_p}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}-p,h) \\
\hat{\sigma}^2(i) &= \hat{\sigma}^2(i|\text{boot}-p,h) - \frac{1}{n_p - 1}\hat{\sigma}^2(pi|\text{boot}-p,h) \\
&\quad - \frac{1}{n_h - 1}\hat{\sigma}^2(ih)|\text{boot}-p,h) + \frac{1}{(n_p-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}-p,h) \\
\hat{\sigma}^2(h) &= \frac{n_h}{n_h - 1}\hat{\sigma}^2(h|\text{boot}-p,h) - \frac{n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}-p,h) \\
\hat{\sigma}^2(pi) &= \frac{n_p}{n_p - 1}\hat{\sigma}^2(pi|\text{boot}-p,h) - \frac{n_p}{(n_p-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}-p,h) \\
\hat{\sigma}^2(ph) &= \frac{n_p n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}-p,h)
\end{aligned}
$$

$$\hat{\sigma}^2(ih) = \frac{n_h}{n_h-1}\hat{\sigma}^2(ih|\text{boot}{-}p,h) - \frac{n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}p,h)$$
$$\hat{\sigma}^2(pih) = \frac{n_p n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}p,h)$$

Adjusted Estimates Based on boot-i,h

$$\hat{\sigma}^2(p) = \hat{\sigma}^2(p|\text{boot}{-}i,h) - \frac{1}{n_i-1}\hat{\sigma}^2(pi|\text{boot}{-}i,h) - \frac{1}{n_h-1}\hat{\sigma}^2(ph)|\text{boot}{-}i,h) + \frac{1}{(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}i,h)$$
$$\hat{\sigma}^2(i) = \frac{n_i}{n_i-1}\hat{\sigma}^2(i|\text{boot}{-}i,h) - \frac{n_i}{(n_i-1)(n_h-1)}\hat{\sigma}^2(ih|\text{boot}{-}i,h)$$
$$\hat{\sigma}^2(h) = \frac{n_h}{n_h-1}\hat{\sigma}^2(h|\text{boot}{-}i,h) - \frac{n_h}{(n_i-1)(n_h-1)}\hat{\sigma}^2(ih|\text{boot}{-}i,h)$$
$$\hat{\sigma}^2(pi) = \frac{n_i}{n_i-1}\hat{\sigma}^2(pi|\text{boot}{-}i,h) - \frac{n_i}{(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}i,h)$$
$$\hat{\sigma}^2(ph) = \frac{n_h}{n_h-1}\hat{\sigma}^2(ph|\text{boot}{-}i,h) - \frac{n_h}{(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}i,h)$$
$$\hat{\sigma}^2(ih) = \frac{n_i n_h}{(n_i-1)(n_h-1)}\hat{\sigma}^2(ih|\text{boot}{-}i,h)$$
$$\hat{\sigma}^2(pih) = \frac{n_i n_h}{(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}i,h)$$

Adjusted Estimates Based on boot-p,i,h

$$\hat{\sigma}^2(p) = \frac{n_p}{n_p-1}\hat{\sigma}^2(p|\text{boot}{-}p,i,h) - \frac{n_p}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi|\text{boot}{-}p,i,h) - \frac{n_p}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph)|\text{boot}{-}p,i,h) + \frac{n_p}{(n_p-1)(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}p,i,h)$$
$$\hat{\sigma}^2(i) = \frac{n_i}{n_i-1}\hat{\sigma}^2(i|\text{boot}{-}p,i,h) - \frac{n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi|\text{boot}{-}p,i,h) - \frac{n_i}{(n_i-1)(n_h-1)}\hat{\sigma}^2(ih)|\text{boot}{-}p,i,h) + \frac{n_i}{(n_p-1)(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}p,i,h)$$
$$\hat{\sigma}^2(h) = \frac{n_h}{n_h-1}\hat{\sigma}^2(h|\text{boot}{-}p,i,h) - \frac{n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}{-}p,i,h) - \frac{n_h}{(n_i-1)(n_h-1)}\hat{\sigma}^2(ih)|\text{boot}{-}p,i,h) + \frac{n_h}{(n_p-1)(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}p,i,h)$$
$$\hat{\sigma}^2(pi) = \frac{n_p n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi|\text{boot}{-}p,i,h) - \frac{n_p n_i}{(n_p-1)(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}p,i,h)$$

$$
\begin{aligned}
\hat{\sigma}^2(ph) &= \frac{n_p n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}{-}p,i,h) \\
&\quad - \frac{n_p n_h}{(n_p-1)(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}p,i,h) \\
\hat{\sigma}^2(ih) &= \frac{n_i n_h}{(n_i-1)(n_h-1)}\hat{\sigma}^2(ih|\text{boot}{-}p,i,h) \\
&\quad - \frac{n_i n_h}{(n_p-1)(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}p,i,h) \\
\hat{\sigma}^2(pih) &= \frac{n_p n_i n_h}{(n_p-1)(n_i-1)(n_h-1)}\hat{\sigma}^2(pih|\text{boot}{-}p,i,h)
\end{aligned}
$$

9 Appendix C: Bias-Correction Factors for the Nested Design

Adjusted Estimates Based on boot-p

$$
\begin{aligned}
\hat{\sigma}^2(p) &= \frac{n_p}{n_p-1}\hat{\sigma}^2(p|\text{boot}{-}p) \\
\hat{\sigma}^2(h) &= \hat{\sigma}^2(h|\text{boot}{-}p) - \frac{1}{n_p-1}\hat{\sigma}^2(ph|\text{boot}{-}p) \\
\hat{\sigma}^2(i{:}h) &= \hat{\sigma}^2(i{:}h|\text{boot}{-}p) - \frac{1}{n_p-1}\hat{\sigma}^2(pi{:}h|\text{boot}{-}p) \\
\hat{\sigma}^2(ph) &= \frac{n_p}{n_p-1}\hat{\sigma}^2(ph|\text{boot}{-}p) \\
\hat{\sigma}^2(pi{:}h) &= \frac{n_p}{n_p-1}\hat{\sigma}^2(pi{:}h|\text{boot}{-}p)
\end{aligned}
$$

Adjusted Estimates Based on boot-h

$$
\begin{aligned}
\hat{\sigma}^2(p) &= \hat{\sigma}^2(p|\text{boot}{-}h) - \frac{1}{n_h-1}\hat{\sigma}^2(ph|\text{boot}{-}h) \\
\hat{\sigma}^2(h) &= \frac{n_h}{n_h-1}\hat{\sigma}^2(h|\text{boot}{-}h) \\
\hat{\sigma}^2(i{:}h) &= \hat{\sigma}^2(i{:}h|\text{boot}{-}h) \\
\hat{\sigma}^2(ph) &= \frac{n_h}{n_h-1}\hat{\sigma}^2(ph|\text{boot}{-}h) \\
\hat{\sigma}^2(pi{:}h) &= \hat{\sigma}^2(pi{:}h|\text{boot}{-}h)
\end{aligned}
$$

Adjusted Estimates Based on boot-i, h

$$
\begin{aligned}
\hat{\sigma}^2(p) &= \hat{\sigma}^2(p|\text{boot}{-}i,h) - \frac{1}{n_h(n_i-1)}\hat{\sigma}^2(pi{:}h|\text{boot}{-}i,h) \\
\hat{\sigma}^2(h) &= \frac{n_h}{n_h-1}\hat{\sigma}^2(h|\text{boot}{-}i,h) - \frac{1}{n_i-1}\hat{\sigma}^2(i{:}h|\text{boot}{-}i,h)
\end{aligned}
$$

$$\begin{aligned}
\hat{\sigma}^2(i{:}h) &= \frac{n_i}{n_i-1}\hat{\sigma}^2(i{:}h|\text{boot}{-}i,h)\\
\hat{\sigma}^2(ph) &= \frac{n_h}{n_h-1}\hat{\sigma}^2(ph|\text{boot}{-}i,h) - \frac{1}{n_i-1}\hat{\sigma}^2(pi{:}h|\text{boot}{-}i,h)\\
\hat{\sigma}^2(pi{:}h) &= \frac{n_i}{n_i-1}\hat{\sigma}^2(pi{:}h|\text{boot}{-}i,h)
\end{aligned}$$

Adjusted Estimates Based on boot-p,h

$$\begin{aligned}
\hat{\sigma}^2(p) &= \frac{n_p}{n_p-1}\hat{\sigma}^2(p|\text{boot}{-}p,h) - \frac{n_p}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}{-}p,h)\\
\hat{\sigma}^2(h) &= \frac{n_h}{n_h-1}\hat{\sigma}^2(h|\text{boot}{-}p,h) - \frac{n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}{-}p,h)\\
\hat{\sigma}^2(i{:}h) &= \hat{\sigma}^2(i{:}h|\text{boot}{-}p,h) - \frac{1}{n_p-1}\hat{\sigma}^2(pi{:}h|\text{boot}{-}p,h)\\
\hat{\sigma}^2(ph) &= \frac{n_p n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}{-}p,h)\\
\hat{\sigma}^2(pi{:}h) &= \frac{n_p}{n_p-1}\hat{\sigma}^2(pi{:}h|\text{boot}{-}p,h)
\end{aligned}$$

Adjusted Estimates Based on boot-p,i,h

$$\begin{aligned}
\hat{\sigma}^2(p) &= \frac{n_p}{n_p-1}\hat{\sigma}^2(p|\text{boot}{-}p,i,h) - \frac{n_p}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph)|\text{boot}{-}p,i,h)\\
\hat{\sigma}^2(h) &= \frac{n_h}{n_h-1}\hat{\sigma}^2(h|\text{boot}{-}p,i,h) - \frac{n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}{-}p,i,h)\\
&\quad -\frac{1}{n_i-1}\hat{\sigma}^2(i{:}h)|\text{boot}{-}p,i,h)\\
&\quad +\frac{1}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi{:}h|\text{boot}{-}p,i,h)\\
\hat{\sigma}^2(i{:}h) &= \frac{n_i}{n_i-1}\hat{\sigma}^2(i{:}h|\text{boot}{-}p,i,h)\\
&\quad -\frac{n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi{:}h|\text{boot}{-}p,i,h)\\
\hat{\sigma}^2(ph) &= \frac{n_p n_h}{(n_p-1)(n_h-1)}\hat{\sigma}^2(ph|\text{boot}{-}p,i,h)\\
&\quad -\frac{n_p}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi{:}h|\text{boot}{-}p,i,h)\\
\hat{\sigma}^2(pi{:}h) &= \frac{n_p n_i}{(n_p-1)(n_i-1)}\hat{\sigma}^2(pi{:}h|\text{boot}{-}p,i,h)
\end{aligned}$$