United States
Department of
Agriculture

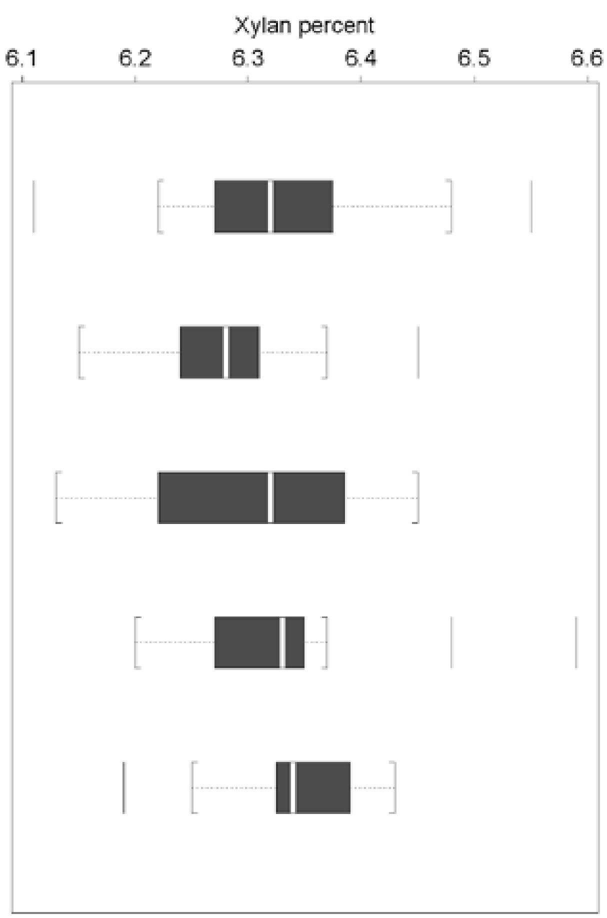
Forest Service

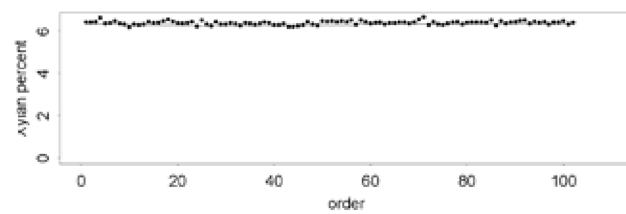
Forest
Products
Laboratory

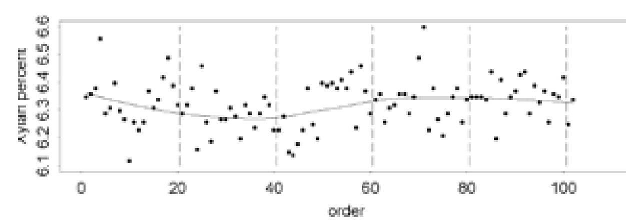
Research
Paper
FPL–RP–638

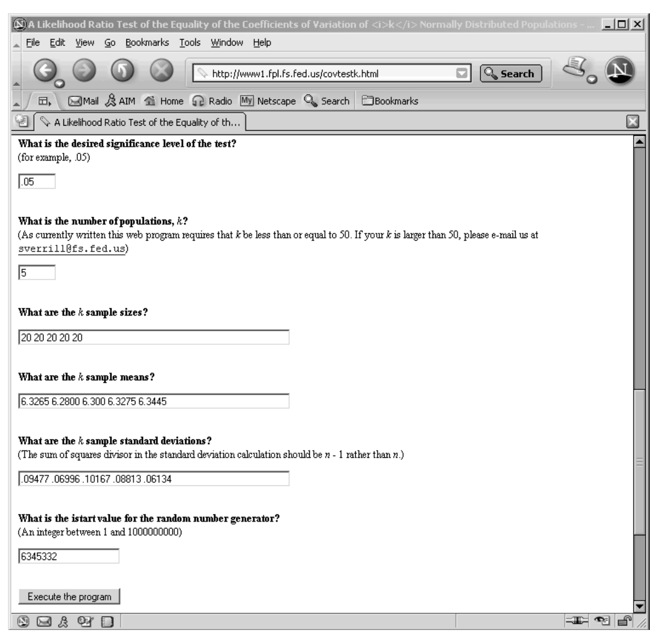
Confidence Bounds and Hypothesis Tests for Normal Distribution Coefficients of Variation

Steve P. Verrill
Richard A. Johnson









Abstract

For normally distributed populations, we obtain confidence bounds on a ratio of two coefficients of variation, provide a test for the equality of *k* coefficients of variation, and provide confidence bounds on a coefficient of variation shared by *k* populations.

To develop these confidence bounds and test, we first establish that estimators based on Newton steps from $\sqrt{n}$-consistent estimators may be used in place of efficient solutions of the likelihood equations in likelihood ratio, Wald, and Rao tests. Taking a quadratic mean differentiability approach, Lehmann and Romano have outlined proofs of similar results. We take a Cramér condition approach and make the conditions and their use explicit.

Keywords: coefficient of variation, signal to noise ratio, risk to return ratio, one-step Newton estimators, Newton's method, $\sqrt{n}$-consistent estimators, efficient likelihood estimators, Cramér conditions, quadratic mean differentiability, likelihood ratio test, Wald test, Rao test, asymptotics

October 2007

Verrill, Steve P.; Johnson, Richard A. 2007. Confidence bounds and hypothesis tests for normal distribution coefficients of variation. Research Paper FPL-RP-638. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 57 p.

A limited number of free copies of this publication are available to the public from the Forest Products Laboratory, One Gifford Pinchot Drive, Madison, WI 53726–2398. This publication is also available online at www.fpl.fs.fed.us. Laboratory publications are sent to hundreds of libraries in the United States and elsewhere.

The Forest Products Laboratory is maintained in cooperation with the University of Wisconsin.

The use of trade or firm names in this publication is for reader information and does not imply endorsement by the United States Department of Agriculture (USDA) of any product or service.

The USDA prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or a part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720–2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250–9410, or call (800) 795–3272 (voice) or (202) 720–6382 (TDD). USDA is an equal opportunity provider and employer.

Contents

Page

Confidence Bounds and Hypothesis Tests for Normal Distribution Coefficients of Variation

Steve Verrill, Mathematical Statistician
USDA Forest Products Laboratory, Madison, Wisconsin
Richard A. Johnson, Professor of Statistics
University of Wisconsin, Madison, Wisconsin

1 Introduction

The coefficent of variation (COV) of a distribution with mean μ and variance σ^2 is defined as the noise to signal ratio or σ/μ. (Sometimes this ratio is multiplied by 100 and reported as a percentage.) Building materials are often evaluated not only on the basis of mean strength but also on relative variability. Laboratory techniques are often compared on the basis of their coefficients of variation. Financial managers treat coefficients of variation as measures of risk to return. Thus, scientists, engineers, and portfolio managers are interested in obtaining confidence intervals on population coefficients of variation and in testing for the equality of coefficients of variation.

For normally distributed populations, Vangel (1996) and Verrill (2003) focused on techniques for obtaining confidence intervals on single coefficients of variation. In this paper, for normally distributed populations, we use likelihood-based methods to obtain large sample solutions to three additional problems:

1. Obtain a confidence interval on the ratio of two coefficients of variation

2. Perform a test of the hypothesis that the coefficients of variation associated with k populations are all equal

3. Obtain a confidence interval on a coefficient of variation that is shared by k populations

A likelihood ratio test of the hypothesis that k coefficients of variation are equal has previously appeared in the literature (Doornbos and Dijkstra 1983, Nairy and Rao 2003). However, no rigorous demonstration of its asymptotic distribution has previously appeared. In addition, several other tests of the equality of k coefficients of variation have been proposed (Bennett 1976, Shafer and Sullivan 1986, Miller 1991, Feltz and Miller 1996, Nairy and Rao 2003). Simulation studies (Doornbos and Dijkstra 1983, Shafer and Sullivan 1986, Fung and Tsang 1998, Nairy and Rao 2003) have suggested that the tests of Bennett (1976), Miller (1991) (and their modifications — Shafer and Sullivan 1986, Feltz and Miller 1996), and the ICV Wald test of Nairy and Rao (2003) have statistical sizes that are near nominal for small sample sizes while the statistical size of the likelihood ratio test is overly liberal. This has led some authors to recommend the tests of Bennett and Miller over the likelihood ratio test. However, we provide Web-based simulation tools that protect a user from an overly liberal test. We illustrate the use of one of these tools in the next section.

In a subsequent paper we will compare the power properties of the tests when actual statistical sizes for all tests are near nominal.

A confidence interval on a ratio of two coefficients of variation has not previously appeared in the literature. Tian (2005) has taken a Weerahandi (1993) approach to obtain a generalized confidence interval on a coefficient of variation that is shared by k populations.

In Section 2 we present an example of the application of our methods to a real data set. In Sections 3 through 8, we present our coefficient of variation theorems. In Appendices A through G, we provide the details of the proofs of these theorems. In Appendix H, we present the proofs of general theorems that provide the basis for the specific coefficient of variation theorems. In particular, in Appendix H we derive the asymptotic distributions of general likelihood ratio, Wald, and Rao test statistics. Versions of these asymptotic results have been reported previously in the literature, but portions of our proofs are novel, and we make the conditions needed to establish our asymptotic results so explicit that we can readily apply them to the special case of coefficients of variation.

In addition, in Appendix H we demonstrate that a solution of the likelihood equations can be replaced in likelihood ratio, Wald, and Rao tests by a Newton step refinement of a $\sqrt{n}$-consistent estimator.[1] Taking a quadratic mean differentiability approach, Lehmann and Romano (2005) have outlined proofs of similar results.

We make use of likelihood ratio tests in Theorems 2, 4, and 6. The asymptotic distributions of our test statistics are given by Theorems H.1 and H.4. We establish the conditions that permit us to invoke these theorems in Appendices B, E, and G. These conditions would also have permitted us to invoke Theorems H.2, H.3, H.5, and H.6 to obtain the asymptotic distributions of the corresponding Wald and Rao test statistics.

2 Application of Our Results to Laboratory Quality Control Data

In this section we illustrate the use of one of our programs to test the hypothesis that five coefficients of variation are equal. The data set consists of quality control measurements of the xylan percentage (xylan is a polysaccharide) in a standard measured in the USDA Forest Products Laboratory's Analytical Chemistry Laboratory over the course of 2 years.[2] The data are presented in Table 1. The measurements were time sequential. For the purposes of illustration, we have grouped observations 1 through 20 into Group 1, observations 21 through 40 into Group 2, and so on. Ideally the quality control measurements represent a steady state process, so the coefficient of variation should not be changing. We plot the data in Figure 1. The curve in the figure was produced by a locally weighted regression (loess) smoother.

From the top plot in Figure 1, it is clear that for practical purposes, the xylan measurement is quite stable. However, the bottom plot in Figure 1 suggests that at a micro level, there might be statistically significant local trends in the data. (There were 102 observations in the full data set. The dashed lines in the bottom plot of Figure 1 demarcate the five groups of 20.)

We plot box plots of the data in Figure 2. The sample means and standard deviations are presented in Table 2. Normal probability plots and formal tests of normality indicated that Groups 1, 2, and 5 do not violate the normality assumption. Group 3 appears to be bimodal and nonnormal. Group 4 appears to have two outliers that cause the normality assumption to be violated. We have performed three analyses. For the first we accepted all the data. For the second we removed two "outliers." For the third we removed all "outliers." The Web program that we used to analyze the three cases is available at `http://www1.fpl.fs.fed.us/covtestk.html`. It is illustrated in Figure 3. As indicated in that figure a user need only provide the program with the number of groups, the sample sizes, means, and standard deviations of the groups, and an integer starting value for the random number generator that is used in the small-sample simulation. The program returns the p-value calculated from the asymptotic test and a simulation-based estimated p-value.

[1] $\hat{a}$ is a $\sqrt{n}$-consistent estimator of a if $\sqrt{n}(\hat{a} - a) = O_p(1)$

[2] We thank Dr. Mark Davis for providing this data.

For the data presented in Table 1, the asymptotic p-value is 0.144, and the small sample simulation test is not significant at a 0.10 level. However, if the two "outliers" of Group 4 are removed, the asymptotic p-value is 0.016 and the estimated p-value from the small sample simulation is 0.025. If all "outliers" in all groups are removed, the asymptotic p-value is 0.006 and the estimated p-value from the small sample simulation is 0.012. A user should give greater credence to the results of the small sample tests. (See Section 9.)

We emphasize that we have presented this example solely for the purpose of illustrating the use of the program. A fully defensible analysis in this case would have to consider issues of serial correlation, non-normal data, and the legitimacy of discarding outliers.

We have also developed two additional Web-based programs. A program that calculates a confidence interval on a ratio of two coefficients of variation can be found at `http://www1.fpl.fs.fed.us/covratio.html`. A program that calculates a confidence interval on a coefficient of variation that is shared by k normally distributed populations can be found at `http://www1.fpl.fs.fed.us/covconfk.html`.

In Sections 3 through 8 we develop the statistical theory that underlies these programs. Readers who are not interested in this theory should skip to Section 9.

3 Confidence Interval on the Ratio of Two Coefficients of Variation

In Sections 3 and 4 we present two approaches to obtaining a confidence interval on a ratio of two coefficients of variation. In Section 3 we obtain the asymptotic distribution of the maximum likelihood estimate of the ratio. A 95% (for example) confidence interval is then just the usual "estimate plus or minus two (1.96) standard deviations." See below for details. In Section 4 we take a likelihood ratio test approach. In this case, a 95% (for example) confidence interval is simply the collection of those ratio values that are not rejected at a 0.05 significance level by the test. We have found via simulations that actual confidence levels approach nominal confidence levels more rapidly when we take the likelihood ratio approach. Thus our Web programs implement only the material in Section 4. However, we include the Section 3 material for completeness.

We assume that we have n_1 observations, $x_{11}, \ldots, x_{n_1 1}$, from a $\mathrm{N}(\mu_1, \sigma_1^2)$ population, and n_2 observations, $x_{12}, \ldots, x_{n_2 2}$, from a $\mathrm{N}(\mu_2, \sigma_2^2)$ population, and that $\mu_1, \mu_2 > 0$. We assume that all these observations are statistically independent. Let $n \equiv n_1 + n_2$. We further assume that $n_1/n \to \lambda_1 > 0$ and $n_2/n \to \lambda_2 > 0$ as $n \to \infty$. We denote the coefficient of variation of the first population by

$$c \equiv \sigma_1/\mu_1$$

so

$$\mu_1 = \sigma_1/c$$

We denote the ratio of the coefficient of variation of the second population to the coefficient of variation of the first population by r. Thus

$$r \equiv (\sigma_2/\mu_2)/c$$

and

$$\mu_2 = \sigma_2/(rc)$$

Then, we have the following theorem.
Theorem 1

$$\sqrt{n}\left(\begin{pmatrix}\hat{\sigma}_1\\ \hat{\sigma}_2\\ \hat{c}\\ \hat{r}\end{pmatrix} - \begin{pmatrix}\sigma_1\\ \sigma_2\\ c\\ r\end{pmatrix}\right) \stackrel{D}{\rightarrow} \mathrm{N}(\mathbf{0}, I(\boldsymbol{\theta})^{-1})$$

where $\boldsymbol{\theta} \equiv (\sigma_1, \sigma_2, c, r)^T$,

$$\begin{aligned}
\bar{x}_{\cdot j} &\equiv \sum_{i=1}^{n_j} x_{ij}/n_j \text{ for } j = 1,2\\
\hat{\sigma}_j &\equiv \sqrt{\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{\cdot j})^2/n_j} \text{ for } j = 1,2\\
\hat{c} &\equiv \hat{\sigma}_1/\bar{x}_{\cdot 1}\\
\hat{r} &\equiv (\hat{\sigma}_2/\bar{x}_{\cdot 2})/\hat{c}
\end{aligned}$$

and Fisher's information matrix is given by

$$I(\boldsymbol{\theta}) = \lambda_1 I_1(\boldsymbol{\theta}) + \lambda_2 I_2(\boldsymbol{\theta})$$

where

$$I_1(\boldsymbol{\theta}) = \begin{pmatrix} (2+1/c^2)/\sigma_1^2 & 0 & -1/(c^3\sigma_1) & 0\\ 0 & 0 & 0 & 0\\ -1/(c^3\sigma_1) & 0 & 1/c^4 & 0\\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$I_2(\boldsymbol{\theta}) = \begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & (2+1/(rc)^2)/\sigma_2^2 & -1/(r^2c^3\sigma_2) & -1/(r^3c^2\sigma_2)\\ 0 & -1/(r^2c^3\sigma_2) & 1/(r^2c^4) & 1/(r^3c^3)\\ 0 & -1/(r^3c^2\sigma_2) & 1/(r^3c^3) & 1/(r^4c^2) \end{pmatrix}$$

Proof

The proof is a standard application of theorem 6.1 of chapter 6 of Lehmann (1983). In Appendix A we establish the conditions needed to invoke Lehmann's theorem. ■

Based on Theorem 1, an approximate $1-\alpha$ confidence interval on r is given by

$$\hat{r} \pm z_{\alpha/2}\sqrt{\hat{d}_{44}/n}$$

where $z_{\alpha/2}$ is the appropriate critical value from a standard normal distribution, $\hat{d}_{44}$ is the 4th diagonal element of $I(\hat{\boldsymbol{\theta}})^{-1}$, and $I(\hat{\boldsymbol{\theta}})$ is $I(\boldsymbol{\theta})$ with σ_1, σ_2, c, and r replaced by $\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{c}$, and $\hat{r}$.

4 Likelihood-Ratio–Based Confidence Interval on the Ratio of Two Coefficients of Variation

In Section 3 we obtained a confidence interval on the ratio of two coefficients of variation by establishing the asymptotic normality of the estimated parameter vector. In this section we take a likelihood ratio approach to this problem. That is, a 95% (for example) confidence interval is simply the collection of those ratio values that are not rejected at a 0.05 significance level by the test. We have found via simulations that actual confidence levels approach nominal confidence levels more rapidly when we take this likelihood ratio approach. Thus our Web programs implement only the material in Section 4. However, we include the Section 3 material for completeness.

We make the same assumptions as those made in Section 3. Then, in the notation of Section 18.3 of Appendix H, we have the following theorem.

Theorem 2

Provided that the ratio of coefficients of variation, r, equals r_0,

$$2(\ln L(\hat{\boldsymbol{\theta}}_n) - \ln L(\mathbf{g}(\hat{\boldsymbol{\nu}}_n))) \xrightarrow{D} \chi_1^2$$

where $\boldsymbol{\theta} \equiv (\sigma_1, \sigma_2, c, r)^T$, $\boldsymbol{\nu} \equiv (\sigma_1, \sigma_2, c)^T$, up to a constant

$$\ln L(\boldsymbol{\theta}) = -n_1 \ln(\sigma_1) - \sum_{i=1}^{n_1}(x_{i1} - \sigma_1/c)^2/(2\sigma_1^2) - n_2 \ln(\sigma_2) - \sum_{i=1}^{n_2}(x_{i2} - \sigma_2/(rc))^2/(2\sigma_2^2)$$

$\hat{\boldsymbol{\theta}}_n = (\hat{\sigma}_1, \hat{\sigma}_2, \hat{c}, \hat{r})^T$ is the solution of the unconstrained likelihood equations described in connection with Theorem 1,

$$\mathbf{g}(\boldsymbol{\nu}) = \begin{pmatrix} g_1(\nu_1, \nu_2, \nu_3) \\ g_2(\nu_1, \nu_2, \nu_3) \\ g_3(\nu_1, \nu_2, \nu_3) \\ g_4(\nu_1, \nu_2, \nu_3) \end{pmatrix} = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ r_0 \end{pmatrix} = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ c \\ r_0 \end{pmatrix}$$

and $\hat{\boldsymbol{\nu}}_n$ is the solution to the likelihood equations obtained in Appendix C.

Proof

Because two probability density functions are involved, Theorem 2 is a slight extension of a standard likelihood ratio result. To prove Theorem 2, we invoke Theorem H.1 of Appendix H. In Appendix B we establish the conditions needed to invoke Theorem H.1. ■

We have written a FORTRAN program that uses Theorem 2 to perform a test of the hypothesis that $r = r_0$. The program uses this test to obtain a confidence interval for r — those r_0 that are not rejected at an α significance level constitute a $1-\alpha$ confidence interval for r. A user of the program need only supply $n_1, n_2, \bar{x}_{\cdot 1}, \bar{x}_{\cdot 2}$, $\sqrt{\sum_{i=1}^{n_1}(x_{i1} - \bar{x}_{\cdot 1})^2/(n_1-1)}$, and $\sqrt{\sum_{i=1}^{n_2}(x_{i2} - \bar{x}_{\cdot 2})^2/(n_2-1)}$. The program can be run over the Web at `http://www1.fpl.fs.fed.us/covratio.html`

5 Confidence Interval on a Coefficient of Variation that Is Shared by Two Normally Distributed Populations

In this section, we obtain a confidence interval on a coefficient of variation that is shared by two normally distributed populations. (Lohrding (1969) attacked a related problem by parametrizing by μ_1, μ_2, and c, rather than by σ_1, σ_2, and c as we do here.) In Section 7 we generalize this problem and obtain a confidence interval on a coefficient of variation that is shared by k normally distributed populations. The special case of two populations is still worth investigating because in this case we can obtain a closed form solution for the estimate of the coefficient of variation. Further, the arguments needed to handle the k-population case are simple extensions of those needed to handle the two-population case, and these arguments are most simply presented in the two-population case.

We assume that we have n_1 observations, $x_{11}, \ldots, x_{n_1 1}$, from a $\mathrm{N}(\mu_1, \sigma_1^2)$ population, and n_2 observations, $x_{12}, \ldots, x_{n_2 2}$, from a $\mathrm{N}(\mu_2, \sigma_2^2)$ population, and that $\mu_1, \mu_2 > 0$. We assume that all these observations are statistically independent. Let $n \equiv n_1 + n_2$. We further assume that $n_1/n \to \lambda_1 > 0$ and $n_2/n \to \lambda_2 > 0$ as $n \to \infty$. We denote the shared coefficient of variation of the two populations by

$$c = \sigma_1/\mu_1 = \sigma_2/\mu_2$$

Then, we have the following theorem.

Theorem 3

$$\sqrt{n}\left(\begin{pmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \hat{c} \end{pmatrix} - \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ c \end{pmatrix}\right) \xrightarrow{D} \mathrm{N}(\mathbf{0}, I(\boldsymbol{\theta})^{-1})$$

where $\boldsymbol{\theta} \equiv (\sigma_1, \sigma_2, c)^T$, $\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{c}$ are derived in Appendix C (set r_0 in Appendix C to 1),

$$I(\boldsymbol{\theta}) = \lambda_1 I_1(\boldsymbol{\theta}) + \lambda_2 I_2(\boldsymbol{\theta})$$

and

$$I_1(\boldsymbol{\theta}) = \begin{pmatrix} (2 + 1/c^2)/\sigma_1^2 & 0 & -1/(c^3\sigma_1) \\ 0 & 0 & 0 \\ -1/(c^3\sigma_1) & 0 & 1/c^4 \end{pmatrix}$$

$$I_2(\boldsymbol{\theta}) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & (2 + 1/c^2)/\sigma_2^2 & -1/(c^3\sigma_2) \\ 0 & -1/(c^3\sigma_2) & 1/c^4 \end{pmatrix}$$

Proof

The proof is a standard application of theorem 6.1 of chapter 6 of Lehmann (1983). In Appendix C we establish the conditions needed to invoke Lehmann's theorem. ■

Based on Theorem 3, an approximate $1 - \alpha$ confidence interval on c is given by

$$\hat{c} \pm z_{\alpha/2}\sqrt{\hat{d}_{33}/n}$$

where $z_{\alpha/2}$ is the appropriate critical value from a standard normal distribution, $\hat{d}_{33}$ is the third diagonal element of $I(\hat{\boldsymbol{\theta}})^{-1}$, and $I(\hat{\boldsymbol{\theta}})$ is $I(\boldsymbol{\theta})$ with σ_1, σ_2, and c replaced by $\hat{\sigma}_1$, $\hat{\sigma}_2$, and $\hat{c}$. In Appendix I we establish that $d_{33} = c^4 + c^2/2$.

6 Likelihood Ratio Test of the Hypothesis that k Normally Distributed Populations Share the Same Coefficient of Variation

We assume that we have n_1 observations, $x_{11}, \ldots, x_{n_1 1}$, from a $\mathrm{N}(\mu_1, \sigma_1^2)$ population, n_2 observations, $x_{12}, \ldots, x_{n_2 2}$, from a $\mathrm{N}(\mu_2, \sigma_2^2)$ population, ..., and n_k observations, $x_{1k}, \ldots, x_{n_k k}$, from a $\mathrm{N}(\mu_k, \sigma_k^2)$ population, and that $\mu_1, \ldots, \mu_k > 0$. We assume that all these observations are statistically independent. Let $n \equiv n_1 + \ldots + n_k$. We further assume that $n_j/n \to \lambda_j > 0$ as $n \to \infty$ for $j = 1, \ldots, k$. We denote the shared coefficient of variation of the k populations by

$$c = \sigma_1/\mu_1 = \cdots = \sigma_k/\mu_k$$

Then, in the notation of Section 18.3 of Appendix H, we have the following theorem.

Theorem 4

Provided that $\sigma_1/\mu_1 = \cdots = \sigma_k/\mu_k$,

$$2(\ln L(\hat{\boldsymbol{\theta}}_n) - \ln L(\mathbf{g}(\boldsymbol{\nu}_{n,\text{Newt}}))) \stackrel{D}{\to} \chi^2_{k-1}$$

where $\boldsymbol{\theta} \equiv (\mu_1, \sigma_1, \ldots, \mu_k, \sigma_k)^T$, $\boldsymbol{\nu} \equiv (\sigma_1, \ldots, \sigma_k, c)^T$ [note that in this section the parameter vector is given by $\boldsymbol{\theta} \equiv (\mu_1, \sigma_1, \ldots, \mu_k, \sigma_k)^T$ while in Sections 5, 7, and 8 $\boldsymbol{\theta} \equiv (\sigma_1, \ldots, \sigma_k, c)^T$], up to a constant

$$\ln L(\boldsymbol{\theta}) = \sum_{j=1}^{k} \left(-n_j \ln(\sigma_j) - \sum_{i=1}^{n_j} (x_{ij} - \mu_j)^2 / (2\sigma_j^2) \right)$$

$\hat{\boldsymbol{\theta}}_n$ is the standard solution of the unconstrained likelihood equations, that is

$$\hat{\boldsymbol{\theta}}_n = \begin{pmatrix} \bar{x}_{\cdot 1} \\ s_1 \\ \vdots \\ \bar{x}_{\cdot k} \\ s_k \end{pmatrix}$$

where

$$s_j = \sqrt{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{\cdot j})^2 / n_j}$$

$$\mathbf{g}(\boldsymbol{\nu}) = \begin{pmatrix} \nu_1/\nu_{k+1} \\ \nu_1 \\ \vdots \\ \nu_k/\nu_{k+1} \\ \nu_k \end{pmatrix} = \begin{pmatrix} \sigma_1/c \\ \sigma_1 \\ \vdots \\ \sigma_k/c \\ \sigma_k \end{pmatrix}$$

$\boldsymbol{\nu}_{n,\text{Newt}}$ is the Newton estimator of $(\sigma_1, \ldots, \sigma_k, c)^T$ given by

$$\boldsymbol{\nu}_{n,\text{Newt}} = -\left[\frac{\partial^2 \ln L}{\partial \nu_l \partial \nu_m} \right]^{-1} |_{\boldsymbol{\nu}_{n,c}} \begin{pmatrix} \partial \ln L / \partial \nu_1 \\ \vdots \\ \partial \ln L / \partial \nu_{k+1} \end{pmatrix} |_{\boldsymbol{\nu}_{n,c}} + \boldsymbol{\nu}_{n,c} \qquad (1)$$

where $\boldsymbol{\nu}_{n,c}$ is any $\sqrt{n}$-consistent estimator of $(\sigma_1, \ldots, \sigma_k, c)^T$ (such as $(s_1, \ldots, s_k, \hat{c})^T$ where $\hat{c}$ is described in connection with Equation 78 in Subsection 15.7). The partial derivatives in Equation 1 (actually, the partials of the $\ln f_j(x; \boldsymbol{\theta})$'s where $\ln L = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \ln f_j(x_{ij}; \boldsymbol{\theta})$) are listed in Appendix C, and a simple technique for solving the equation is provided in Appendix I.

Proof

Because k probability density functions are involved, and because we are dealing with a Newton one-step estimator, Theorem 4 is a slight extension of a standard likelihood ratio result. To prove Theorem 4, we invoke Theorem H.4 of Appendix H. (Note that in applying Theorem H.4 we can use

$\hat{\boldsymbol{\theta}}_n$ as our $\sqrt{n}$-consistent estimator in the unconstrained case. Then the Newton 1-step estimator is again $\hat{\boldsymbol{\theta}}_n$.) In Appendix E we establish the conditions needed to invoke Theorem H.4. ■

We have written a FORTRAN program[3] that uses Theorem 4 to perform a test of the hypothesis that $\sigma_1/\mu_1 = \ldots = \sigma_k/\mu_k$. The user need only supply $n_1, \ldots, n_k$, $\bar{x}_{\cdot 1}, \ldots, \bar{x}_{\cdot k}$, $\sqrt{\sum_{i=1}^{n_1}(x_{i1} - \bar{x}_{\cdot 1})^2/(n_1 - 1)}, \ldots, \sqrt{\sum_{i=1}^{n_k}(x_{ik} - \bar{x}_{\cdot k})^2/(n_k - 1)}$. The program can be run over the Web at `http://www1.fpl.fs.fed.us/covtestk.html`.

7 Confidence Interval on a Coefficient of Variation that Is Shared by k Normally Distributed Populations

In Sections 7 and 8 we present two approaches to obtaining a confidence interval on a coefficient of variation that is shared by k normally distributed populations. In Section 7 we obtain the asymptotic distribution of the estimate of the coefficient of variation. A 95% (for example) confidence interval is then just the usual "estimate plus or minus two (1.96) standard deviations." See below for details. In Section 8 we take a likelihood ratio test approach. In this case, a 95% (for example) confidence interval is simply the collection of those coefficient of variation values that are not rejected at a 0.05 significance level by the test. We have found via simulations that actual confidence levels approach nominal confidence levels more rapidly when we take the likelihood ratio approach. Thus our Web programs implement only the material in Section 8. However, we include the Section 7 material for completeness.

We make the same assumptions as those made in Section 6. However, here the parameter vector is given by $\boldsymbol{\theta} \equiv (\sigma_1, \ldots, \sigma_k, c)^T$ (as opposed to the vector $(\mu_1, \sigma_1, \ldots, \mu_k, \sigma_k)^T$ of Section 6).

Then, we have the following theorem.

Theorem 5

$$\sqrt{n}(\boldsymbol{\theta}_{n,\mathrm{Newt}} - \boldsymbol{\theta}) \xrightarrow{D} \mathrm{N}(\mathbf{0}, I(\boldsymbol{\theta})^{-1})$$

where $\boldsymbol{\theta}_{n,\mathrm{Newt}}$ is the Newton estimator of $(\sigma_1, \ldots, \sigma_k, c)^T$ given by

$$\boldsymbol{\theta}_{n,\mathrm{Newt}} \equiv -\left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m}\right]^{-1} |_{\boldsymbol{\theta}_{n,c}} \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_{k+1} \end{pmatrix} |_{\boldsymbol{\theta}_{n,c}} + \boldsymbol{\theta}_{n,c} \tag{2}$$

$\boldsymbol{\theta}_{n,c}$ is any $\sqrt{n}$-consistent estimator of $(\sigma_1, \ldots, \sigma_k, c)^T$ (such as $(s_1, \ldots, s_k, \hat{c})^T$ where $\hat{c}$ is described in connection with Equation 78 in Subsection 15.7), and

$$I(\boldsymbol{\theta}) = \sum_{j=1}^{k} \lambda_j I_j(\boldsymbol{\theta})$$

where the j,j th element of $I_j(\boldsymbol{\theta})$ is $(2 + 1/c^2)/\sigma_j^2$, the $j, k+1$ th and $k+1, j$ th elements of $I_j(\boldsymbol{\theta})$ are $-1/(c^3\sigma_j)$, the $k+1, k+1$ th element is $1/c^4$, and the remaining elements are 0. The partial derivatives in Equation 2 (actually, the partials of the $\ln f_j(x;\boldsymbol{\theta})$'s where $\ln L = \sum_{j=1}^{k}\sum_{i=1}^{n_j} \ln f_j(x_{ij};\boldsymbol{\theta})$) are listed in Appendix C, and a simple technique for solving the equation is provided in Appendix I.

[3]We note that in this program, we start with the $\sqrt{n}$-consistent estimator given by $(s_1, \ldots, s_k, \hat{c})$, perform a limited number of backtracking Newton steps (which still leaves us with a $\sqrt{n}$-consistent estimator), and then do a final full Newton step.

Proof

Because k probability density functions are involved, and because we are dealing with a Newton one-step estimator, Theorem 5 is a slight extension of a standard efficient likelihood estimator result. To prove Theorem 5, we invoke Corollary 1 to Lemma H.8 of Appendix H. In Appendix F we establish the conditions needed to invoke the corollary. ■

Based on Theorem 5, an approximate $1-\alpha$ confidence interval on c is given by

$$\hat{c} \pm z_{\alpha/2}\sqrt{\hat{d}_{k+1,k+1}/n}$$

where $\hat{c}$ is the $k+1$ th element of $\boldsymbol{\theta}_{n,\text{Newt}}$, $z_{\alpha/2}$ is the appropriate critical value from a standard normal distribution, $\hat{d}_{k+1,k+1}$ is the $k+1$ th diagonal element of $I(\hat{\boldsymbol{\theta}})^{-1}$, and $I(\hat{\boldsymbol{\theta}})$ is $I(\boldsymbol{\theta})$ with $(\sigma_1,\ldots,\sigma_k,c)^T$ replaced by $\boldsymbol{\theta}_{n,\text{Newt}}$. In Appendix I we establish that $d_{k+1,k+1} = c^4 + c^2/2$.

8 Likelihood-Ratio-Based Confidence Interval on a Coefficient of Variation that Is Shared by k Normally Distributed Populations

In Section 7, we obtained a confidence interval on a coefficient of variation shared by k normally distributed populations by establishing the asymptotic normality of the Newton one-step estimator of the parameter vector. In this section we take a likelihood ratio approach to this problem. That is, a 95% (for example) confidence interval is simply the collection of those coefficient of variation values that are not rejected at a 0.05 significance level by the test. We have found via simulations that actual confidence levels approach nominal confidence levels more rapidly when we take this likelihood ratio approach. Thus our Web programs implement only the material in Section 8. However, we include the Section 7 material for completeness.

We make the same assumptions as those made in Section 7. Then, in the notation of Section 18.3 of Appendix H, we have the following theorem.

Theorem 6

Provided that $c = c_0$,

$$2(\ln L(\boldsymbol{\theta}_{n,\text{Newt}}) - \ln L(\mathbf{g}(\hat{\boldsymbol{\nu}}_n))) \xrightarrow{D} \chi^2_1$$

where $\boldsymbol{\theta} \equiv (\sigma_1,\ldots,\sigma_k,c)^T$, $\boldsymbol{\nu} \equiv (\sigma_1,\ldots,\sigma_k)^T$, up to a constant

$$\ln L(\boldsymbol{\theta}) = \sum_{j=1}^{k}\left(-n_j\ln(\sigma_j) - \sum_{i=1}^{n_j}(x_{ij}-\sigma_j/c)^2/(2\sigma_j^2)\right)$$

$\boldsymbol{\theta}_{n,\text{Newt}}$ is the Newton estimator of $(\sigma_1,\ldots,\sigma_k,c)^T$ given by

$$\boldsymbol{\theta}_{n,\text{Newt}} \equiv -\left[\frac{\partial^2 \ln L}{\partial\theta_l\theta_m}\right]^{-1}\Big|_{\boldsymbol{\theta}_{n,c}} \begin{pmatrix} \partial\ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_{k+1}\end{pmatrix}\Big|_{\boldsymbol{\theta}_{n,c}} + \boldsymbol{\theta}_{n,c} \tag{3}$$

where $\boldsymbol{\theta}_{n,c}$ is any $\sqrt{n}$-consistent estimator of $(\sigma_1,\ldots,\sigma_k,c)^T$ (such as $(s_1,\ldots,s_k,\hat{c})^T$ where $\hat{c}$ is described in connection with Equation 78 in Subsection 15.7), and $\hat{\boldsymbol{\nu}}_n$ is the solution of the constrained likelihood equations obtained in Appendix G. (The constraint is $c = c_0$.)

The partial derivatives in Equation 3 (actually, the partials of the $\ln f_j(x;\boldsymbol{\theta})$'s where $\ln L = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\ln f_j(x_{ij};\boldsymbol{\theta})$) are listed in Appendix C, and a simple technique for solving the equation is provided in Appendix I.

Proof

Because k probability density functions are involved, and because we are dealing with a Newton one-step estimator, Theorem 6 is a slight extension of a standard likelihood ratio result. To prove Theorem 6, we invoke Theorem H.4 of Appendix H. (Note that in applying Theorem H.4 we can use $\hat{\boldsymbol{\nu}}_n$ as our $\sqrt{n}$-consistent estimator in the constrained case. Then the Newton 1-step estimator is again $\hat{\boldsymbol{\nu}}_n$.) In Appendix G we establish the conditions needed to invoke Theorem H.4. ■

We have written a FORTRAN program[4] that uses Theorem 6 to perform a test of the hypothesis that $c = c_0$. The program uses this test to obtain a confidence interval for c — those c_0 that are not rejected at a α significance level constitute a $1-\alpha$ confidence interval for c. A user of the program need only supply $n_1, \ldots, n_k$, $\bar{x}_{\cdot 1}, \ldots, \bar{x}_{\cdot k}$, $\sqrt{\sum_{i=1}^{n_1}(x_{i1}-\bar{x}_{\cdot 1})^2/(n_1-1)}$, …, $\sqrt{\sum_{i=1}^{n_k}(x_{ik}-\bar{x}_{\cdot k})^2/(n_k-1)}$. The program can be run over the Web at `http://www1.fpl.fs.fed.us/covconfk.html`.

9 Small Sample Tools and Further Research

The theory that underlies Theorems 1 through 6 is asymptotic theory. That is, it yields good approximations for large data sets but poorer approximations for small data sets. In our case, for small data sets, it leads to confidence intervals that are too narrow and to tests of hypotheses that reject true null hypotheses too frequently. We are currently engaged in research that should lead to improved small sample approximations. In the interim, however, we have provided a simulation-based fix to the problem. Our Web-based programs (see `http://www1.fpl.fs.fed.us/covconfk.html`) perform the theoretical calculations needed to obtain confidence intervals or to test hypotheses. However, they also perform tests and calculate confidence intervals that are based on simulations. In particular small sample critical values are obtained via simulations in which 10,000 samples are generated from the normal distributions estimated from the original data. The value of the appropriate likelihood ratio statistic is calculated for each of these samples. Empirical estimates of the 90th, 95th, and 99th percentiles of the distribution of the likelihood ratio statistic are then obtained from these 10,000 values and used to perform small sample tests and to calculate small-sample confidence intervals.

These simulations might be suspect because they make use of estimated population parameters rather than the true (and unknown) population parameters. However we have performed "simulations of simulations" that indicate that this small sample approach works quite well. In particular for a variety of cases we have performed 10,000 trial simulations in which we generated a sample data set from known normal distributions, estimated population parameters, drew 10,000 samples from the estimated populations, calculated estimates of the percentiles of the likelihood ratio statistic, and then used these to perform tests and obtain confidence intervals. Test sizes and confidence interval coverages were very near nominal for the small sample simulation approach. A subset of results from this simulation of simulations appears in Tables 3 – 5. For these three tables, $k = 2$ and $n_1 = n_2$. The tables suggest that the small sample approach works well.

Our current Web programs report both asymptotic results and simulation results. As we have noted, simulation works well even for small samples. However, for larger samples, simulations can become quite time consuming. (For 10 samples of size 60, the hypothesis test program takes approximately 6.9 seconds to report. The confidence interval program takes about 7.3 seconds to report.) Hence the need for asymptotic results. We are currently engaged in performing a wide-

[4]We note that in this program, we start with the $\sqrt{n}$-consistent estimator given by $(s_1, \ldots, s_k, \hat{c})$, perform a limited number of backtracking Newton steps (which still leaves us with a $\sqrt{n}$-consistent estimator), and then do a final full Newton step.

ranging set of size/power studies. These studies suggest that for $n_j > 30$, the two approaches essentially coincide. After we have become convinced that this holds generally, we will modify the Web program so that simulations are not performed for larger sample sizes. Instead, for these larger sample sizes, only the asymptotic results will be reported. This will improve the program's performance.

10 Summary

We have developed asymptotic theory that permits us to address three normal distribution coefficient of variation estimation or testing problems: obtain a confidence interval on the ratio of two coefficients of variation, perform a test of the hypothesis that the coefficients of variation associated with k populations are all equal, and obtain a confidence interval on a coefficient of variation that is shared by k populations. We have developed Web-based computer programs that implement these large sample techniques, and also provide simulation results that are valid for small samples. These programs can be accessed at the following web addresses: `http://www1.fpl.fs.fed.us/covratio.html`, `http://www1.fpl.fs.fed.us/covtestk.html`, and `http://www1.fpl.fs.fed.us/covconfk.html`.

REFERENCES

Bennett, B.M. (1976), "On an Approximate Test for Homogeneity of Coefficients of Variation," in *Contributions to Applied Statistics: Dedicated to Professor Arthur Linder*, edited by Walter John Ziegler, Birkhauser, Stuttgart, 169 – 171.

Dennis, J.E. and Schnabel, R.B. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey.

Doornbos, R. and Dijkstra, J.B. (1983), "A Multi Sample Test for the Equality of Coefficients of Variation in Normal Populations," *Communications in Statistics — Simulation and Computation*, 12, 147–158.

Feltz, C.J. and Miller, G.E. (1996), "An Asymptotic Test for the Equality of Coefficients of Variation from k Populations," *Statistics in Medicine*, 15, 647–658.

Fung, W.K. and Tsang, T.S. (1998), "A Simulation Study Comparing Tests for the Equality of Coefficients of Variation," *Statistics in Medicine*, 17, 2003–2014.

Lehmann, E.L. (1983), *Theory of Point Estimation*, John Wiley, New York.

Lehmann, E.L. and Romano, J.P. (2005), *Testing Statistical Hypotheses, Third Edition*, Springer, New York.

Lohrding, R.K. (1969), "A Test of Equality of Two Normal Population Means Assuming Homogeneous Coefficients of Variation," *Annals of Mathematical Statistics*, 40, 1374–1385.

Miller, G.E. (1991), "Asymptotic Test Statistics for Coefficients of Variation," *Communications in Statistics — Theory and Methods*, 20, 3351–3363.

Nairy, K.S. and Rao, K.A. (2003), "Tests of Coefficients of Variation of Normal Population," *Communications in Statistics — Simulation and Computation*, 32, 641–661.

Rao, C.R. (1973), *Linear Statistical Inference and Its Applications, Second Edition*, John Wiley, New York.

Searle, S.R. (1982), *Matrix Algebra Useful for Statistics*, John Wiley, New York.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley, New York.

Shafer, N.J. and Sullivan, J.A. (1986), "A Simulation Study of a Test for the Equality of the Coefficients of Variation," *Communications in Statistics — Simulation and Computation*, 15, 681–695.

Silvey, S.D. (1959), "The Lagrangian Multiplier Test," *The Annals of Mathematical Statistics*, 30, 389–407.

Vangel, M.G. (1996), "Confidence Intervals for a Normal Coefficient of Variation," *The American Statistician*, 50, 21–26.

Tian, L. (2005), "Inferences on the Common Coefficient of Variation," *Statistics in Medicine*, 24, 2213–2220.

Verrill, S. (2003), "Confidence Bounds for Normal and Lognormal Distribution Coefficients of Variation," USDA Forest Products Laboratory Research Paper FPL-RP-609.

Wald, A. (1943), "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426–482.

Weerahandi, S. (1993), "Generalized Confidence Intervals," *Journal of the American Statistical Association*, 88, 899–905.

Wilks, S.S. (1938), "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *The Annals of Mathematical Statistics*, 9, 60–62.

11 Appendix A — Verification of the Conditions Needed to Establish Theorem 1

To invoke Lehmann's (1983) theorem 6.1 to prove our Theorem 1, we must establish conditions **(A0)** through **(A2)** and **(A)** through **(D)** of Appendix H.

In the notation of Section 18.1, in the case under consideration, $k = 2$, $f_1(x;\boldsymbol{\theta}) = \exp\left(-(x-\sigma_1/c)^2/(2\sigma_1^2)\right) / \left(\sigma_1\sqrt{2\pi}\right)$, and $f_2(x;\boldsymbol{\theta}) = \exp\left(-(x-\sigma_2/(rc))^2/(2\sigma_2^2)\right) / \left(\sigma_2\sqrt{2\pi}\right)$.

It is clear that conditions **(A0)** through **(A2)** and **(A)** hold.

11.1 Condition (B)

We have

$$\frac{\partial \ln f_1(x;\boldsymbol{\theta})}{\partial \sigma_1} = \frac{-1}{\sigma_1} + \frac{x-\sigma_1/c}{c\sigma_1^2} + \frac{(x-\sigma_1/c)^2}{\sigma_1^3} \quad (4)$$

$$\frac{\partial \ln f_1(x;\boldsymbol{\theta})}{\partial c} = \frac{-(x-\sigma_1/c)}{c^2\sigma_1} \quad (5)$$

and

$$\frac{\partial \ln f_1(x;\boldsymbol{\theta})}{\partial \sigma_2} = \frac{\partial \ln f_1(x;\boldsymbol{\theta})}{\partial r} = 0$$

Thus

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial \sigma_1}\right) = E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial c}\right) = E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial \sigma_2}\right) = E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial r}\right) = 0 \quad (6)$$

Also,

$$\frac{\partial \ln f_2(x;\boldsymbol{\theta})}{\partial \sigma_2} = \frac{-1}{\sigma_2} + \frac{x-\sigma_2/(rc)}{rc\sigma_2^2} + \frac{(x-\sigma_2/(rc))^2}{\sigma_2^3} \quad (7)$$

$$\frac{\partial \ln f_2(x;\boldsymbol{\theta})}{\partial c} = \frac{-(x-\sigma_2/(rc))}{rc^2\sigma_2} \quad (8)$$

$$\frac{\partial \ln f_2(x;\boldsymbol{\theta})}{\partial r} = \frac{-(x-\sigma_2/(rc))}{r^2c\sigma_2} \quad (9)$$

$$\frac{\partial \ln f_2(x;\boldsymbol{\theta})}{\partial \sigma_1} = 0$$

Thus

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial \sigma_2}\right) = E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial c}\right) = E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial r}\right) = E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial \sigma_1}\right) = 0 \quad (10)$$

Next, we have

$$\frac{\partial^2 \ln f_1(x;\boldsymbol{\theta})}{\partial \sigma_1^2} = \frac{1}{\sigma_1^2} - \frac{1}{c^2\sigma_1^2} - \frac{4(x-\sigma_1/c)}{c\sigma_1^3} - \frac{3(x-\sigma_1/c)^2}{\sigma_1^4} \quad (11)$$

$$\frac{\partial^2 \ln f_1(x;\boldsymbol{\theta})}{\partial c\partial \sigma_1} = \frac{1}{c^3\sigma_1} + \frac{x-\sigma_1/c}{c^2\sigma_1^2} \quad (12)$$

$$\frac{\partial^2 \ln f_1(x;\boldsymbol{\theta})}{\partial c^2} = -\frac{1}{c^4} + \frac{2(x-\sigma_1/c)}{c^3\sigma_1} \quad (13)$$

$$\frac{\partial^2 \ln f_2(x;\boldsymbol{\theta})}{\partial \sigma_2^2} = \frac{1}{\sigma_2^2} - \frac{1}{r^2c^2\sigma_2^2} - \frac{4(x-\sigma_2/(rc))}{rc\sigma_2^3} - \frac{3(x-\sigma_2/(rc))^2}{\sigma_2^4} \quad (14)$$

$$\frac{\partial^2 \ln f_2(x;\boldsymbol{\theta})}{\partial c\partial \sigma_2} = \frac{1}{r^2c^3\sigma_2} + \frac{x-\sigma_2/(rc)}{rc^2\sigma_2^2} \quad (15)$$

$$\frac{\partial^2 \ln f_2(x;\boldsymbol{\theta})}{\partial c^2} = -\frac{1}{r^2c^4} + \frac{2(x-\sigma_2/(rc))}{rc^3\sigma_2} \quad (16)$$

$$\frac{\partial^2 \ln f_2(x;\boldsymbol{\theta})}{\partial r^2} = -\frac{1}{r^4c^2} + \frac{2(x-\sigma_2/(rc))}{r^3c\sigma_2} \quad (17)$$

$$\frac{\partial^2 \ln f_2(x;\boldsymbol{\theta})}{\partial r\partial \sigma_2} = \frac{1}{r^3c^2\sigma_2} + \frac{x-\sigma_2/(rc)}{r^2c\sigma_2^2} \quad (18)$$

$$\frac{\partial^2 \ln f_2(x;\boldsymbol{\theta})}{\partial r\partial c} = -\frac{1}{r^3c^3} + \frac{x-\sigma_2/(rc)}{r^2c^2\sigma_2} \quad (19)$$

Thus,

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_1(X;\boldsymbol{\theta})}{\partial\sigma_1^2}\right) = -\left(2+1/c^2\right)/\sigma_1^2 \quad (20)$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_1(X;\boldsymbol{\theta})}{\partial c\partial\sigma_1}\right) = \frac{1}{c^3\sigma_1} \quad (21)$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_1(X;\boldsymbol{\theta})}{\partial c^2}\right) = -\frac{1}{c^4} \quad (22)$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_2(X;\boldsymbol{\theta})}{\partial\sigma_2^2}\right) = -\left(2+1/(rc)^2\right)/\sigma_2^2 \quad (23)$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_2(X;\boldsymbol{\theta})}{\partial c\partial\sigma_2}\right) = \frac{1}{r^2c^3\sigma_2} \quad (24)$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_2(X;\boldsymbol{\theta})}{\partial c^2}\right) = -\frac{1}{r^2c^4} \quad (25)$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_2(X;\boldsymbol{\theta})}{\partial r^2}\right) = -\frac{1}{r^4c^2} \quad (26)$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_2(X;\boldsymbol{\theta})}{\partial r\partial\sigma_2}\right) = \frac{1}{r^3c^2\sigma_2} \quad (27)$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_2(X;\boldsymbol{\theta})}{\partial r\partial c}\right) = -\frac{1}{r^3c^3} \quad (28)$$

We also have

$$\begin{aligned} E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial\sigma_1}\times\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial\sigma_1}\right) &= E_{\boldsymbol{\theta}}\left(\frac{1}{\sigma_1^2}-\frac{2(X-\sigma_1/c)}{c\sigma_1^3}-\frac{2(X-\sigma_1/c)^2}{\sigma_1^4}\right. \quad (29)\\ &\quad \left.+\frac{(X-\sigma_1/c)^2}{c^2\sigma_1^4}+\frac{2(X-\sigma_1/c)^3}{c\sigma_1^5}+\frac{(X-\sigma_1/c)^4}{\sigma_1^6}\right)\\ &= \left(2+1/c^2\right)/\sigma_1^2 \end{aligned}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial c}\times\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial\sigma_1}\right) = E_{\boldsymbol{\theta}}\left(\frac{(X-\sigma_1/c)}{c^2\sigma_1^2}-\frac{(X-\sigma_1/c)^2}{c^3\sigma_1^3}-\frac{(X-\sigma_1/c)^3}{c^2\sigma_1^4}\right) = -\frac{1}{c^3\sigma_1} \quad (30)$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial c}\times\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial c}\right) = E_{\boldsymbol{\theta}}\left(\frac{(X-\sigma_1/c)^2}{c^4\sigma_1^2}\right) = \frac{1}{c^4} \quad (31)$$

$$\begin{aligned} E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial\sigma_2}\times\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial\sigma_2}\right) &= E_{\boldsymbol{\theta}}\left(\frac{1}{\sigma_2^2}-\frac{2(X-\sigma_2/(rc))}{rc\sigma_2^3}-\frac{2(X-\sigma_2/(rc))^2}{\sigma_2^4}\right.\\ &\quad \left.+\frac{(X-\sigma_2/(rc))^2}{r^2c^2\sigma_2^4}+\frac{2(X-\sigma_2/(rc))^3}{rc\sigma_2^5}+\frac{(X-\sigma_2/(rc))^4}{\sigma_2^6}\right)\\ &= \left(2+1/(rc)^2\right)/\sigma_2^2 \quad (32) \end{aligned}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial c} \times \frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial \sigma_2}\right) = E_{\boldsymbol{\theta}}\left(\frac{(X-\sigma_2/(rc))}{rc^2\sigma_2^2} - \frac{(X-\sigma_2/(rc))^2}{r^2c^3\sigma_2^3} - \frac{(X-\sigma_2/(rc))^3}{rc^2\sigma_2^4}\right) = -\frac{1}{r^2c^3\sigma_2} \tag{33}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial c} \times \frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial c}\right) = E_{\boldsymbol{\theta}}\left(\frac{(X-\sigma_2/(rc))^2}{r^2c^4\sigma_2^2}\right) = \frac{1}{r^2c^4} \tag{34}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial r} \times \frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial r}\right) = E_{\boldsymbol{\theta}}\left(\frac{(X-\sigma_2/(rc))^2}{r^4c^2\sigma_2^2}\right) = \frac{1}{r^4c^2} \tag{35}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial r} \times \frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial \sigma_2}\right) = E_{\boldsymbol{\theta}}\left(\frac{(X-\sigma_2/(rc))}{r^2c\sigma_2^2} - \frac{(X-\sigma_2/(rc))^2}{r^3c^2\sigma_2^3} - \frac{(X-\sigma_2/(rc))^3}{r^2c\sigma_2^4}\right) = -\frac{1}{r^3c^2\sigma_2} \tag{36}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial r} \times \frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial c}\right) = E_{\boldsymbol{\theta}}\left(\frac{(X-\sigma_2/(rc))^2}{r^3c^3\sigma_2^2}\right) = \frac{1}{r^3c^3} \tag{37}$$

Results 6, 10, and 20 through 37 establish condition **(B)**.

11.2 Condition (C)

Let $\mathbf{x}^T \equiv (x_1, x_2, x_3, x_4)$. Then (from results 29 through 37)

$$\begin{aligned}\mathbf{x}^T I(\boldsymbol{\theta})\mathbf{x} &= \lambda_1 x_1^2(2+1/c^2)/\sigma_1^2 + \lambda_2 x_2^2(2+1/(rc)^2)/\sigma_2^2 + (\lambda_1/c^4 + \lambda_2/(r^2c^4))x_3^2 + \lambda_2 x_4^2/(r^4c^2) \\ &\quad -2\lambda_1 x_1 x_3/(c^3\sigma_1) - 2\lambda_2 x_2 x_3/(r^2c^3\sigma_2) - 2\lambda_2 x_2 x_4/(r^3c^2\sigma_2) + 2\lambda_2 x_3 x_4/(r^3c^3) \\ &= 2\lambda_1 x_1^2/\sigma_1^2 + \lambda_1\left(x_1/(c\sigma_1) - x_3/c^2\right)^2 + 2\lambda_2 x_2^2/\sigma_2^2 + \lambda_2\left(x_2/(rc\sigma_2) - x_3/(rc^2) - x_4/(r^2c)\right)^2\end{aligned}$$

which is clearly positive unless x_1, x_2, x_3, and x_4 are all 0. Thus condition **(C)** is established.

11.3 Condition (D)

From results 11 through 19 (actually from the third-order derivatives that are based on them), the fact that σ_1, σ_2, c, and r are bounded away from 0 if they are in a sufficiently small open neighborhood of $\boldsymbol{\theta}_0$, and the fact that normal distributions have finite absolute moments, it is clear that condition **(D)** holds.

11.4 A Subtlety

Lehmann's theorem only guarantees the existence of a vector of solutions of the likelihood equations that is asymptotically normal. In our Theorem 1 we are dealing with a vector of maximum likelihood estimates. It is well-known that $\bar{x}_{\cdot 1}$, $\sqrt{\sum_{i=1}^{n_1}(x_{i1} - \bar{x}_{\cdot 1})^2/n_1}$, $\bar{x}_{\cdot 2}$, and $\sqrt{\sum_{i=1}^{n_2}(x_{i2} - \bar{x}_{\cdot 2})^2/n_2}$ are the unique solutions to the likelihood equations when the likelihood is parametrized by μ_1, σ_1, μ_2, and

σ_2. The fact that $\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{c}$, $\hat{r}$ must be the unique solution of the likelihood equations when we parametrize by σ_1, σ_2, c, and r then follows from the relations among the partial derivatives. For example,

$$\begin{aligned}\partial \ln L/\partial c &= \partial \ln L/\partial \mu_1 \times \partial \mu_1/\partial c \\ &+\partial \ln L/\partial \sigma_1 \times \partial \sigma_1/\partial c + \partial \ln L/\partial \mu_2 \times \partial \mu_2/\partial c + \partial \ln L/\partial \sigma_2 \times \partial \sigma_2/\partial c\end{aligned} \tag{38}$$

and

$$\begin{aligned}\partial \ln L/\partial \mu_1 &= \partial \ln L/\partial \sigma_1 \times \partial \sigma_1/\partial \mu_1 \\ &+\partial \ln L/\partial \sigma_2 \times \partial \sigma_2/\partial \mu_1 + \partial \ln L/\partial c \times \partial c/\partial \mu_1 + \partial \ln L/\partial r \times \partial r/\partial \mu_1\end{aligned} \tag{39}$$

Equation 38 and the related equations for $\partial \ln L/\partial \sigma_1$, $\partial \ln L/\partial \sigma_2$, and $\partial \ln L/\partial r$ assure us that $(\hat{\sigma}_1, \hat{\sigma}_2, \hat{c}, \hat{r})^T$ is indeed a solution of the likelihood equations. Equation 39 and the related equations for $\partial \ln L/\partial \sigma_1$, $\partial \ln L/\partial \mu_2$, and $\partial \ln L/\partial \sigma_2$ assure us that $(\hat{\sigma}_1, \hat{\sigma}_2, \hat{c}, \hat{r})^T$ must be the unique solution of the likelihood equations.

12 Appendix B — Verification of the Conditions Needed to Establish Theorem 2

In the notation of Section 18.3, $\boldsymbol{\theta}^T = (\sigma_1, \sigma_2, c, r)$ and $\boldsymbol{\nu}^T = (\sigma_1, \sigma_2, c)$ where the σ's are the standard deviations of the two populations, c is the coefficient of variation of the first population, and rc is the coefficient of variation of the second population. Under the null hypothesis, $r = r_0$.

To invoke Theorem H.1 to prove our Theorem 2, we must establish conditions **(A0)** through **(A2)**, **(A)** through **(D)**, and **(E1)** through **(I)** of Appendix H.

In the notation of Section 18.1, in the case under consideration,
$f_1(x;\boldsymbol{\theta}) = \exp\left(-(x-\sigma_1/c)^2/(2\sigma_1^2)\right) / \left(\sigma_1\sqrt{2\pi}\right)$ and $f_2(x;\boldsymbol{\theta}) = \exp\left(-(x-\sigma_2/(rc))^2/(2\sigma_2^2)\right) / \left(\sigma_2\sqrt{2\pi}\right)$.

Conditions **(A0)** through **(A2)** and **(A)** through **(D)** are established in Appendix A.

12.1 Condition (E1)

In the notation of Section 18.3, we have

$$R_1(\boldsymbol{\theta}) = r - r_0$$

and

$$\begin{aligned}\theta_1 &= \sigma_1 = g_1(\nu_1, \nu_2, \nu_3) = \nu_1 \\ \theta_2 &= \sigma_2 = g_2(\nu_1, \nu_2, \nu_3) = \nu_2 \\ \theta_3 &= c = g_3(\nu_1, \nu_2, \nu_3) = \nu_3 \\ \theta_4 &= r = g_4(\nu_1, \nu_2, \nu_3) = r_0\end{aligned}$$

The equivalence of these two forms of the constraints is clear. They both permit the σ's and c to vary freely from 0 to ∞ and restrict r to the single value r_0.

12.2 Condition (E2)

Clear.

12.3 Condition (E3)

Clear.

12.4 Condition (E4)

$$\mathbf{C}(\boldsymbol{\theta}_0) = (0, 0, 0, 1)_{1\times 4}$$

which is clearly of rank 1, which equals r in the notation of Section 18.3. (Again, please do not be confused by our awkward notation; that is, we are using r to denote both a parameter of the f_2 density and, in Section 18.3, the rank of a particular matrix.)

12.5 Condition (E5)

We have

$$\mathbf{D}(\boldsymbol{\nu}_0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}_{4\times 3}$$

which is clearly of rank $3 = 4 - 1$, which equals $s - r$ in the notation of Section 18.3. (Please do not be confused by our awkward notation; that is, we are using r to denote both a parameter of the f_2 density and, in Section 18.3, the rank of a particular matrix.)

12.6 Condition (F)

Clear.

12.7 Conditions (G) through (I)

These conditions are established (as conditions **(B)** through **D**) in Appendix C.

12.8 The solution to the likelihood equations in the constrained case

See Appendix C.

13 Appendix C — Verification of the Conditions Needed to Establish Theorem 3

To invoke Lehmann's (1983) theorem 6.1 to prove our Theorem 3, we must establish conditions **(A0)** through **(A2)** and **(A)** through **(D)** of Appendix H.

In the notation of Section 18.1, in the case under consideration, $k = 2$, $f_1(x; \boldsymbol{\theta}) = \exp\left(-(x - \sigma_1/c)^2/(2\sigma_1^2)\right) / \left(\sigma_1\sqrt{2\pi}\right)$, and $f_2(x; \boldsymbol{\theta}) = \exp\left(-(x - \sigma_2/c)^2/(2\sigma_2^2)\right) / \left(\sigma_2\sqrt{2\pi}\right)$. Because we also want to use the material below to establish Theorem 2, we generalize this slightly by taking $f_2(x; \boldsymbol{\theta}) = \exp(-(x - \sigma_2/(r_0 c))^2)/(\sigma_2\sqrt{2\pi})$ where $r_0 > 0$ is fixed. (Theorem 3 then corresponds to the $r_0 = 1$ case.)

It is clear that conditions **(A0)** through **(A2)** and **(A)** hold.

Conditions **(B)** through **(D)** follow as in Appendix A. However, for clarity and because the proofs of Theorems 2, 4, 5, and 6 refer to them, we provide the details here.

13.1 Condition (B)

We have

$$\frac{\partial \ln f_1(x;\boldsymbol{\theta})}{\partial \sigma_1} = \frac{-1}{\sigma_1} + \frac{x-\sigma_1/c}{c\sigma_1^2} + \frac{(x-\sigma_1/c)^2}{\sigma_1^3} \tag{40}$$

$$\frac{\partial \ln f_1(x;\boldsymbol{\theta})}{\partial c} = \frac{-(x-\sigma_1/c)}{c^2\sigma_1} \tag{41}$$

$$\frac{\partial \ln f_1(x;\boldsymbol{\theta})}{\partial \sigma_2} = 0$$

Thus

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial \sigma_1}\right) = E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial c}\right) = E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial \sigma_2}\right) = 0 \tag{42}$$

Similarly,

$$\frac{\partial \ln f_2(x;\boldsymbol{\theta})}{\partial \sigma_2} = \frac{-1}{\sigma_2} + \frac{x-\sigma_2/(r_0c)}{r_0c\sigma_2^2} + \frac{(x-\sigma_2/(r_0c))^2}{\sigma_2^3} \tag{43}$$

$$\frac{\partial \ln f_2(x;\boldsymbol{\theta})}{\partial c} = \frac{-(x-\sigma_2/(r_0c))}{r_0c^2\sigma_2} \tag{44}$$

$$\frac{\partial \ln f_2(x;\boldsymbol{\theta})}{\partial \sigma_1} = 0$$

Thus

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial \sigma_2}\right) = E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial c}\right) = E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial \sigma_1}\right) = 0 \tag{45}$$

Next, we have

$$\frac{\partial^2 \ln f_1(x;\boldsymbol{\theta})}{\partial \sigma_1^2} = \frac{1}{\sigma_1^2} - \frac{1}{c^2\sigma_1^2} - \frac{4(x-\sigma_1/c)}{c\sigma_1^3} - \frac{3(x-\sigma_1/c)^2}{\sigma_1^4} \tag{46}$$

$$\frac{\partial^2 \ln f_1(x;\boldsymbol{\theta})}{\partial c \partial \sigma_1} = \frac{1}{c^3\sigma_1} + \frac{x-\sigma_1/c}{c^2\sigma_1^2} \tag{47}$$

$$\frac{\partial^2 \ln f_1(x;\boldsymbol{\theta})}{\partial c^2} = -\frac{1}{c^4} + \frac{2(x-\sigma_1/c)}{c^3\sigma_1} \tag{48}$$

$$\frac{\partial^2 \ln f_2(x;\boldsymbol{\theta})}{\partial \sigma_2^2} = \frac{1}{\sigma_2^2} - \frac{1}{r_0^2c^2\sigma_2^2} - \frac{4(x-\sigma_2/(r_0c))}{r_0c\sigma_2^3} - \frac{3(x-\sigma_2/(r_0c))^2}{\sigma_2^4} \tag{49}$$

$$\frac{\partial^2 \ln f_2(x;\boldsymbol{\theta})}{\partial c \partial \sigma_2} = \frac{1}{r_0^2c^3\sigma_2} + \frac{x-\sigma_2/(r_0c)}{r_0c^2\sigma_2^2} \tag{50}$$

$$\frac{\partial^2 \ln f_2(x;\boldsymbol{\theta})}{\partial c^2} = -\frac{1}{r_0^2c^4} + \frac{2(x-\sigma_2/(r_0c))}{r_0c^3\sigma_2} \tag{51}$$

Thus,

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_1(X;\boldsymbol{\theta})}{\partial\sigma_1^2}\right) = -\left(2+1/c^2\right)/\sigma_1^2 \tag{52}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_1(X;\boldsymbol{\theta})}{\partial c\partial\sigma_1}\right) = \frac{1}{c^3\sigma_1} \tag{53}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_1(X;\boldsymbol{\theta})}{\partial c^2}\right) = -\frac{1}{c^4} \tag{54}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_2(X;\boldsymbol{\theta})}{\partial\sigma_2^2}\right) = -\left(2+1/(r_0^2c^2)\right)/\sigma_2^2 \tag{55}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_2(X;\boldsymbol{\theta})}{\partial c\partial\sigma_2}\right) = \frac{1}{r_0^2c^3\sigma_2} \tag{56}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial^2 \ln f_2(X;\boldsymbol{\theta})}{\partial c^2}\right) = -\frac{1}{r_0^2c^4} \tag{57}$$

We also have

$$\begin{aligned}E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial\sigma_1}\times\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial\sigma_1}\right) &= E_{\boldsymbol{\theta}}\left(\frac{1}{\sigma_1^2}-\frac{2(x-\sigma_1/c)}{c\sigma_1^3}-\frac{2(X-\sigma_1/c)^2}{\sigma_1^4}\right.\\ &\quad\left.+\frac{(X-\sigma_1/c)^2}{c^2\sigma_1^4}+\frac{2(X-\sigma_1/c)^3}{c\sigma_1^5}+\frac{(X-\sigma_1/c)^4}{\sigma_1^6}\right)\\ &= \left(2+1/c^2\right)/\sigma_1^2\end{aligned} \tag{58}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial c}\times\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial\sigma_1}\right) = E_{\boldsymbol{\theta}}\left(\frac{(x-\sigma_1/c)}{c^2\sigma_1^2}-\frac{(X-\sigma_1/c)^2}{c^3\sigma_1^3}-\frac{(X-\sigma_1/c)^3}{c^2\sigma_1^4}\right) = -\frac{1}{c^3\sigma_1} \tag{59}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial c}\times\frac{\partial \ln f_1(X;\boldsymbol{\theta})}{\partial c}\right) = E_{\boldsymbol{\theta}}\left(\frac{(x-\sigma_1/c)^2}{c^4\sigma_1^2}\right) = \frac{1}{c^4} \tag{60}$$

$$\begin{aligned}E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial\sigma_2}\times\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial\sigma_2}\right) &= E_{\boldsymbol{\theta}}\left(\frac{1}{\sigma_2^2}-\frac{2(X-\sigma_2/(r_0c))}{r_0c\sigma_2^3}-\frac{2(X-\sigma_2/(r_0c))^2}{\sigma_2^4}\right.\\ &\quad\left.+\frac{(X-\sigma_2/(r_0c))^2}{r_0^2c^2\sigma_2^4}+\frac{2(X-\sigma_2/(r_0c))^3}{r_0c\sigma_2^5}+\frac{(X-\sigma_2/(r_0c))^4}{\sigma_2^6}\right)\\ &= \left(2+1/(r_0^2c^2)\right)/\sigma_2^2\end{aligned} \tag{61}$$

$$\begin{aligned}E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial c}\times\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial\sigma_2}\right) &= E_{\boldsymbol{\theta}}\left(\frac{(X-\sigma_2/(r_0c))}{r_0c^2\sigma_2^2}-\frac{(X-\sigma_2/(r_0c))^2}{r_0^2c^3\sigma_2^3}-\frac{(X-\sigma_2/(r_0c))^3}{r_0c^2\sigma_2^4}\right)\\ &= -\frac{1}{r_0^2c^3\sigma_2}\end{aligned} \tag{62}$$

$$E_{\boldsymbol{\theta}}\left(\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial c}\times\frac{\partial \ln f_2(X;\boldsymbol{\theta})}{\partial c}\right) = E_{\boldsymbol{\theta}}\left(\frac{(X-\sigma_2/(r_0c))^2}{r_0^2c^4\sigma_2^2}\right) = \frac{1}{r_0^2c^4} \tag{63}$$

Results 42, 45, and 52 through 63 establish condition **(B)**.

13.2 Condition (C)

Let $\mathbf{x}^T \equiv (x_1, x_2, x_3)$. Then, from results 58 through 63 (recall that $I(\boldsymbol{\theta}) = \lambda_1 I_1(\boldsymbol{\theta}) + \lambda_2 I_2(\boldsymbol{\theta})$)

$$\begin{aligned}\mathbf{x}^T I(\boldsymbol{\theta})\mathbf{x} &= \lambda_1 x_1^2(2+1/c^2)/\sigma_1^2 + \lambda_2 x_2^2(2+1/(r_0^2c^2))/\sigma_2^2 + (\lambda_1/c^4 + \lambda_2/(r_0^2c^4))x_3^2 \\ &\quad -2\lambda_1 x_1 x_3/(c^3\sigma_1) - 2\lambda_2 x_2 x_3/(r_0^2 c^3 \sigma_2) \\ &= 2\lambda_1 x_1^2/\sigma_1^2 + \lambda_1\left(x_1/(c\sigma_1) - x_3/c^2\right)^2 + 2\lambda_2 x_2^2/\sigma_2^2 + \lambda_2\left(x_2/(r_0 c\sigma_2) - x_3/(r_0c^2)\right)^2\end{aligned}$$

which is clearly positive unless x_1, x_2, and x_3 are all 0. Thus condition **(C)** is established.

13.3 Condition (D)

From results 46 through 51 (actually from the third-order derivatives that are based on them), the fact that σ_1, σ_2, and c are bounded away from 0 if they are in a sufficiently small open neighborhood of $\boldsymbol{\theta}_0$, and the fact that normal distributions have finite absolute moments, it is clear that condition **(D)** holds.

13.4 The solution to the likelihood equations

Up to a constant, the log likelihood in this case is

$$\ln L = -n_1\ln(\sigma_1) - \sum_{i=1}^{n_1}(x_{i1} - \sigma_1/c)^2/(2\sigma_1^2) - n_2\ln(\sigma_2) - \sum_{i=1}^{n_2}(x_{i2} - \sigma_2/(r_0c))^2/(2\sigma_2^2)$$

We have

$$\frac{\partial \ln L}{\partial c} = -\sum_{i=1}^{n_1}(x_{i1} - \sigma_1/c)/(c^2\sigma_1) - \sum_{i=1}^{n_2}(x_{i2} - \sigma_2/(r_0c))/(r_0c^2\sigma_2)$$

Setting $\frac{\partial \ln L}{\partial c} = 0$, we obtain

$$\sum_{i=1}^{n_1}(x_{i1} - \sigma_1/c)/\sigma_1 + \sum_{i=1}^{n_2}(x_{i2} - \sigma_2/(r_0c))/(r_0\sigma_2) = 0$$

or

$$\left(n_1\bar{x}_{\cdot 1}/\sigma_1 + n_2\bar{x}_{\cdot 2}/(r_0\sigma_2)\right) / \left(n_1 + n_2/r_0^2\right) = 1/c \tag{64}$$

Next,

$$\frac{\partial \ln L}{\partial \sigma_1} = -n_1/\sigma_1 + \sum_{i=1}^{n_1}(x_{i1} - \sigma_1/c)/(c\sigma_1^2) + \sum_{i=1}^{n_1}(x_{i1} - \sigma_1/c)^2/\sigma_1^3$$

Setting $\frac{\partial \ln L}{\partial \sigma_1} = 0$, we obtain

$$\begin{aligned}\sigma_1^2 &= \sum_{i=1}^{n_1}(x_{i1} - \sigma_1/c)(\sigma_1/c)/n_1 + \sum_{i=1}^{n_1}(x_{i1} - \sigma_1/c)^2/n_1 \\ &= (\bar{x}_{\cdot 1} - \sigma_1/c)(\sigma_1/c) + \sum_{i=1}^{n_1} x_{i1}^2/n_1 - 2\bar{x}_{\cdot 1}\sigma_1/c + (\sigma_1/c)^2 \\ &= -\bar{x}_{\cdot 1}\sigma_1/c + \sum_{i=1}^{n_1} x_{i1}^2/n_1\end{aligned}$$

so

$$\sum_{i=1}^{n_1} x_{i1}^2/n_1 - \sigma_1^2 = \bar{x}_{\cdot 1}\sigma_1/c$$

or (assuming that $\bar{x}_{\cdot 1} \neq 0$)

$$\left(\sum_{i=1}^{n_1} x_{i1}^2/n_1 - \sigma_1^2\right)/(\bar{x}_{\cdot 1}\sigma_1) = 1/c \tag{65}$$

Similarly, setting $\frac{\partial \ln L}{\partial \sigma_2} = 0$, we obtain (assuming that $\bar{x}_{\cdot 2} \neq 0$)

$$\left(\sum_{i=1}^{n_2} x_{i2}^2/n_2 - \sigma_2^2\right)/(\bar{x}_{\cdot 2}\sigma_2/r_0) = 1/c \tag{66}$$

From Equations 64, 65, and 66, we have

$$\left(\sum_{i=1}^{n_1} x_{i1}^2/n_1 - \sigma_1^2\right)/(\bar{x}_{\cdot 1}\sigma_1) = (n_1\bar{x}_{\cdot 1}/\sigma_1 + n_2\bar{x}_{\cdot 2}/(r_0\sigma_2))/\left(n_1 + n_2/r_0^2\right) \tag{67}$$

and

$$\left(\sum_{i=1}^{n_2} x_{i2}^2/n_2 - \sigma_2^2\right)/(\bar{x}_{\cdot 2}\sigma_2/r_0) = (n_1\bar{x}_{\cdot 1}/\sigma_1 + n_2\bar{x}_{\cdot 2}/(r_0\sigma_2))/\left(n_1 + n_2/r_0^2\right) \tag{68}$$

Now define

$$y \equiv \sigma_1/\sigma_2$$

so

$$\sigma_2 = \sigma_1/y \tag{69}$$

Equation 67 becomes

$$\sum_{i=1}^{n_1} x_{i1}^2/n_1 - \sigma_1^2 = \left(n_1\bar{x}_{\cdot 1}^2 + n_2\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}y/r_0\right)/\left(n_1 + n_2/r_0^2\right)$$

or

$$\sum_{i=1}^{n_1} x_{i1}^2/n_1 - n_1\bar{x}_{\cdot 1}^2/\left(n_1 + n_2/r_0^2\right) - (n_2\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}y/r_0)/\left(n_1 + n_2/r_0^2\right) = \sigma_1^2 \tag{70}$$

Equation 68 becomes

$$\sum_{i=1}^{n_2} x_{i2}^2/n_2 - \sigma_1^2/y^2 = \left(n_1\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}/(r_0 y) + n_2\bar{x}_{\cdot 2}^2/r_0^2\right)/\left(n_1 + n_2/r_0^2\right)$$

or

$$y^2\left(\sum_{i=1}^{n_2} x_{i2}^2/n_2 - (n_2\bar{x}_{\cdot 2}^2/r_0^2)/\left(n_1 + n_2/r_0^2\right)\right) - y(n_1\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}/r_0)/\left(n_1 + n_2/r_0^2\right) = \sigma_1^2 \tag{71}$$

From Equations 70 and 71 we have

$$ay^2 + by + d = 0$$

(we use d here rather than the standard c because we have already used c to denote the coefficient of variation) where

$$a \equiv \sum_{i=1}^{n_2} x_{i2}^2/n_2 - n_2(\bar{x}_{\cdot 2}^2/r_0^2)/\left(n_1 + n_2/r_0^2\right) \quad (72)$$

$$b \equiv ((n_2 - n_1)\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}/r_0)/\left(n_1 + n_2/r_0^2\right) \quad (73)$$

$$d \equiv -\left(\sum_{i=1}^{n_1} x_{i1}^2/n_1 - n_1\bar{x}_{\cdot 1}^2/\left(n_1 + n_2/r_0^2\right)\right) \quad (74)$$

The possible solutions for $y = \sigma_1/\sigma_2$ are, of course, $\left(-b \pm \sqrt{b^2 - 4ad}\right)/(2a)$.

Now note that

$$a = \sum_{i=1}^{n_2} x_{i2}^2/n_2 - n_2(\bar{x}_{\cdot 2}^2/r_0^2)/\left(n_1 + n_2/r_0^2\right) \geq \sum_{i=1}^{n_2} x_{i2}^2/n_2 - \bar{x}_{\cdot 2}^2 \geq 0$$

with $a = 0$ only if all the x_{i2}'s are zero. Similarly, $d \leq 0$ with equality only if all the x_{i1}'s equal zero. Thus, unless all the x_{i1}'s equal 0 or all the x_{i2}'s equal 0,

$$-4ad > 0 \text{ and } \sqrt{b^2 - 4ad} > |b|$$

So if $b > 0$,

$$\left(-b - \sqrt{b^2 - 4ad}\right)/(2a) < 0$$

and

$$\left(-b + \sqrt{b^2 - 4ad}\right)/(2a) > (-|b| + |b|)/(2a) = 0$$

If $b < 0$,

$$\left(-b - \sqrt{b^2 - 4ad}\right)/(2a) < (|b| - |b|)/(2a) = 0$$

and

$$\left(-b + \sqrt{b^2 - 4ad}\right)/(2a) > 0$$

Thus $\left(-b + \sqrt{b^2 - 4ad}\right)/(2a)$ is the unique positive solution for $y = \sigma_1/\sigma_2$. The estimates of σ_1^2 and σ_2^2 can then be obtained from Equations 70 and 69. (See Appendix D for a proof that the estimate of σ_1^2 obtained from Equation 70 is positive.)

14 Appendix D — A Proof that the Estimate of σ_1^2 Obtained from Equation 70 Is Positive

Assume that we do not have

1. $x_{11} = \cdots = x_{n_1,1}$ and $x_{12} = \cdots = x_{n_2,2}$, or
2. $\bar{x}_{\cdot 1} = 0$, or
3. $\bar{x}_{\cdot 2} = 0$

If $\bar{x}_{\cdot 1}$, $\bar{x}_{\cdot 2}$ are of opposite sign, then Equation 70 implies that $\hat{\sigma}_1^2 > 0$ so assume that $\bar{x}_{\cdot 1}\bar{x}_{\cdot 2} > 0$. Then, given Equation 70, to establish that $\hat{\sigma}_1^2 > 0$, we need to show that

$$L > \left(-b + \sqrt{b^2 - 4ad}\right)/(2a)$$

or

$$L + b/(2a) > \sqrt{b^2 - 4ad}/(2a) \tag{75}$$

where

$$L \equiv \left(\sum_{i=1}^{n_1} x_{i1}^2/n_1 - n_1\bar{x}_{\cdot 1}^2/(n_1 + n_2/r_0^2)\right) / \left((n_2\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}/r_0)/(n_1 + n_2/r_0^2)\right)$$

and a, b, and d are given by Equations 72 through 74. Now suppose that

$$(L + b/(2a))^2 > \left(\sqrt{b^2 - 4ad}/(2a)\right)^2 \tag{76}$$

Then either result 75 holds or

$$L + b/(2a) < -\sqrt{b^2 - 4ad}/(2a)$$

which is equivalent to

$$L < \left(-b - \sqrt{b^2 - 4ad}\right)/(2a) \tag{77}$$

But L is positive and, as we saw at the end of Section 13.4, $\left(-b - \sqrt{b^2 - 4ad}\right)/(2a)$ is negative so inequality 77 cannot hold. Thus, result 76 implies result 75.

We establish result 76 by showing that

$$\begin{aligned}
&\left(\sum_{i=1}^{n_1} x_{i1}^2/n_1 - n_1\bar{x}_{\cdot 1}^2/(n_1 + n_2/r_0^2)\right)^2 / \left((n_2\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}/r_0)/(n_1 + n_2/r_0^2)\right)^2 \\
&+ \left[\left(\sum_{i=1}^{n_1} x_{i1}^2/n_1 - n_1\bar{x}_{\cdot 1}^2/(n_1 + n_2/r_0^2)\right) / \left((n_2\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}/r_0)/(n_1 + n_2/r_0^2)\right)\right] \\
&\times \left[((n_2 - n_1)\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}/r_0)/(n_1 + n_2/r_0^2)\right] / \left(\sum_{i=1}^{n_2} x_{i2}^2/n_2 - n_2(\bar{x}_{\cdot 2}^2/r_0^2)/(n_1 + n_2/r_0^2)\right) \\
> &\left(\sum_{i=1}^{n_1} x_{i1}^2/n_1 - n_1\bar{x}_{\cdot 1}^2/(n_1 + n_2/r_0^2)\right) / \left(\sum_{i=1}^{n_2} x_{i2}^2/n_2 - n_2(\bar{x}_{\cdot 2}^2/r_0^2)/(n_1 + n_2/r_0^2)\right)
\end{aligned}$$

or

$$\begin{aligned}
&\left(\sum_{i=1}^{n_1} x_{i1}^2/n_1 - n_1\bar{x}_{\cdot 1}^2/(n_1+n_2/r_0^2)\right) \\
&\times\left(\sum_{i=1}^{n_2} x_{i2}^2/n_2 - n_2(\bar{x}_{\cdot 2}^2/r_0^2)/(n_1+n_2/r_0^2)\right) \Big/ \left((n_2\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}/r_0)/(n_1+n_2/r_0^2)\right)^2 \\
&+(n_2-n_1)/n_2 \\
> \; & 1
\end{aligned}$$

or

$$\left(\sum_{i=1}^{n_1} x_{i1}^2/n_1 - n_1\bar{x}_{\cdot 1}^2/(n_1+n_2/r_0^2)\right)\left(\sum_{i=1}^{n_2} x_{i2}^2/n_2 - n_2(\bar{x}_{\cdot 2}^2/r_0^2)/(n_1+n_2/r_0^2)\right)$$

$$> (n_1/n_2)\left((n_2\bar{x}_{\cdot 1}\bar{x}_{\cdot 2}/r_0)/(n_1+n_2/r_0^2)\right)^2$$

This will follow if

$$\left(\sum_{i=1}^{n_1}(x_{i1}-\bar{x}_{\cdot 1})^2/n_1 + \bar{x}_{\cdot 1}^2\left(1-n_1/(n_1+n_2/r_0^2)\right)\right)\left(\sum_{i=1}^{n_2}(x_{i2}-\bar{x}_{\cdot 2})^2/n_2 + \bar{x}_{\cdot 2}^2\left(1-(n_2/r_0^2)/(n_1+n_2/r_0^2)\right)\right)$$

$$> (n_1n_2/r_0^2)\bar{x}_{\cdot 1}^2\bar{x}_{\cdot 2}^2/(n_1+n_2/r_0^2)^2$$

or

$$\begin{aligned}
&\left(\sum_{i=1}^{n_1}(x_{i1}-\bar{x}_{\cdot 1})^2/n_1\right)\left(\sum_{i=1}^{n_2}(x_{i2}-\bar{x}_{\cdot 2})^2/n_2\right) \\
&+\left(\sum_{i=1}^{n_1}(x_{i1}-\bar{x}_{\cdot 1})^2/n_1\right)\bar{x}_{\cdot 2}^2 n_1/(n_1+n_2/r_0^2) \\
&+\left(\sum_{i=1}^{n_2}(x_{i2}-\bar{x}_{\cdot 2})^2/n_2\right)\bar{x}_{\cdot 1}^2(n_2/r_0^2)/(n_1+n_2/r_0^2) \\
&+(n_1n_2/r_0^2)\bar{x}_{\cdot 1}^2\bar{x}_{\cdot 2}^2/(n_1+n_2/r_0^2)^2 \\
> \; & (n_1n_2/r_0^2)\bar{x}_{\cdot 1}^2\bar{x}_{\cdot 2}^2/(n_1+n_2/r_0^2)^2
\end{aligned}$$

which clearly holds.

15 Appendix E — Verification of the Conditions Needed to Establish Theorem 4

In the notation of Section 18.3, $\boldsymbol{\theta}^T = (\mu_1, \sigma_1, \ldots, \mu_k, \sigma_k)$ and $\boldsymbol{\nu}^T = (\sigma_1, \ldots, \sigma_k, c)$ where the μ's are the means and the σ's are the standard deviations of the k populations. Under the null hypothesis, the shared coefficient of variation is c.

To invoke Theorem H.4 to prove our Theorem 4, we must establish conditions **(A0)** through **(A2)**, **(A)** through **(D)**, and **(E1)** through **(I)** of Appendix H.

In the notation of Section 18.1, in the case under consideration,
$f_j(x;\boldsymbol{\theta}) = \exp\left(-(x-\mu_j)^2/(2\sigma_j^2)\right) / \left(\sigma_j\sqrt{2\pi}\right)$ for $j = 1, \ldots, k$.

It is clear that conditions **(A0)** through **(A2)** and **(A)** through **(D)** hold. (The proofs are well known in the $2k$ parameter case.)

15.1 Condition (E1)

In the notation of Section 18.3, we have

$$\begin{aligned} R_1(\boldsymbol{\theta}) &= \sigma_1/\mu_1 - \sigma_2/\mu_2 \\ &\vdots \\ R_{k-1}(\boldsymbol{\theta}) &= \sigma_1/\mu_1 - \sigma_k/\mu_k \end{aligned}$$

and

$$\begin{aligned} \theta_1 &= \mu_1 = \sigma_1/c = \nu_1/\nu_{k+1} = g_1(\nu_1, \ldots, \nu_{k+1}) \\ \theta_2 &= \sigma_1 = \nu_1 = g_2(\nu_1, \ldots, \nu_{k+1}) \\ &\vdots \\ \theta_{2k-1} &= \mu_k = \sigma_k/c = \nu_k/\nu_{k+1} = g_{2k-1}(\nu_1, \ldots, \nu_{k+1}) \\ \theta_{2k} &= \sigma_k = \nu_k = g_{2k}(\nu_1, \ldots, \nu_{k+1}) \end{aligned}$$

The equivalence of these two forms of the constraints follows from the fact that $\sigma_j/\mu_j = c \Leftrightarrow \sigma_j/c = \mu_j$

15.2 Condition (E2)

Clear.

15.3 Condition (E3)

Straightforward calculations demonstrate that

$$\begin{aligned} \partial R_j/\partial \mu_1 &= -\sigma_1/\mu_1^2 \\ \partial R_j/\partial \sigma_1 &= 1/\mu_1 \\ \partial R_j/\partial \mu_{j+1} &= \sigma_{j+1}/\mu_{j+1}^2 \\ \partial R_j/\partial \sigma_{j+1} &= -1/\mu_{j+1} \end{aligned}$$

and the remaining first-order partials are all 0.

15.4 Conditions (E4) and (E5)

We have

$$\mathbf{C}(\boldsymbol{\theta}_0)^T = \left[\frac{\partial R_i(\boldsymbol{\theta})}{\partial \theta_j}\right]^T_{2k\times(k-1)} = \begin{pmatrix} -\sigma_1/\mu_1^2 & \ldots & -\sigma_1/\mu_1^2 \\ 1/\mu_1 & \ldots & 1/\mu_1 \\ \sigma_2/\mu_2^2 & & 0 \\ -1/\mu_2 & & 0 \\ & \ddots & \\ 0 & & \sigma_k/\mu_k^2 \\ 0 & & -1/\mu_k \end{pmatrix}$$

which is clearly of rank $k-1 = r$ (in the notation of Section 18.3).

Also

$$\mathbf{D}(\boldsymbol{\nu}_0) = \left[\frac{\partial g_i(\boldsymbol{\nu}_0)}{\partial \nu_j}\right]_{2k\times(k+1)} = \begin{pmatrix} 1/c & 0 & \dots & 0 & -\sigma_1/c^2 \\ 1 & 0 & \dots & 0 & 0 \\ & & \ddots & & \vdots \\ 0 & \dots & 0 & 1/c & -\sigma_k/c^2 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$$

which is clearly of rank $k+1 = 2k-(k-1) = s-r$ (in the notation of Section 18.3).

15.5 Condition (F)

Clear.

15.6 Conditions (G) through (I)

The necessary arguments are straightforward extensions of the arguments used to establish conditions **(B)** through **(D)** for the $k=2$ case treated in Appendix C.

15.7 The solution to one of the likelihood equations

Setting $\frac{\partial \ln L}{\partial c} = 0$, we obtain (as a straightforward extension of the manner in which we obtained Equation 64)

$$\left(\sum_{j=1}^{k} n_j \bar{x}_{\cdot j}/\sigma_j\right) / (n_1 + \ldots + n_k) = 1/c \qquad (78)$$

If we replace σ_j in Equation 78 by $s_j \equiv \sqrt{\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_{\cdot j})^2/n_j}$, we obtain a $\sqrt{n}$-consistent[5] estimator of $1/c$. Further, the inverse of this estimator is a $\sqrt{n}$-consistent estimator of c.

This follows from the relations

$$\sqrt{n}(\bar{x}_{\cdot j}/s_j - \mu_j/\sigma_j) = \sqrt{n}(\bar{x}_{\cdot j}/s_j - \mu_j/s_j + \mu_j/s_j - \mu_j/\sigma_j) = \sqrt{n}(\bar{x}_{\cdot j} - \mu_j)/s_j + \sqrt{n}\mu_j(\sigma_j - s_j)/(s_j\sigma_j)$$

and

$$\sqrt{n}(1/\hat{a} - 1/a) = \sqrt{n}(a-\hat{a})/(\hat{a}a)$$

16 Appendix F — Verification of the Conditions Needed to Establish Theorem 5

To invoke Corollary 1 to Lemma H.8 in Appendix H to prove Theorem 5, we must establish conditions **(A0)** through **(A2)** and **(A)** through **(D)** of Appendix H.

In the notation of Section 18.1, in the case under consideration, $f_j(x;\boldsymbol{\theta}) = \exp\left(-(x-\sigma_j/c)^2/(2\sigma_j^2)\right) / \left(\sigma_j\sqrt{2\pi}\right)$ for $j = 1, \ldots, k$.

It is clear that conditions **(A0)** through **(A2)** and **(A)** hold. Conditions **(B)** through **(D)** can be established by straightforward extensions of the arguments used in the $k=2$ case addressed in Appendix C.

[5] $\hat{a}$ is a $\sqrt{n}$-consistent estimator of a if $\sqrt{n}(\hat{a}-a) = O_p(1)$

17 Appendix G — Verification of the Conditions Needed to Establish Theorem 6

In the notation of Section 18.3, $\boldsymbol{\theta}^T = (\sigma_1, \ldots, \sigma_k, c)$ and $\boldsymbol{\nu}^T = (\sigma_1, \ldots, \sigma_k)$ where the σ's are the standard deviations of the k populations and c is the shared coefficient of variation. Under the null hypothesis, this shared coefficient of variation is c_0.

To invoke Theorem H.4 to prove our Theorem 6, we must establish conditions **(A0)** through **(A2)**, **(A)** through **(D)**, and **(E1)** through **(I)** of Appendix H.

In the notation of Section 18.1, in the case under consideration,
$f_j(x; \boldsymbol{\theta}) = \exp\left(-(x - \sigma_j/c)^2/(2\sigma_j^2)\right) / \left(\sigma_j\sqrt{2\pi}\right)$ for $j = 1, \ldots, k$.

The fact that conditions **(A0)** through **(A2)** and **(A)** hold is clear.

Conditions **(B)** through **(D)** can be established by straightforward extensions of the arguments used to establish conditions **(B)** through **(D)** for the $k = 2$ case in Appendix C.

17.1 Condition (E1)

In the notation of Section 18.3, we have

$$R_1(\boldsymbol{\theta}) = c - c_0$$

and

$$\begin{aligned}
\theta_1 &= \sigma_1 = g_1(\nu_1, \ldots, \nu_k) = \nu_1 \\
&\vdots \\
\theta_k &= \sigma_k = g_k(\nu_1, \ldots, \nu_k) = \nu_k \\
\theta_{k+1} &= c = g_{k+1}(\nu_1, \ldots, \nu_k) = c_0
\end{aligned}$$

The equivalence of these two forms of the constraints is clear. They both permit the σ's to vary freely from 0 to ∞ and restrict c to the single value c_0.

17.2 Condition (E2)

Clear.

17.3 Condition (E3)

Clear.

17.4 Condition (E4)

$$\mathbf{C}(\boldsymbol{\theta}_0) = (0, \ldots, 0, 1)_{1 \times (k+1)}$$

which is clearly of rank $1 = r$ (in the notation of Section 18.3).

17.5 Condition (E5)

We have

$$\mathbf{D}(\boldsymbol{\nu}_0) = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}_{(k+1)\times k}$$

which is clearly of rank $k = (k+1) - 1 = s - r$ (in the notation of Section 18.3).

17.6 Condition (F)

Clear.

17.7 Conditions (G) through (I)

The necessary arguments are straightforward extensions of the arguments used to establish conditions **(B)** through **(D)** for the $k = 2$ case treated in Appendix C. (Note that, in fact, we do not have to do as much work here as in Appendix C because here c is fixed at c_0.)

17.8 The solution to the likelihood equations in the constrained case

We have

$$\frac{\partial \ln L}{\partial \sigma_j} = \frac{-n_j}{\sigma_j} + \frac{\sum_{i=1}^{n_j}(x_{ij} - \sigma_j/c_0)}{c_0\sigma_j^2} + \frac{\sum_{i=1}^{n_j}(x_{ij} - \sigma_j/c_0)^2}{\sigma_j^3} \tag{79}$$

After setting $\frac{\partial \ln L}{\partial \sigma_j} = 0$ and performing some simple algebra, we obtain

$$\sigma_j^2 + (\bar{x}_{\cdot j}/c_0)\sigma_j - \sum_{i=1}^{n_j} x_{ij}^2/n_j = 0 \tag{80}$$

whose unique positive solution is given by

$$\hat{\sigma}_j = \left(-\bar{x}_{\cdot j}/c_0 + \sqrt{\bar{x}_{\cdot j}^2/c_0^2 + 4\sum_{i=1}^{n_j} x_{ij}^2/n_j} \right) /2 \tag{81}$$

18 Appendix H — Some Asymptotic Results

This appendix is based on material that appears in chapter 6 of Lehmann (1983) and chapter 4 of Serfling (1980). In section 4.4.4 of his book, Serfling investigates the asymptotic properties of the likelihood ratio test for the case in which the null hypothesis is composite and the number of samples, k, equals one. For our purposes we need to extend his proofs to the case in which $k > 1$. This is a relatively trivial task and we might wave our hands were it not for the fact that there is a flaw in Serfling's proof. Here we correct that flaw.

We note that related results have been established under related conditions by Silvey (1959). An outline of an alternative proof of results closely related to our Theorems H.1 through H.3 is provided in section 6e.3[6] of Rao (1973). Taking a quadratic mean differentiability approach, Lehmann and Romano (2005) outline the proof of results (see their sections 12.4.2 through 12.4.4) closely related to our Theorems H.1 through H.6. We take a Cramér condition approach and make the conditions and their use explicit. The original work on asymptotic tests of composite hypotheses was due to Wilks (1938) and Wald (1943).

Before we proceed with our proof it is useful to recall Lehmann's versions of the Cramér conditions, and one of Lehmann's asymptotic results.

18.1 Lehmann's (1983) version of the Cramér conditions

Let the parameter space be denoted by $\Theta \subset R^s$. Let $\boldsymbol{\theta}_0 \in \Theta$ denote the true parameter value.

Suppose that we have k independent samples. Let $P_j(\boldsymbol{\theta})$ denote the distribution associated with the jth sample. Define $n \equiv n_1 + \ldots + n_k$ and further suppose that for $j = 1, \ldots, k$, $n_j/n \to \lambda_j > 0$ as $n \to \infty$.

(A0) Distinct $\boldsymbol{\theta}$'s correspond to distinct $P_j(\boldsymbol{\theta})$'s.

(A1) The distributions $P_j(\boldsymbol{\theta})$ have common support.

(A2) For $j = 1, \ldots, k$, the observations are $\mathbf{X}_j = (X_{1j} \ldots X_{n_j j})^T$ where the X_{ij} are iid with probability density $f_j(x; \boldsymbol{\theta})$ (and the observations from the k different samples are independent).

(A) There exists an open subset T of Θ that contains the true parameter value $\boldsymbol{\theta}_0$ such that for all $j \in \{1, \ldots, k\}$ and almost all x, the densities $f_j(x; \boldsymbol{\theta})$ have continuous third derivatives, $\partial^3 f_j(x; \boldsymbol{\theta})/\partial\theta_l \partial\theta_m \partial\theta_p$ for all $\boldsymbol{\theta} \in T$.

(B) For all $\boldsymbol{\theta} \in T$, the first and second logarithmic derivatives of f_j satisfy the equations

$$E_{\boldsymbol{\theta}}(\partial \ln f_j(X; \boldsymbol{\theta})/\partial\theta_l) = 0$$

for $j = 1, \ldots, k$; $l = 1, \ldots, s$ and

$$I_{lm,j}(\boldsymbol{\theta}) \equiv E_{\boldsymbol{\theta}}(\partial \ln f_j(X; \boldsymbol{\theta})/\partial\theta_l \times \partial \ln f_j(X; \boldsymbol{\theta})/\partial\theta_m) = E_{\boldsymbol{\theta}}(-\partial^2 \ln f_j(X; \boldsymbol{\theta})/\partial\theta_l \partial\theta_m)$$

for $j = 1, \ldots, k$; $l, m = 1, \ldots, s$. The $I_{lm,j}$ are finite.

(C)

$$I(\boldsymbol{\theta}) \equiv \lambda_1 I_1(\boldsymbol{\theta}) + \ldots + \lambda_k I_k(\boldsymbol{\theta})$$

is positive definite for all $\boldsymbol{\theta}$ in T. Here $I_j(\boldsymbol{\theta}) \equiv [I_{lm,j}]_{s \times s}$.

(D) For all l, m, p, j, $\partial^3 \ln f_j(x; \boldsymbol{\theta})/\partial\theta_l \partial\theta_m \partial\theta_p$ is a continuous function of $\boldsymbol{\theta}$ for $\boldsymbol{\theta} \in T$. There exist integrable functions $M_{lmp,j}(x)$ such that

$$\left|\partial^3 \ln f_j(x; \boldsymbol{\theta})/\partial\theta_l \partial\theta_m \partial\theta_p\right| \le M_{lmp,j}(x)$$

for all $\boldsymbol{\theta} \in T$, and

$$m_{lmp,j} \equiv E_{\boldsymbol{\theta}_0}(M_{lmp,j}(X)) < \infty$$

for all l, m, p, j.

[6] Note that there is an error above (6e.3.9) in Rao. The line above the "Therefore," should be, in essence, $2[l(\hat{\boldsymbol{\theta}}) - l(\boldsymbol{\beta})] - \mathbf{V}^T I(\boldsymbol{\theta})^{-1}\mathbf{V} \xrightarrow{p} 0$.

18.2 One of Lehmann's results

Given conditions **(A0)** through **(D)**, Lehmann (1983) establishes (theorem 6.1 of his chapter 6) that with probability tending to 1 as $n \to \infty$, there exists a solution of the likelihood equations, $\hat{\boldsymbol{\theta}}_n$, that satisfies

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, I(\boldsymbol{\theta}_0)^{-1}) \tag{82}$$

where

$$I(\boldsymbol{\theta}_0) = \lambda_1 I_1(\boldsymbol{\theta}_0) + \ldots + \lambda_k I_k(\boldsymbol{\theta}_0)$$

18.3 An extension of Serfling's likelihood ratio theorem to the case of k samples

We need two additional assumptions:

(E1) There are two equivalent methods for characterizing the nature of the space that constitutes the null hypothesis. It can be specified by the constraints:

$$\begin{aligned} R_1(\boldsymbol{\theta}) &= 0 \\ &\vdots \\ R_r(\boldsymbol{\theta}) &= 0 \end{aligned}$$

or by the equations

$$\begin{aligned} \theta_1 &= g_1(\nu_1, \ldots, \nu_{s-r}) \\ &\vdots \\ \theta_s &= g_s(\nu_1, \ldots, \nu_{s-r}) \end{aligned}$$

In both cases the effective dimension of the parameter space is $s - r$ (we assume that $r < s$) rather than s.

(E2) $\mathbf{g}$ takes R^{s-r} 1 to 1 into R^s.

Under assumptions **(E1)** and **(E2)**, there exists a unique $\boldsymbol{\nu}_0$ such that

$$\boldsymbol{\nu}_0 = \mathbf{g}^{-1}(\boldsymbol{\theta}_0).$$

We need to establish a relationship between the two modes of specifying the null hypothesis parameter space. Define

$$\mathbf{C}(\boldsymbol{\theta}) \equiv \left[\frac{\partial R_i(\boldsymbol{\theta})}{\partial \theta_j}\right]_{r \times s}$$

and

$$\mathbf{D}(\boldsymbol{\nu}) \equiv \left[\frac{\partial g_i(\boldsymbol{\nu})}{\partial \nu_j}\right]_{s \times (s-r)}$$

Lemma H.1

Assume that conditions **(E1)** and **(E2)** hold. Further assume that for $l \in \{1, \ldots, s\}$, $i \in \{1, \ldots, r\}$, $\partial R_i / \partial \theta_l$ is continuous within an open neighborhood of $\boldsymbol{\theta}_0$, and that for $l \in \{1, \ldots, s-r\}$, $i \in \{1, \ldots, s\}$, $\partial g_i / \partial \nu_l$ is continuous within an open neighborhood of $\boldsymbol{\nu}_0$.

Then

$$\mathbf{C}(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\nu}_0) = \mathbf{0}_{r \times (s-r)} \tag{83}$$

Proof

For $\boldsymbol{\theta}, \boldsymbol{\theta}_0$ in the $s - r$ dimensional null hypothesis space, by Taylor's theorem we have

$$\begin{pmatrix} R_1(\boldsymbol{\theta}) \\ \vdots \\ R_r(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} R_1(\boldsymbol{\theta}_0) \\ \vdots \\ R_r(\boldsymbol{\theta}_0) \end{pmatrix} + \begin{pmatrix} [\partial R_1(\boldsymbol{\theta}_{*,1})/\partial \theta_l] \\ \vdots \\ [\partial R_r(\boldsymbol{\theta}_{*,r})/\partial \theta_l] \end{pmatrix} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \tag{84}$$

where $\boldsymbol{\theta}_{*,i}$ lies on the line segment between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$.

We also have

$$\boldsymbol{\theta} - \boldsymbol{\theta}_0 = \mathbf{g}(\boldsymbol{\nu}) - \mathbf{g}(\boldsymbol{\nu}_0) = \begin{pmatrix} [\partial g_1(\boldsymbol{\nu}_{*,1})/\partial \nu_l] \\ \vdots \\ [\partial g_s(\boldsymbol{\nu}_{*,s})/\partial \nu_l] \end{pmatrix} (\boldsymbol{\nu} - \boldsymbol{\nu}_0) \tag{85}$$

where $\boldsymbol{\nu}_{*,i}$ lies on the line segment between $\boldsymbol{\nu}$ and $\boldsymbol{\nu}_0$.

Now suppose that $\mathbf{C}(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\nu}_0) \neq \mathbf{0}_{r \times (s-r)}$. Then $\mathbf{C}(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\nu}_0)$ contains some non-zero column. Assume that the column is the first column. (The proof is essentially the same in the other cases.) Take

$$\boldsymbol{\nu} - \boldsymbol{\nu}_0 = h \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{(s-r) \times 1}$$

Then, by substituting Equation 85 into Equation 84, and noting that both the vector on the left-hand side of Equation 84 and the first vector on the right-hand side of Equation 84 must equal $\mathbf{0}$ (since $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ are in the null hypothesis space), we have

$$\mathbf{0}_{r \times 1} = h \times \left[\text{column 1 of} \begin{pmatrix} [\partial R_1(\boldsymbol{\theta}_{*,1})/\partial \theta_l] \\ \vdots \\ [\partial R_r(\boldsymbol{\theta}_{*,r})/\partial \theta_l] \end{pmatrix} \begin{pmatrix} [\partial g_1(\boldsymbol{\nu}_{*,1})/\partial \nu_l] \\ \vdots \\ [\partial g_s(\boldsymbol{\nu}_{*,s})/\partial \nu_l] \end{pmatrix} \right]$$

Dividing by h and then letting h converge to 0, we obtain a contradiction (that the first column of $\mathbf{C}(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\nu}_0)$ equals $\mathbf{0}_{r \times 1}$). ■

Next we need to establish a relationship between the solution to the likelihood equations in the restricted case, $\hat{\boldsymbol{\nu}}_n$, and the solution in the unrestricted case, $\hat{\boldsymbol{\theta}}_n$. To do so we first list an additional series of assumptions:

(E3) For $l \in \{1, \ldots, s\}$, $i \in \{1, \ldots, r\}$, $\partial R_i / \partial \theta_l$ is continuous within an open neighborhood of $\boldsymbol{\theta}_0$.

(E4) $\mathbf{C}(\boldsymbol{\theta}_0)_{r \times s}$ is of rank r.

(E5) $\mathbf{D}(\boldsymbol{\nu}_0)_{s\times(s-r)}$ is of rank $s-r$.

(F) There exists an open neighborhood, S, of $\boldsymbol{\nu}_0$ in R^{s-r} such that for all ν in this neighborhood, $\partial^3 g_i(\nu)/\partial\nu_l\partial\nu_m\partial\nu_p$ is continuous for $i\in\{1,\ldots,s\}$; $l,m,p\in\{1,\ldots,s-r\}$.

(G) For all $\boldsymbol{\nu}$ in S, the first and second logarithmic derivatives of f_j satisfy the equations

$$E_{\boldsymbol{\nu}}(\partial\ln f_j(X;\mathbf{g}(\boldsymbol{\nu}))/\partial\nu_l)=0$$

for $j\in\{1,\ldots,k\}$; $l\in\{1,\ldots,s-r\}$, and

$$J_{lm,j}(\boldsymbol{\nu})\equiv E_{\boldsymbol{\nu}}(\partial\ln f_j(X;\mathbf{g}(\boldsymbol{\nu}))/\partial\nu_l\times\partial\ln f_j(X;\mathbf{g}(\boldsymbol{\nu}))/\partial\nu_m)=E_{\boldsymbol{\nu}}(-\partial^2\ln f_j(X;\mathbf{g}(\boldsymbol{\nu}))/\partial\nu_l\partial\nu_m)$$

for $j\in\{1,\ldots,k\}$; $l,m\in\{1,\ldots,s-r\}$. The $J_{lm,j}$ are finite. (Note that, for example,

$$\partial f_j(x;\mathbf{g}(\boldsymbol{\nu}))/\partial\nu_l=\partial f_j(x;\boldsymbol{\theta})/\partial\theta_1|_{\boldsymbol{\theta}=\mathbf{g}(\boldsymbol{\nu})}\partial g_1(\boldsymbol{\nu})/\partial\nu_l+\ldots+\partial f_j(x;\boldsymbol{\theta})/\partial\theta_s|_{\boldsymbol{\theta}=\mathbf{g}(\boldsymbol{\nu})}\partial g_s(\boldsymbol{\nu})/\partial\nu_l$$

so this condition places restrictions on the behavior of the partials of the g_i's as well as the partials of the f_j's.)

(H)

$$J(\boldsymbol{\nu})\equiv\lambda_1 J_1(\boldsymbol{\nu})+\ldots+\lambda_k J_k(\boldsymbol{\nu})$$

is positive definite for all $\boldsymbol{\nu}$ in S. Here $J_j(\boldsymbol{\nu})\equiv[J_{lm,j}]_{(s-r)\times(s-r)}$.

(I) For all l,m,p,j, $\partial^3\ln f_j(x;\mathbf{g}(\boldsymbol{\nu}))/\partial\nu_l\partial\nu_m\partial\nu_p$ is a continuous function of $\boldsymbol{\nu}$, there exist integrable functions $N_{lmp,j}(x)$ such that

$$\left|\partial^3\ln f_j(x;\mathbf{g}(\boldsymbol{\nu}))/\partial\nu_l\partial\nu_m\partial\nu_p\right|\le N_{lmp,j}(x)$$

for all $\boldsymbol{\nu}\in S$, and

$$n_{lmp,j}\equiv E_{\boldsymbol{\nu}_0}(N_{lmp,j}(X))<\infty$$

for all l,m,p,j.

Now we can state and prove the following lemma.

Lemma H.2

Assume that conditions **(A0)** through **(A)**, **(E1)**, **(E2)**, and **(F)** through **(I)** hold. Then there exists a solution of the likelihood equations, $\hat{\boldsymbol{\nu}}_n$, that satisfies

$$\sqrt{n}(\hat{\boldsymbol{\nu}}_n-\boldsymbol{\nu}_0)\xrightarrow{D}\mathrm{N}(\mathbf{0},\mathbf{J}(\boldsymbol{\nu}_0)^{-1})\tag{86}$$

(Note that in this context, a solution of the likelihood equations is a solution of $\partial\ln L(\mathbf{g}(\boldsymbol{\nu}))/\partial\nu_1=\ldots=\partial\ln L(\mathbf{g}(\boldsymbol{\nu}))/\partial\nu_{s-r}=0$.)

Proof

The result follows from theorem 6.1 in chapter 6 of Lehmann (1983). ■

Note that

$$
\begin{aligned}
\frac{\partial^2 \ln f_j}{\partial \nu_l \partial \nu_m} &= \frac{\partial}{\partial \nu_l}\left(\frac{\partial \ln f_j}{\partial \theta_1}\frac{\partial g_1}{\partial \nu_m} + \ldots + \frac{\partial \ln f_j}{\partial \theta_s}\frac{\partial g_s}{\partial \nu_m}\right) \\
&= \left(\frac{\partial^2 \ln f_j}{\partial \theta_1 \partial \theta_1}\frac{\partial g_1}{\partial \nu_l} + \ldots + \frac{\partial^2 \ln f_j}{\partial \theta_s \partial \theta_1}\frac{\partial g_s}{\partial \nu_l}\right)\frac{\partial g_1}{\partial \nu_m} \\
&\quad + \\
&\quad \vdots \\
&\quad + \\
&\quad \left(\frac{\partial^2 \ln f_j}{\partial \theta_1 \partial \theta_s}\frac{\partial g_1}{\partial \nu_l} + \ldots + \frac{\partial^2 \ln f_j}{\partial \theta_s \partial \theta_s}\frac{\partial g_s}{\partial \nu_l}\right)\frac{\partial g_s}{\partial \nu_m} \\
&\quad + \frac{\partial \ln f_j}{\partial \theta_1}\frac{\partial^2 g_1}{\partial \nu_l \partial \nu_m} + \ldots + \frac{\partial \ln f_j}{\partial \theta_s}\frac{\partial^2 g_s}{\partial \nu_l \partial \nu_m}
\end{aligned}
$$

Thus, given condition **(B)** (note that the expectation of the last line above is 0),

$$
\begin{aligned}
J(\boldsymbol{\nu}_0) &= \sum_{j=1}^{k} \lambda_j J_j(\boldsymbol{\nu}_0) = \sum_{j=1}^{k} \lambda_j \left[-E_{\boldsymbol{\nu}_0}\left(\frac{\partial^2 \ln f_j}{\partial \nu_l \partial \nu_m}\right)\right] \\
&= \sum_{j=1}^{k} \lambda_j \mathbf{D}(\boldsymbol{\nu}_0)^T I_j(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\nu}_0) = \mathbf{D}(\boldsymbol{\nu}_0)^T I(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\nu}_0) \qquad (87)
\end{aligned}
$$

Lemma H.3

Under condition **(B)** and the conditions needed to establish Lemma H.2,

$$
\sqrt{n}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_0) \xrightarrow{D} \mathrm{N}\left(\mathbf{0}, \mathbf{D}(\boldsymbol{\nu}_0)\left(\mathbf{D}(\boldsymbol{\nu}_0)^T I(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\nu}_0)\right)^{-1}\mathbf{D}(\boldsymbol{\nu}_0)^T\right)
$$

where

$$
\boldsymbol{\theta}_n^* \equiv \mathbf{g}(\hat{\boldsymbol{\nu}}_n).
$$

Proof

By Taylor's theorem and assumption **(F)**, we have

$$
\sqrt{n}(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_0) = \sqrt{n}\begin{pmatrix} \partial g_1/\partial \nu_1 \ldots \partial g_1/\partial \nu_{s-r} \\ \vdots \\ \partial g_s/\partial \nu_1 \ldots \partial g_s/\partial \nu_{s-r} \end{pmatrix}(\hat{\boldsymbol{\nu}}_n - \boldsymbol{\nu}_0)
$$

where the partials are evaluated at $\boldsymbol{\nu}$'s that lie on the line segment between $\hat{\boldsymbol{\nu}}_n$ and $\boldsymbol{\nu}_0$. Then the Lemma follows from Lemma H.2, result 87, and assumption **(F)**.■

Corollary

Under conditions **(A0)** through **(D)**, **(E1)**, **(E2)**, and **(F)** through **(I)**,

$$
\sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) = \mathbf{O}_p(1) \qquad (88)
$$

Proof

Under conditions **(A0)** through **(D)**, Lehmann's theorem implies

$$
\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \mathbf{O}_p(1)
$$

This result and Lemma H.3 yield the corollary. ■

Next we need the following lemma.

Lemma H.4

Recall that $\boldsymbol{\theta}_n^* \equiv \mathbf{g}(\hat{\boldsymbol{\nu}}_n)$. Assuming that the partials exist, we have

$$\mathbf{D}(\hat{\boldsymbol{\nu}}_n)^T \begin{pmatrix} \partial \ln L(\boldsymbol{\theta}_n^*)/\partial\theta_1 \\ \vdots \\ \partial \ln L(\boldsymbol{\theta}_n^*)/\partial\theta_s \end{pmatrix} = \mathbf{0}_{(s-r)\times 1}$$

Proof

We have

$$\begin{aligned}
\mathbf{0}_{(s-r)\times 1} &= \begin{pmatrix} \partial \ln L(\mathbf{g}(\hat{\boldsymbol{\nu}}_n))/\partial\nu_1 \\ \vdots \\ \partial \ln L(\mathbf{g}(\hat{\boldsymbol{\nu}}_n))/\partial\nu_{s-r} \end{pmatrix} \\
&= \begin{pmatrix} \partial \ln L(\mathbf{g}(\hat{\boldsymbol{\nu}}_n))/\partial\theta_1 \times \partial g_1(\hat{\boldsymbol{\nu}}_n)/\partial\nu_1 + \ldots + \partial \ln L(\mathbf{g}(\hat{\boldsymbol{\nu}}_n))/\partial\theta_s \times \partial g_s(\hat{\boldsymbol{\nu}}_n)/\partial\nu_1 \\ \vdots \\ \partial \ln L(\mathbf{g}(\hat{\boldsymbol{\nu}}_n))/\partial\theta_1 \times \partial g_1(\hat{\boldsymbol{\nu}}_n)/\partial\nu_{s-r} + \ldots + \partial \ln L(\mathbf{g}(\hat{\boldsymbol{\nu}}_n))/\partial\theta_s \times \partial g_s(\hat{\boldsymbol{\nu}}_n)/\partial\nu_{s-r} \end{pmatrix} \\
&= \mathbf{D}(\hat{\boldsymbol{\nu}}_n)^T \begin{pmatrix} \partial \ln L(\boldsymbol{\theta}_n^*)/\partial\theta_1 \\ \vdots \\ \partial \ln L(\boldsymbol{\theta}_n^*)/\partial\theta_s \end{pmatrix}
\end{aligned}$$

We are now in a position to establish a closer relationship between $\boldsymbol{\theta}_n^*$ and $\hat{\boldsymbol{\theta}}_n$.

Lemma H.5

Under conditions **(A0)** through **(D)**, **(E1)**, **(E2)**, and **(F)** through **(I)**,

$$\mathbf{D}(\boldsymbol{\nu}_0)^T I(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) = \mathbf{o}_p(1)$$

Proof

By Taylor's theorem we have

$$\begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix} |_{\boldsymbol{\theta}_n^*} = \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix} |_{\hat{\boldsymbol{\theta}}_n} + \left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m}\right]_{s\times s} (\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n)$$

where the second partials are evaluated at $\boldsymbol{\theta}$'s along a line segment from $\boldsymbol{\theta}_n^*$ to $\hat{\boldsymbol{\theta}}_n$. Thus

$$\mathbf{D}(\hat{\boldsymbol{\nu}}_n)^T \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix} |_{\boldsymbol{\theta}_n^*} = \mathbf{D}(\hat{\boldsymbol{\nu}}_n)^T \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix} |_{\hat{\boldsymbol{\theta}}_n} + \mathbf{D}(\hat{\boldsymbol{\nu}}_n)^T \left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m}\right]_{s\times s} (\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) \tag{89}$$

By assumptions **(D)** (this assumption implies that the first partials of $\ln L$ exist in a neighborhood of $\boldsymbol{\theta}_0$) and **(F)** (implicit in this assumption is the conclusion that $D(\boldsymbol{\nu})$ exists in a neighborhood of $\boldsymbol{\nu}_0$), and Lemmas H.2, H.3, and H.4, the left hand side of Equation 89 is $\mathbf{o}_p(1)$. By the definition of $\hat{\boldsymbol{\theta}}_n$, the first term on the right hand side of Equation 89 is $\mathbf{o}_p(1)$. Thus we have

$$\mathbf{D}(\hat{\boldsymbol{\nu}}_n)^T \left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m}\right]_{s\times s} (\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) = \mathbf{o}_p(1)$$

and

$$\mathbf{D}(\hat{\boldsymbol{\nu}}_n)^T \left(- \left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m} \right]_{s\times s} /n \right) \sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) = \mathbf{o}_p(1) \tag{90}$$

where the second partials are evaluated at $\boldsymbol{\theta}$'s along a line segment from $\boldsymbol{\theta}_n^*$ to $\hat{\boldsymbol{\theta}}_n$.

By Lemma H.2, $\hat{\boldsymbol{\nu}}_n - \boldsymbol{\nu}_0 = \mathbf{o}_p(1)$. Thus since (by **(F)**) the partial derivatives in $\mathbf{D}(\hat{\boldsymbol{\nu}}_n)$ are continuous functions of the argument, the elements of $\mathbf{D}(\hat{\boldsymbol{\nu}}_n)$ converge in probability to the corresponding elements of $\mathbf{D}(\boldsymbol{\nu}_0)$.

Next we know (by **(A2)**, **(B)**, and the strong law of large numbers) that the elements of

$$- \left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m} \right]_{s\times s} |_{\boldsymbol{\theta}_0}/n$$

converge almost surely to those of $I(\boldsymbol{\theta}_0)$.

By Lemma H.3 and Lehmann's theorem 6.1, $\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_0 = \mathbf{o}_p(1)$ and $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = \mathbf{o}_p(1)$, so assumption **(D)** implies that the differences in the elements of

$$- \left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m} \right]_{s\times s} /n$$

evaluated at $\boldsymbol{\theta}$'s along a line segment from $\boldsymbol{\theta}_n^*$ to $\hat{\boldsymbol{\theta}}_n$ and the elements evaluated at $\boldsymbol{\theta}_0$ converge in probability to zero.

Finally, by the corollary to Lemma H.3,

$$\sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) = \mathbf{O}_p(1)$$

Thus, the lemma follows from result 90. ■

We next need to establish the following lemma.

Lemma H.6

Assume that $I(\boldsymbol{\theta}_0)$ is positive definite, and that conditions **(E1)** through **(E5)**, and **(F)** hold. Then

$$\mathbf{I}_{s\times s} = \mathbf{A}_1 + \mathbf{A}_2 \tag{91}$$

where

$$\mathbf{A}_1 \equiv I(\boldsymbol{\theta}_0)^{1/2}\mathbf{D}(\boldsymbol{\nu}_0)[\mathbf{D}(\boldsymbol{\nu}_0)^T I(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\nu}_0)]^{-1}\mathbf{D}(\boldsymbol{\nu}_0)^T I(\boldsymbol{\theta}_0)^{1/2}$$

and

$$\mathbf{A}_2 \equiv I(\boldsymbol{\theta}_0)^{-1/2}\mathbf{C}(\boldsymbol{\theta}_0)^T[\mathbf{C}(\boldsymbol{\theta}_0)I(\boldsymbol{\theta}_0)^{-1}\mathbf{C}(\boldsymbol{\theta}_0)^T]^{-1}\mathbf{C}(\boldsymbol{\theta}_0)I(\boldsymbol{\theta}_0)^{-1/2}$$

Proof

By assumption **(E5)**, $\mathbf{D}(\boldsymbol{\nu}_0)_{s\times(s-r)}$ is of full rank (rank $s-r$). Thus, since $I(\boldsymbol{\theta}_0)_{s\times s}$ is positive definite, $I(\boldsymbol{\theta}_0)^{1/2}$ is of full rank and

$$\mathbf{B}_1 \equiv I(\boldsymbol{\theta}_0)^{1/2}\mathbf{D}(\boldsymbol{\nu}_0)_{s\times(s-r)}$$

is of full rank. Thus

$$\mathbf{A}_1 = \mathbf{B}_1(\mathbf{B}_1^T\mathbf{B}_1)^{-1}\mathbf{B}_1^T$$

is the projection matrix onto the linear span of the columns of $\mathbf{B}_1$. Let $\mathbf{u}_1, \ldots, \mathbf{u}_{s-r}$ be an orthonormal basis of this linear span. Then

$$\mathbf{A}_1 = \mathbf{u}_1\mathbf{u}_1^T + \ldots + \mathbf{u}_{s-r}\mathbf{u}_{s-r}^T$$

Similarly, by assumption **(E4)**,

$$\mathbf{B}_2 \equiv I(\boldsymbol{\theta}_0)^{-1/2}\mathbf{C}(\boldsymbol{\theta}_0)^T_{s\times r}$$

is of full rank, and

$$\mathbf{A}_2 = \mathbf{B}_2(\mathbf{B}_2^T\mathbf{B}_2)^{-1}\mathbf{B}_2^T$$

is the projection matrix onto the linear span of the columns of $\mathbf{B}_2$. Let $\mathbf{u}_{s-r+1}, \ldots, \mathbf{u}_s$ be an orthonormal basis of this linear span. Then

$$\mathbf{A}_2 = \mathbf{u}_{s-r+1}\mathbf{u}_{s-r+1}^T + \ldots + \mathbf{u}_s\mathbf{u}_s^T$$

Finally, by **(E1)** through **(E3)**, and **(F)**, Lemma H.1 implies that $A_2A_1 = \mathbf{0}_{s\times s}$ so all the $\mathbf{u}$'s are orthonormal and

$$\mathbf{I}_{s\times s} = \mathbf{u}_1\mathbf{u}_1^T + \ldots + \mathbf{u}_s\mathbf{u}_s^T = \mathbf{A}_1 + \mathbf{A}_2 \blacksquare$$

Lemma H.7

Given conditions **(A0)** through **(D)**, **(E1)** through **(E4)**, and **(F)** through **(I)**, we have

$$n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^*)^T\mathbf{C}(\boldsymbol{\theta}_n^*)^T\left(\mathbf{C}(\boldsymbol{\theta}_0)I(\boldsymbol{\theta}_0)^{-1}\mathbf{C}(\boldsymbol{\theta}_0)^T\right)^{-1}\mathbf{C}(\boldsymbol{\theta}_n^*)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^*) \stackrel{D}{\rightarrow} \chi_r^2 \tag{92}$$

Proof

By Taylor's theorem, we have

$$\begin{pmatrix} R_1(\hat{\boldsymbol{\theta}}_n) \\ \vdots \\ R_r(\hat{\boldsymbol{\theta}}_n) \end{pmatrix} = \begin{pmatrix} R_1(\boldsymbol{\theta}_0) \\ \vdots \\ R_r(\boldsymbol{\theta}_0) \end{pmatrix} + \begin{pmatrix} [\partial R_1(\boldsymbol{\theta}_{*,1})/\partial\theta_l] \\ \vdots \\ [\partial R_r(\boldsymbol{\theta}_{*,r})/\partial\theta_l] \end{pmatrix}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \tag{93}$$

where $\boldsymbol{\theta}_{*,i}$ lies on the line segment between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$.

By the definition of the R_i's, the first term on the right-hand side of Equation 93 equals $\mathbf{0}_{r\times 1}$. Thus

$$\sqrt{n}\begin{pmatrix} R_1(\hat{\boldsymbol{\theta}}_n) \\ \vdots \\ R_r(\hat{\boldsymbol{\theta}}_n) \end{pmatrix} = \begin{pmatrix} [\partial R_1(\boldsymbol{\theta}_{*,1})/\partial\theta_l] \\ \vdots \\ [\partial R_r(\boldsymbol{\theta}_{*,r})/\partial\theta_l] \end{pmatrix}\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \tag{94}$$

Assumptions **(E3)** and **(E4)** and results 82 and 94 imply

$$\left(\mathbf{C}(\boldsymbol{\theta}_0)I(\boldsymbol{\theta}_0)^{-1}\mathbf{C}(\boldsymbol{\theta}_0)^T\right)^{-1/2}\sqrt{n}\begin{pmatrix} R_1(\hat{\boldsymbol{\theta}}_n) \\ \vdots \\ R_r(\hat{\boldsymbol{\theta}}_n) \end{pmatrix} \stackrel{D}{\rightarrow} \mathrm{N}(\mathbf{0}_{r\times 1}, \mathbf{I}_{r\times r}) \tag{95}$$

Next note that by assumption **(E1)**, because $\boldsymbol{\theta}_n^* = \mathbf{g}(\hat{\boldsymbol{\nu}}_n)$, $R_1(\boldsymbol{\theta}_n^*) = \ldots = R_r(\boldsymbol{\theta}_n^*) = 0$, so by Taylor's theorem we have

$$\sqrt{n}\begin{pmatrix} R_1(\hat{\boldsymbol{\theta}}_n) \\ \vdots \\ R_r(\hat{\boldsymbol{\theta}}_n) \end{pmatrix} = \begin{pmatrix} [\partial R_1(\boldsymbol{\theta}_{*,1})/\partial\theta_l] \\ \vdots \\ [\partial R_r(\boldsymbol{\theta}_{*,r})/\partial\theta_l] \end{pmatrix} \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^*) \tag{96}$$

where $\boldsymbol{\theta}_{*,i}$ lies on the line segment between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_n^*$. By assumption **(E3)** and results 82, 88, and 96

$$\sqrt{n}\begin{pmatrix} R_1(\hat{\boldsymbol{\theta}}_n) \\ \vdots \\ R_r(\hat{\boldsymbol{\theta}}_n) \end{pmatrix} - \mathbf{C}(\boldsymbol{\theta}_n^*)\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^*) = \mathbf{o}_p(1) \tag{97}$$

From results 95 and 97, we have

$$\left(\mathbf{C}(\boldsymbol{\theta}_0)I(\boldsymbol{\theta}_0)^{-1}\mathbf{C}(\boldsymbol{\theta}_0)^T\right)^{-1/2}\mathbf{C}(\boldsymbol{\theta}_n^*)\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^*) \xrightarrow{D} \mathrm{N}(\mathbf{0}_{r\times 1}, \mathbf{I}_{r\times r})$$

or

$$n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^*)^T\mathbf{C}(\boldsymbol{\theta}_n^*)^T\left(\mathbf{C}(\boldsymbol{\theta}_0)I(\boldsymbol{\theta}_0)^{-1}\mathbf{C}(\boldsymbol{\theta}_0)^T\right)^{-1}\mathbf{C}(\boldsymbol{\theta}_n^*)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^*) \xrightarrow{D} \chi_r^2$$

which completes the proof. ■

We are now prepared to establish our main results.

Theorem H.1 (likelihood ratio statistic)

Under assumptions **(A0)** through **(D)** and **(E1)** through **(I)**, we have

$$2(\ln(L(\hat{\boldsymbol{\theta}}_n)) - \ln(L(\boldsymbol{\theta}_n^*))) \xrightarrow{D} \chi_r^2$$

Proof

By Taylor's theorem we have

$$\begin{aligned} 2(\ln(L(\boldsymbol{\theta}_n^*)) - \ln(L(\hat{\boldsymbol{\theta}}_n))) &= 2 \times \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix}^T |_{\hat{\boldsymbol{\theta}}_n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) \\ &\quad + \sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n)^T\left(\left[\frac{\partial^2 \ln L}{\partial\theta_l\partial\theta_m}\right]/n\right)\sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) \end{aligned} \tag{98}$$

where the second partials are evaluated at a point on the line segment between $\boldsymbol{\theta}_n^*$ and $\hat{\boldsymbol{\theta}}_n$.

Now by the definition of $\hat{\boldsymbol{\theta}}_n$ and the corollary to Lemma H.3, the first term on the right of Equation 98 is $o_p(1)$. By **(A2)**, **(B)**, **(D)**, Lemma H.3, result 82, and the law of large numbers, the matrix in the second term on the right of the equation converges in probability to $-I(\boldsymbol{\theta}_0)$. Thus, by the corollary to Lemma H.3 we will be done if we can establish that

$$Q \equiv \sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n)^T I(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) \xrightarrow{D} \chi_r^2 \tag{99}$$

From Lemma H.6 we have

$$\begin{aligned} Q &= \sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n)^T I(\boldsymbol{\theta}_0)^{1/2}\mathbf{I}_{s\times s}I(\boldsymbol{\theta}_0)^{1/2}\sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) \\ &= \sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n)^T I(\boldsymbol{\theta}_0)^{1/2}\mathbf{A}_1 I(\boldsymbol{\theta}_0)^{1/2}\sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) \\ &\quad + \sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n)^T I(\boldsymbol{\theta}_0)^{1/2}\mathbf{A}_2 I(\boldsymbol{\theta}_0)^{1/2}\sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) \\ &= \sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n)^T I(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\nu}_0)[\mathbf{D}(\boldsymbol{\nu}_0)^T I(\boldsymbol{\theta}_0)\mathbf{D}(\boldsymbol{\nu}_0)]^{-1}\mathbf{D}(\boldsymbol{\nu}_0)^T I(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) \\ &\quad + \sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n)^T\mathbf{C}(\boldsymbol{\theta}_0)^T[\mathbf{C}(\boldsymbol{\theta}_0)I(\boldsymbol{\theta}_0)^{-1}\mathbf{C}(\boldsymbol{\theta}_0)^T]^{-1}\mathbf{C}(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n^* - \hat{\boldsymbol{\theta}}_n) \end{aligned}$$

By Lemma H.5 the $\mathbf{D}(\boldsymbol{\nu}_0)$ term converges in probability to 0. By Lemma H.7, assumption **(E3)**, and Lemma H.3 and its corollary, the $\mathbf{C}(\boldsymbol{\theta}_0)$ term converges in distribution to a χ^2_r random variable. ■

As one would expect we can obtain asymptotic distributions of the Wald and Rao statistics as corollaries to the development leading to Theorem H.1.

Theorem H.2 (Wald's statistic)

Assume that $I(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$. Then under assumptions **(A0)** through **(D)**, and **(E1)** through **(E4)**, we have

$$n \begin{pmatrix} R_1(\hat{\boldsymbol{\theta}}_n) \\ \vdots \\ R_r(\hat{\boldsymbol{\theta}}_n) \end{pmatrix}^T \left(\mathbf{C}(\hat{\boldsymbol{\theta}}_n) I(\hat{\boldsymbol{\theta}}_n)^{-1} \mathbf{C}(\hat{\boldsymbol{\theta}}_n)^T\right)^{-1} \begin{pmatrix} R_1(\hat{\boldsymbol{\theta}}_n) \\ \vdots \\ R_r(\hat{\boldsymbol{\theta}}_n) \end{pmatrix} \stackrel{D}{\to} \chi^2_r \tag{100}$$

Proof

By result (95) we will be done if we establish

$$\left(\mathbf{C}(\hat{\boldsymbol{\theta}}_n) I(\hat{\boldsymbol{\theta}}_n)^{-1} \mathbf{C}(\hat{\boldsymbol{\theta}}_n)^T\right)^{-1} - \left(\mathbf{C}(\boldsymbol{\theta}_0) I(\boldsymbol{\theta}_0)^{-1} \mathbf{C}(\boldsymbol{\theta}_0)^T\right)^{-1} = [o_p(1)]_{r\times r} \tag{101}$$

Result 101 follows from result 82, assumptions **(C)**, **(E3)**, and **(E4)**, and the assumption that $I(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$. ■

Theorem H.3 (Rao's statistic)

Assume that $I(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$. Then under assumptions **(A0)** through **(D)** and **(E1)** through **(I)**, we have

$$\frac{1}{n} \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix}^T_{|\boldsymbol{\theta}^*_n} I(\boldsymbol{\theta}^*_n)^{-1} \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix}_{|\boldsymbol{\theta}^*_n} \stackrel{D}{\to} \chi^2_r \tag{102}$$

Proof

We have

$$\begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix}_{|\boldsymbol{\theta}^*_n} = \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix}_{|\hat{\boldsymbol{\theta}}_n} + \left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m}\right]_{s\times s} (\boldsymbol{\theta}^*_n - \hat{\boldsymbol{\theta}}_n)$$

where the second partials are evaluated at points on the line segment between $\boldsymbol{\theta}^*_n$ and $\hat{\boldsymbol{\theta}}_n$. Thus the expression on the left hand side of (102) is equal to

$$\sqrt{n}(\boldsymbol{\theta}^*_n - \hat{\boldsymbol{\theta}}_n)^T \left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m}/n\right] I(\boldsymbol{\theta}^*_n)^{-1} \left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m}/n\right] \sqrt{n}(\boldsymbol{\theta}^*_n - \hat{\boldsymbol{\theta}}_n)$$

so by the corollary to Lemma H.3 and result 99, we will be done if we can establish that

$$\left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m}/n\right] I(\boldsymbol{\theta}^*_n)^{-1} \left[\frac{\partial^2 \ln L}{\partial\theta_l \partial\theta_m}/n\right] - I(\boldsymbol{\theta}_0) = [o_p(1)]_{s\times s} \tag{103}$$

where the second partials are evaluated at points on the line segment between $\boldsymbol{\theta}^*_n$ and $\hat{\boldsymbol{\theta}}_n$. Equality 103 follows from the assumed continuity of $I(\boldsymbol{\theta})$, result 82, Lemma H.3, assumptions **(A2)**, **(B)** and **(D)**, and the law of large numbers. ■

18.4 Using Newton method estimators rather than solutions of the likelihood equations

We sometimes find it difficult to obtain closed form solutions to the likelihood equations or to establish that the solutions are unique. In this case, if we begin with a $\sqrt{n}$-consistent estimator and take a Newton step we will be led to an estimator that is asymptotically efficient. (See, for example, theorems 3.1 and 4.2 in chapter 6 of Lehmann (1983).) Here we establish that in likelihood ratio, Wald, and Rao tests, we can replace solutions of the likelihood equations with Newton step estimators.

Lemma H.8

Assume that conditions **(A0)** through **(D)** hold.

Let $\boldsymbol{\theta}_{n,c}$ be a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_0$. That is, assume that

$$\sqrt{n}(\boldsymbol{\theta}_{n,c} - \boldsymbol{\theta}_0) = \mathbf{O}_p(1) \tag{104}$$

Then, 1) with probability approaching one as $n \to \infty$, the Newton estimator,

$$\boldsymbol{\theta}_{n,\text{Newt}} \equiv - \left[\frac{\partial^2 \ln L}{\partial\theta_l \theta_m}\right]^{-1}_{s\times s} |_{\boldsymbol{\theta}_{n,c}} \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix} |_{\boldsymbol{\theta}_{n,c}} + \boldsymbol{\theta}_{n,c} \tag{105}$$

is well-defined (that is the partials exist and the matrix is invertible), and 2)

$$\boldsymbol{\theta}_{n,\text{Newt}} - \hat{\boldsymbol{\theta}}_n = \mathbf{O}_p(n^{-1}) \tag{106}$$

where $\hat{\boldsymbol{\theta}}_n$ is a consistent solution of the likelihood equations guaranteed by theorem 6.1 in chapter 6 of Lehmann (1983) (see result 82).

Proof

We will be making use of the fact that the Newton method yields quadratic convergence. In particular, we will verify the conditions of theorem 5.2.1 in Dennis and Schnabel (1983).

By assumption **(D)** we can define

$$J_n(\boldsymbol{\theta}) \equiv - \left(\left[\frac{\partial^2 \ln L}{\partial\theta_l \theta_m}\right]_{s\times s} |_{\boldsymbol{\theta}} \right) /n$$

We have

$$J_n(\hat{\boldsymbol{\theta}}_n) - J_n(\boldsymbol{\theta}_0) = - \left[\sum_{j=1}^{k} (n_j/n) \sum_{i=1}^{n_j} \left(\frac{\partial^2 \ln f(X_{ij};\boldsymbol{\theta})}{\partial\theta_l \partial\theta_m} |_{\hat{\boldsymbol{\theta}}_n} - \frac{\partial^2 \ln f(X_{ij};\boldsymbol{\theta})}{\partial\theta_l \partial\theta_m} |_{\boldsymbol{\theta}_0} \right) /n_j \right]_{s\times s}$$

and, making use of assumption **(D)**, by Taylor's theorem

$$\frac{\partial^2 \ln f(X_{ij};\boldsymbol{\theta})}{\partial\theta_l \partial\theta_m} |_{\hat{\boldsymbol{\theta}}_n} - \frac{\partial^2 \ln f(X_{ij};\boldsymbol{\theta})}{\partial\theta_l \partial\theta_m} |_{\boldsymbol{\theta}_0} = \left(\frac{\partial^3 \ln f(X_{ij};\boldsymbol{\theta})}{\partial\theta_l \partial\theta_m \partial\theta_1}, \ldots, \frac{\partial^3 \ln f(X_{ij};\boldsymbol{\theta})}{\partial\theta_l \partial\theta_m \partial\theta_s} \right) |_{\boldsymbol{\theta}^*_{lm,n}} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

where $\boldsymbol{\theta}^*_{lm,n}$ lies on the line segment between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$.

Thus by assumption **(D)**, for $\hat{\boldsymbol{\theta}}_n \in T$ (an open neighborhood of $\boldsymbol{\theta}_0$), the absolute value of the lmth element of $J_n(\hat{\boldsymbol{\theta}}_n) - J_n(\boldsymbol{\theta}_0)$ is bounded by

$$\sum_{j=1}^{k} (n_j/n) \sum_{i=1}^{n_j} \sum_{p=1}^{s} M_{lmp,j}(X_{ij}) \left| \hat{\theta}_{pn} - \theta_{p0} \right| /n_j \tag{107}$$

Since (by assumptions **(A2)** and **(D)** and the strong law of large numbers)

$$\sum_{i=1}^{n_j} M_{lmp,j}(X_{ij})/n_j \stackrel{a.s.}{\to} m_{lmp,j} < \infty$$

for $l, m, p \in \{1, \ldots, s\}$, $j \in \{1, \ldots, k\}$, results 82 and 107 imply that

$$||J_n(\hat{\boldsymbol{\theta}}_n) - J_n(\boldsymbol{\theta}_0)||_F \stackrel{p}{\to} 0 \tag{108}$$

where $||M||_F$ denotes the Frobenius norm of the matrix M.

Now by assumptions **(A2)** and **(B)** and the strong law of large numbers, we know that

$$J_n(\boldsymbol{\theta}_0) \stackrel{a.s.}{\to} I(\boldsymbol{\theta}_0) \tag{109}$$

Results 108 and 109 imply that

$$||J_n(\hat{\boldsymbol{\theta}}_n) - I(\boldsymbol{\theta}_0)||_F \stackrel{p}{\to} 0 \tag{110}$$

By assumption **(C)**, $I(\boldsymbol{\theta}_0)$ is positive definite. Since the inverse and norm of a matrix are continuous functions of the elements of the matrix, this implies that given any $\delta > 0$, we can find an $N_{\delta,1}$ such that $n > N_{\delta,1}$ implies that

$$\text{Prob}\left(||J_n(\hat{\boldsymbol{\theta}}_n)^{-1}||_F < 2||I(\boldsymbol{\theta}_0)^{-1}||_F\right) > 1 - \delta \tag{111}$$

Since (see, for example, theorem 3.1.3 of Dennis and Schnabel (1983))

$$||M_{s\times s}||_F/\sqrt{s} \leq ||M_{s\times s}||_2 \leq ||M_{s\times s}||_F$$

where $||M||_2$ denotes the l_2 induced matrix norm of M (see, for example, pages 43 and 44 of Dennis and Schnabel (1983)), result 111 implies that for $n > N_{\delta,1}$

$$\text{Prob}\left(||J_n(\hat{\boldsymbol{\theta}}_n)^{-1}||_F < 2\sqrt{s}||I(\boldsymbol{\theta}_0)^{-1}||_2\right) > 1 - \delta$$

or

$$\text{Prob}\left(||J_n(\hat{\boldsymbol{\theta}}_n)^{-1}||_F < \beta\right) > 1 - \delta \tag{112}$$

where $\beta \equiv 2\sqrt{s}/\lambda$ and λ is the smallest eigenvalue of $I(\boldsymbol{\theta}_0)$.

Let $r > 0$ be such that $D(\boldsymbol{\theta}_0; 2r) \subset T$, the open neighborhood of $\boldsymbol{\theta}_0$ in assumptions **(A)** through **(D)**. (Here, $D(\boldsymbol{\theta}_0; 2r)$ denotes the open ball of radius $2r$ centered at $\boldsymbol{\theta}_0$.) Since (result 82) $\hat{\boldsymbol{\theta}}_n \stackrel{p}{\to} \boldsymbol{\theta}_0$, given any $\delta > 0$, we can find an $N_{\delta,2}$ such that $n > N_{\delta,2}$ implies that $\text{Prob}(\hat{\boldsymbol{\theta}}_n \in D(\boldsymbol{\theta}_0; r)) > 1 - \delta$.

Now, provided that $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in D(\boldsymbol{\theta}_0; 2r)$,

$$||J_n(\boldsymbol{\theta}_1) - J_n(\boldsymbol{\theta}_2)||_F = ||[a_{lm}]_{s\times s}||_F$$

where

$$\begin{aligned} a_{lm} &\equiv \sum_{j=1}^{k}(n_j/n)\sum_{i=1}^{n_j}\left(\frac{\partial^2 \ln f(X_{ij};\boldsymbol{\theta})}{\partial\theta_l\partial\theta_m}|_{\boldsymbol{\theta}_1} - \frac{\partial^2 \ln f(X_{ij};\boldsymbol{\theta})}{\partial\theta_l\partial\theta_m}|_{\boldsymbol{\theta}_2}\right)/n_j \\ &= \sum_{j=1}^{k}(n_j/n)\sum_{i=1}^{n_j}\left(\frac{\partial^3 \ln f(X_{ij};\boldsymbol{\theta})}{\partial\theta_l\partial\theta_m\partial\theta_1}, \ldots, \frac{\partial^3 \ln f(X_{ij};\boldsymbol{\theta})}{\partial\theta_l\partial\theta_m\partial\theta_s}\right)|_{\boldsymbol{\theta}^*_{lm,n}}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)/n_j \end{aligned}$$

where $\boldsymbol{\theta}^*_{lm,n}$ lies on the line segment between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$.

Thus, by assumption **(D)**, if $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ are within r of $\hat{\boldsymbol{\theta}}_n$, then for $n > N_{\delta,2}$, with probability greater than $1 - \delta$, we have

$$||J_n(\boldsymbol{\theta}_1) - J_n(\boldsymbol{\theta}_2)||_F^2 = \sum_{l=1}^{s}\sum_{m=1}^{s} a_{lm}^2 \leq \left(\sum_{l=1}^{s}\sum_{m=1}^{s} |a_{lm}|\right)^2 \tag{113}$$

$$\leq \left(\sum_{l=1}^{s}\sum_{m=1}^{s}\sum_{j=1}^{k}(n_j/n)\sum_{i=1}^{n_j}\sum_{p=1}^{s}\left(M_{lmp,j}(X_{ij})/n_j\right)|\theta_{p1} - \theta_{p2}|\right)^2$$

Since (by assumption **(D)**)

$$\sum_{i=1}^{n_j} M_{lmp,j}(X_{ij})/n_j \overset{a.s.}{\to} m_{lmp,j} < \infty$$

for $l, m, p \in \{1, \ldots, s\}$, $j \in \{1, \ldots, k\}$, result 113 implies that given any $\delta > 0$, we can find an $N_{\delta,3}$ such that $n > N_{\delta,3}$ implies that, if $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ are within r of $\hat{\boldsymbol{\theta}}_n$, then with probability greater than $1 - \delta$,

$$||J_n(\boldsymbol{\theta}_1) - J_n(\boldsymbol{\theta}_2)||_F^2 \leq \left(\sum_{l=1}^{s}\sum_{m=1}^{s}\sum_{j=1}^{k}(n_j/n)\sum_{p=1}^{s}(m_{lmp,j} + 1)|\theta_{p1} - \theta_{p2}|\right)^2$$

$$\leq \gamma^2||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||^2$$

where

$$\gamma \equiv \left(\sum_{l=1}^{s}\sum_{m=1}^{s}\sum_{j=1}^{k}(\lambda_j + 1)\sum_{p=1}^{s}(m_{lmp,j} + 1)\right) < \infty$$

(Recall that $n_j/n \to \lambda_j$ as $n \to \infty$.) That is, for $n > N_{\delta,3}$, with probability greater than $1 - \delta$,

$$J_n \in \mathrm{Lip}_\gamma(D(\hat{\boldsymbol{\theta}}_n, r)) \tag{114}$$

Results 112 and 114 permit us to invoke Dennis and Schnabel's (1983) theorem 5.2.1 to conclude that given any $\delta > 0$, we can find an $N_{\delta,4}$ such that $n > N_{\delta,4}$ implies that with probability greater than $1 - \delta$

$$\boldsymbol{\theta}_{n,\mathrm{Newt}} \equiv -\left[\frac{\partial^2 \ln L}{\partial\theta_l\partial\theta_m}\right]^{-1}|_{\boldsymbol{\theta}_{n,c}} \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix}|_{\boldsymbol{\theta}_{n,c}} + \boldsymbol{\theta}_{n,c}$$

is well-defined (that is the partials exist and the matrix is invertible), and

$$||\boldsymbol{\theta}_{n,\mathrm{Newt}} - \hat{\boldsymbol{\theta}}_n|| \leq \beta \times \gamma \times ||\boldsymbol{\theta}_{n,c} - \hat{\boldsymbol{\theta}}_n||^2 \tag{115}$$

provided that

$$||\boldsymbol{\theta}_{n,c} - \hat{\boldsymbol{\theta}}_n|| < \epsilon \equiv \min(r, \frac{1}{2\beta\gamma})$$

But by result 82 and the fact that $\boldsymbol{\theta}_{n,c}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_0$, we have $\sqrt{n}(\boldsymbol{\theta}_{n,c} - \hat{\boldsymbol{\theta}}_n) = \mathbf{O}_p(1)$ so given any $\delta > 0$ we can find a K_δ and an $N_{\delta,5}$ such that $n > N_{\delta,5}$ implies

$$\text{Prob}(\sqrt{n}||\boldsymbol{\theta}_{n,c} - \hat{\boldsymbol{\theta}}_n|| \leq K_\delta) \geq 1 - \delta \tag{116}$$

If we require that $N_{\delta,5} > K_\delta^2/\epsilon^2$, then $n > N_{\delta,5}$ also implies

$$\text{Prob}(||\boldsymbol{\theta}_{n,c} - \hat{\boldsymbol{\theta}}_n|| < \epsilon) \geq 1 - \delta \tag{117}$$

Results 115, 116, and 117 imply that given any $\delta > 0$, we can find an N such that $n > N$ implies that with probability greater than $1 - \delta$,

$$||\boldsymbol{\theta}_{n,\text{Newt}} - \hat{\boldsymbol{\theta}}_n|| \leq \beta \times \gamma \times K_\delta^2/n \tag{118}$$

which completes the proof of the lemma. ■

Corollary 1

Assume that conditions **(A0)** through **(D)** hold. Then

$$\sqrt{n}(\boldsymbol{\theta}_{n,\text{Newt}} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, I(\boldsymbol{\theta}_0)^{-1})$$

Proof

Given conditions **(A0)** through **(D)**, Lehmann's theorem 6.1 implies that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, I(\boldsymbol{\theta}_0)^{-1}) \tag{119}$$

Results 106 and 119 yield the corollary.

Corollary 2

Assume that conditions **(A0)** through **(A)**, **(E1)**, **(E2)**, and **(F)** through **(I)** hold.

Let $\boldsymbol{\nu}_{n,c}$ be a $\sqrt{n}$-consistent estimator of $\boldsymbol{\nu}_0$. That is, assume that

$$\sqrt{n}(\boldsymbol{\nu}_{n,c} - \boldsymbol{\nu}_0) = \mathbf{O}_p(1) \tag{120}$$

Then, 1) with probability approaching one as $n \to \infty$, the Newton estimator,

$$\boldsymbol{\nu}_{n,\text{Newt}} \equiv -\left[\frac{\partial^2 \ln L}{\partial \nu_l \partial \nu_m}\right]^{-1} |_{\boldsymbol{\nu}_{n,c}} \begin{pmatrix} \partial \ln L/\partial \nu_1 \\ \vdots \\ \partial \ln L/\partial \nu_{s-r} \end{pmatrix} |_{\boldsymbol{\nu}_{n,c}} + \boldsymbol{\nu}_{n,c} \tag{121}$$

is well-defined (that is the partials exist and the matrix is invertible), and 2)

$$\boldsymbol{\nu}_{n,\text{Newt}} - \hat{\boldsymbol{\nu}}_n = \mathbf{O}_p(n^{-1})$$

where $\hat{\boldsymbol{\nu}}_n$ is the consistent solution of the constrained likelihood equations guaranteed by Lemma H.2.

We are now in a position to establish the following three useful theorems.

Theorem H.4 (likelihood ratio statistic, Newton step version)
Under assumptions **(A0)** through **(I)**, we have

$$2(\ln L(\boldsymbol{\theta}_{n,\text{Newt}}) - \ln L(\mathbf{g}(\boldsymbol{\nu}_{n,\text{Newt}}))) \xrightarrow{D} \chi^2_r$$

where $\boldsymbol{\theta}_{n,\text{Newt}}$ is defined by Equation 105 and $\boldsymbol{\nu}_{n,\text{Newt}}$ is defined by Equation 121.
Proof
We have

$$\begin{aligned}\ln L(\boldsymbol{\theta}_{n,\text{Newt}}) - \ln L(\hat{\boldsymbol{\theta}}_n) &= \left(\frac{\partial \ln L}{\partial \theta_1} \cdots \frac{\partial \ln L}{\partial \theta_s}\right)|_{\hat{\boldsymbol{\theta}}_n}(\boldsymbol{\theta}_{n,\text{Newt}} - \hat{\boldsymbol{\theta}}_n) \\ &+\frac{1}{2}\sqrt{n}(\boldsymbol{\theta}_{n,\text{Newt}} - \hat{\boldsymbol{\theta}}_n)^T\left[\frac{\partial^2 \ln L}{\partial \theta_l \partial \theta_m}|_{\boldsymbol{\theta}_{n,b}}/n\right]\sqrt{n}(\boldsymbol{\theta}_{n,\text{Newt}} - \hat{\boldsymbol{\theta}}_n)\end{aligned} \tag{122}$$

where $\boldsymbol{\theta}_{n,b}$ lies on the line segment between $\boldsymbol{\theta}_{n,\text{Newt}}$ and $\hat{\boldsymbol{\theta}}_n$.
We know (by **(A2)**, **(B)**, and the strong law of large numbers) that the elements of

$$-\left[\frac{\partial^2 \ln L}{\partial \theta_l \partial \theta_m}\right]_{s\times s}|_{\boldsymbol{\theta}_0}/n$$

converge in probability to those of $I(\boldsymbol{\theta}_0)$.
Since (by Lemma H.8) $\boldsymbol{\theta}_{n,\text{Newt}} - \hat{\boldsymbol{\theta}}_n = o_p(1)$ and (by result 82) $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_p(1)$, assumption **(D)** implies that the differences in the elements of

$$-\left[\frac{\partial^2 \ln L}{\partial \theta_l \partial \theta_m}\right]_{s\times s}/n$$

evaluated at $\boldsymbol{\theta}_{n,b}$ and the elements evaluated at $\boldsymbol{\theta}_0$ converge in probability to zero. Thus, by Equation 122 and Lemma H.8

$$\ln L(\boldsymbol{\theta}_{n,\text{Newt}}) - \ln L(\hat{\boldsymbol{\theta}}_n) = o_p(1) \tag{123}$$

Similarly,

$$\ln L(\mathbf{g}(\boldsymbol{\nu}_{n,\text{Newt}})) - \ln L(\mathbf{g}(\hat{\boldsymbol{\nu}}_n)) = \left(\frac{\partial \ln L(\mathbf{g}(\boldsymbol{\nu})}{\partial \nu_1} \cdots \frac{\partial \ln L(\mathbf{g}(\boldsymbol{\nu})}{\partial \nu_{s-r}}\right)|_{\hat{\boldsymbol{\nu}}_n}(\boldsymbol{\nu}_{n,\text{Newt}} - \hat{\boldsymbol{\nu}}_n) \tag{124}$$

$$+\frac{1}{2}\sqrt{n}(\boldsymbol{\nu}_{n,\text{Newt}} - \hat{\boldsymbol{\nu}}_n)^T\left[\frac{\partial^2 \ln L(\mathbf{g}(\boldsymbol{\nu})}{\partial \nu_l \partial \nu_m}|_{\boldsymbol{\nu}_{n,b}}/n\right]\sqrt{n}(\boldsymbol{\nu}_{n,\text{Newt}} - \hat{\boldsymbol{\nu}}_n)$$

where $\boldsymbol{\nu}_{n,b}$ lies on the line segment between $\boldsymbol{\nu}_{n,\text{Newt}}$ and $\hat{\boldsymbol{\nu}}_n$, and since (as above) the matrix converges in probability to $-J(\boldsymbol{\nu}_0)$, Corollary 2 to Lemma H.8 implies that

$$\ln L(\mathbf{g}(\boldsymbol{\nu}_{n,\text{Newt}})) - \ln L(\mathbf{g}(\hat{\boldsymbol{\nu}}_n)) = o_p(1) \tag{125}$$

Theorem H.1, and results 123 and 125 (recall that $\boldsymbol{\theta}^*_n \equiv \mathbf{g}(\hat{\boldsymbol{\nu}}_n)$) complete the proof. ■
Theorem H.5 (Wald's Statistic, Newton step version)
Assume that $I(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$. Then under assumptions **(A0)** through **(D)**, and **(E1)** through **(E4)**, we have

$$n\begin{pmatrix} R_1(\boldsymbol{\theta}_{n,\text{Newt}}) \\ \vdots \\ R_r(\boldsymbol{\theta}_{n,\text{Newt}})\end{pmatrix}^T \left(\mathbf{C}(\boldsymbol{\theta}_{n,\text{Newt}})I(\boldsymbol{\theta}_{n,\text{Newt}})^{-1}\mathbf{C}(\boldsymbol{\theta}_{n,\text{Newt}})^T\right)^{-1}\begin{pmatrix} R_1(\boldsymbol{\theta}_{n,\text{Newt}}) \\ \vdots \\ R_r(\boldsymbol{\theta}_{n,\text{Newt}})\end{pmatrix} \xrightarrow{D} \chi^2_r \tag{126}$$

where $\boldsymbol{\theta}_{n,\text{Newt}}$ is defined by Equation 105.

Proof

For $i \in \{1, \ldots, r\}$ we have

$$R_i(\boldsymbol{\theta}_{n,\text{Newt}}) - R_i(\hat{\boldsymbol{\theta}}_n) = \left(\frac{\partial R_i}{\partial \theta_1}, \ldots, \frac{\partial R_i}{\partial \theta_s}\right)\left(\boldsymbol{\theta}_{n,\text{Newt}} - \hat{\boldsymbol{\theta}}_n\right) \tag{127}$$

where the partials are evaluated on a line segment between $\boldsymbol{\theta}_{n,\text{Newt}}$ and $\hat{\boldsymbol{\theta}}_n$.

By results 127 and 82, Corollary 1 to Lemma H.8, assumption **(E3)**, and Lemma H.8, for $i \in \{1, \ldots, r\}$ we have

$$\sqrt{n}\left(R_i(\boldsymbol{\theta}_{n,\text{Newt}}) - R_i(\hat{\boldsymbol{\theta}}_n)\right) = o_p(1) \tag{128}$$

Result 128 implies that result 95 holds with $\hat{\boldsymbol{\theta}}_n$ replaced by $\boldsymbol{\theta}_{n,\text{Newt}}$. Thus we will be done if we can establish

$$\left(\mathbf{C}(\boldsymbol{\theta}_{n,\text{Newt}})I(\boldsymbol{\theta}_{n,\text{Newt}})^{-1}\mathbf{C}(\boldsymbol{\theta}_{n,\text{Newt}})^T\right)^{-1} - \left(\mathbf{C}(\boldsymbol{\theta}_0)I(\boldsymbol{\theta}_0)^{-1}\mathbf{C}(\boldsymbol{\theta}_0)^T\right)^{-1} = [o_p(1)]_{r\times r} \tag{129}$$

Result 129 follows from Corollary 1 to Lemma H.8, assumptions **(C)**, **(E3)**, and **(E4)**, and the assumption that $I(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$. ■

Theorem H.6 (Rao's Statistic, Newton step version)

Define

$$\boldsymbol{\theta}^*_{n,\text{Newt}} \equiv \mathbf{g}(\boldsymbol{\nu}_{n,\text{Newt}})$$

where $\mathbf{g}$ is defined in connection with assumption **(E1)** and $\boldsymbol{\nu}_{n,\text{Newt}}$ is defined by Equation 121. Assume that $I(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$. Then under assumptions **(A0)** through **(D)** and **(E1)** through **(I)**, we have

$$\frac{1}{n}\left(\begin{array}{c}\partial \ln L/\partial\theta_1\\ \vdots \\ \partial \ln L/\partial\theta_s\end{array}\right)^T_{|\boldsymbol{\theta}^*_{n,\text{Newt}}} I(\boldsymbol{\theta}^*_{n,\text{Newt}})^{-1}\left(\begin{array}{c}\partial \ln L/\partial\theta_1\\ \vdots \\ \partial \ln L/\partial\theta_s\end{array}\right)_{|\boldsymbol{\theta}^*_{n,\text{Newt}}} \xrightarrow{D} \chi^2_r \tag{130}$$

Proof

We have

$$\mathbf{g}(\boldsymbol{\nu}_{n,\text{Newt}}) - \mathbf{g}(\hat{\boldsymbol{\nu}}_n) = \left[\frac{\partial g_i}{\partial \nu_j}\right]_{s\times(s-r)}(\boldsymbol{\nu}_{n,\text{Newt}} - \hat{\boldsymbol{\nu}}_n)$$

where the partials are evaluated on a line segment between $\boldsymbol{\nu}_{n,\text{Newt}}$ and $\hat{\boldsymbol{\nu}}_n$. Thus, by Lemma H.2, Corollary 2 to Lemma H.8, and assumption **(F)**, we have

$$\sqrt{n}(\boldsymbol{\theta}^*_{n,\text{Newt}} - \boldsymbol{\theta}^*_n) = \sqrt{n}(\mathbf{g}(\boldsymbol{\nu}_{n,\text{Newt}}) - \mathbf{g}(\hat{\boldsymbol{\nu}}_n)) = \mathbf{o}_p(1) \tag{131}$$

Result 131 and Lemma H.3 imply

$$\sqrt{n}(\boldsymbol{\theta}^*_{n,\text{Newt}} - \boldsymbol{\theta}_0) = \mathbf{O}_p(1) \tag{132}$$

Result 131 and the corollary to Lemma H.3 imply

$$\sqrt{n}(\boldsymbol{\theta}^*_{n,\text{Newt}} - \hat{\boldsymbol{\theta}}_n) = \mathbf{O}_p(1) \tag{133}$$

Results 131, 133, and 99, and the corollary to Lemma H.3 imply

$$\sqrt{n}(\boldsymbol{\theta}^*_{n,\mathrm{Newt}} - \hat{\boldsymbol{\theta}}_n)^T I(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}^*_{n,\mathrm{Newt}} - \hat{\boldsymbol{\theta}}_n) \stackrel{D}{\to} \chi^2_r \tag{134}$$

Now we have

$$\begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix}_{|\boldsymbol{\theta}^*_{n,\mathrm{Newt}}} = \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_s \end{pmatrix}_{|\hat{\boldsymbol{\theta}}_n} + \left[\frac{\partial^2 \ln L}{\partial\theta_l\partial\theta_m}\right]_{s\times s} (\boldsymbol{\theta}^*_{n,\mathrm{Newt}} - \hat{\boldsymbol{\theta}}_n)$$

where the second partials are evaluated at points on the line segment between $\boldsymbol{\theta}^*_{n,\mathrm{Newt}}$ and $\hat{\boldsymbol{\theta}}_n$. Thus the expression on the left hand side of (130) is equal to

$$\sqrt{n}(\boldsymbol{\theta}^*_{n,\mathrm{Newt}} - \hat{\boldsymbol{\theta}}_n)^T \left[\frac{\partial^2 \ln L}{\partial\theta_l\partial\theta_m}/n\right] I(\boldsymbol{\theta}^*_{n,\mathrm{Newt}})^{-1} \left[\frac{\partial^2 \ln L}{\partial\theta_l\partial\theta_m}/n\right] \sqrt{n}(\boldsymbol{\theta}^*_{n,\mathrm{Newt}} - \hat{\boldsymbol{\theta}}_n)$$

so by results 133 and 134, we will be done if we can establish that

$$\left[\frac{\partial^2 \ln L}{\partial\theta_l\partial\theta_m}/n\right] I(\boldsymbol{\theta}^*_{n,\mathrm{Newt}})^{-1} \left[\frac{\partial^2 \ln L}{\partial\theta_l\partial\theta_m}/n\right] - I(\boldsymbol{\theta}_0) = [o_p(1)]_{s\times s} \tag{135}$$

where the second partials are evaluated at points on the line segment between $\boldsymbol{\theta}^*_{n,\mathrm{Newt}}$ and $\hat{\boldsymbol{\theta}}_n$. Equality 135 follows from the assumed continuity of $I(\boldsymbol{\theta})$, result 82, result 132, assumptions **(A2)**, **(B)** and **(D)**, and the law of large numbers. ■

19 Appendix I — Two Useful Matrix Calculations

19.1 The $c^4 + c^2/2$ result

The necessary calculation is just an instance of a Schur complement result (see, for example, Searle (1982), or page 33 of Rao (1973)). The result is

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^T & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}^T & \mathbf{E}^{-1} \end{pmatrix}$$

where $\mathbf{E} = \mathbf{D} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$.

For our Theorems 3 and 5 we have

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{pmatrix}$$

where

$$\mathbf{A} = \begin{pmatrix} \lambda_1(2+1/c^2)/\sigma_1^2 & 0 & \dots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \dots & 0 & \lambda_k(2+1/c^2)/\sigma_k^2 \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} -\lambda_1/(c^3\sigma_1) \\ \vdots \\ -\lambda_k/(c^3\sigma_k) \end{pmatrix}$$

and

$$\mathbf{D} = 1/c^4$$

To obtain an asymptotic confidence interval on $\hat{c}$ we need $\mathbf{E}^{-1}$ where

$$\mathbf{E} = \mathbf{D} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B} = 1/c^4 - \sum_{j=1}^{k}(\lambda_j^2/(c^6\sigma_j^2))\sigma_j^2/(\lambda_j(2+1/c^2)) = 1/(c^4 + c^2/2)$$

(Recall that $\sum_{j=1}^{k}\lambda_j = 1$.)

19.2 The Newton step in Theorems 4, 5, and 6

To take this step we must calculate

$$\boldsymbol{\Delta} = -\left[\frac{\partial^2 \ln L}{\partial\theta_l\partial\theta_m}\right]^{-1}|_{\boldsymbol{\theta}_{n,c}} \begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_{k+1} \end{pmatrix}|_{\boldsymbol{\theta}_{n,c}}$$

We do this by obtaining an analytic solution to the equivalent set of equations

$$\left[\frac{\partial^2 \ln L}{\partial\theta_l\partial\theta_m}\right]|_{\boldsymbol{\theta}_{n,c}}\boldsymbol{\Delta} = -\begin{pmatrix} \partial \ln L/\partial\theta_1 \\ \vdots \\ \partial \ln L/\partial\theta_{k+1} \end{pmatrix}|_{\boldsymbol{\theta}_{n,c}}$$

The equations have the special form

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{pmatrix}\boldsymbol{\Delta} = \mathbf{L}$$

where

$$\begin{aligned}
\mathbf{A} &= \begin{pmatrix} a_1 & 0 & \dots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \dots & 0 & a_k \end{pmatrix} \\
\mathbf{B} &= \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix} \\
\mathbf{D} &= d \\
\boldsymbol{\Delta} &= \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_{k+1} \end{pmatrix} \\
\mathbf{L} &= \begin{pmatrix} l_1 \\ \vdots \\ l_{k+1} \end{pmatrix}
\end{aligned}$$

These equations can be rewritten as

$$\begin{aligned} l_1 &= a_1\delta_1 + b_1\delta_{k+1} \\ &\vdots \\ l_k &= a_k\delta_k + b_k\delta_{k+1} \\ l_{k+1} &= b_1\delta_1 + \ldots + b_k\delta_k + d\delta_{k+1} \end{aligned} \tag{136}$$

Now if, for $j = 1, \ldots, k$, we multiply the jth equation by b_j/a_j (which we can do if $\mathbf{A}$ is of full rank) and then subtract the sum of the resulting k equations from the original $k+1$th equation we obtain

$$(l_{k+1} - (b_1 l_1/a_1 + \ldots + b_k l_k/a_k))/(d - (b_1^2/a_1 + \ldots + b_k^2/a_k)) = \delta_{k+1}$$

(The denominator must be non-zero if

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{pmatrix}$$

is of full rank. If the denominator were zero, the last row of the matrix would be a linear combination of the first k rows.) Given δ_{k+1} we can then solve for the remaining δ's via

$$\delta_j = (l_j - b_j\delta_{k+1})/a_j$$

Table 1: Measurement order and xylan content

Order	Xylan content (%)	Order	Xylan content (%)	Order	Xylan content (%)
1	6.34	46	6.22	91	6.42
2	6.35	47	6.37	92	6.43
3	6.37	48	6.24	93	6.28
4	6.55	49	6.19	94	6.38
5	6.28	50	6.39	95	6.32
6	6.30	51	6.38	96	6.36
7	6.39	52	6.39	97	6.25
8	6.29	53	6.37	98	6.35
9	6.26	54	6.40	99	6.34
10	6.11	55	6.37	100	6.41
11	6.25	56	6.43	101	6.24
12	6.22	57	6.23	102	6.33
13	6.25	58	6.45		
14	6.36	59	6.36		
15	6.30	60	6.28		
16	6.33	61	6.33		
17	6.41	62	6.35		
18	6.48	63	6.25		
19	6.38	64	6.30		
20	6.31	65	6.31		
21	6.28	66	6.35		
22	6.31	67	6.35		
23	6.37	68	6.28		
24	6.15	69	6.34		
25	6.45	70	6.48		
26	6.25	71	6.59		
27	6.18	72	6.22		
28	6.36	73	6.37		
29	6.26	74	6.26		
30	6.26	75	6.20		
31	6.30	76	6.28		
32	6.27	77	6.34		
33	6.19	78	6.37		
34	6.31	79	6.25		
35	6.28	80	6.33		
36	6.23	81	6.34		
37	6.28	82	6.34		
38	6.34	83	6.34		
39	6.31	84	6.33		
40	6.22	85	6.43		
41	6.22	86	6.19		
42	6.27	87	6.40		
43	6.14	88	6.28		
44	6.13	89	6.34		
45	6.17	90	6.36		

Table 2: Sample means and standard deviations of the xylan data

Group	"Outliers" in		"Outliers" out	
	Mean	SD	Mean	SD
1	6.3265	0.09477	6.3261	0.06590
2	6.2800	0.06996	6.2711	0.05896
3	6.3000	0.10167	6.3000	0.10167
4	6.3275	0.08813	6.3044	0.05193
5	6.3445	0.06134	6.3526	0.05075

Table 3: Confidence interval coverage of the ratio[6] of two coefficients of variation (based on Theorem 2), $n = n_1 = n_2$

COV	n	Nominal coverage	Asymptotic procedure: Estimate[7] of coverage	Asymptotic procedure: 95% CI[7] on coverage	Simulation procedure: Estimate[7] of coverage	Simulation procedure: 95% CI[7] on coverage
0.01	5	0.90	0.8330	[0.8256,0.8402]	0.8987	[0.8927,0.9045]
		0.95	0.9006	[0.8947,0.9064]	0.9496	[0.9452,0.9538]
		0.99	0.9711	[0.9677,0.9743]	0.9886	[0.9864,0.9906]
	10	0.90	0.8709	[0.8643,0.8774]	0.9007	[0.8948,0.9065]
		0.95	0.9324	[0.9274,0.9372]	0.9529	[0.9487,0.9570]
		0.99	0.9836	[0.9810,0.9860]	0.9915	[0.9896,0.9932]
	20	0.90	0.8844	[0.8781,0.8906]	0.8995	[0.8935,0.9053]
		0.95	0.9413	[0.9366,0.9458]	0.9500	[0.9456,0.9542]
		0.99	0.9864	[0.9840,0.9886]	0.9904	[0.9884,0.9922]
	30	0.90	0.8945	[0.8884,0.9004]	0.9022	[0.8963,0.9079]
		0.95	0.9459	[0.9414,0.9502]	0.9512	[0.9469,0.9553]
		0.99	0.9881	[0.9859,0.9901]	0.9903	[0.9883,0.9921]
0.05	5	0.90	0.8372	[0.8299,0.8444]	0.8985	[0.8925,0.9043]
		0.95	0.9007	[0.8948,0.9065]	0.9490	[0.9446,0.9532]
		0.99	0.9723	[0.9690,0.9754]	0.9885	[0.9863,0.9905]
	10	0.90	0.8732	[0.8666,0.8796]	0.9033	[0.8974,0.9090]
		0.95	0.9314	[0.9264,0.9363]	0.9518	[0.9475,0.9559]
		0.99	0.9830	[0.9804,0.9854]	0.9908	[0.9888,0.9926]
	20	0.90	0.8863	[0.8800,0.8924]	0.8990	[0.8930,0.9048]
		0.95	0.9403	[0.9356,0.9449]	0.9484	[0.9440,0.9526]
		0.99	0.9853	[0.9828,0.9876]	0.9892	[0.9871,0.9911]
	30	0.90	0.8880	[0.8817,0.8941]	0.8956	[0.8895,0.9015]
		0.95	0.9398	[0.9351,0.9444]	0.9455	[0.9410,0.9499]
		0.99	0.9847	[0.9822,0.9870]	0.9880	[0.9858,0.9900]

[6] Ratio = 1 in these trials

[7] Based on 10,000 trials (and 10,000 trials within each of these to determine the simulation-based critical values used to construct small-sample confidence intervals)

Table 3 continued: Confidence interval coverage of the ratio[6] of two coefficients of variation (based on Theorem 2), $n = n_1 = n_2$

COV	n	Nominal coverage	Asymptotic procedure: Estimate[7] of coverage	Asymptotic procedure: 95% CI[7] on coverage	Simulation procedure: Estimate[7] of coverage	Simulation procedure: 95% CI[7] on coverage
0.15	5	0.90	0.8288	[0.8214,0.8361]	0.8990	[0.8930,0.9048]
		0.95	0.9009	[0.8950,0.9067]	0.9504	[0.9461,0.9546]
		0.99	0.9718	[0.9685,0.9750]	0.9913	[0.9894,0.9930]
	10	0.90	0.8715	[0.8649,0.8780]	0.9009	[0.8950,0.9067]
		0.95	0.9294	[0.9243,0.9343]	0.9490	[0.9446,0.9532]
		0.99	0.9825	[0.9798,0.9850]	0.9911	[0.9892,0.9928]
	20	0.90	0.8869	[0.8806,0.8930]	0.9001	[0.8941,0.9059]
		0.95	0.9427	[0.9381,0.9472]	0.9507	[0.9464,0.9549]
		0.99	0.9868	[0.9845,0.9889]	0.9902	[0.9882,0.9920]
	30	0.90	0.8891	[0.8829,0.8952]	0.8965	[0.8905,0.9024]
		0.95	0.9466	[0.9421,0.9509]	0.9514	[0.9471,0.9555]
		0.99	0.9891	[0.9870,0.9910]	0.9912	[0.9893,0.9929]
0.25	5	0.90	0.8349	[0.8276,0.8421]	0.8956	[0.8895,0.9015]
		0.95	0.8974	[0.8914,0.9033]	0.9496	[0.9452,0.9538]
		0.99	0.9700	[0.9666,0.9733]	0.9900	[0.9880,0.9919]
	10	0.90	0.8670	[0.8603,0.8736]	0.8966	[0.8906,0.9025]
		0.95	0.9257	[0.9205,0.9308]	0.9491	[0.9447,0.9533]
		0.99	0.9828	[0.9802,0.9853]	0.9900	[0.9880,0.9919]
	20	0.90	0.8896	[0.8834,0.8957]	0.9039	[0.8980,0.9096]
		0.95	0.9439	[0.9393,0.9483]	0.9538	[0.9496,0.9578]
		0.99	0.9855	[0.9831,0.9877]	0.9886	[0.9864,0.9906]
	30	0.90	0.8913	[0.8851,0.8973]	0.9007	[0.8948,0.9065]
		0.95	0.9470	[0.9425,0.9513]	0.9526	[0.9483,0.9567]
		0.99	0.9877	[0.9854,0.9898]	0.9896	[0.9875,0.9915]
0.40	5	0.90	0.8291	[0.8217,0.8364]	0.8977	[0.8917,0.9036]
		0.95	0.8985	[0.8925,0.9043]	0.9499	[0.9455,0.9541]
		0.99	0.9714	[0.9680,0.9746]	0.9904	[0.9884,0.9922]
	10	0.90	0.8692	[0.8625,0.8757]	0.8956	[0.8895,0.9015]
		0.95	0.9273	[0.9221,0.9323]	0.9486	[0.9442,0.9528]
		0.99	0.9821	[0.9794,0.9846]	0.9899	[0.9878,0.9918]
	20	0.90	0.8799	[0.8735,0.8862]	0.8949	[0.8888,0.9008]
		0.95	0.9377	[0.9329,0.9424]	0.9477	[0.9433,0.9520]
		0.99	0.9846	[0.9821,0.9869]	0.9892	[0.9871,0.9911]
	30	0.90	0.8967	[0.8907,0.9026]	0.9043	[0.8985,0.9100]
		0.95	0.9450	[0.9404,0.9494]	0.9505	[0.9462,0.9547]
		0.99	0.9890	[0.9869,0.9910]	0.9916	[0.9897,0.9933]

[6] Ratio = 1 in these trials

[7] Based on 10,000 trials (and 10,000 trials within each of these to determine the simulation-based critical values used to construct small-sample confidence intervals)

Table 4: Size of a test that k coefficients of variation are equal (based on Theorem 4), $k = 2$, $n = n_1 = n_2$

COV	n	Nominal size	Asymptotic procedure: Estimate[7] of size	Asymptotic procedure: 95% CI[7] on size	Simulation procedure: Estimate[7] of size	Simulation procedure: 95% CI[7] on size
0.01	5	0.10	0.1639	[0.1567,0.1712]	0.1013	[0.0955,0.1073]
		0.05	0.0991	[0.0933,0.1050]	0.0525	[0.0482,0.0570]
		0.01	0.0295	[0.0263,0.0329]	0.0092	[0.0074,0.0112]
	10	0.10	0.1329	[0.1263,0.1396]	0.1044	[0.0985,0.1105]
		0.05	0.0728	[0.0678,0.0780]	0.0520	[0.0477,0.0564]
		0.01	0.0174	[0.0149,0.0201]	0.0098	[0.0080,0.0118]
	20	0.10	0.1159	[0.1097,0.1222]	0.1039	[0.0980,0.1100]
		0.05	0.0594	[0.0549,0.0641]	0.0509	[0.0467,0.0553]
		0.01	0.0139	[0.0117,0.0163]	0.0095	[0.0077,0.0115]
	30	0.10	0.1116	[0.1055,0.1178]	0.1040	[0.0981,0.1101]
		0.05	0.0599	[0.0553,0.0646]	0.0532	[0.0489,0.0577]
		0.01	0.0130	[0.0109,0.0153]	0.0113	[0.0093,0.0135]
0.05	5	0.10	0.1639	[0.1567,0.1712]	0.0981	[0.0923,0.1040]
		0.05	0.0961	[0.0904,0.1020]	0.0477	[0.0436,0.0520]
		0.01	0.0291	[0.0259,0.0325]	0.0112	[0.0092,0.0134]
	10	0.10	0.1334	[0.1268,0.1401]	0.1054	[0.0995,0.1115]
		0.05	0.0742	[0.0691,0.0794]	0.0520	[0.0477,0.0564]
		0.01	0.0184	[0.0159,0.0211]	0.0113	[0.0093,0.0135]
	20	0.10	0.1163	[0.1101,0.1227]	0.1009	[0.0951,0.1069]
		0.05	0.0589	[0.0544,0.0636]	0.0503	[0.0461,0.0547]
		0.01	0.0148	[0.0125,0.0173]	0.0119	[0.0099,0.0141]
	30	0.10	0.1064	[0.1004,0.1125]	0.1006	[0.0948,0.1066]
		0.05	0.0563	[0.0519,0.0609]	0.0504	[0.0462,0.0548]
		0.01	0.0122	[0.0101,0.0144]	0.0100	[0.0081,0.0120]

[7]Based on 10,000 trials (and 10,000 trials within each of these to determine the simulation-based critical values)

Table 4 continued: Size of a test that k coefficients of variation are equal (based on Theorem 4), $k = 2$, $n = n_1 = n_2$

			Asymptotic procedure		Simulation procedure	
COV	n	Nominal size	Estimate[7] of size	95% CI[7] on size	Estimate[7] of size	95% CI[7] on size
0.15	5	0.10	0.1678	[0.1605,0.1752]	0.1026	[0.0967,0.1086]
		0.05	0.1009	[0.0951,0.1069]	0.0514	[0.0472,0.0558]
		0.01	0.0290	[0.0258,0.0324]	0.0094	[0.0076,0.0114]
	10	0.10	0.1287	[0.1222,0.1353]	0.0964	[0.0907,0.1023]
		0.05	0.0675	[0.0627,0.0725]	0.0498	[0.0456,0.0541]
		0.01	0.0169	[0.0145,0.0195]	0.0109	[0.0090,0.0130]
	20	0.10	0.1128	[0.1067,0.1191]	0.0996	[0.0938,0.1055]
		0.05	0.0587	[0.0542,0.0634]	0.0497	[0.0455,0.0540]
		0.01	0.0146	[0.0123,0.0170]	0.0101	[0.0082,0.0122]
	30	0.10	0.1064	[0.1004,0.1125]	0.1001	[0.0943,0.1061]
		0.05	0.0570	[0.0525,0.0616]	0.0509	[0.0467,0.0553]
		0.01	0.0124	[0.0103,0.0147]	0.0096	[0.0078,0.0116]
0.25	5	0.10	0.1643	[0.1571,0.1716]	0.1010	[0.0952,0.1070]
		0.05	0.0994	[0.0936,0.1053]	0.0515	[0.0473,0.0559]
		0.01	0.0298	[0.0266,0.0332]	0.0103	[0.0084,0.0124]
	10	0.10	0.1275	[0.1210,0.1341]	0.0984	[0.0926,0.1043]
		0.05	0.0701	[0.0652,0.0752]	0.0497	[0.0455,0.0540]
		0.01	0.0167	[0.0143,0.0193]	0.0094	[0.0076,0.0114]
	20	0.10	0.1101	[0.1040,0.1163]	0.0966	[0.0909,0.1025]
		0.05	0.0580	[0.0535,0.0627]	0.0467	[0.0427,0.0509]
		0.01	0.0121	[0.0101,0.0143]	0.0090	[0.0072,0.0109]
	30	0.10	0.1124	[0.1063,0.1187]	0.1048	[0.0989,0.1109]
		0.05	0.0589	[0.0544,0.0636]	0.0529	[0.0486,0.0574]
		0.01	0.0120	[0.0100,0.0142]	0.0091	[0.0073,0.0111]
0.40	5	0.10	0.1662	[0.1590,0.1736]	0.0997	[0.0939,0.1056]
		0.05	0.0988	[0.0930,0.1047]	0.0511	[0.0469,0.0555]
		0.01	0.0296	[0.0264,0.0330]	0.0093	[0.0075,0.0113]
	10	0.10	0.1321	[0.1255,0.1388]	0.1041	[0.0982,0.1102]
		0.05	0.0742	[0.0691,0.0794]	0.0513	[0.0471,0.0557]
		0.01	0.0180	[0.0155,0.0207]	0.0101	[0.0082,0.0122]
	20	0.10	0.1099	[0.1038,0.1161]	0.0972	[0.0915,0.1031]
		0.05	0.0609	[0.0563,0.0657]	0.0511	[0.0469,0.0555]
		0.01	0.0142	[0.0120,0.0166]	0.0105	[0.0086,0.0126]
	30	0.10	0.1047	[0.0988,0.1108]	0.0976	[0.0919,0.1035]
		0.05	0.0546	[0.0502,0.0591]	0.0490	[0.0449,0.0533]
		0.01	0.0124	[0.0103,0.0147]	0.0103	[0.0084,0.0124]

[7]Based on 10,000 trials (and 10,000 trials within each of these to determine the simulation-based critical values)

Table 5: Confidence interval on a coefficient of variation shared by k normally distributed populations (based on Theorem 6), $k = 2$, $n = n_1 = n_2$

COV	n	Nominal coverage	Asymptotic procedure: Estimate[7] of coverage	Asymptotic procedure: 95% CI[7] on coverage	Simulation procedure: Estimate[7] of coverage	Simulation procedure: 95% CI[7] on coverage
0.01	5	0.90	0.8091	[0.8013,0.8167]	0.8981	[0.8921,0.9040]
		0.95	0.8850	[0.8787,0.8912]	0.9478	[0.9434,0.9521]
		0.99	0.9629	[0.9591,0.9665]	0.9903	[0.9883,0.9921]
	10	0.90	0.8583	[0.8514,0.8651]	0.8975	[0.8915,0.9034]
		0.95	0.9223	[0.9170,0.9275]	0.9500	[0.9456,0.9542]
		0.99	0.9784	[0.9755,0.9812]	0.9897	[0.9876,0.9916]
	20	0.90	0.8833	[0.8769,0.8895]	0.9009	[0.8950,0.9067]
		0.95	0.9374	[0.9326,0.9421]	0.9485	[0.9441,0.9527]
		0.99	0.9860	[0.9836,0.9882]	0.9897	[0.9876,0.9916]
	30	0.90	0.8872	[0.8809,0.8933]	0.9000	[0.8940,0.9058]
		0.95	0.9439	[0.9393,0.9483]	0.9515	[0.9472,0.9556]
		0.99	0.9882	[0.9860,0.9902]	0.9904	[0.9884,0.9922]
0.05	5	0.90	0.8134	[0.8057,0.8210]	0.9026	[0.8967,0.9083]
		0.95	0.8889	[0.8827,0.8950]	0.9506	[0.9463,0.9548]
		0.99	0.9653	[0.9616,0.9688]	0.9912	[0.9893,0.9929]
	10	0.90	0.8574	[0.8505,0.8642]	0.8984	[0.8924,0.9042]
		0.95	0.9197	[0.9143,0.9249]	0.9496	[0.9452,0.9538]
		0.99	0.9799	[0.9771,0.9826]	0.9893	[0.9872,0.9912]
	20	0.90	0.8789	[0.8724,0.8852]	0.8973	[0.8913,0.9032]
		0.95	0.9331	[0.9281,0.9379]	0.9477	[0.9433,0.9520]
		0.99	0.9848	[0.9823,0.9871]	0.9887	[0.9865,0.9907]
	30	0.90	0.8864	[0.8801,0.8925]	0.9008	[0.8949,0.9066]
		0.95	0.9417	[0.9370,0.9462]	0.9492	[0.9448,0.9534]
		0.99	0.9878	[0.9856,0.9899]	0.9914	[0.9895,0.9931]

[7] Based on 10,000 trials (and 10,000 trials within each of these to determine the simulation-based critical values used to construct small-sample confidence intervals)

Table 5 continued: Confidence interval on a coefficient of variation shared by k normally distributed populations (based on Theorem 6), $k = 2$, $n = n_1 = n_2$

COV	n	Nominal coverage	Asymptotic procedure: Estimate[7] of coverage	Asymptotic procedure: 95% CI[7] on coverage	Simulation procedure: Estimate[7] of coverage	Simulation procedure: 95% CI[7] on coverage
0.15	5	0.90	0.8146	[0.8069,0.8222]	0.9012	[0.8953,0.9070]
		0.95	0.8883	[0.8821,0.8944]	0.9499	[0.9455,0.9541]
		0.99	0.9650	[0.9613,0.9685]	0.9900	[0.9880,0.9919]
	10	0.90	0.8596	[0.8527,0.8663]	0.9010	[0.8951,0.9068]
		0.95	0.9238	[0.9185,0.9289]	0.9517	[0.9474,0.9558]
		0.99	0.9826	[0.9799,0.9851]	0.9909	[0.9889,0.9927]
	20	0.90	0.8853	[0.8790,0.8915]	0.9035	[0.8976,0.9092]
		0.95	0.9391	[0.9343,0.9437]	0.9516	[0.9473,0.9557]
		0.99	0.9875	[0.9852,0.9896]	0.9908	[0.9888,0.9926]
	30	0.90	0.8900	[0.8838,0.8961]	0.9031	[0.8972,0.9088]
		0.95	0.9427	[0.9381,0.9472]	0.9506	[0.9463,0.9548]
		0.99	0.9869	[0.9846,0.9890]	0.9893	[0.9872,0.9912]
0.25	5	0.90	0.8163	[0.8086,0.8238]	0.8985	[0.8925,0.9043]
		0.95	0.8883	[0.8821,0.8944]	0.9479	[0.9435,0.9522]
		0.99	0.9658	[0.9621,0.9693]	0.9904	[0.9884,0.9922]
	10	0.90	0.8603	[0.8534,0.8670]	0.8987	[0.8927,0.9045]
		0.95	0.9231	[0.9178,0.9282]	0.9512	[0.9469,0.9553]
		0.99	0.9804	[0.9776,0.9830]	0.9890	[0.9869,0.9910]
	20	0.90	0.8888	[0.8826,0.8949]	0.9070	[0.9012,0.9126]
		0.95	0.9422	[0.9375,0.9467]	0.9540	[0.9498,0.9580]
		0.99	0.9877	[0.9854,0.9898]	0.9911	[0.9892,0.9928]
	30	0.90	0.8886	[0.8824,0.8947]	0.8996	[0.8936,0.9054]
		0.95	0.9441	[0.9395,0.9485]	0.9521	[0.9478,0.9562]
		0.99	0.9871	[0.9848,0.9892]	0.9895	[0.9874,0.9914]
0.40	5	0.90	0.8350	[0.8277,0.8422]	0.9112	[0.9055,0.9167]
		0.95	0.9032	[0.8973,0.9089]	0.9562	[0.9521,0.9601]
		0.99	0.9721	[0.9688,0.9752]	0.9921	[0.9903,0.9937]
	10	0.90	0.8655	[0.8587,0.8721]	0.8975	[0.8915,0.9034]
		0.95	0.9246	[0.9193,0.9297]	0.9482	[0.9438,0.9525]
		0.99	0.9793	[0.9764,0.9820]	0.9887	[0.9865,0.9907]
	20	0.90	0.8881	[0.8818,0.8942]	0.9019	[0.8960,0.9077]
		0.95	0.9420	[0.9373,0.9465]	0.9537	[0.9495,0.9577]
		0.99	0.9882	[0.9860,0.9902]	0.9911	[0.9892,0.9928]
	30	0.90	0.8925	[0.8864,0.8985]	0.9021	[0.8962,0.9078]
		0.95	0.9446	[0.9400,0.9490]	0.9508	[0.9465,0.9550]
		0.99	0.9883	[0.9861,0.9903]	0.9902	[0.9882,0.9920]

[7] Based on 10,000 trials (and 10,000 trials within each of these to determine the simulation-based critical values used to construct small-sample confidence intervals)

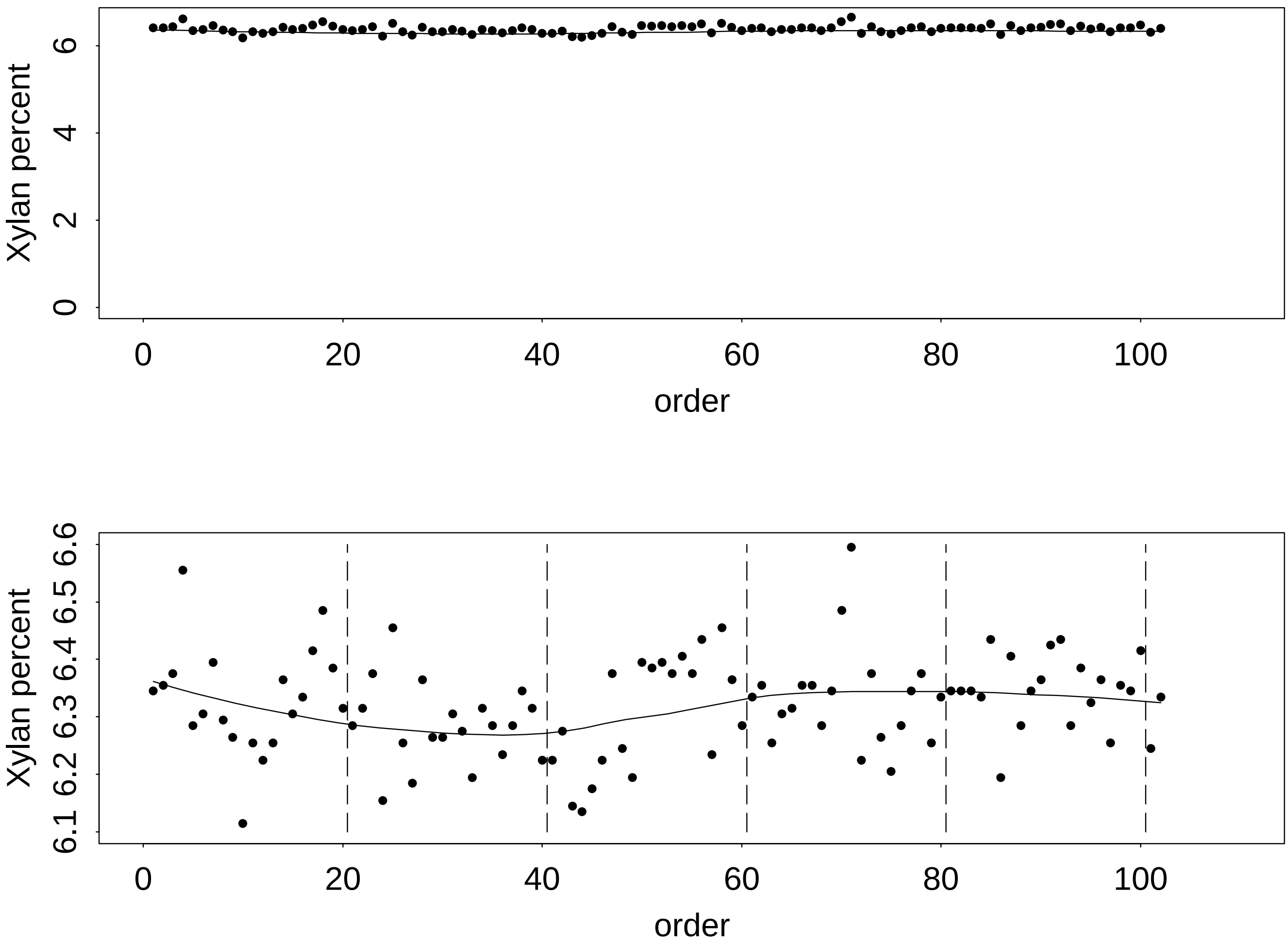


Figure 1: Plots of xylan content (%) versus time order

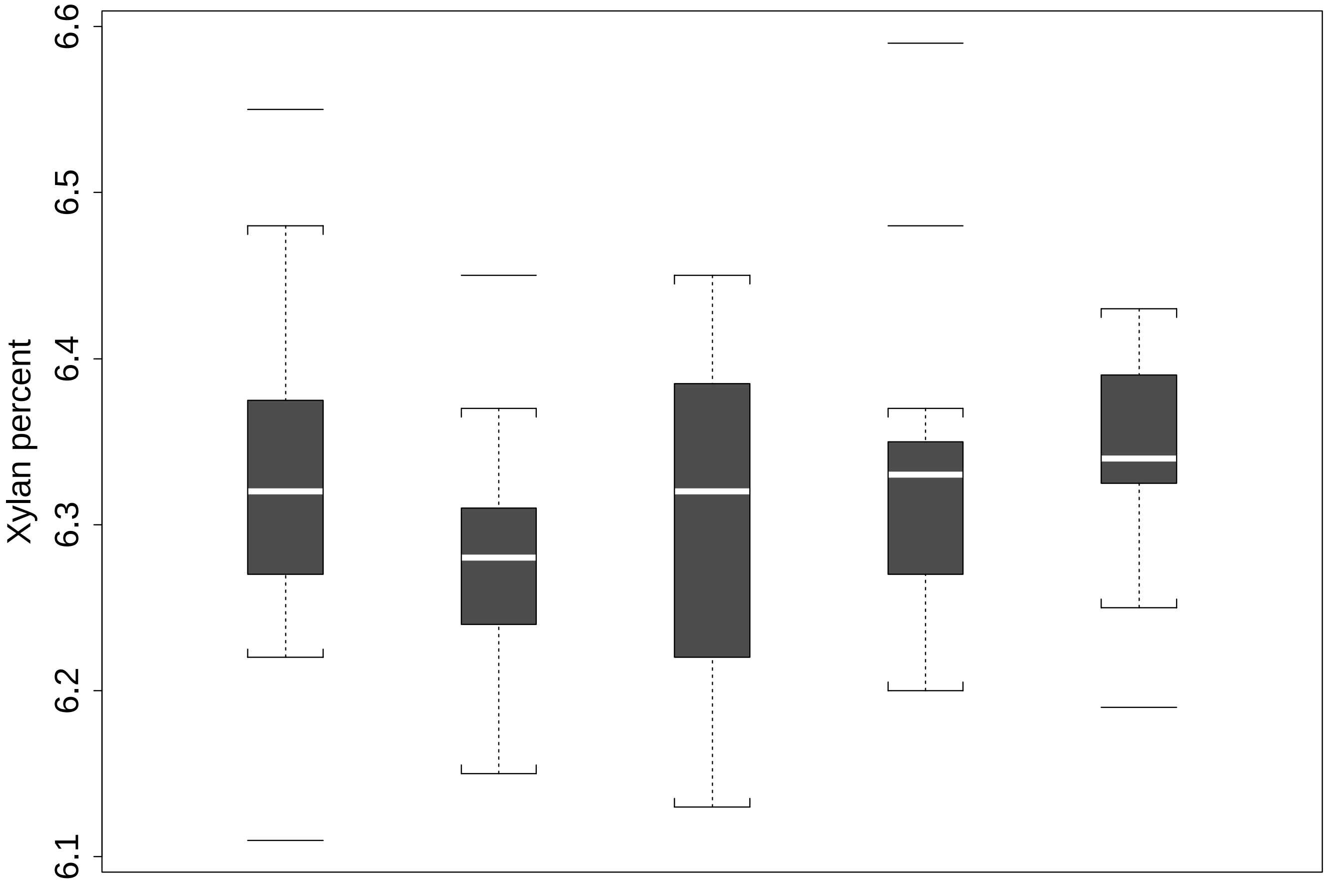


Figure 2: Boxplots of the data from the five groups

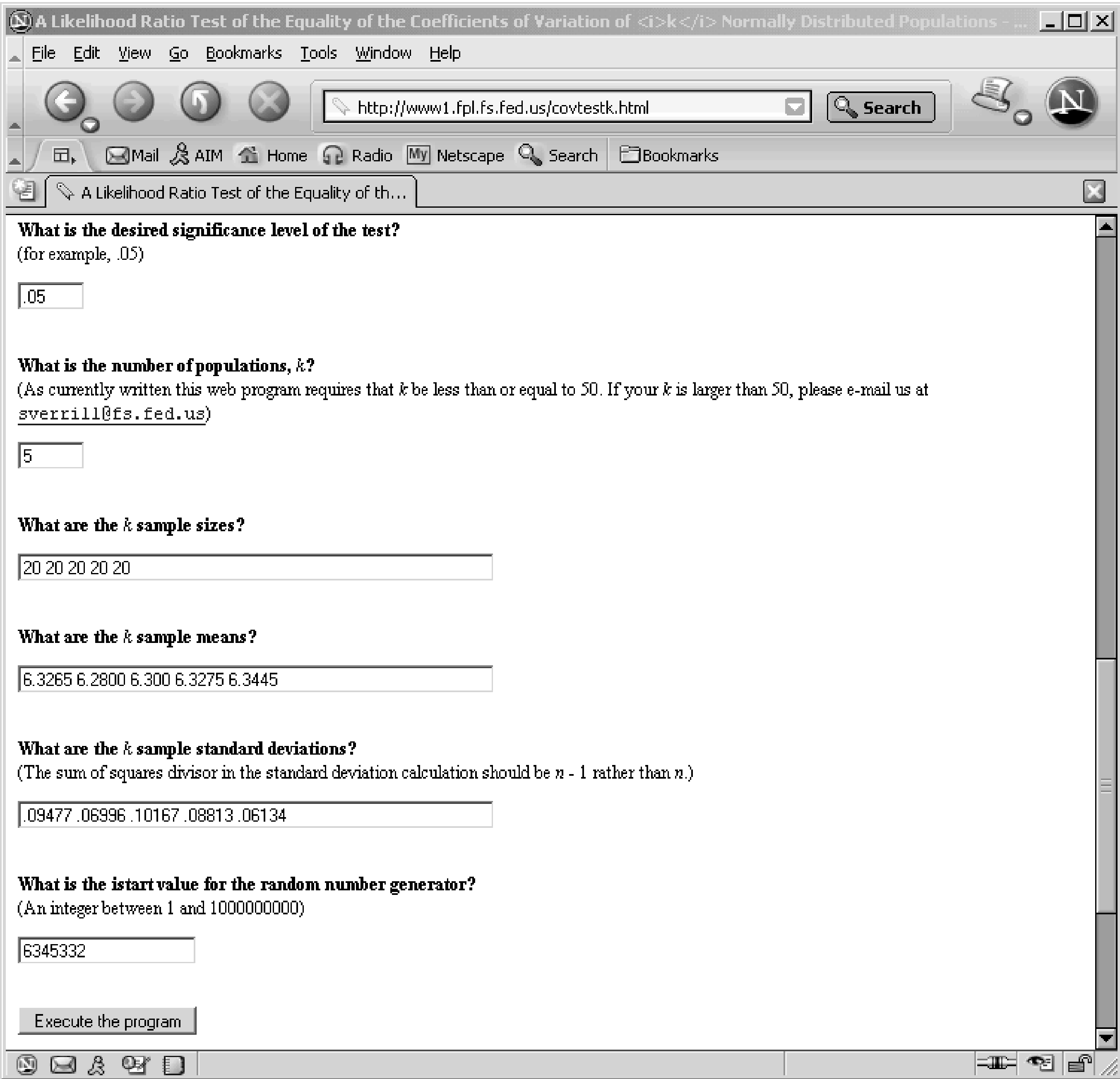


Figure 3: Web page for a test of the equality of k coefficients of variation