



UNIVERSITY
of HAWAI'I®

MĀNOA

University of Hawai`i at Mānoa
Department of Economics
Working Paper Series

Saunders Hall 542, 2424 Maile Way,
Honolulu, HI 96822
Phone: (808) 956 -8496
www.economics.hawaii.edu

Working Paper No. 12-9

A Note on the Asymptotic Variance of Sample Roots

By

Timothy J Halliday

May 2012

A Note on the Asymptotic Variance of Sample Roots

Timothy J Halliday*

University of Hawai'i at Mānoa[†]and IZA

May 25, 2012

Abstract

We derive the asymptotic distribution of the eigenvalues of a sample covariance matrix with distinct roots. Our theorem can accommodate the situation in which the population covariance matrix is estimated via its sample analogue as well as the more general case in which it is estimated via a \sqrt{N} -consistent extremum estimator. The sample roots will have a Normal distribution in

*Address: 2424 Maile Way; 533 Saunders Hall; Honolulu, HI 96822. Tele: (808) 956-8615. E-mail: halliday@hawaii.edu. WWW: www2.hawaii.edu/~halliday. I thank Bo Honoré for useful feedback.

[†]Department of Economics

a large sample with a covariance matrix that is easy to compute. We conduct Monte Carlo experiments that show that standard errors based on our derived asymptotic distribution accurately approximate standard errors in the empirical distribution.

JEL Codes: C01, C02

Key Words: Principal Components Analysis, Asymptotic Distribution, Extremum Estimation

1 Introduction

As pointed out by Lawley and Maxwell (1971), Principal Components Analysis is a useful method for decomposing the total variance of a set of psychometric measurements into a set of more fundamental constituents that is widely applicable throughout the social and behavioral sciences. As it turns out, the variances of each of these more fundamental components will be the eigenvalues or roots of the population covariance matrix. When these eigenvalues are estimated in a sample, they will be random variables and so, it becomes necessary to compute standard errors so that inferences can be made about the population roots.

There has been a substantial amount of work on this topic in statistics. The first derivation of the asymptotic distribution of the sample roots was provided by

Girshick (1939). He considered a situation in which the eigenvalues were computed from a sample covariance matrix for the case of a Normal parent population when the population roots are all distinct. This work was generalized by Anderson (1963) to allow subsets of the roots to be equal. Work by Waternaux (1976) provides an elegant generalization of Girshick. When the parent population is non-Normal, she shows that the asymptotic covariance matrix deviates from the Normal case by an amount that depends on fourth-order cummulants. When the parent distribution is Normal, these cummulants are zero and so her result reduces to Girshick's.

In this note, we prove a theorem that generalizes Waternaux (1976) to account for any \sqrt{N} -consistent estimator of a covariance matrix. This includes situations such as in Girshick (1939) and Waternaux (1976) in which the covariance matrix is estimated via its sample analogue but it can also accommodate more general frameworks in which a covariance matrix of latent variables is estimated via an extremum estimator such as Maximum Likelihood Estimation or the Generalized Method of Moments. We show that the asymptotic distribution of the sample roots is Normal with a covariance matrix that is easy for practitioners to compute. We conduct a Monte Carlo simulation that shows that standard errors that are based on the asymptotic distribution accurately approximate standard errors computed from the empirical distribution.

2 The Result

We consider an $M \times M$ covariance matrix denoted by Σ . This matrix can be decomposed as

$$\Sigma = \Gamma \Delta \Gamma'$$

where the matrix Γ is orthogonal (*i.e.* $\Gamma' \Gamma = \mathbf{I}_M$). The matrix, Δ , is a diagonal matrix with the eigenvalues of Σ along its diagonal. We denote these eigenvalues by $\delta_1, \dots, \delta_M$. The matrix, Γ , contains the corresponding eigenvectors of Σ . We normalize the decomposition so that the eigenvalues are in descending order.

Suppose that we have a random sample of size N that we use to compute $\widehat{\Sigma}$ with the property that

$$\sqrt{N} \left(\text{vec} \left(\widehat{\Sigma} \right) - \text{vec} \left(\Sigma \right) \right) = \mathbf{A} \frac{1}{\sqrt{N}} \sum_{i=1}^N s_i \left(\Sigma \right) + o_p \left(1 \right)$$

where \mathbf{A} is an $M^2 \times M^2$ non-stochastic matrix, $s_i \left(\Sigma \right)$ is an $M^2 \times 1$ function of observation i , and

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N s_i \left(\Sigma \right) \widetilde{\rightarrow} N \left(\mathbf{0}_{M^2}, E \left(s_i \left(\Sigma \right) s_i \left(\Sigma \right)' \right) \right).$$

This is a fairly general environment that applies to any \sqrt{N} -consistent estimator of Σ .

For example, it can accompany a Maximum Likelihood Estimator of the covariance matrix of a set of M latent variables in a binary choice model with multiple outcomes. In this scenario, $s_i(\boldsymbol{\Sigma})$ would be the $M^2 \times 1$ vector of scores and $\mathbf{A} = E[s'_i(\boldsymbol{\Sigma})]$ would be the corresponding Hessian Matrix. It could also accompany the situation in Waternaux (1976) in which $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix of a set of M observed covariates denoted by \mathbf{x}_i . For the case of $M = 2$, we would then have that $\mathbf{A} = \mathbf{I}_4$ and

$$s_i(\boldsymbol{\Sigma}) = \begin{bmatrix} (x_{1i} - \mu_1)^2 - \sigma_1, \\ (x_{1i} - \mu_1)(x_{2i} - \mu_2) - \sigma_{12}, \\ (x_{1i} - \mu_1)(x_{2i} - \mu_2) - \sigma_{12}, \\ (x_{2i} - \mu_2)^2 - \sigma_2 \end{bmatrix}$$

where μ_1 and μ_2 are the population means of x_{1i} and x_{2i} . If we let $\widehat{\boldsymbol{\Delta}}$ denote the matrix of eigenvalues of $\widehat{\boldsymbol{\Sigma}}$, then the following theorem describes the asymptotic behavior of sample roots in $\widehat{\boldsymbol{\Delta}}$.

Proposition 1 *If we partition $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_M]$ where $\boldsymbol{\gamma}_m$ is the eigenvector corresponding to δ_m and if*

$$\sqrt{N} \left(\text{vec} \left(\widehat{\boldsymbol{\Sigma}} \right) - \text{vec} \left(\boldsymbol{\Sigma} \right) \right) \rightsquigarrow N \left(\mathbf{0}_{M^2}, \mathbf{V}_{\boldsymbol{\Sigma}} \right)$$

where $\delta_1 > \delta_2 > \dots > \delta_M$, then we will obtain that

$$\sqrt{N} \left(\text{diag} \left(\widehat{\Delta} \right) - \text{diag} \left(\Delta \right) \right) \rightsquigarrow N \left(\mathbf{0}_M, \mathbf{V}_\Delta \right)$$

where

$$\mathbf{V}_\Delta \equiv \begin{pmatrix} (\gamma'_1 \otimes \gamma'_1) \mathbf{V}_\Sigma (\gamma_1 \otimes \gamma_1) & \cdots & (\gamma'_1 \otimes \gamma'_1) \mathbf{V}_\Sigma (\gamma_M \otimes \gamma_M) \\ \vdots & \ddots & \vdots \\ (\gamma'_M \otimes \gamma'_M) \mathbf{V}_\Sigma (\gamma_1 \otimes \gamma_1) & \cdots & (\gamma'_M \otimes \gamma'_M) \mathbf{V}_\Sigma (\gamma_M \otimes \gamma_M) \end{pmatrix}.$$

3 Monte Carlo Evidence

In this section, we provide Monte Carlo evidence showing that standard errors based on the asymptotic variance in the above theorem closely approximate the standard errors from the empirical distribution. We consider the case in which $\widehat{\Sigma}$ is the sample analogue of Σ . Following Waternaux (1976), we consider a Normal, short-tailed and long-tailed distribution. Throughout, we suppose that $M = 2$. Note that when the parent population is Normal, Girshick (1939) showed that $\mathbf{V}_\Delta = 2\Delta^2$. Throughout this exercise, we will compare asymptotic standard errors based on our general formula to Girshick's to show how failure to account for non-normal parent

populations might matter. For the short-tailed population, we assume

$$\begin{aligned}x_1 &= u + v \\x_2 &= \frac{3}{2}(u + w)\end{aligned}$$

where u, v , and w are all independent and distributed uniformly on $[-1, 1]$. The fat-tailed population is given by

$$\begin{aligned}x_1 &= a + bc \\x_2 &= a + bd\end{aligned}$$

where a, b, c , and d are all independent, mean zero Normal random variables with $\sigma_a = \frac{1}{2}$, $\sigma_b = \frac{1}{6}$, $\sigma_c = 1$, and $\sigma_d = 6$. Note that for the short and fat tailed distributions, we have that

$$\Sigma = \begin{bmatrix} \frac{2}{3} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{bmatrix}$$

and so we use this same covariance matrix for the Normal case. This ensures that the population roots are the same for all three cases.

We report the results of this exercise in Tables 1, 2, and 3 for each of the three cases. In each table, we report $V_{\widehat{F}}(\widehat{\delta}_i)$, $\widehat{\mathbf{V}}_{\Delta}^{ii}/N$, and $2\delta_i^2/N$ for $i \in \{1, 2\}$. The

statistic, $V_{\widehat{F}}(\widehat{\delta}_i)$, is the variance of the sample root in the empirical distribution. We use the “hat-notation” for $V_{\widehat{F}}(\widehat{\delta}_i)$ to denote that we estimate the empirical variance by sampling from the parent population a total of $S = 2000$ times and computing the variance of these S sample roots. The quantity, $\widehat{\mathbf{V}}_{\Delta}^{ii}/N$, is based on the asymptotic variance of the sample root that we derived above. As with the empirical variance, we computed the sample analogue of $\mathbf{V}_{\Delta}^{ii}/N$ a total of $S = 2000$ times and then we took the average of this number across all of the simulations. In both the case of $V_{\widehat{F}}(\widehat{\delta}_i)$ and $\widehat{\mathbf{V}}_{\Delta}^{ii}/N$, if S is large then these statistics should be arbitrarily close to $V_{\widehat{F}}(\widehat{\delta}_i)$ and $\mathbf{V}_{\Delta}^{ii}/N$. Finally, $2\delta_i^2/N$ is the asymptotic variance of the sample roots when the parent population is Normal. We also report $C_{\widehat{F}}(\widehat{\delta}_1, \widehat{\delta}_2)$, which is the empirical covariance of the sample roots, and $\widehat{\mathbf{V}}_{\Delta}^{12}/N$ which is computed analogous to $\widehat{\mathbf{V}}_{\Delta}^{ii}/N$. Note that in the case of a Normal parent population, the covariance is zero, but as shown in Waternaux (1976), the covariance may be non-zero when the population is not Normal.

For all three cases considered, we see that the empirical variance is very close to asymptotic variance computed using our formula. For the Normal case, all three variances are very close for both roots. We also see that both of the covariances are very close to zero as they should be. For the other two cases, we see that $V_{\widehat{F}}(\widehat{\delta}_i)$ and $\widehat{\mathbf{V}}_{\Delta}^{ii}/N$ are very close, particularly, for large samples, but both quantities deviate

substantially from Girshick’s formula. In particular, assuming Normality overstates the variance in the short-tailed case, but understates it in the long-tailed case. In both of these cases, the two covariances are non-zero. Accordingly, consistent with Waternaux (1976), we show that failure to account for non-Normal parent populations will result in incorrect inferences.

4 Conclusions

We derived the asymptotic distribution of the sample roots of a covariance matrix. Our result applies to any \sqrt{N} -consistent estimate of a covariance matrix with distinct roots. This work generalizes previous work by Girshick (1939) and Waternaux (1976) that considered the case in which the covariance matrix was estimated via its sample analogue in Normal and non-Normal populations, respectively. Monte Carlo evidence indicates that standard errors based on the derived asymptotic distribution accurately approximate standard errors computed from the empirical distribution.

5 Appendix

Proof. We define

$$V \equiv \sqrt{N} \left(\mathbf{\Gamma}' \widehat{\mathbf{\Sigma}} \mathbf{\Gamma} - \mathbf{\Delta} \right).$$

Note that each element of this object will be asymptotically Normal because of the Central Limit Theorem. Employing results from Konishi (1979), we can write

$$\widehat{\delta}_i = \delta_i + \frac{1}{\sqrt{N}}v_{ii} + \frac{1}{N}\sum_{j \neq i}^M (\delta_i - \delta_j)^{-1} v_{ij}^2 + O_p(N^{-3/2}).$$

and so, following Konishi and Rao (1992), we can write

$$\sqrt{N}(\widehat{\delta}_i - \delta_i) = v_{ii} + o_p(1).$$

Thus, we obtain that

$$\sqrt{N} \begin{pmatrix} \widehat{\delta}_1 - \delta_1 \\ \vdots \\ \widehat{\delta}_M - \delta_M \end{pmatrix} = \text{diag}(V) + o_p(1) = \sqrt{N} \begin{pmatrix} \boldsymbol{\gamma}'_1 \widehat{\boldsymbol{\Sigma}} \boldsymbol{\gamma}_1 - \delta_1 \\ \vdots \\ \boldsymbol{\gamma}'_M \widehat{\boldsymbol{\Sigma}} \boldsymbol{\gamma}_M - \delta_M \end{pmatrix} + o_p(1).$$

Next we note that

$$\boldsymbol{\gamma}'_m \mathbf{B} \boldsymbol{\gamma}_m = (\boldsymbol{\gamma}'_m \otimes \boldsymbol{\gamma}'_m) \text{vec}(\mathbf{B})$$

for an $M \times M$ matrix \mathbf{B} . Hence, we obtain that

$$\text{diag}(\mathbf{\Delta}) = \begin{pmatrix} \boldsymbol{\gamma}'_1 \otimes \boldsymbol{\gamma}'_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{\gamma}'_M \otimes \boldsymbol{\gamma}'_M \end{pmatrix} \begin{pmatrix} \text{vec}(\boldsymbol{\Sigma}) \\ \vdots \\ \text{vec}(\boldsymbol{\Sigma}) \end{pmatrix}$$

where the second matrix on the right-hand side stacks $\text{vec}(\boldsymbol{\Sigma})$ a total of M times.

So, noting that

$$\sqrt{N} \begin{pmatrix} \text{vec}(\widehat{\boldsymbol{\Sigma}}) - \text{vec}(\boldsymbol{\Sigma}) \\ \vdots \\ \text{vec}(\widehat{\boldsymbol{\Sigma}}) - \text{vec}(\boldsymbol{\Sigma}) \end{pmatrix} \rightsquigarrow N \left(\mathbf{0}_{M^3}, \begin{pmatrix} \mathbf{V}_\Sigma & \cdots & \mathbf{V}_\Sigma \\ \vdots & \ddots & \vdots \\ \mathbf{V}_\Sigma & \cdots & \mathbf{V}_\Sigma \end{pmatrix} \right),$$

we obtain that

$$\sqrt{N} \left(\text{diag}(\widehat{\boldsymbol{\Delta}}) - \text{diag}(\boldsymbol{\Delta}) \right) \rightsquigarrow N(\mathbf{0}_M, V_\Delta).$$

This proves the result. ■

References

ANDERSON, T. (1963): “Asymptotic Theory for Principal Components,” *Annals of Mathematical Statistics*, 34(1), 122–148.

- GIRSHICK, M. (1939): “On the Sampling Theory of Roots of Determinantal Equations,” *Annals of Mathematical Statistics*, 10(3), 203–224.
- KONISHI, S. (1979): “Asymptotic Expansions for the Distributions of Statistics Based on the Sample Correlation Matrix in Principal Components Analysis,” *Hiroshima Mathematical Journal*, 6(3), 647–700.
- KONISHI, S., AND R. RAO (1992): “Principal Components Analysis for Multivariate Familial Data,” *Biometrika*, 79(3), 631–641.
- LAWLEY, D., AND A. MAXWELL (1971): *Factor Analysis as a Statistical Method*. Butterworth and Co. Ltd., London.
- WATERNAUX, C. (1976): “Asymptotic Distribution of the Sample Roots for a Non-normal Population,” *Biometrika*, 63(3), 639–645.

Table 1: Monte Carlo Simulations: Normal Case

N	$V_{\widehat{F}}(\widehat{\delta}_1)$	$\widehat{\mathbf{V}}_{\Delta}^{11}/N$	$2\delta_1^2/N$	$V_{\widehat{F}}(\widehat{\delta}_2)$	$\widehat{\mathbf{V}}_{\Delta}^{22}/N$	$2\delta_2^2/N$	$C_{\widehat{F}}(\widehat{\delta}_1, \widehat{\delta}_2)$	$\widehat{\mathbf{V}}_{\Delta}^{12}/N$
100	0.0620	0.0594	0.0601	0.0038	0.0035	0.0037	-0.0002	0.0009
250	0.0237	0.0236	0.0241	0.0015	0.0015	0.0015	-0.0000	-0.0000
500	0.0128	0.0120	0.0120	0.0007	0.0007	0.0007	-0.0001	-0.0000
1000	0.0063	0.0060	0.0060	0.0004	0.0004	0.0004	0.0000	-0.0000
5000	0.0011	0.0012	0.0012	0.0001	0.0001	0.0001	-0.0000	0.0000

Table 2: Monte Carlo Simulations: Short-Tail Case

N	$V_{\widehat{F}}(\widehat{\delta}_1)$	$\widehat{\mathbf{V}}_{\Delta}^{11}/N$	$2\delta_1^2/N$	$V_{\widehat{F}}(\widehat{\delta}_2)$	$\widehat{\mathbf{V}}_{\Delta}^{22}/N$	$2\delta_2^2/N$	$C_{\widehat{F}}(\widehat{\delta}_1, \widehat{\delta}_2)$	$\widehat{\mathbf{V}}_{\Delta}^{12}/N$
100	0.0434	0.0415	0.0601	0.0027	0.0026	0.0037	-0.0010	-0.0015
250	0.0159	0.0167	0.0241	0.0011	0.0010	0.0015	-0.0005	-0.0006
500	0.0089	0.0084	0.0120	0.0005	0.0005	0.0007	-0.0002	-0.0003
1000	0.0042	0.0042	0.0060	0.0003	0.0003	0.0004	-0.0001	-0.0001
5000	0.0009	0.0008	0.0012	0.00005	0.00005	0.0001	-0.00004	-0.00003

Table 3: Monte Carlo Simulations: Long-Tail Case

N	$V_{\widehat{F}}(\widehat{\delta}_1)$	$\widehat{\mathbf{V}}_{\Delta}^{11}/N$	$2\delta_1^2/N$	$V_{\widehat{F}}(\widehat{\delta}_2)$	$\widehat{\mathbf{V}}_{\Delta}^{22}/N$	$2\delta_2^2/N$	$C_{\widehat{F}}(\widehat{\delta}_1, \widehat{\delta}_2)$	$\widehat{\mathbf{V}}_{\Delta}^{12}/N$
100	0.1115	0.1062	0.0601	0.0089	0.0082	0.0037	0.0094	0.0077
250	0.0429	0.0421	0.0241	0.0038	0.0035	0.0015	0.0037	0.0034
500	0.0194	0.0205	0.0120	0.0020	0.0018	0.0007	0.0017	0.0017
1000	0.0106	0.0104	0.0060	0.0010	0.0010	0.0004	0.0009	0.0009
5000	0.0020	0.0021	0.0012	0.0002	0.0002	0.0001	0.0002	0.0002