Statistics of eigenvalue dispersion indices: quantifying the magnitude of phenotypic integration

Junya Watanabe

Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EQ, United Kingdom

jw2098@cam.ac.uk

https://orcid.org/0000-0002-9810-5286

Abstract

Quantification of the magnitude of covariation plays a major role in the studies of phenotypic integration, for which statistics based on dispersion of eigenvalues of a covariance or correlation matrix—eigenvalue dispersion indices—are commonly used. However, their use has been hindered by a lack of clear understandings on their statistical meaning and sampling properties such as the magnitude of sampling bias and error. This study remedies these issues by investigating properties of these statistics with both analytic and simulation-based approaches. The relative eigenvalue variance of a covariance matrix is known in the statistical literature as a test statistic for sphericity, thus is an appropriate measure of eccentricity of variation. The same of a correlation matrix is exactly equal to the average squared correlation, thus is a clear measure of overall integration. Exact and approximate expressions for the mean and variance of these statistics are analytically derived for the null and arbitrary conditions under multivariate normality, clarifying the effects of sample size N, number of variables p, and parameters on the sampling bias and error. Accuracy of the approximate expressions are evaluated with simulations, confirming that most of them work reasonably well with a moderate sample size ($N \geq 16$–64). Importantly, sampling properties of these indices are not adversely affected by high p:N ratio, promising their utility in high-dimensional phenotypic analyses. These statistics can potentially be applied to shape variables and phylogenetically structured data, for which necessary assumptions and modifications are presented.

Keywords: covariance matrix; evolutionary constraint; morphometrics; phenotypic integration; quantitative genetics; Wishart distribution.

Introduction

Analysis of trait covariation plays a central role in investigations into evolution of quantitative traits. The well-known quantitative genetic theory of correlated traits predicts that evolutionary response in a population under selection is dictated by the additive genetic covariance matrix **G** as well as the selection gradient (Lande, 1979; Lande & Arnold, 1983). Short-term evolutionary changes of a population are expected to be concentrated in major axes of the **G** matrix (Schluter, 1996). Arguably, the structure of the **G** matrix can be approximated by that of the phenotypic covariance matrix for certain types of traits (Cheverud, 1988, 1996; Roff, 1995; Dochtermann, 2011; Sodini et al., 2018), so the latter could potentially be analyzed when accurate estimation of the **G** matrix is not feasible. These theories and conjectures spurred extensive theoretical and empirical explorations on character covariation as an evolutionary constraint (e.g., Steppan et al., 2002; Chenoweth et al., 2010; Hansen et al., 2019 and references therein). Partly fueled by these developments, the study of phenotypic integration has developed as an active field of research, where various aspects of character covariation are investigated with diverse motivations and scopes (e.g., Olson & Miller, 1958; Cheverud, 1982; Goswami, 2006; Hallgrímsson et al., 2009; Armbruster et al., 2014; Felice et al., 2018). In the latter context, many different levels of organismal variation can be subjects of research, such as static, ontogenetic, and evolutionary levels (Klingenberg, 2014). For example, relationships between within-population integration and evolutionary rate and/or trajectories have attained much attention as potential links between micro- and macroevolutionary phenomena (e.g., Klingenberg et al., 2012; Renaud & Auffray, 2013; Goswami et al., 2015; Haber, 2015, 2016).

An obvious target of investigation in these contexts is quantitative analysis of magnitude of constraint or integration entailed in covariance structures. In particular, this paper concerns methodology for quantifying the overall magnitude of covariation within a set of traits. Quantification of relative (in)dependence between multiple sets of traits—the modularity–integration spectrum—is another major way of studying integration which has separate methodological frameworks (e.g., Goswami & Polly, 2010; Adams, 2016; Goswami & Finarelli, 2016; Adams & Collyer, 2019a). Demonstrating the presence of integration with a statistically justified measure can be the scope of an empirical analysis, sometimes as a part of testing combined hypotheses (e.g., Brommer, 2014; Watanabe, 2018). A univariate summary statistic for magnitude of integration can conveniently be used in comparative analyses across developmental stages, populations, or phylogeny (e.g., Marroig et al., 2009; Porto et al., 2009; Haber, 2016). A plethora of statistics have been proposed for such purposes from various standpoints (e.g., Van Valen, 1974, 2005; Cheverud et al., 1983, 1989; Wagner, 1984; Cane, 1993; Hansen & Houle, 2008; Agrawal & Stinchcombe, 2009; Kirkpatrick, 2009; Pavlicev et al., 2009; Armbruster et al., 2009, 2014; Haber, 2011; Pitchers et al., 2014). One of the most popular class of such statistics is based on the dispersion of eigenvalues of a covariance or correlation matrix. These statistics have the forms

$$V = \frac{1}{p}\sum_{i=1}^{p}\left(\lambda_i - \bar{\lambda}\right)^2,$$

$$V_{\mathrm{rel}} = \frac{\sum_{i=1}^{p}\left(\lambda_i - \bar{\lambda}\right)^2}{p(p-1)\,\bar{\lambda}^2},$$

where p is the number of variables (traits), λ_i is the ith eigenvalue of the covariance or correlation matrix under analysis, and $\bar{\lambda}$ is the average of the eigenvalues. Here, V is the most naïve form of eigenvalue dispersion, and V_{rel} is a scaled version which ranges between 0 and 1. Formal definitions are given below with distinction between population and sample quantities. Some authors use square root or a constant multiple of these forms, but such variants essentially bear identical information when calculated from the same matrix. Alternative terms for this class of statistics include the tightness (for V_{rel}; Van Valen, 1974;

later used for $\sqrt{V_{\mathrm{rel}}}$ by Van Valen, 2005), integration coefficient of variation (for $\sqrt{(p-1)V_{\mathrm{rel}}}$; Shirai & Marroig, 2010), and phenotypic integration index (for V; Torices & Muñoz-Pajares, 2015). In this paper, V and V_{rel}.are called the eigenvalue variance and relative eigenvalue variance, respectively, to take a balance between brevity and descriptiveness. These quantities are not to be confused with the sampling variance associated with eigenvalues in a sample (see below).

Since eigenvalues of a covariance or correlation matrix correspond to the variance along the corresponding eigenvectors (principal components), these statistics are supposed to represent eccentricity of variation across directions in a trait space (Fig. 1; Wagner, 1984). Cheverud et al. (1983) and Wagner (1984) were the first to propose using V of a correlation matrix for quantifying magnitude of integration. Pavlicev et al. (2009) devised V_{rel} of a correlation matrix, and explored its relationships to correlation structures in certain biologically relevant conditions. Haber (2011) pointed out similarity between these indices and Van Valen's (1974) tightness index for a covariance matrix, and proposed that these indices can be applied to either covariance or correlation matrices with slightly different interpretations. Eigenvalue dispersion indices are frequently used in empirical analyses of phenotypic integration at various levels of organismal variation, from phenotypic covariance at the population level to evolutionary covariance at the interspecific level (e.g., Ordano et al., 2008; Torices & Mendez, 2014; Haber, 2016; Haber & Dworkin, 2017; Watanabe, 2018; Arlegi et al., 2020). However, use of these indices has been criticized for a lack of clear statistical justifications; it has not been known—or not widely appreciated by biologists—exactly what they are designed to measure, beyond the intuitive allusion to eccentricity mentioned above (Hansen & Houle, 2008; Hansen et al., 2019).

Another fundamental issue over the eigenvalue dispersion indices is a virtual lack of systematic understanding of their sampling properties. In empirical analyses, eigenvalue dispersion indices are calculated from sample covariance or correlation matrices, but interests will be in making inferences for the underlying populations. For example, interest may be in detecting the presence of bias in a population, i.e., testing the null hypothesis of sphericity (no eccentricity). As detailed below, however, sample eigenvalues are always estimated with error, so that V and V_{rel} calculated from them take a positive value, even if the corresponding population values are 0. In other words, empirical eigenvalue dispersion indices are biased estimators of the corresponding population values under the null hypothesis. For statistically justified inferences, it is crucial to capture essential aspects of their sampling distributions, e.g., expectation and variance.

The presence of estimation or sampling bias in eigenvalue dispersion indices has been well known in the literature (Wagner, 1984; Cheverud et al., 1989; Grabowski & Porto, 2017; see also Marroig et al., 2012). Simulation-based approaches have been taken to sketch sampling distributions of eigenvalue dispersion indices and related statistics (Haber, 2011; Grabowski & Porto, 2017; Machado et al., 2019; Jung et al. 2020). However, these approaches hardly give any systematic insight beyond the specific conditions considered. Analytic results should preferably be sought to comprehend the sampling bias and error. In this regard, it is notable that Wagner (1984) derived the first two moments of eigenvalues of sample covariance and correlation matrices under the null conditions, proposing to use the variance of sample eigenvalues obtained from these moments as an estimate of sampling bias in these conditions. Strictly speaking, however, the variance of a sample eigenvalue is fundamentally different from the expectation of the eigenvalue variance V. These quantities are identical for correlation matrices under the null hypothesis, but this is not the case for covariance matrices where the covariances between sample eigenvalues cannot be ignored (see below). Furthermore, Wagner's (1984) results have a few restrictive conditions: variables to have the means of 0, or equivalently, to be centered at the population mean rather than the sample mean as is done in most empirical analyses (although this was probably appropriate in the strict context of his theoretical model); and the sample size N to be equal to or larger than the number of variables p, so their applicability to $p > N$ conditions has not been demonstrated.

In addition to the naïve null condition of no integration, moments under arbitrary conditions are also desired. Such would be useful in testing hypotheses about the magnitude (rather than the mere presence/absence) of integration (Harder, 2009; Fornoni et al., 2009) and comparing the magnitudes between different samples (Cheverud et al., 1989). Also, the assumption of no covariation is intrinsically inappropriate as a null hypothesis for shape variables where raw data are transformed in such a way that individual "variables" are necessarily dependent on one another (e.g., Mitteroecker et al., 2012). For this type of data, a covariance matrix with an appropriate structure needs to be specified as the null model representing the intrinsic covariation.

This paper addresses the issues over the eigenvalue dispersion indices mentioned above. It first gives a theoretical overview of these statistics to clarify their statistical justifications, particularly in connection to the sphericity test in multivariate analysis. Then exact and approximate expressions are analytically derived for the expectation and variance of V and V_{rel} of sample covariance and correlation matrices under the null and arbitrary conditions, assuming the multivariate normality of original variables. These expressions are derived without any assumption on p or N, except for the variance of V and V_{rel} of a correlation matrix under arbitrary conditions, which is based on a strict large-sample asymptotic theory. Simulations were subsequently conducted to obtain systematic insights into sampling properties and to evaluate the accuracy of the approximate expressions. Potential extensions into shape variables and phylogenetically structured data are briefly discussed.

Theory

Preliminaries

For the purpose here, the distinction between population and sample quantities is essential. Corresponding Greek and Latin letters are used as symbols for the former and latter, respectively. Let $\mathbf{\Sigma}$ be the $p \times p$ population covariance matrix, whose (i, j)-th component σ_{ij} is the population variance ($i = j$) or covariance ($i \neq j$). It is a symmetric, nonnegative definite matrix with the eigendecomposition

$$\mathbf{\Sigma} = \mathbf{\Upsilon}\mathbf{\Lambda}\mathbf{\Upsilon}^T, \tag{1}$$

where the superscript ^{T} denotes matrix transposition, $\mathbf{\Upsilon}$ is an orthogonal matrix of eigenvectors ($\mathbf{\Upsilon}\mathbf{\Upsilon}^T = \mathbf{\Upsilon}^T\mathbf{\Upsilon} = \mathbf{I}_p$ where $\mathbf{I}_p$ is the $p \times p$ identity matrix), and $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ of $\mathbf{\Sigma}$ (population eigenvalues). For convenience, the eigenvalues are arranged in the non-increasing order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Let $\boldsymbol{\mu}$ be the $p \times 1$ population mean vector.

Let $\mathbf{X}$ be an $N \times p$ observation matrix consisting of p-variate observations, which are individually denoted as $\mathbf{x}_i$ ($p \times 1$ vector; transposed in the rows of $\mathbf{X}$). (No strict notational distinction is made between a random variable and its realization.) At this point, N observations are assumed to be identically and independently distributed (i.i.d.). The sample covariance matrix $\mathbf{S}$ and cross-product matrix $\mathbf{A}$ are defined as

$$\mathbf{S} = \frac{1}{n_*}\mathbf{A} = \frac{1}{n_*}(\mathbf{X} - \mathbf{1}_N\bar{\mathbf{x}}^T)^T(\mathbf{X} - \mathbf{1}_N\bar{\mathbf{x}}^T), \tag{2}$$

where $\mathbf{1}_N$ is a $N \times 1$ column vector of 1's, $\bar{\mathbf{x}} = \sum_{i=1}^{N} \mathbf{x}_i / N$ is the sample mean vector, and n_* denotes an appropriate devisor; e.g., $n_* = N - 1$ for the ordinary unbiased estimator, and $n_* = N$ for the maximum likelihood estimator under the normal distribution. The (i, j)-th component of $\mathbf{S}$, denoted s_{ij}, is the sample variance or covariance. The eigendecomposition of $\mathbf{S}$ is constructed in the same way as above:

$$\mathbf{S} = \mathbf{ULU}^T, \tag{3}$$

where $\mathbf{U}$ is an orthogonal matrix of sample eigenvectors and $\mathbf{L}$ is a diagonal matrix whose elements are the sample eigenvalues $l_1, l_2, ..., l_p$.

In what follows, the following identity entailed by the orthogonality of $\mathbf{U}$ is frequently utilized:

$$\sum_{i=1}^{p} s_{ii}^r = \mathrm{tr}(\mathbf{S}^r) = \mathrm{tr}(\mathbf{ULU}^T\mathbf{ULU}^T \dots \mathbf{ULU}^T) = \mathrm{tr}(\mathbf{L}^r) = \sum_{i=1}^{p} l_i^r\,, \; r = 1, 2, ..., \tag{4}$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace operator, i.e., summation of the diagonal elements; the parentheses are omitted for visual clarity when little ambiguity exists. The sum of variances $\mathrm{tr}\,\mathbf{S} = \mathrm{tr}\,\mathbf{L}$ is called total variance. Note that equation 4 holds even when $n < p$, in which case $l_i = 0$ for some i. In other words, when $n < p$, the sample total variance is in a way concentrated in a subspace with fewer dimensions than the full space.

The population and sample correlation matrices $\mathbf{P}$ and $\mathbf{R}$, whose (i, j)-th components are the population and sample correlation coefficients ρ_{ij} and r_{ij}, respectively, are obtained by standardizing $\mathbf{\Sigma}$ and $\mathbf{S}$:

$$\mathbf{P} = \mathrm{diag}\left(\sigma_{ii}^{-1/2}\right) \mathbf{\Sigma}\, \mathrm{diag}\left(\sigma_{ii}^{-1/2}\right),$$

$$\mathbf{R} = \mathrm{diag}\left(s_{ii}^{-1/2}\right) \mathbf{S}\, \mathrm{diag}\left(s_{ii}^{-1/2}\right), \tag{5}$$

where $\mathrm{diag}(\cdot)$ stands for the $p \times p$ diagonal matrix with the designated ith elements. Their eigendecomposition is defined as for covariance matrices, and the eigenvalues are denoted with the same symbols here. For any i, $\rho_{ii} = r_{ii} = 1$, and hence, for correlation matrices

$$\mathrm{tr}\,\mathbf{P} = \mathrm{tr}\,\mathbf{\Lambda} = \mathrm{tr}\,\mathbf{R} = \mathrm{tr}\,\mathbf{L} = p. \tag{6}$$

In what follows, the notations $\mathrm{E}(\cdot)$, $\mathrm{Var}(\cdot)$, and $\mathrm{Cov}(\cdot,\cdot)$ are used for the expectation (mean), variance, and covariance of random variables, respectively.

Eigenvalue dispersion

The eigenvalue variance V is defined as:

$$V(\mathbf{\Sigma}) = \frac{1}{p}\sum_{i=1}^{p}\left(\lambda_i - \bar{\lambda}\right)^2,$$

$$V(\mathbf{S}) = \frac{1}{p}\sum_{i=1}^{p}\left(l_i - \bar{l}\right)^2, \qquad (7)$$

where $\bar{\lambda}$ and $\bar{l}$ are the averages of the population and sample eigenvalues, respectively ($\bar{\lambda} = \sum_{i=1}^{p}\lambda_i/p = \operatorname{tr}\mathbf{\Lambda}/p$, $\bar{l} = \sum_{i=1}^{p} l_i/p = \operatorname{tr}\mathbf{L}/p$). Note that $V(\mathbf{\Sigma})$ is a quantity pertaining to the population, whereas $V(\mathbf{S})$ is a sample statistic. The definition here follows the convention in the literature that p, rather than $p - 1$, is used as the divisor (e.g., Cheverud et al., 1983; Pavlicev et al., 2009; Haber, 2011). The latter might be more suitable for $V(\mathbf{S})$ because the sum of squares is taken around the average sample eigenvalue which is a random variable. After all, however, the choice of $p - 1$ is not so useful because $V(\mathbf{S})$ cannot be an unbiased estimator of $V(\mathbf{\Sigma})$ even with that choice (below).

Note that the average and sum of squares are taken across all p eigenvalues, even if some eigenvalues are zero due to the condition $n < p$. This is reasonable given that sums of moments across all p sample eigenvalues are comparable in magnitude to those of population eigenvalues (see below). One could alternatively use eigenvalue standard deviation $\sqrt{V}$ (Pavlicev et al., 2009; Haber, 2011), but this study concentrates on V rather than $\sqrt{V}$, because the former is much more tractable for the purposes of characterizing distributions.

It is obvious that $V(\mathbf{\Sigma})$ takes a single minimum of 0 at $\left(\lambda_1, \lambda_2, \ldots, \lambda_p\right) = (\bar{\lambda}, \bar{\lambda}, \ldots, \bar{\lambda})$. On the other hand, for a fixed $\bar{\lambda}$, it takes a single maximum of $(p-1)\bar{\lambda}^2$ at $(p\bar{\lambda}, 0, \ldots, 0)$ (e.g., Van Valen, 1974; Machado et al., 2019). Hence, not only is $V(\mathbf{\Sigma})$ scale-variant, but also its range depends on $p - 1$. Therefore, it is often useful to standardize V by division with this maximum to obtain the relative eigenvalue variance V_{rel}:

$$V_{\mathrm{rel}}(\mathbf{\Sigma}) = \frac{\sum_{i=1}^{p}(\lambda_i - \bar{\lambda})^2}{p(p-1)\,\bar{\lambda}^2},$$

$$V_{\mathrm{rel}}(\mathbf{S}) = \frac{\sum_{i=1}^{p}(l_i - \bar{l})^2}{p(p-1)\,\bar{l}^2}. \quad (8)$$

Because of the standardization, V_{rel} ranges between 0 and 1. This is a heuristic introduction of V_{rel} from V, but it will be seen below that $V_{\mathrm{rel}}(\mathbf{S})$ has a clearer theoretical justification.

These indices are similarly defined for correlation matrices. By noting $\bar{\lambda} = \bar{l} = 1$ (eq. 6), these are

$$V(\mathbf{P}) = \frac{1}{p}\sum_{i=1}^{p}(\lambda_i - 1)^2,$$

$$V(\mathbf{R}) = \frac{1}{p}\sum_{i=1}^{p}(l_i - 1)^2.$$

$$V_{\mathrm{rel}}(\mathbf{P}) = \frac{\sum_{i=1}^{p}(\lambda_i - 1)^2}{p(p-1)},$$

$$V_{\mathrm{rel}}(\mathbf{R}) = \frac{\sum_{i=1}^{p}(l_i - 1)^2}{p(p-1)}. \quad (9)$$

In most of the following discussions, we will concentrate on V_{rel} for correlation matrices, because $V(\mathbf{R})$ and $V_{\mathrm{rel}}(\mathbf{R})$ are proportional to each other by the factor $p - 1$, and hence their distributions are identical up to this scaling. This is in contrast to those of covariance matrices, where $\bar{l}$ in the denominator in $V_{\mathrm{rel}}(\mathbf{S})$ is a random variable and affects sampling properties.

Importantly, a single value of V_{rel} in general corresponds to multiple combinations of eigenvalues even if the average eigenvalue is fixed, except when $p = 2$ or under the extreme conditions $V_{\mathrm{rel}} = 0$ and $V_{\mathrm{rel}} = 1$ (Fig. 1). As such, it is not always straightforward to discern how intermediate values of V_{rel} are translated into actual covariance structures when $p > 2$. Nevertheless, it is possible to show that $V_{\mathrm{rel}} > 0.5$ cannot happen when multiple leading eigenvalues are of the same magnitude (Appendix A); in other words, such a large value indicates dominance of the first principal component.

As would be obvious from the definition, V and V_{rel} of covariance matrices only describe the (relative) magnitudes of eigenvalues—proportions of the axes of variation—and do not reflect any information of eigenvectors—directions of the axes. A large eigenvalue of a covariance matrix can represent, e.g., strong covariation between equally varying traits or large variation of a single trait uncorrelated with others; either of these cases describes eccentricity of variation in the multivariate space. By contrast, a large eigenvalue of a correlation matrix can only happen in the presence of correlation. Therefore, a large eigenvalue dispersion in a correlation matrix constrains conformation of eigenvectors to a certain extent. The correlations can nevertheless be realized in various ways depending on eigenvectors, whose conformation does influence the sampling distribution of $V_{\mathrm{rel}}(\mathbf{R})$ (see below).

For covariance matrices, $V_{\mathrm{rel}}(\mathbf{S})$ has a natural relation to the test of sphericity, i.e., test of the null hypothesis that $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$ for an arbitrary positive constant σ^2. Simple transformations from equation 8 lead to

$$V_{\mathrm{rel}}(\mathbf{S}) = \frac{1}{p-1}\left(p \frac{\sum l_i^2}{(\sum l_i)^2} - 1\right). \tag{10}$$

By noting $\sum l_i^2 / (\sum l_i)^2 = \mathrm{tr}(\mathbf{S}^2)/(\mathrm{tr}\,\mathbf{S})^2 = \mathrm{tr}(\mathbf{A}^2)/(\mathrm{tr}\,\mathbf{A})^2$ (see eqs. 2 and 4), $V_{\mathrm{rel}}(\mathbf{S})$ in the form of equation 10 is exactly John's (1972) T statistic for the test of sphericity (see also Ledoit & Wolf, 2002). Beyond the intuition that it measures eccentricity of variation along principal components, this statistic (and its linear functions) can be justified as the most powerful test statistic in the proximity of the null hypothesis under multivariate normality, among the class of such statistics that are invariant against translation by a constant vector, uniform scaling, and orthogonal rotation (John, 1971, 1972; Sugiura, 1972; Nagao, 1973). On the other hand, $V(\mathbf{S})$ does not seem to have as much theoretical justification, but rather has a practical advantage in the tractability of its moments and ease of correcting sampling bias (see below).

For a correlation matrix, V_{rel} is a measure of association between variables. Following similar transformations, it is straightforward to see

$$V_{\mathrm{rel}}(\mathbf{R}) = \frac{\mathrm{tr}(\mathbf{R}^2) - p}{p(p-1)}$$

$$= \frac{2}{p(p-1)} \sum_{i<j}^{p} r_{ij}^2, \tag{11}$$

because $r_{ii}^2 = 1$ for all i. This relationship has been known in the statistical literature (e.g., Gleason & Staelin, 1975; Durand & Le Roux, 2017), and empirically confirmed by Haber (2011). This statistic is used as a measure of overall association between variables (e.g., Schott, 2005; Durand & Le Roux, 2017), with the corresponding null hypothesis being $\mathbf{P} = \mathbf{I}_p$.

Sampling properties of eigenvalues

The distribution of eigenvalues of **S**, or equivalently those of **A** (which are n_* times those of **S**), has been extensively investigated in the literature of multivariate analysis (see, e.g., Jolliffe, 2002; Anderson, 2003). Unfortunately, however, most of such results are of limited value for the present purposes. On the one hand, forms of the exact joint distribution of the eigenvalues of **A** are known under certain assumptions on the population eigenvalues (e.g., Muirhead, 1982: pp. 107, 388), but they do not allow for much intuitive interpretation (let alone direct evaluation of moments), apart from the following points: 1) sample eigenvalues are *not* stochastically independent from one another; and 2) the distribution of sample eigenvalues are only dependent on the population eigenvalues, but not on the population eigenvectors. On the other hand, a substantial body of results is available for large-sample asymptotic distributions of sample eigenvalues (assuming $n \to \infty$, p being constant; e.g., Anderson, 1963, 2003), but their accuracy under finite n conditions is questionable. For example, a well-known result under a certain simple condition states that $l_i \sim N(\lambda_i, 2\lambda_i^2/n)$ and $\mathrm{Cov}(l_i, l_j) \approx 0$ for $i \neq j$, assuming all population eigenvalues to be distinct and $n_* = n$ (Girshick, 1939; Anderson, 1963; Srivastava & Khatri, 1979). However, these expressions ignore terms of order $O(n^{-1})$—that is, all terms with n or its higher power in the denominator—whose magnitude can be substantial for a finite n. Indeed, with further evaluation of higher-order terms, it becomes evident that sample eigenvalues are biased estimators of the population equivalents, where large eigenvalues are prone to overestimation and small ones are prone to underestimation, and that $\mathrm{Cov}(l_i, l_j) = 2\lambda_i\lambda_j/\left[(\lambda_i - \lambda_j)n\right]^2 + O(n^{-3})$ for $i \neq j$ (Lawley, 1956; Srivastava & Khatri, 1979). An important insight is that covariance between sample eigenvalues is nonzero. When all population eigenvalues are equal, then $\mathrm{Cov}(l_i, l_j) = -\sigma^4/n$ (Girshick, 1939).

Much less is known about eigenvalues of a sample correlation matrix **R** (Jolliffe, 2002). Their distribution seems intractable except under certain special conditions (Anderson, 1963). Asymptotic results indicate that the limiting distribution ($n \to \infty$) of an eigenvalue of **R** is normal with the mean coinciding with the corresponding population eigenvalue, but that its variance depends on population eigenvectors (Anderson, 1963; Konishi, 1979), unlike that of a covariance matrix where the distribution does not depend on population eigenvectors (above).

It is often of practical interest to detect the presence of eccentricity or integration, i.e., to test the null hypothesis of sphericity $\mathbf{\Sigma} = \sigma^2\mathbf{I}_p$ or no correlation $\mathbf{P} = \mathbf{I}_p$. These hypotheses are equivalent to $V(\mathbf{\Sigma}) = V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0$ and $V_{\mathrm{rel}}(\mathbf{P}) = 0$, respectively. Even under these conditions, nonzero sampling variance in sample eigenvalues renders $V(\mathbf{S}) > 0$, $V_{\mathrm{rel}}(\mathbf{S}) > 0$, and $V_{\mathrm{rel}}(\mathbf{R}) > 0$ with probability 1, because these statistics are calculated from sum of squares. The primary aim here is to derive explicit expressions for this sampling bias (expectation), as well as sampling variance.

It should be remembered that the expectation of the eigenvalue variance $\mathrm{E}[V(\mathbf{S})]$ is fundamentally different from the variance of eigenvalues $\mathrm{Var}(l_i)$. This point will be clarified by the following transformation:

$$\mathrm{E}[V(\mathbf{S})] = \frac{1}{p}\mathrm{E}\left(\sum_{i=1}^{p} l_i^2\right) - \frac{1}{p^2}\mathrm{E}\left[\left(\sum_{i=1}^{p} l_i\right)^2\right]$$

$$= \frac{p-1}{p^2}\sum_{i=1}^{p}\mathrm{E}(l_i^2) - \frac{1}{p^2}\sum_{i\neq j}^{p}\left[\mathrm{E}(l_i)\mathrm{E}(l_j) + \mathrm{Cov}(l_i, l_j)\right]. \qquad (12)$$

Under the null hypothesis, the moments are equal across all i, and the above simplifies into

$$\frac{p-1}{p}\left[\mathrm{Var}(l_i) - \mathrm{Cov}(l_i, l_j)\right], i \neq j. \qquad (13)$$

If $\mathrm{Cov}(l_i, l_j)$ were zero, the expectation would coincide with $(p-1)\mathrm{Var}(l_i)/p$, which can be evaluated from, e.g., Wagner's (1984) results. As already mentioned, however, this covariance is nonzero and hence cannot be ignored for covariance matrices. This is unlike the case for correlation matrices, where $\mathrm{E}[V(\mathbf{R})] = \mathrm{Var}(l_i)$ holds under the null hypothesis, because $\bar{l}$ is a constant and equals $\mathrm{E}(l_i) = 1$.

In the following discussions on moments of eigenvalue dispersion indices, observations are assumed to be i.i.d. multivariate normal variables. If the $N \times p$ matrix $\mathbf{X}$ consists of N i.i.d. p-variate normal variables $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the distribution of the sample-mean-centered cross product matrix $\mathbf{A}$ (eq. 2) is said to be the (central) Wishart distribution $W_p(\boldsymbol{\Sigma}, n)$, where $n = N - 1$ is the degree of freedom. It is well known that this is identical to the distribution of $\mathbf{Z}^T\mathbf{Z}$, where the $n \times p$ matrix $\mathbf{Z}$ consists of n i.i.d. p-variate normal variables $\mathbf{z}_i \sim N_p(\mathbf{0}_p, \boldsymbol{\Sigma})$ with $\mathbf{0}_p$ being the $p \times 1$ column vector of 0's (e.g., Anderson, 2003). Therefore, for analyzing statistics associated with sample covariance or correlation matrices, we can conveniently consider

$$\mathbf{S} = \frac{1}{n_*}\mathbf{Z}^T\mathbf{Z}. \tag{14}$$

without loss of generality, by bearing in mind the distinction between the degree of freedom n and sample size N. From elementary moments of the normal distribution, the following general relationships can be easily confirmed

$$\mathrm{E}(s_{ij}) = \frac{1}{n_*}\sum_{k=1}^{n}\mathrm{E}(z_{ki}z_{kj}) = \frac{n}{n_*}\sigma_{ij},$$

$$\mathrm{E}(s_{ij}s_{km}) = \frac{n^2}{n_*^2}\left[\sigma_{ij}\sigma_{km} + \frac{1}{n}(\sigma_{ik}\sigma_{jm} + \sigma_{im}\sigma_{jk})\right]. \tag{15}$$

where z_{ij}, s_{ij}, σ_{ij} and the like are the (i,j)-th elements of $\mathbf{Z}$, $\mathbf{S}$, and $\boldsymbol{\Sigma}$, respectively.

Moments under null hypotheses

Covariance matrix

Before proceeding to arbitrary covariance structures, let us consider the null hypothesis of sphericity: $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_p$, where σ^2 is the population variance of arbitrary magnitude. For the expectation of $V(\mathbf{S})$, we need $\mathrm{Var}(l_i)$ and $\mathrm{Cov}(l_i, l_j)$, or equivalently $\mathrm{E}\left(\sum_{i=1}^{p} l_i^2\right)$ and $\mathrm{E}\left[\left(\sum_{i=1}^{p} l_i\right)^2\right]$ (see eqs. 12 and 13); we will proceed with the latter here. By use of equations 4 and 15, we have

$$\mathrm{E}\left(\sum_{i=1}^{p} l_i^2\right) = \mathrm{E}\left(\sum_{i,j=1}^{p} s_{ij}^2\right)$$

$$= \mathrm{E}\left(\sum_{i=1}^{p} s_{ii}^2 + \sum_{i \neq j}^{p} s_{ij}^2\right)$$

$$= \left[p\mathrm{E}(s_{ii}^2) + p(p-1)\mathrm{E}(s_{ij}^2)\right]$$

$$= \frac{pn}{n_*^2}(p + n + 1)\sigma^4, \tag{16}$$

and similarly

$$\mathrm{E}\left[\left(\sum_{i=1}^{p} l_i\right)^2\right] = \mathrm{E}\left[\left(\sum_{i=1}^{p} s_{ii}\right)^2\right]$$

$$= \mathrm{E}\left[\sum_{i=1}^{p} s_{ii}^2 + \sum_{i \neq j}^{p} s_{ii}s_{jj}\right]$$

$$= \left[p\mathrm{E}(s_{ii}^2) + p(p-1)\mathrm{E}(s_{ii}s_{jj})\right]$$

$$= \frac{pn}{n_*^2}(pn + 2)\sigma^4. \tag{17}$$

Then, inserting these results into equation 12,

$$\mathrm{E}[V(\mathbf{S})] = \frac{n}{pn_*^2}(p-1)(p+2)\sigma^4. \tag{18}$$

Alternatively, it could be seen that $\mathrm{Var}(l_i) = n(p+1)\sigma^2/n_*^2$ and $\mathrm{Cov}(l_i, l_j) = -n\sigma^4/n_*^2$ for $i \neq j$, with which equation 13 yields the identical result.

The variance of $V(\mathbf{S})$ is, by equation 12,

$$\mathrm{Var}[V(\mathbf{S})] = \frac{1}{p^2}\mathrm{Var}\left[\sum_{i=1}^{p} l_i^2\right] + \frac{1}{p^4}\mathrm{Var}\left[\left(\sum_{i=1}^{p} l_i\right)^2\right] - 2\frac{1}{p^3}\mathrm{Cov}\left[\sum_{i=1}^{p} l_i^2, \left(\sum_{i=1}^{p} l_i\right)^2\right]. \tag{19}$$

The relevant moments can most conveniently be found as a special case of general expressions under arbitrary $\mathbf{\Sigma}$ (see below and Appendix B), although direct derivation using normal moments is possible:

$$\mathrm{E}\left[\left(\sum l_i^2\right)^2\right] = \frac{pn}{n_*^4}(p^3n + pn^3 + 2p^2n^2 + 2p^2n + 2pn^2 + 8p^2 + 8n^2 + 21pn + 20p + 20n + 20)\sigma^8;$$

$$\mathrm{E}\left[\left(\sum l_i\right)^4\right] = \frac{pn}{n_*^4}(pn+2)(pn+4)(pn+6)\sigma^8;$$

$$\mathrm{E}\left[\left(\sum l_i^2\right)\cdot\left(\sum l_i\right)^2\right] = \frac{pn}{n_*^4}(pn+2)(pn+4)(p+n+1)\sigma^8;$$

$$\mathrm{Var}\left[\sum l_i^2\right] = \frac{4pn}{n_*^4}(2p^2 + 2n^2 + 5pn + 5p + 5n + 5)\sigma^8;$$

$$\mathrm{Var}\left[\left(\sum l_i\right)^2\right] = \frac{8pn}{n_*^4}(pn+2)(pn+3)\sigma^8;$$

$$\mathrm{Cov}\left[\sum l_i^2, \left(\sum l_i\right)^2\right] = \frac{8pn}{n_*^4}(p+n+1)(pn+3)\sigma^8. \tag{20}$$

Inserting these into equation 19 yields

$$\mathrm{Var}[V(\mathbf{S})] = \frac{4n}{p^3n_*^4}(p-1)(p+2)(2p^2 + pn + 3p - 6)\sigma^8. \tag{21}$$

Next, consider the first two moments of $V_{\mathrm{rel}}(\mathbf{S})$ under the null hypothesis (which have previously been derived by John [1972]). Recalling the form of equation 10,

$$\mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})] = \frac{1}{p-1}\left(p\mathrm{E}\left[\frac{\sum l_i^2}{(\sum l_i)^2}\right] - 1\right),$$

and

$$\mathrm{Var}[V_{\mathrm{rel}}(\mathbf{S})] = \left(\frac{p}{p-1}\right)^2 \mathrm{Var}\left[\frac{\sum l_i^2}{(\sum l_i)^2}\right]. \tag{22}$$

In general, moments of the ratio $\sum l_i^2/(\sum l_i)^2$ do not coincide with the ratio of the moments of the numerator and denominator. Specifically under the null hypothesis, however,

$$\mathrm{E}\left[\frac{\sum l_i^2}{(\sum l_i)^2}\right] = \frac{\mathrm{E}[\sum l_i^2]}{\mathrm{E}[(\sum l_i)^2]}$$

and

$$\mathrm{E}\left[\frac{(\sum l_i^2)^2}{(\sum l_i)^4}\right] = \frac{\mathrm{E}\left[(\sum l_i^2)^2\right]}{\mathrm{E}[(\sum l_i)^4]} \tag{23}$$

hold because of the stochastic independence between $\sum l_i^2/(\sum l_i)^2$ and $\sum l_i$ in this special condition (this point requires inspection of the density; John, 1972). Therefore, by use of equations 16, 17, and 20,

$$\mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})] = \frac{1}{p-1}\left(p\frac{\mathrm{E}[\sum l_i^2]}{\mathrm{E}[(\sum l_i)^2]} - 1\right)$$

$$= \frac{p+2}{pn+2}, \tag{24}$$

and

$$\mathrm{Var}[V_{\mathrm{rel}}(\mathbf{S})] = \left(\frac{p}{p-1}\right)^2 \left(\frac{\mathrm{E}[(\sum l_i^2)^2]}{\mathrm{E}[(\sum l_i)^4]} - \left\{\frac{\mathrm{E}[\sum l_i^2]}{\mathrm{E}[(\sum l_i)^2]}\right\}^2\right)$$

$$= \frac{4(p-1)(p+2)(n-1)(n+2)}{(pn+2)^2(pn+4)(pn+6)}. \tag{25}$$

These results are non-asymptotic (valid across any p and n) and exact under multivariate normality.

Correlation matrix

Consider the null hypothesis $\mathbf{P} = \mathbf{I}_p$ or $\rho_{ij} = 0$ for $i \neq j$. The moments can conveniently be obtained from the form of average squared correlation (eq. 11). It is well known that, under the assumptions of normality and $\rho_{ij} = 0$ for $i \neq j$, r_{ij}^2 is distributed as $\mathrm{Beta}(1/2, (n-1)/2)$, where n is the degree of freedom (e.g., Anderson, 2003). Therefore, under the null hypothesis,

$$\mathrm{E}\left(r_{ij}^2\right) = \frac{1}{n},\ i \neq j,$$

$$\mathrm{Var}\left(r_{ij}^2\right) = \frac{2(n-1)}{n^2(n+2)},\ i \neq j. \tag{26}$$

The expectation of $V_{\mathrm{rel}}(\mathbf{R})$ is simply the average:

$$\mathrm{E}[V_{\mathrm{rel}}(\mathbf{R})] = \frac{1}{n}. \tag{27}$$

This expression is identical to $(p-1)^{-1}\mathrm{Var}(l_i)$ obtainable from Wagner's (1984) results, except for having the degree of freedom n rather than the sample size N in the denominator. This is because Wagner (1984) considered N uncentered observations with mean 0 without explicitly distinguishing n and N. Most practical analyses would concern data centered at the sample mean, thus should use n rather than N.

Derivation of the variance is more complicated than it may seem, because, in principle,

$$\mathrm{Var}[V_{\mathrm{rel}}(\mathbf{R})] = \frac{4}{p^2(p-1)^2}\left[\sum_{i<j}\mathrm{Var}(r_{ij}^2) + \sum_{\substack{i<j,k<l,\\(i,j)\neq(k,l)}} 2\mathrm{Cov}(r_{ij}^2, r_{kl}^2)\right]. \tag{28}$$

However, it is possible to show $\mathrm{Cov}(r_{ij}^2, r_{kl}^2) = 0$ under the null hypothesis (Appendix C). Therefore, from equations 26 and 28,

$$\mathrm{Var}[V_{\mathrm{rel}}(\mathbf{R})] = \frac{4(n-1)}{p(p-1)n^2(n+2)}. \tag{29}$$

These expressions are non-asymptotic and exact for any p and n. Schott (2005) proposed a test for independence between sets of normal variables based on these moments.

Moments under arbitrary conditions

Covariance matrix

This section considers moments of eigenvalue dispersion indices under arbitrary covariance/correlation structures and multivariate normality. It is straightforward to obtain the first two moments of $V(\mathbf{S})$ under arbitrary $\mathbf{\Sigma}$, provided that the expectations of relevant terms in equation 12 are available. The results are

$$\mathrm{E}[V(\mathbf{S})] = \frac{n}{p^2 n_*^2}[(p-n)(\operatorname{tr}\mathbf{\Lambda})^2 + (pn+p-2)\operatorname{tr}(\mathbf{\Lambda}^2)]$$

$$= \frac{n}{p n_*^2}[(pn+p-2)V(\mathbf{\Sigma}) + (p-1)(p+2)(\operatorname{tr}\mathbf{\Lambda})^2/p^2],$$

$$\mathrm{Var}[V(\mathbf{S})] = \frac{4n}{p^4 n_*^4}\{2(p-n)^2 \operatorname{tr}(\mathbf{\Lambda}^2)(\operatorname{tr}\mathbf{\Lambda})^2 + (p^2 n + p^2 - 4p + 2n)[\operatorname{tr}(\mathbf{\Lambda}^2)]^2$$

$$+ 4(p-n)(pn+p-2)\operatorname{tr}(\mathbf{\Lambda}^3)\operatorname{tr}\mathbf{\Lambda}$$

$$+ (2p^2n^2 + 5p^2 n + 5p^2 - 12pn - 12p + 12)\operatorname{tr}(\mathbf{\Lambda}^4)\}. \quad (30)$$

The derivations are given in Appendix B. The second expression for the expectation comes from the fact $V(\mathbf{\Sigma}) = [p\operatorname{tr}(\mathbf{\Lambda}^2) - (\operatorname{tr}\mathbf{\Lambda})^2]/p^2$, and clarifies that the expectation is a linear function of $V(\mathbf{\Sigma})$. These results are exact, and it can be easily verified that they reduce to equations 18 and 19 under the null hypothesis. Profiles of $\mathrm{E}[V(\mathbf{S})]$ across a range of $V(\mathbf{\Sigma})$ are shown in Figure 2 (top row), under single large eigenvalue conditions with varying p and N and a fixed $\operatorname{tr}\mathbf{\Sigma}$ (details are described under simulation settings below).

Moments of $V_{\mathrm{rel}}(\mathbf{S})$ are more difficult to obtain, as moments of the ratio $\sum l_i^2/(\sum l_i)^2$ do not coincide with the ratio of moments under arbitrary $\mathbf{\Sigma}$. Here we utilize the following approximations based on the delta method (e.g., Stuart & Ord, 1994: chapter 10):

$$\mathrm{E}\left(\frac{X}{Y}\right) \approx \frac{\mathrm{E}(X)}{\mathrm{E}(Y)} - \frac{\mathrm{Cov}(X,Y)}{\mathrm{E}(Y)^2} + \frac{\mathrm{E}(X)\,\mathrm{Var}(Y)}{\mathrm{E}(Y)^3},$$

and

$$\mathrm{Var}\left(\frac{X}{Y}\right) \approx \frac{\mathrm{E}(X)^2}{\mathrm{E}(Y)^2}\left[\frac{\mathrm{Var}(X)}{\mathrm{E}(X)^2} + \frac{\mathrm{Var}(Y)}{\mathrm{E}(Y)^2} - 2\frac{\mathrm{Cov}(X,Y)}{\mathrm{E}(X)\,\mathrm{E}(Y)}\right]. \quad (32)$$

The approximate moments are (Appendix B):

$$\mathrm{E}\left[\frac{\sum l_i^2}{(\sum l_i)^2}\right] \approx \frac{(\operatorname{tr}\mathbf{\Lambda})^2 + (n+1)\operatorname{tr}(\mathbf{\Lambda}^2)}{n(\operatorname{tr}\mathbf{\Lambda})^2 + 2\operatorname{tr}(\mathbf{\Lambda}^2)} - \frac{8(n-1)(n+2)}{n[n(\operatorname{tr}\mathbf{\Lambda})^2 + 2\operatorname{tr}(\mathbf{\Lambda}^2)]^3}$$

$$\times \{n(\operatorname{tr}\mathbf{\Lambda})^3\operatorname{tr}(\mathbf{\Lambda}^3) - n(\operatorname{tr}\mathbf{\Lambda})^2[\operatorname{tr}(\mathbf{\Lambda}^2)]^2 - 2[\operatorname{tr}(\mathbf{\Lambda}^2)]^3 - 2\operatorname{tr}\mathbf{\Lambda}\operatorname{tr}(\mathbf{\Lambda}^2)\operatorname{tr}(\mathbf{\Lambda}^3)$$

$$+ 3(\operatorname{tr}\mathbf{\Lambda})^2\operatorname{tr}(\mathbf{\Lambda}^4)\};$$

$$\mathrm{Var}\left[\frac{\sum l_i^2}{(\sum l_i)^2}\right] \approx \frac{4(n-1)(n+2)}{n[n(\mathrm{tr}\,\boldsymbol{\Lambda})^2 + 2\,\mathrm{tr}(\boldsymbol{\Lambda}^2)]^4}$$
$$\times \{n(\mathrm{tr}\,\boldsymbol{\Lambda})^4[\mathrm{tr}(\boldsymbol{\Lambda}^2)]^2 + 2n(n+1)(\mathrm{tr}\,\boldsymbol{\Lambda})^2[\mathrm{tr}(\boldsymbol{\Lambda}^2)]^3 + 2(n+1)[\mathrm{tr}(\boldsymbol{\Lambda}^2)]^4$$
$$- 4(n-1)(n+2)(\mathrm{tr}\,\boldsymbol{\Lambda})^3\,\mathrm{tr}(\boldsymbol{\Lambda}^2)\,\mathrm{tr}(\boldsymbol{\Lambda}^3) + (2n^2+3n-6)(\mathrm{tr}\,\boldsymbol{\Lambda})^4\,\mathrm{tr}(\boldsymbol{\Lambda}^4)$$
$$- 4n(\mathrm{tr}\,\boldsymbol{\Lambda})^2\,\mathrm{tr}(\boldsymbol{\Lambda}^2)\,\mathrm{tr}(\boldsymbol{\Lambda}^4) - 4(\mathrm{tr}\,\boldsymbol{\Lambda})^4[\mathrm{tr}(\boldsymbol{\Lambda}^2)]^2\}. \qquad (33)$$

Inserting these into equation 22 yields the desired moments. The approximate expectation reduces to equation 24 under the null hypothesis, as the higher-order terms cancel out, whereas this is not the case for the approximate variance. Because these expressions are specified only by the population eigenvalues regardless of eigenvectors, they are invariant with respect to orthogonal rotations, as expected from theoretical considerations above. Also, it is easily discerned that these expressions are invariant with respect to uniform scaling of the variables. The accuracy of these approximations will be examined in simulations below.

Profiles of the approximation of $\mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})]$ across a range of $V_{\mathrm{rel}}(\boldsymbol{\Sigma})$ are shown in Figure 2 (middle row) for the same conditions as explained above. The profiles are nonlinear; $V_{\mathrm{rel}}(\mathbf{S})$ tends to overestimate $V_{\mathrm{rel}}(\boldsymbol{\Sigma})$ when the latter is small, but tends to slightly underestimate when the latter is large. The initial decrease of $\mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})]$ observed in some profiles appears to be an artifact of the approximation.

Correlation matrix

The expectation of $V_{\mathrm{rel}}(\mathbf{R})$ under arbitrary conditions can be obtained from the expression of equation 11 with r_{ij}^2 replaced by its expectations, which is known to be (e.g., Ghosh, 1966; Muirhead, 1982)

$$\mathrm{E}(r_{ij}^2) = 1 - \frac{(n-1)(1-\rho_{ij}^2)}{n}\,{}_2F_1\left(1, 1; \frac{n+2}{2}; \rho_{ij}^2\right), \; i \neq j, \qquad (34)$$

where

$$ {}_2F_1(a,b;c;z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!} \quad (35) $$

is the hypergeometric function, with $(x)_k = x(x+1)\dots(x+k-1)$ denoting rising factorial (defined to be 1 when $k = 0$). Taking the average of equation 34 gives the desired expectation. This result is non-asymptotic and exact. It is easily seen that equation 34 reduces to equation 26 under the null hypothesis.

When $p = 2$, the exact variance of $V_{\mathrm{rel}}(\mathbf{R})$ is equal to that of the single squared correlation coefficient, thus can be obtained from known results (Ghosh, 1966) as follows:

$$ \mathrm{Var}[V_{\mathrm{rel}}(\mathbf{R})] = \mathrm{Var}(r^2) = \frac{(n-1)(n+1)(1-\rho^2)}{2n}\left[F - nF' - \frac{2(n-1)(1-\rho^2)}{n(n+1)}F^2\right], \quad (36) $$

where $F = {}_2F_1(1,1;(n+2)/2\,;\rho^2)$ and $F' = (F-1)/2\rho^2 = {}_2F_1(1,2;(n+4)/2\,;\rho^2)/(n+2)$; this last form is preferred to avoid numerical instability when ρ^2 is close to 0. This expression reduces to equation 26 under the null hypothesis. When $p > 2$, we cannot ignore the covariance between squared correlation coefficients (see eq. 28), which appears to be nonzero. Unfortunately, no exact expression seems available for this in the literature, so we resort to asymptotic results. The following asymptotic expression based on Konishi's (1979) theory may potentially be used (see Appendix D for derivation):

$$ \mathrm{Var}[V_{\mathrm{rel}}(\mathbf{R})] \approx \frac{8}{p^2(p-1)^2 n} \sum_{\alpha,\beta=1}^{p} \lambda_\alpha^2 \lambda_\beta^2 \left[\delta_{\alpha\beta} - (\lambda_\alpha + \lambda_\beta)\sum_{i=1}^{p} \upsilon_{i\alpha}^2 \upsilon_{i\beta}^2 + \sum_{i,j=1}^{p} \rho_{ij}^2 \upsilon_{i\alpha}^2 \upsilon_{j\beta}^2\right], \quad (37) $$

where δ_{ij} is the Kronecker delta (equals 1 for $i = j$ and 0 otherwise) and $\upsilon_{i\alpha}$ is the (i, α)-th element of the population eigenvector matrix $\boldsymbol{\Upsilon}$. For $p = 2$, the accuracy of this expression can be compared with the exact expression (Fig. 3); visual inspection of the profiles suggest that the accuracy is satisfactory past $N = 32$–64, except around $V_{\mathrm{rel}}(\mathbf{P}) = 0$ where the asymptotic expression diminishes to 0. For $p > 2$, the accuracy is to be evaluated with simulations below.

Importantly, the expectation of $V_{\mathrm{rel}}(\mathbf{R})$ is functions of ρ^2's rather than $\mathbf{\Lambda}$, and cannot be specified by the latter alone in general. For instance, consider $\begin{pmatrix} 1 & 0.9 & 0 \\ 0.9 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0.9/\sqrt{2} & 0 \\ 0.9/\sqrt{2} & 1 & 0.9/\sqrt{2} \\ 0 & 0.9/\sqrt{2} & 1 \end{pmatrix}$, both of which are valid correlation matrices. These matrices have identical eigenvalues $\mathbf{\Lambda} = \mathrm{diag}(1.9, 1.0, 0.1)$ and hence an identical value of $V_{\mathrm{rel}}(\mathbf{P})$ (= 0.27), but $\mathrm{E}[V_{\mathrm{rel}}(\mathbf{R})]$ with $n = 10$ are 0.3326 and 0.3156, respectively. Although the difference in the expectations decreases as n increases, this example highlights that the distribution of $V_{\mathrm{rel}}(\mathbf{R})$ is also dependent on population eigenvectors.

Profiles of $\mathrm{E}[V_{\mathrm{rel}}(\mathbf{R})]$ across a range of $V_{\mathrm{rel}}(\mathbf{P})$ are shown in Figure 2 (bottom row), under the same conditions as above. These conditions with single large eigenvalues are special cases in which $\mathrm{E}[V_{\mathrm{rel}}(\mathbf{R})]$ can be specified by $V_{\mathrm{rel}}(\mathbf{P})$ regardless of eigenvectors (detailed in Appendix A). Indeed, the profiles of the expectations are invariant across p in these special conditions. In some way similar to $V_{\mathrm{rel}}(\mathbf{S})$, $V_{\mathrm{rel}}(\mathbf{R})$ tends to overestimate and underestimate small and large values of $V_{\mathrm{rel}}(\mathbf{P})$, respectively.

Bias correction

Some authors (Cheverud et al., 1989; Torices & Muñoz-Pajares, 2015) have suggested correcting the sampling bias in eigenvalue dispersion indices by means of subtracting Wagner's (1984) null expectation from empirical values (but see also Armbruster et al., 2009). This method could potentially be used for V and V_{rel} with the correct null expectations derived above, to obtain estimators that is unbiased under the null hypothesis. For $V_{\mathrm{rel}}(\mathbf{S})$ and $V_{\mathrm{rel}}(\mathbf{R})$, however, the subtraction truncates the upper end of the range, potentially compromising interpretability. To avoid this, it might be desirable to scale these indices in a way analogous to the adjusted coefficient of determination in regression analysis (e.g., Cramer, 1987):

$$\bar{V}_{\mathrm{rel}}(\mathbf{S}) = 1 - \frac{1 - V_{\mathrm{rel}}(\mathbf{S})}{1 - \mathrm{E}_{\mathrm{null}}[V_{\mathrm{rel}}(\mathbf{S})]} = \frac{pn+2}{p(n-1)} V_{\mathrm{rel}}(\mathbf{S}) - \frac{p+2}{p(n-1)},$$

$$\bar{V}_{\mathrm{rel}}(\mathbf{R}) = 1 - \frac{1 - V_{\mathrm{rel}}(\mathbf{R})}{1 - \mathrm{E}_{\mathrm{null}}[V_{\mathrm{rel}}(\mathbf{R})]} = \frac{n}{n-1} V_{\mathrm{rel}}(\mathbf{R}) - \frac{1}{n-1}, \tag{38}$$

where $\mathrm{E}_{\mathrm{null}}(\cdot)$ denotes expectation under the appropriate null hypothesis (eqs. 24 and 27). This adjustment inflates the variance by the factor of $1/[1 - \mathrm{E}_{\mathrm{null}}(V_{\mathrm{rel}})]^2$. Furthermore, these adjusted indices are unbiased only under the null hypothesis (and trivially the case of complete integration), and uniformly underestimate the corresponding population values otherwise (Fig. S2). As the population value gets away from 0, the adjusted index is outperformed by the unadjusted one in both precision and bias (Fig. S3). It should also be borne in mind that the profiles of expectations are nonlinear and dependent on N (Fig. 2). As the adjusted indices will be increasingly conservative for small N, it is questionable whether they can be used for comparing samples with different N, as originally intended by Cheverud et al. (1989). For these reasons, use of this adjustment would be restricted to estimation of the population value near 0 (up to 0.1–0.2, depending on p and N).

On the other hand, a global unbiased estimator of $V(\mathbf{\Sigma})$ can be derived from above results:

$$\tilde{V}(\mathbf{S}) = \frac{n_*^2}{n(pn+p-2)} \left(pV(\mathbf{S}) - \frac{(p-1)(p+2)}{p^2(n-1)(n+2)} [(n+1)(\operatorname{tr}\mathbf{S})^2 - 2\operatorname{tr}(\mathbf{S}^2)] \right)$$

$$= \frac{1}{p^2 n(n-1)(n+2)} [(pn+2)(\operatorname{tr}\mathbf{A})^2 - (p+n+1)\operatorname{tr}(\mathbf{A}^2)]. \tag{39}$$

The unbiasedness $\mathrm{E}[\tilde{V}(\mathbf{S})] = V(\mathbf{\Sigma})$ can be easily confirmed. Its variance can be similarly obtained:

$$\begin{aligned}\mathrm{Var}[\tilde{V}(\mathbf{S})] = \frac{4}{p^4 n(n-1)(n+2)}\{&2(n-1)(n+2)\,\mathrm{tr}(\mathbf{\Lambda}^2)\,(\mathrm{tr}\,\mathbf{\Lambda})^2 \\ &+ (p^2 n + 4p + 2n + 2)[\mathrm{tr}(\mathbf{\Lambda}^2)]^2 - 4p(n-1)(n+2)\,\mathrm{tr}(\mathbf{\Lambda}^3)\,\mathrm{tr}\,\mathbf{\Lambda} \\ &+ (2p^2n^2 + 3p^2n - 6p^2 - 4pn - 4)\,\mathrm{tr}(\mathbf{\Lambda}^4)\},\end{aligned} \tag{40}$$

which reduces to $4(p-1)(p+2)\sigma^8/p^3 n(n-1)(n+2)$ under the null hypothesis. Comparison with equations 21 and 30 suggests that this variance is smaller than that of $V(\mathbf{S})$, especially under the null hypothesis. Therefore, $\tilde{V}(\mathbf{S})$ seems superior in both precision and bias and can be used when estimation of $V(\mathbf{\Sigma})$ is desired. It can be used to compare multiple samples, provided that its sensitivity to overall scaling is not of concern, e.g., comparison between closely related taxa.

Simulation

Methods

Simulations were conducted under various conditions in order to understand sampling properties of the eigenvalue dispersion indices. All simulations were done assuming multivariate normality, with varying population covariance matrix $\mathbf{\Sigma}$, number of variables p (= 2, 4, 8, 16, 32, 64, 128, 256, and 1024), and sample size N (= 4, 8, 16, 32, 64, 128, and 256).

For every p, the following population eigenvalue conformations were considered: 1) the null condition, 2) q-large λ conditions, 3) a linearly decreasing λ condition, and 4) a quadratically decreasing λ condition (see Fig. 4 for examples). The null condition is where all population eigenvalues are equal in magnitude ($\lambda_1 = \lambda_2 = \cdots = \lambda_p = \bar{\lambda}$; $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0$), corresponding to the null hypothesis of sphericity (Fig. 4A). The q-large λ conditions are where the first q (= 1, 2, and 4, provided $p > q$) population eigenvalues are equally large and the remaining $p - q$ ones are equally small ($\lambda_1 = \cdots = \lambda_q > \lambda_{q+1} = \cdots = \lambda_p$), with varying $V_{\mathrm{rel}}(\boldsymbol{\Sigma})$ (= 0.1, 0.2, 0.4, 0.6, and 0.8; Fig. 4B–G). The necessary condition $\lambda_p \geq 0$ constrains possible combinations of q and $V_{\mathrm{rel}}(\boldsymbol{\Sigma})$: the possible choices of $V_{\mathrm{rel}}(\boldsymbol{\Sigma})$ are 0.1–0.8, 0.1–0.4, and 0.1–0.2 for q = 1, 2, and 4, respectively (Appendix A). These conditions are intended to represent hypothetical situations where only a few components of meaningful signals are present in the covariance structure. Individual eigenvalues were calculated for each combination of p, q, and $V_{\mathrm{rel}}(\boldsymbol{\Sigma})$ as described in Appendix A. The linearly and quadratically decreasing λ conditions are where the population eigenvalues are linearly and quadratically, respectively, decreasing in magnitude (Fig. 4H; Appendix A), in which cases the value of $V_{\mathrm{rel}}(\boldsymbol{\Sigma})$ is fixed for a given p. These conditions are intended to represent covariance structures with gradually decreasing signals. One might claim that some of these situations, especially q-large λ conditions, are too simplistic and biologically unrealistic, but these simple settings enable us to clarify systematic relationships between parameters and sampling properties. The primary aim here is to explore sampling properties across a wide range of parameters, rather than confined to a biologically "realistic" region (which would depend on specific organismal systems). It should also be recalled that sampling error alone can yield gradually decreasing patterns of sample eigenvalues typically observed in empirical datasets (see above and below).

For sake of simplicity, all population covariance matrices were scaled to ensure $V(\boldsymbol{\Sigma}) = V_{\mathrm{rel}}(\boldsymbol{\Sigma})$; that is, $\operatorname{tr}\boldsymbol{\Sigma} = p(p-1)^{-1/2}$. This scaling also makes the magnitude of $V(\boldsymbol{\Sigma})$ comparable across varying p. In addition, a population covariance matrix $\boldsymbol{\Sigma}$ was constructed from a predefined set of eigenvalues such that its diagonal elements are equal: $\sigma_{ii} = \bar{\lambda} = (p-1)^{-1/2}$ for all i, thereby enforcing $\boldsymbol{\Sigma} = (p-1)^{-1/2}\mathbf{P}$. This construction allows for examining both covariance and correlation matrices with the same population V_{rel} from a single simulated dataset, saving computational resources. $\boldsymbol{\Sigma}$ was constructed from $\boldsymbol{\Lambda}$ by the iterative Givens rotation algorithm of Davies & Higham (2000), which is guaranteed to converge within $p-1$ iterations. This algorithm was implemented as coded by Waller (2020), but with the following modifications for reproducibility: no random orthogonal rotation was involved in the initial stage, and rotation axes were chosen in a fixed order. It should be noted that the rotations involved—choice of eigenvectors—would in general influence distributions of $V_{\mathrm{rel}}(\mathbf{R})$, except for certain special cases including the 1-large λ condition (see Appendix A). It is impractical to exhaustively examine numerous possible conformations of eigenvectors, so only the single conformation generated by this algorithm was used for each combination of parameters.

The eigenvalues of a sample covariance matrix were obtained from singular value decomposition of the data matrix, as the singular values squared and then divided by n_* (see, e.g., Jolliffe, 2002). When $p > N-1$, 0's were appended to this vector so that p sample eigenvalues were present. Data were centered at the sample mean before the decomposition, therefore $n = N-1$. It was chosen that $n_* = n$. The eigenvalues for a sample correlation matrix were obtained similarly from the sample-mean-centered data matrix scaled with the sample standard deviation for each variable.

To summarize, each set of simulations consists of the following steps: 1) define a desired set of eigenvalues $\mathbf{\Lambda}$; 2) construct the population covariance matrix $\mathbf{\Sigma}$ with the rotation algorithm explained above; 3) generate N i.i.d. normal observations from $N_p(\mathbf{0}, \mathbf{\Sigma})$; 4) eigenvalues of sample covariance and correlation matrices were obtained from singular value decomposition of the sample-mean-centered data; 5) $V(\mathbf{S})$, $V_{\mathrm{rel}}(\mathbf{S})$, and $V_{\mathrm{rel}}(\mathbf{R})$ were calculated from the eigenvalues; 6) the steps 3–5 were iterated for 5,000 times in total with the same N and $\mathbf{\Sigma}$. The simulations were conducted on the R environment version 3.5.3 (R Core Team, 2019). The function "genhypergeo" of the package "hypergeo" was used to evaluate the hypergeometric function in the moments of $V_{\mathrm{rel}}(\mathbf{R})$. The codes are provided as Supplementary Material.

Results

Examined individually, sample eigenvalues were biased estimators of population eigenvalues, as expected. Examples of eigenvalue distribution of sample covariance and correlation matrices are shown in Figures 4 and S1, respectively. Typically, the first few eigenvalues were overestimated, with the rest being underestimated. Note that gradually decreasing scree-like profiles of sample eigenvalues typical of empirical datasets can arise even when most population eigenvalues are identical in magnitude. The sampling biases decreased as N increases. These overall trends were similarly observed for correlation matrices, although the upper tail of the largest eigenvalue tended to be truncated for correlation matrices because of the constraint $\operatorname{tr} \mathbf{R} = p$, effectively cancelling the tendency of overestimation in this eigenvalue (Fig. S1).

Sampling distributions of $V(\mathbf{S})$ are shown in Figures 5 and S4–S6, and their summary statistics are shown in Tables 1 and S1. Distributions were unimodal but highly skewed with long upper tails, especially when N or p is small. As expected, sampling dispersion decreases consistently with increasing N, with skewness decreasing at the same time. Interestingly, the shape of distribution does not visibly change with increasing p, at least with moderately large N ($\geq$ 32, say). In all conditions, $V(\mathbf{S})$ tended to overestimate the population value $V(\mathbf{\Sigma})$. Increasing $V(\mathbf{\Sigma})$ drastically increased sampling dispersion and skewness, whereas increasing q with a fixed $V(\mathbf{\Sigma})$ decreased sampling dispersion without affecting the mean as much. Sampling distributions of $V(\mathbf{S})$ under linearly and quadratically decreasing λ conditions look similar to those under q-large λ conditions with similar $V(\mathbf{\Sigma})$ values for the respective p. The expressions of the expectation and variance of $V(\mathbf{S})$ almost always coincided with the sampling mean and variance within a reasonable range of random fluctuations (as expected, since those results are exact).

Results for $V_{\mathrm{rel}}(\mathbf{S})$ are summarized in Figures 6 and S6–S8 and Tables 2 and S2. Distributions were unimodal within the range (0, 1), except when $N = 4$ and $p = 2$ where the distribution was essentially uniform. As was the case for $V(\mathbf{S})$, the sampling dispersion of $V_{\mathrm{rel}}(\mathbf{S})$ decreased drastically with increasing N, and to some extent with increasing p, and the shape of distribution does not seem to change drastically with increasing p past certain N. $V_{\mathrm{rel}}(\mathbf{S})$ tended to overestimate the population value $V_{\mathrm{rel}}(\mathbf{\Sigma})$, except when the latter is rather large (= 0.8) where slight underestimation was observed. With increasing q for a fixed $V_{\mathrm{rel}}(\mathbf{\Sigma})$, the distributions tended to shrink, but the sampling bias remained virtually unchanged or slightly increased. In the null conditions, the exact expressions of the expectation and variance performed perfectly (as expected). The approximate expectation for arbitrary conditions derived above yielded substantially smaller values than the empirical means when N is small; however, the approximation worked satisfactorily with moderate N (≥ 16–32), with the deviations from empirical means mostly falling within 2 standard error units. In addition, the approximate expectation worked rather well, even with small N, under either A) the q-large λ conditions with $q = 2$ and $V_{\mathrm{rel}}(\mathbf{\Sigma})$ = 0.4, B) same with $q = 4$, or C) linearly and quadratically decreasing λ conditions with moderately large p (≥ 16). Other conditions held constant, the accuracy of the approximate expectation in absolute scale tended to slightly improve with increasing p, effectively balancing with the decreasing sampling dispersion, so that the relative bias in standard error unit remains almost invariant across varying p. The approximate variance for arbitrary conditions derived above yielded substantially larger values than the empirical variance, except under the q-large λ conditions with $q = 1$ and $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.8$ where it yielded smaller values. Even with the moderately large N of 64, the expression yielded values inaccurate by ~5% in the scale of standard deviation (SD scale hereafter), except under the linearly and quadratically decreasing λ conditions with moderately large p (≥ 16), where it worked largely satisfactorily.

Results for $V_{\text{rel}}(\mathbf{R})$ are summarized in Figures 7, S6, S9, and S10 and Tables 3 and S3. Distributions were unimodal within the range (0, 1), except when $p = 2$ and $N \leq 8$ where an additional peak is usually present near 0. The overall response to varying p and N is largely similar to that of $V_{\text{rel}}(\mathbf{S})$, although the shape of distribution was substantially different for small N. As expected from the theoretical expectations noted above, $V_{\text{rel}}(\mathbf{R})$ tends to overestimate the population value $V_{\text{rel}}(\mathbf{P})$ when the latter is small but tends to underestimate it when $V_{\text{rel}}(\mathbf{P}) = 0.8$. The expressions of expectation for the null and arbitrary conditions and variance for the null condition derived above showed almost perfect match with the empirical means and variances (as expected). The asymptotic variance for arbitrary conditions with $p > 2$ derived above behaved somewhat idiosyncratically. It yielded larger values than the empirical variances under A) the q-large λ conditions with $q = 1$ and $V_{\text{rel}}(\mathbf{P}) = 0.1$– 0.6, B) same with $q = 2$ and $V_{\text{rel}}(\mathbf{P}) = 0.1$– 0.2 except when $p = 4$, and C) the quadratically decreasing λ conditions with $p = 4$; whereas it yielded smaller values under a) the q-large λ conditions with $q = 1$ and $V_{\text{rel}}(\mathbf{P}) = 0.8$, b) same with $q = 2$ and $V_{\text{rel}}(\mathbf{P}) = 0.4$, c) same with $q = 4$, d) same with $q = 2$ and $p = 4$, e) the linearly decreasing λ conditions, and f) the quadratically decreasing λ conditions except when $p = 4$. This latter underestimation of sampling dispersion seems to happen when the smallest population eigenvalue is smaller than ~0.125, although this is not true for the case d. In all cases, the accuracy of the asymptotic expression tends to improve with increasing N. Relative error decreases to 3–10% in SD scale with large N (≥ 64) under the q-large λ conditions with $q = 1$ (all cases), $q = 2$ and $V_{\text{rel}}(\mathbf{P}) = 0.1$– 0.2, or $q = 4$ and $V_{\text{rel}}(\mathbf{P}) = 0.1$. However, in other conditions, the relative error can be extremely large (10–300% in SD scale with $N = 256$), especially when the smallest population eigenvalue is small (<0.1).

Discussion

Eigenvalue dispersion indices can be calculated for covariance or correlation matrices in similar ways, but implications are rather different. On the one hand, the relative eigenvalue variance of a sample covariance matrix $V_{\mathrm{rel}}(\mathbf{S})$ is a test statistic for the sphericity (John, 1972; Sugiura, 1972; Nagao, 1973), and is thus interpreted as a measure of eccentricity of variation, be it due to large variation of a single trait or covariation between traits. Interpretation of the unstandardized eigenvalue variance of a sample covariance matrix $V(\mathbf{S})$ is less straightforward, but it can potentially be useful in comparing eccentricity between samples when the sensitivity to overall scaling is not of concern, primarily for the presence of an unbiased estimator of the corresponding population value with a known variance (eq. 39). On the other hand, the relative eigenvalue variance of a sample correlation matrix $V_{\mathrm{rel}}(\mathbf{R})$ is identical to the average of the squared correlation coefficients across all pairs of traits (Durand & Le Roux, 2017; see above). The average squared correlation is another commonly used index of phenotypic integration (e.g., Cheverud et al., 1983), but its equivalence to $V_{\mathrm{rel}}(\mathbf{R})$ seems to have been overlooked, apart from an empirical confirmation by Haber's (2011) simulations. Obviously, the choice between covariance and correlation should be made according to the scope of individual analyses (Klingenberg, 1996; Hansen & Houle, 2008; Pavlicev et al., 2009; Goswami & Polly, 2010; see also Machado et al., 2019 for an interesting discussion). Usual caveats for the choice between covariance and correlation is also pertinent here (Jolliffe, 2002): covariance between traits have clear interpretability only if all traits are in the same unit and dimension. These are despite that $V_{\mathrm{rel}}(\mathbf{S})$ is dimensionless and independent of the overall scaling of traits.

Perhaps the most remarkable finding of this study is that the distributions of $V_{\mathrm{rel}}(\mathbf{S})$ and $V_{\mathrm{rel}}(\mathbf{R})$ do not seem to vary much with the number of variables p itself. The above expressions for the (approximate) mean and variance can be readily calculated for any p, and simulation results indicate that their accuracy are not compromised by large p (Figs. 5–7 and S2–S8; Tables 1–3 and S1–S3). This finding highlights potential applicability of these measures to high-dimensional phenotypic data. Nevertheless, it should be remembered that, when p exceeds the degree of freedom n, $p - n$ of the sample eigenvalues are 0 and hence the corresponding population eigenvalues are not estimable. In addition, the first sample eigenvector tends to be consistently diverged from the first population eigenvector in high-dimensional settings (Johnstone & Paul, 2018).

Applications and limitations

The present analytic results assume simple independent sampling from a multivariate normal population and the Wishart-ness of the cross-product matrix. For some biological datasets, certain modifications would be required. A simple example is data consisting of multiple groups with potentially heterogeneous means, e.g., intraspecific variation calculated from multiple geographic populations or sexes. If uniform $\mathbf{\Sigma}$ across groups can be assumed, cross-product matrices from the data centered at the respective group's sample mean can be summed across groups to obtain a pooled cross-product matrix, which is, by the additivity of Wishart variables, distributed as $W_p(\mathbf{\Sigma}, N - g)$, where N is the total sample size and g is the number of groups. That is, all above expressions can be applied by simply using the degree of freedom $N - g$. A similar correction is required when eigenvalue dispersion indices are applied to partial correlation matrices (Torices & Méndez, 2014; Torices & Muñoz-Pajares, 2015). The distribution of sample partial correlation coefficients in p_1 variables conditionalized on p_2 other variables based on N observations is the same as that of ordinary correlation coefficients based on $N - p_2$ observations with the same corresponding parameters (e.g., Anderson, 2003: p. 143), so the appropriate degree of freedom is $n - p_2$.

These procedures are essentially to examine the covariance/correlation matrix of residuals after conditionalizing on covariates.

Present analytical results may not be applicable to those empirical covariance or correlation matrices that are not based on a Wishart matrix. Primary examples are the empirical **G** matrices estimated from variance components in MANOVA designs or as (restricted) maximum likelihood estimators in mixed models (e.g., Lynch & Walsh, 1998). Mean-standardization, a method recommended for analyzing **G** matrices (e.g., Houle, 1992; Hereford et al., 2004; Hansen & Houle, 2008; Haber, 2016), can also violate the distributional assumption if sample means are used in the standardization. If eigenvalue dispersion indices are to be used with any of these methods, their sampling properties need to be critically assessed.

The assumption of multivariate normality may be intrinsically inappropriate for some types of data. Examples include meristic (count) data, compositional or proportional data, angles, and directional data. Application of eigenvalue dispersion indices (or indeed covariance/correlation itself) to such data types would require special treatments, which are beyond the scope of this paper. Needless to say, the appropriateness of multivariate normality should be critically assessed in every empirical dataset when the present analytic results are to be applied, even if the data type is conformable with normality. Robustness of the above results against non-normality may deserve some investigations.

Shape variables

The application to traditional morphometric datasets, in which all variables are typically measured in the same unit, is rather straightforward, as covariance/correlation in such variables has full interpretability in the Euclidean trait space. Quite often, component(s) of little interest, e.g., size, are removed by transforming raw data, inducing covariation in resultant variables that needs to be taken into account in hypothesis tests. The most typical transformation is the division by an isometric or allometric size variable (Jolicoeur, 1963; Mosimann, 1970; Mosimann & James, 1979; Darroch & Mosimann, 1985; Klingenberg, 1996, 2016), which can conveniently be done by orthogonal projection in the space of log-transformed variables. The projection of objects onto the hyperplane orthogonal to a subspace, say, the column space of $\mathbf{H}$ ($p \times k$ full-column-rank matrix; for the isometric size vector, $\mathbf{H} = p^{-1/2}\mathbf{1}_p$), can be done by right-multiplying the data by the projection matrix (e.g., Burnaby, 1966):

$$\mathbf{I}_p - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T.$$

Therefore, the covariance matrix in the resultant space can be obtained from that in the original space $\mathbf{\Sigma}$ as

$$\left[\mathbf{I}_p - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right]\mathbf{\Sigma}\left[\mathbf{I}_p - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right].$$

Under the null condition ($\mathbf{\Sigma} = \sigma^2\mathbf{I}_p$) specifically, this becomes

$$\sigma^2\left[\mathbf{I}_p - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right],$$

because the projection matrix is symmetric and idempotent. This transformation renders k eigenvalues to be 0 by construction. When the focus is on covariance rather than correlation, these null eigenvalues can optionally be dropped from calculation of eigenvalue mean and dispersion, so that the resultant dispersion index quantifies eccentricity of variation in the subspace of interest. Theories derived above can be applied with minimal modifications, although the asymptotic variance of $V_{\mathrm{rel}}(\mathbf{R})$ may not work well due to the singularity. These discussions assume that independence between the raw variables can at least hypothetically be conceived, e.g., when measurements are taken from non-overlapping parts of an organism. If measurements are taken from overlapping parts of an organism, then there will be dependence between variables due to the geometric configuration, which needs to be taken into consideration on case-by-case basis (Mitteroecker et al., 2012). It may even be inappropriate to assume multivariate normality for this last type of data.

Application to landmark-based geometric morphometric data is more complicated, primarily because the shape space of Procrustes-aligned landmark configurations is (typically a restricted region of) the surface of a hyper(hemi)sphere. In practice, however, empirical analyses are usually conducted on a Euclidean tangent space instead of the shape space itself, assuming that the former gives a satisfactory metric approximation of the latter (e.g., Rohlf, 1999; Marcus et al., 2000; Klingenberg, 2020). It will in principle be possible to obtain an approximate population covariance matrix of landmark coordinates in this tangent space from that of raw landmark coordinates before alignment, by using the orthogonal projection method mentioned above with such an $\mathbf{H}$ whose columns represent the non-shape components. Such a set of vectors can be obtained either as a basis of the complement of the tangent space (see Rohlf & Bookstein, 2003) or directly from the consensus configuration (Klingenberg, 2020). The stereographic projection might potentially be preferred over the orthogonal projection in projecting aligned empirical configurations in the shape space to the tangent space—not to be confused with the projection from the raw space to the tangent space—for purposes of analysing eccentricity of variation. This is because the stereographic projection tends to approximately preserve multivariate normality of the raw coordinates into the resultant tangent space, provided that the variation in the raw coordinates is sufficiently small and that the mean configuration is taken as the point of tangency (Rohlf, 1999). It should be noted that Procrustes superimposition changes perceived patterns of variation in landmark coordinates, often rather drastically (Rohlf & Slice, 1990; Walker, 2000). Such phenomena are probably to be seen as properties of shape variables, rather than necessarily nuisance artefacts (Klingenberg, 2021). Whether these can be of concern or not would depend on the scope of individual analyses (see also Machado et al., 2019).

Phylogenetic data

So far data were assumed to be i.i.d. multivariate normal variables. Important applications in evolutionary biology involve non-i.i.d. observations, most notably phylogenetically structured data in which N observations (typically species) have covariance due to shared evolutionary histories. Trait covariation at the interspecific level may have interpretations under certain microevolutionary models (Felsenstein, 1988; Hansen & Martins, 1996; Revell & Harmon, 2008; Uyeda & Harmon, 2014; Caetano & Harmon, 2019). Under the assumption that trait evolution along phylogeny can be described by (potentially a mixture of) linear invariant Gaussian models, such as the Brownian motion (BM), accelerating–decelerating (ACDC; or early burst), and Ornstein–Uhlenbeck (OU) processes, the joint distribution of the observations is known to be multivariate normal (Hansen & Martins, 1996; Manceau et al., 2017; Mitov et al., 2020). A brief overview is given below for potential applications of the present analytic results to phylogenetically structured data.

For BM and its modifications, including BM with a trend, Pagel's λ, and ACDC models, the covariance matrix of the $N \times p$ dimensional data $\mathbf{X}$ can be factorized into the intertrait and interspecific components in the form of Kronecker product: $\mathbf{\Sigma} \otimes \mathbf{\Psi}$, where $\mathbf{\Psi}$ is the $N \times N$ interspecific covariance matrix specified by the underlying phylogeny and parameter(s) specific to the evolutionary model (see Hansen & Martins, 1996; Freckleton et al., 2002; Blomberg et al., 2003; Clavel et al., 2015; Mitov et al., 2020). In this case the data can conveniently be considered as a matrix-variate normal variable (see Gupta & Nagar, 1999): $\mathbf{X} \sim N_{N,p}(\mathbf{M}, \mathbf{\Sigma} \otimes \mathbf{\Psi})$, where $\mathbf{M}$ is a $N \times p$ matrix of means. If $\mathbf{\Psi}$ is known a priori—that is, we have an accurate phylogenetic hypothesis and parameters— the change of variables $\mathbf{Y} = \mathbf{\Psi}^{-1/2}\mathbf{X}$ leads to $\mathbf{Y} \sim N_{N,p}(\mathbf{\Psi}^{-1/2}\mathbf{M}, \mathbf{\Sigma} \otimes \mathbf{I}_N)$, thereby essentially avoiding the complication of dependence between observations. This procedure is widely recognized as the (phylogenetic) generalized least squares (GLS; e.g., Grafen, 1989; Martins & Hansen, 1997; Rohlf, 2001; Symonds & Blomberg, 2014). If we know the population mean $\mathbf{M}$ in addition, then the cross product matrix centered at it,

$$\left(\mathbf{Y} - \mathbf{\Psi}^{-1/2}\mathbf{M}\right)^T\left(\mathbf{Y} - \mathbf{\Psi}^{-1/2}\mathbf{M}\right) = (\mathbf{X} - \mathbf{M})^T\mathbf{\Psi}^{-1}(\mathbf{X} - \mathbf{M}),$$

is distributed as $W_p(\mathbf{\Sigma}, N)$. If we don't exactly know $\mathbf{M}$ yet still assume $\mathbf{M} = \mathbf{1}_N\boldsymbol{\mu}^T$ with the unknown but uniform $p \times 1$ mean vector $\boldsymbol{\mu}$, then the GLS estimate of the mean $\hat{\boldsymbol{\mu}} = (\mathbf{1}_N^T\mathbf{\Psi}^{-1}\mathbf{1}_N)^{-1}\mathbf{1}_N^T\mathbf{\Psi}^{-1}\mathbf{X}$ (e.g., Martins & Hansen, 1997) can be used to obtain a sample-mean-centered cross-product matrix

$$\left(\mathbf{Y} - \mathbf{\Psi}^{-1/2}\mathbf{1}_N\hat{\boldsymbol{\mu}}^T\right)^T\left(\mathbf{Y} - \mathbf{\Psi}^{-1/2}\mathbf{1}_N\hat{\boldsymbol{\mu}}^T\right) = (\mathbf{X} - \mathbf{1}_N\hat{\boldsymbol{\mu}}^T)^T\mathbf{\Psi}^{-1}(\mathbf{X} - \mathbf{1}_N\hat{\boldsymbol{\mu}}^T),$$

which can be shown to be distributed as $W_p(\mathbf{\Sigma}, N - 1)$. If there are multiple blocks of species with different means (regimes), then cross-product matrices calculated separately for each of these can be summed to obtained a Wishart matrix with a modified degree of freedom as mentioned above, although it would naturally be asked first whether those regimes share the same $\mathbf{\Sigma}$ (Revell & Collar, 2009; Caetano & Harmon, 2019). Above analytic results can directly be applied to these Wishart matrices. Estimation of $\mathbf{\Sigma}$ based on this transformation has previously been devised (Revell & Harmon, 2008; see also Huelsenbeck & Rannala, 2003, Revell & Harrison, 2008; Adams & Felice, 2014), and has been shown to have superior accuracy in estimating eigenvalues and eigenvectors over estimation ignoring phylogenetic structure under model conditions (Revell, 2009). Variants of this method have already been applied to analyze eccentricity of interspecific covariation (Haber, 2016; Watanabe, 2018). In practice, however, $\mathbf{\Psi}$ is virtually never known accurately because phylogeny and parameters of evolutionary models are generally estimated with error, so empirical cross-product matrices may not be strictly Wishart. This source of error is inherent to any phylogenetic comparative analysis. Unlike the GLS estimate of the mean, which remains unbiased even when $\mathbf{\Psi}$ is misspecified, the GLS estimate of trait (co)variance is in general biased in this case (see Rohlf, 2006). Although there are certain ways to incorporate phylogenetic uncertainty into statistical inferences (e.g., Huelsenbeck & Rannala, 2003; Garamszegi & Mundry, 2014; Nakagawa & de Villemereuil, 2019), potential consequences of the uncertainty over the distributions of derived statistics require further investigation (see also Revell et al., 2018). Nevertheless, the GLS estimation with slightly inaccurate $\mathbf{\Psi}$ is supposed to yield a better estimate of trait (co)variance than the estimation ignoring phylogenetic covariation altogether (Rohlf, 2006). It should be noted that uniform scaling of $\mathbf{\Psi}$ translates to the reciprocal scaling of the cross-product matrix; $V(\mathbf{S})$ is sensitive to this scaling, whereas $V_{\mathrm{rel}}(\mathbf{S})$ and $V_{\mathrm{rel}}(\mathbf{R})$ are not. Therefore, specifically under the BM model, the only major concern for the latter two indices would be the phylogenetic uncertainty.

Unfortunately, the GLS estimation of trait covariance does not seem feasible for multivariate OU models, where the joint covariance matrix cannot in general be factorized into intertrait and interspecific components (Bartoszek et al., 2012; Mitov et al., 2020). This is notably except when the selection strength matrix is spherical and the tree is ultrametric, in which case a factorization of the form $\mathbf{\Sigma} \otimes \mathbf{\Psi}$ is possible (the scalar OU model; Bastide et al., 2018) and hence the GLS cross-product matrix can in principle be calculated, assuming that the relevant parameters are known. Otherwise, the random drift/diffusion matrix of the OU model estimated in one or other criteria can potentially be analyzed, although little is known about its sampling properties under various implementations, other than that accurate estimation is notoriously difficult (e.g., Ho & Ané, 2014; Clavel et al., 2015). Further studies are required on technical aspects of quantifying trait covariation in phylogenetically structured data under such complex models, as well as its biological implications (e.g., Adams & Collyer, 2018, 2019b; Mitov et al., 2019, 2020; Clavel et al., 2019; Clavel & Morlon, 2020).

Concluding remarks

Eigenvalue dispersion indices of covariance or correlation matrices are commonly used as measures of trait covariation, but their statistical implications have not been well appreciated, against which criticism has reasonably been directed (e.g., Hansen & Houle, 2008; Hansen et al., 2019). As discussed above, $V_{\mathrm{rel}}(\mathbf{S})$ and $V_{\mathrm{rel}}(\mathbf{R})$ have clear statistical justifications as test statistics for sphericity and no correlation, respectively. However, sample eigenvalue dispersion indices are biased estimators of the corresponding population values. This paper derived (or restated) exact and approximate expressions for the expectation and variance of $V(\mathbf{S})$, $V_{\mathrm{rel}}(\mathbf{S})$, and $V_{\mathrm{rel}}(\mathbf{R})$ under the respective null and arbitrary conditions, with which empirical values can be compared. All null moments derived are exact, as well as both moments of $V(\mathbf{S})$ and the expectation of $V_{\mathrm{rel}}(\mathbf{R})$ under arbitrary conditions. Moments of $V_{\mathrm{rel}}(\mathbf{S})$ under arbitrary conditions are approximations based on the delta method; the approximate expectation was shown to work reasonably well with a moderate sample size ($N \geq 16$– 32), whereas the approximate variance requires a larger sample size (e.g., $N \geq 64$, depending on other conditions). The variance of $V_{\mathrm{rel}}(\mathbf{R})$ under arbitrary conditions is asymptotic, and was seen to work well with a relatively large sample size ($N \geq 64$) in some conditions, but not so well in others even with an extremely large sample size. Under such conditions where these expressions work, they can be used for (approximate) statistical inferences and tests for arbitrary covariance/correlation structures, as well as for determination of appropriate sample size in empirical analyses, essentially replacing qualitative thresholds proposed earlier (e.g., Haber, 2011; Jung et al., 2020).

There are several conceivable ways for statistical inferences and hypothesis testing for eigenvalue dispersion indices. When sample size is so large that distributions of the indices are virtually symmetric ($N \geq 16$– 128, depending on other conditions), the moments derived above may potentially be used to construct approximate confidence intervals. If multivariate normality (or any other explicit distribution) can be assumed, then it is straightforward to obtain empirical distributions under appropriate null or alternative conditions with Monte Carlo simulations. Critical points of the null distributions and empirical power at $\alpha = 0.05$ and 0.01 based on the present simulations are presented in Table S1–S3 as a quick guide for sampling design. Several limiting and approximate distributions have been proposed for related statistics (e.g., John, 1972; Nagao, 1973; Ledoit & Wolf, 2002; Schott, 2005), which could be used for simple null hypothesis testing with large N. Resampling-based tests are another potential avenue of development. Applicability and performance of these alternative methods would deserve further investigations.

Acknowledgements

The author would like to thank Carmelo Fruciano and Christian P. Klingenberg for encouragements and constructive comments in an early stage of the study. This work was partly supported by the Newton International Fellowships by the Royal Society (NIF\R1\180520) and the Overseas Research Fellowships by the Japan Society for the Promotion of Science.

References

Adams DC. 2016. Evaluating modularity in morphometric data: challenges with the RV coefficient and a new test measure. *Methods in Ecology and Evolution* **7:** 565–572. https://doi.org/10.1111/2041-210X.12511.

Adams DC, Collyer ML. 2018. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. *Systematic Biology* **67:** 14–31. https://doi.org/10.1093/sysbio/syx055.

Adams DC, Collyer ML. 2019a. Comparing the strength of modular signal, and evaluating alternative modular hypotheses, using covariance ratio effect sizes with morphometric data. *Evolution* **73:** 2352–2367. https://doi.org/10.1111/evo.13867.

Adams DC, Collyer ML. 2019b. Phylogenetic comparative methods and the evolution of multivariate phenotypes. *Annual Review of Ecology, Evolution, and Systematics* **50:** 405–425. https://doi.org/10.1146/annurev-ecolsys-110218-024555.

Adams DC, Felice RN. 2014. Assessing trait covariation and morphological integration on phylogenies using evolutionary covariance matrices. *PLoS ONE* **9:** e94335. https://doi.org/10.1371/journal.pone.0094335.

Agrawal AF, Stinchcombe JR. 2009. How much do genetic covariances alter the rate of adaptation? *Proceedings of the Royal Society B: Biological Sciences* **276:** 1183–1191. https://doi.org/10.1098/rspb.2008.1671.

Anderson TW. 1963. Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics* **34:** 122–148. https://doi.org/10.1214/aoms/1177704248.

Anderson TW. 2003. *An Introduction to Multivariate Statistical Analysis*, 3rd edn. Hoboken, New Jersey: John Wiley & Sons.

Arlegi M, Veschambre-Couture C, Gómez-Olivencia A. 2020. Evolutionary selection and morphological integration in the vertebral column of modern humans. *American Journal of Physical Anthropology* **171:** 17–36. https://doi.org/10.1002/ajpa.23950.

Armbruster WS, Hansen TF, Pélabon C, Pérez-Barrales R, Maad J. 2009. The adaptive accuracy of flowers: measurement and microevolutionary patterns. *Annals of Botany* **103:** 1529–1545. https://doi.org/10.1093/aob/mcp095.

Armbruster WS, Pélabon C, Bolstad GH, Hansen TF. 2014. Integrated phenotypes: understanding trait covariation in plants and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369:** 20130245. https://doi.org/10.1098/rstb.2013.0245.

Bartoszek K, Pienaar J, Mostad P, Andersson S, Hansen TF. 2012. A phylogenetic comparative method for studying multivariate adaptation. *Journal of Theoretical Biology* **314:** 204–215. https://doi.org/10.1016/j.jtbi.2012.08.005.

Bastide P, Ané C, Robin S, Mariadassou M. 2018. Inference of adaptive shifts for multivariate correlated traits. *Systematic Biology* **67:** 662–680. https://doi.org/10.1093/sysbio/syy005.

Blomberg SP, Garland T Jr, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57:** 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x.

Brommer JE. 2014. Using average autonomy to test whether behavioral syndromes constrain evolution. *Behavioral Ecology and Sociobiology* **68:** 691–700. https://doi.org/10.1007/s00265-014-1699-6.

Burnaby TP. 1966. Growth-invariant discriminant functions and generalized distances. *Biometrics* **22:** 96–110. https://doi.org/10.2307/2528217.

Caetano DS, Harmon LJ. 2019. Estimating correlated rates of trait evolution with uncertainty. *Systematic Biology* **68:** 412–429. https://doi.org/10.1093/sysbio/syy067.

Cane WP. 1993. The ontogeny of postcranial integration in the common tern, *Sterna hirundo*. *Evolution* **47:** 1138–1151. https://doi.org/10.1111/j.1558-5646.1993.tb02141.x.

Chenoweth SF, Rundle HD, Blows MW. 2010. The contribution of selection and genetic constraints to phenotypic divergence. *American Naturalist* **175:** 186–196. https://doi.org/10.1086/649594.

Cheverud JM. 1982. Phenotypic, genetic, and environmental morphological integration in the cranium. *Evolution* **36:** 499–516. https://doi.org/10.1111/j.1558-5646.1982.tb05070.x.

Cheverud JM. 1988. A comparison of genetic and phenotypic correlations. *Evolution* **42:** 958–968. https://doi.org/10.1111/j.1558-5646.1988.tb02514.x.

Cheverud JM. 1996. Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. *Journal of Evolutionary Biology* **9:** 5–42. https://doi.org/10.1046/j.1420-9101.1996.9010005.x.

Cheverud JM, Rutledge JJ, Atchley WR. 1983. Quantitative genetics of development: genetic correlations among age-specific trait values and the evolution of ontogeny. *Evolution* **37:** 895–905. https://doi.org/10.1111/j.1558-5646.1983.tb05619.x.

Cheverud JM, Wagner GP, Dow MM. 1989. Methods for the comparative analysis of variation patterns. *Systematic Zoology* **38:** 201–213. https://doi.org/10.2307/2992282.

Clavel J, Aristide L, Morlon H. 2019. A penalized likelihood framework for high-dimensional phylogenetic comparative methods and an application to New-World monkeys brain evolution. *Systematic Biology* **68:** 93–116. https://doi.org/10.1093/sysbio/syy045.

Clavel J, Escarguel G, Merceron G. 2015. mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data. *Methods in Ecology and Evolution* **6:** 1311–1319. https://doi.org/10.1111/2041-210X.12420.

Clavel J, Morlon H. 2020. Reliable phylogenetic regressions for multivariate comparative data: illustration with the MANOVA and application to the effect of diet and mandible morphology in phyllostomid bats. *Systematic Biology* **69:** 927–943. https://doi.org/10.1093/sysbio/syaa010.

Constantine AG. 1963. Some non-central distribution problems in multivariate analysis. *Annals of Mathematical Statistics* **34:** 1270–1285. https://doi.org/10.1214/aoms/1177703863.

Cramer JS. 1987. Mean and variance of R^2 in small and moderate samples. *Journal of Econometrics* **35:** 253–266. https://doi.org/10.1016/0304-4076(87)90027-3.

Darroch JN, Mosimann JE. 1985. Canonical and principal components of shape. *Biometrika* **72:** 241–252. https://doi.org/10.1093/biomet/72.2.241.

Davies PI, Higham NJ. 2000. Numerically stable generation of correlation matrices and their factors. *BIT Numerical Mathematics* **40:** 640–651. https://doi.org/10.1023/A:1022384216930.

Dochtermann NA. 2011. Testing Cheverud's conjecture for behavioral correlations and behavioral syndromes. *Evolution* **65:** 1814–1820. https://doi.org/10.1111/j.1558-5646.2011.01264.x.

Durand J-L, Le Roux B. 2017. Linkage index of variables and its relationship with variance of eigenvalues in PCA and MCA. *Statistica Applicata – Italian Journal of Applied Statistics* **29:** 123–135. https://doi.org/10.26398/IJAS.0029-006.

Felice RN, Randau M, Goswami A. 2018. A fly in a tube: macroevolutionary expectations for integrated phenotypes. *Evolution* **72:** 2580–2594. https://doi.org/10.1111/evo.13608.

Felsenstein J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* **19:** 445–471. https://doi.org/10.1146/annurev.es.19.110188.002305.

Fornoni J, Ordano M, Boege K, Domínguez CA. 2009. Phenotypic integration: between zero and how much is too much. *New Phytologist* **183:** 248–250. https://doi.org/10.1111/j.1469-8137.2009.02911.x.

Freckleton RP, Harvey PH, Pagel M. 2002. Phylogenetic analysis of comparative data: a test and review of evidence. *American Naturalist* **160:** 712–726. https://doi.org/10.1086/343873.

Garamszegi LZ, Mundry R. 2014. Multimodel-inference in comparative analyses. In: Garamszegi LZ, ed. *Modern Phylogenetic Comparative Methods and Their Applications in Evolutionary Biology: Concepts and Practice*. Berlin: Springer, 305–331. https://doi.org/10.1007/978-3-662-43550-2_12.

Ghosh BK. 1966. Asymptotic expansions for the moments of the distribution of correlation coefficient. *Biometrika* **53:** 258–262. https://doi.org/10.2307/2334076.

Girshick MA. 1939. On the sampling theory of roots of determinantal equations. *Annals of Mathematical Statistics* **10:** 203–224. https://doi.org/10.1214/aoms/1177732180.

Gleason TC, Staelin R. 1975. A proposal for handling missing data. *Psychometrika* **40:** 229-252. https://doi.org/10.1007/BF02291569.

Goswami A. 2006. Morphological integration in the carnivoran skull. *Evolution* **60:** 169–183. https://doi.org/10.1111/j.0014-3820.2006.tb01091.x.

Goswami A, Binder WJ, Mearchen J, O'Keefe FR. 2015. The fossil record of phenotypic integration and modularity: a deep-time perspective on developmental and evolutionary dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **112:** 4891–4896. https://doi.org/10.1073/pnas.1403667112.

Goswami A, Finnarelli JA. 2016. EMMLi: a maximum likelihood approach to the analysis of modularity. *Evolution* **70:** 1622–1637. https://doi.org/10.1111/evo.12956.

Goswami A, Polly PD. 2010. Methods for studying morphological integration and modularity. In: Alroy J, Hunt G, eds. *Quantitative Methods in Paleobiology*. *Paleontological Society Papers* **16:** 213–243. https://doi.org/10.1017/S1089332600001881.

Grabowski M, Porto A. 2017. How many more? Sample size determination in studies of morphological integration and evolvability. *Methods in Ecology and Evolution* **8:** 592–603. https://doi.org/10.1111/2041-210X.12674.

Grafen A. 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **326:** 119–157. https://doi.org/10.1098/rstb.1989.0106.

Gupta AK, Nagar DK. 1999. *Matrix Variate Distributions*. Boca Raton, Florida: Chapman & Hall/CRC.

Haber A. 2011. A comparative analysis of integration indices. *Evolutionary Biology* **38:** 476–488. https://doi.org/10.1007/s11692-011-9137-4.

Haber A. 2015. The evolution of morphological integration in the ruminant skull. *Evolutionary Biology* **42:** 99–114. https://doi.org/10.1007/s11692-014-9302-7.

Haber A. 2016. Phenotypic covariation and morphological diversification in the ruminant skull. *American Naturalist* **187:** 576–591. https://doi.org/10.1086/685811.

Haber A, Dworkin I. 2017. Disintegrating the fly: a mutational perspective on phenotypic integration and covariation. *Evolution* **71:** 66–80. https://doi.org/10.1111/evo.13100.

Hallgrímsson B, Jamniczky H, Young NM, Rolian C, Parsons TE, Boughner JC, Marcucio RS. 2009. Deciphering the palimpsest: studying the relationships between morphological integration and phenotypic covariation. *Evolutionary Biology* **36:** 355–376. https://doi.org/10.1007/s11692-009-9076-5.

Hansen TF, Houle D. 2008. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* **21:** 1201–1219. https://doi.org/10.1111/j.1420-9101.2008.01573.x.

Hansen TF, Martins E. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* **50:** 1404–1417. https://doi.org/10.1111/j.1558-5646.1996.tb03914.x.

Hansen TF, Solvin TM, Pavlicev M. 2019. Predicting evolutionary potential: a numerical test of evolvability measures. *Evolution* **73:** 689–703. https://doi.org/10.1111/evo.13705.

Harder LD. 2009. Questions about floral (dis)integration. *New Phytologist* **183:** 247–248. https://doi.org/10.1111/j.1469-8137.2009.02881.x.

Hereford J, Hansen TF, Houle D. 2004. Comparing strengths of directional selection: how strong is strong? *Evolution* **58:** 2133-2143. https://doi.org/10.1111/j.0014-3820.2004.tb01592.x.

Ho LST, Ané C. 2014. Intrinsic inference difficulties for trait evolution with Ornstein–Uhlenbeck models. *Methods in Ecology and Evolution* **5:** 1133–1146. https://doi.org/10.1111/2041-210X.12285.

Houle D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics* **130:** 195–204. https://doi.org/10.1093/genetics/130.1.195.

Huelsenbeck JP, Rannala B. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution* **57:** 1237–1247. https://doi.org/10.1111/j.0014-3820.2003.tb00332.x.

John S. 1971. Some optimal multivariate tests. *Biometrika* **58:** 123–127. https://doi.org/10.1093/biomet/58.1.123.

John S. 1972. The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59:** 169–173. https://doi.org/10.1093/biomet/59.1.169.

Johnstone IM, Paul D. 2018. PCA in high dimensions: an orientation. *Proceedings of the IEEE* **106:** 1277–1292. https://doi.org/10.1109/JPROC.2018.2846730.

Jolicoeur P. 1963. The multivariate generalization of the allometry equation. *Biometrics* **19:** 497–499. https://doi.org/10.2307/2527939.

Jolliffe IT. 2002. *Principal Component Analysis*, 2nd edn. New York: Springer.

Jung H, Conaway MA, von Cramon-Taubadel N. 2020. Examination of sample size determination in integration studies based on the integration coefficient of variation (ICV). *Evolutionary Biology* **47:** 293–307. https://doi.org/10.1007/s11692-020-09514-w.

Kirkpatrick M. 2009. Patterns of quantitative genetic variation in multiple dimensions. *Genetica* **136:** 271–284. https://doi.org/10.1007/s10709-008-9302-6.

Klingenberg CP. 1996. Multivariate allometry. In: Marcus LF, ed. *Advances in Morphometrics*. New York: Plenum Press, 23–49. https://doi.org/10.1007/978-1-4757-9083-2_3.

Klingenberg CP. 2014. Studying morphological integration and modularity at multiple levels: concepts and analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369:** 20130249. https://doi.org/10.1098/rstb.2013.0249.

Klingenberg CP. 2016. Size, shape, and form: concepts of allometry in geometric morphometrics. *Development Genes and Evolution* **226:** 113–137. https://doi.org/10.1007/s00427-016-0539-2.

Klingenberg CP. 2020. Walking on Kendall's shape space: understanding shape spaces and their coordinate systems. *Evolutionary Biology* **47:** 334–352. https://doi.org/10.1007/s11692-020-09513-x.

Klingenberg CP. 2021. How exactly did the nose get that long? A critical rethinking of the Pinocchio effect and how shape changes relate to landmarks. *Evolutionary Biology* **48:** 115–127. https://doi.org/10.1007/s11692-020-09520-y.

Klingenberg CP, Duttke S, Whelan S, Kim M. 2012. Developmental plasticity, morphological variation and evolvability: a multilevel analysis of morphometric integration in the shape of compound leaves. *Journal of Evolutionary Biology* **25:** 115–129. https://doi.org/10.1111/j.1420-9101.2011.02410.x.

Konishi S. 1979. Asymptotic expansions for the distributions of statistics based on the sample correlation matrix in principal component analysis. *Hiroshima Mathematical Journal* **9:** 647–700. https://doi.org/10.32917/hmj/1206134750.

Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33:** 402–416. https://doi.org/10.1111/j.1558-5646.1979.tb04694.x.

Lande R, Arnold SJ. 1983. The measurement of selection on correlated characters. *Evolution* **37:** 1210–1226. https://doi.org/10.1111/j.1558-5646.1983.tb00236.x.

Lawley DN. 1956. Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika* **43:** 128–136. https://doi.org/10.2307/2333586.

Ledoit O, Wolf M. 2002. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Annals of Statistics* **30:** 1081–1102. https://doi.org/10.1214/aos/1031689018.

Lynch M, Walsh B. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, Massachusetts: Sinauer Associates.

Machado FA, Hubbe A, Melo D, Porto A, Marroig G. 2019. Measuring the magnitude of morphological integration: the effect of differences in morphometric representations and the inclusion of size. *Evolution* **73:** 2518–2528. https://doi.org/10.1111/evo.13864.

Manceau M, Lambert A, Morlon H. 2017. A unifying comparative phylogenetic framework including traits coevolving across interacting lineages. *Systematic Biology* **66:** 551–568. https://doi.org/10.1093/sysbio/syw115.

Marcus LF, Hingst-Zaher E, Zaher H. 2000. Application of landmark morphometrics to skulls representing the orders of living mammals. *Hystrix* **11:** 27–47. https://doi.org/10.4404/hystrix-11.1-4135.

Marroig G, Melo DAR, Garcia G. 2012. Modularity, noise, and natural selection. *Evolution* **66:** 1506–1524. https://doi.org/10.1111/j.1558-5646.2011.01555.x.

Marroig G, Shirai LT, Porto A, de Oliveira FB, De Conto V. 2009. The evolution of modularity in the mammalian skull II: evolutionary consequences. *Evolutionary Biology* **36:** 136–148. https://doi.org/10.1007/s11692-009-9051-1.

Martins E, Hansen TF. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* **149:** 646–667. https://doi.org/10.1086/286013.

Mitov V, Bartoszek K, Asimomitis G, Stadler T. 2020. Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. *Theoretical Population Biology* **131:** 66–78. https://doi.org/10.1016/j.tpb.2019.11.005.

Mitov V, Bartoszek K, Stadler T. 2019. Automatic generation of evolutionary hypotheses using mixed Gaussian phylogenetic models. *Proceedings of the National Academy of Sciences of the United States of America* **116:** 16921–16926. https://doi.org/10.1073/pnas.1813823116.

Mitteroecker P, Gunz P, Neubauer S, Müller G. 2012. How to explore morphological integration in human evolution and development? *Evolutionary Biology* **39:** 536–553. https://doi.org/10.1007/s11692-012-9178-3.

Mosimann JE. 1970. Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. *Journal of the American Statistical Association* **65:** 930–945. https://doi.org/10.1080/01621459.1970.10481136.

Mosimann JE, James FC. 1979. New statistical methods for allometry with application to Florida red-winged blackbirds. *Evolution* **33:** 444-459. https://doi.org/10.1111/j.1558-5646.1979.tb04697.x.

Muirhead RJ. 1982. *Aspects of Multivariate Statistical Theory*. Hoboken, New Jersey: John Wiley & Sons.

Nagao H. 1973. On some test criteria for covariance matrix. *Annals of Statistics* **1:** 700–709. https://doi.org/10.1214/aos/1176342464.

Nakagawa S, de Villemereuil P. 2019. A general method for simultaneously accounting for phylogenetic and species sampling uncertainty via Rubin's rules in comparative analysis. *Systematic Biology* **68:** 632–641. https://doi.org/10.1093/sysbio/syy089.

Olson EC, Miller RL. 1958. *Morphological Integration*. [Chicago]: University of Chicago Press.

Ordano M, Fornoni J, Boege K, Domínguez CA. 2008. The adaptive value of phenotypic floral integration. *New Phytologist* **179:** 1183–1192. https://doi.org/10.1111/j.1469-8137.2008.02523.x.

Pavlicev M, Cheverud JM, Wagner GP. 2009. Measuring morphological integration using eigenvalue variance. *Evolutionary Biology* **36:** 157–170. https://doi.org/10.1007/s11692-008-9042-7.

Pitchers W, Wolf JB, Tregenza T, Hunt J, Dworkin I. 2014. Evolutionary rates for multivariate traits: the role of selection and genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369:** 20130252. https://doi.org/10.1098/rstb.2013.0252.

Porto A, de Oliveira FB, Shirai LT, De Conto V, Marroig G. 2009. The evolution of modularity in the mammalian skull I: morphological integration patterns and magnitudes. *Evolutionary Biology* **36:** 118–135. https://doi.org/10.1007/s11692-008-9038-3.

R Core Team. 2019. *R: a language and environment for statistical computing, Version 3.5.3*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Renaud S, Auffray J-C. 2013. The direction of main phenotypic variance as a channel to morphological evolution: case studies in murine rodents. *Hystrix* **24:** 85–93. https://doi.org/10.4404/hystrix-24.1-6296.

Revell LJ. 2009. Size-correction and principal components for interspecific comparative studies. *Evolution* **63:** 3258–3268. https://doi.org/10.1111/j.1558-5646.2009.00804.x.

Revell LJ, Collar DC. 2009. Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution* **63:** 1090–1100. https://doi.org/10.1111/j.1558-5646.2009.00616.x.

Revell LJ, González-Valenzuela LE, Alfonso A, Castellanos-García LA, Guarnizo CE, Crawford AJ. 2018. Comparing evolutionary rates between trees, clades and traits. *Methods in Ecology and Evolution* **9:** 994–1005. https://doi.org/10.1111/2041-210X.12977.

Revell LJ, Harmon LJ. 2008. Testing quantitative genetic hypotheses about the evolutionary rate matrix for continuous characters. *Evolutionary Ecology Research* **10:** 311–331.

Revell LJ, Harrison AS. 2008. PCCA: a program for phylogenetic canonical correlation analysis. *Bioinformatics* **24:** 1018–1020. https://doi.org/10.1093/bioinformatics/btn065.

Roff DA. 1995. The estimation of genetic correlations from phenotypic correlations: a test of Cheverud's conjecture. *Heredity* **74:** 481–490. https://doi.org/10.1038/hdy.1995.68.

Rohlf FJ. 1999. Shape statistics: Procrustes superimpositions and tangent spaces. *Journal of Classification* **16:** 197–223. https://doi.org/10.1007/s003579900054.

Rohlf FJ. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution* **55:** 2143–2160. https://doi.org/10.1111/j.0014-3820.2001.tb00731.x.

Rohlf FJ. 2006. A comment on phylogenetic correction. *Evolution* **60:** 1509–1515. https://doi.org/10.1111/j.0014-3820.2006.tb01229.x.

Rohlf FJ, Bookstein FL. 2003. Computing the uniform component of shape variation. *Systematic Biology* **52:** 66–69. https://doi.org/10.1080/10635150390132759.

Rohlf FJ, Slice D. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* **39:** 40–59. https://doi.org/10.2307/2992207.

Schluter D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution* **50:** 1766–1774. https://doi.org/10.1111/j.1558-5646.1996.tb03563.x.

Schott JR. 2005. Testing for complete independence in high dimensions. *Biometrika* **92:** 951–956. https://doi.org/10.1093/biomet/92.4.951.

Shirai LT, Marroig G. 2010. Skull modularity in Neotropical marsupials and monkeys: size variation and evolutionary constraint and flexibility. *Journal of Experimental Zoology, Part B: Molecular and Developmental Evolution* **314B:** 663–683. https://doi.org/10.1002/jez.b.21367.

Sodini SM, Kemper KE, Wray NR, Trzaskowski M. 2018. Comparison of genotypic and phenotypic correlations: Cheverud’s conjecture in humans. *Genetics* **209:** 941–948. https://doi.org/10.1534/genetics.117.300630.

Srivastava MS, Khatri CG. 1979. *An Introduction to Multivariate Statistics*. New York: North Holland.

Srivastava MS, Yanagihara H. 2010. Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis* **101:** 1319–1329. https://doi.org/10.1016/j.jmva.2009.12.010.

Steppan SJ, Phillips PC, Houle D. 2002. Comparative quantitative genetics: evolution of the **G** matrix. *Trends in Ecology and Evolution* **17:** 320–327. https://doi.org/10.1016/S0169-5347(02)02505-3.

Stuart A, Ord JK. 1994. *Kendall’s Advanced Theory of Statistics*, 6th edn, Vol. 1. London: Hodder Education [Reprinted in 2004 by John Wiley & Sons, Chichester].

Sugiura N. 1972. Locally best invariant test for sphericity and the limiting distributions. *Annals of Mathematical Statistics* **43:** 1312–1316. https://doi.org/10.1214/aoms/1177692481.

Symonds MRE, Blomberg SP. 2014. A primer on phylogenetic generalized least squares. In: Garamszegi LZ, ed. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*. Berlin: Springer, 105–130. https://doi.org/10.1007/978-3-662-43550-2_5.

Torices R, Méndez M. 2014. Resource allocation to inflorescence components is highly integrated despite differences between allocation currencies and sites. *International Journal of Plant Sciences* **175:** 713–723. https://doi.org/10.1086/676622.

Torices R, Muñoz-Pajares AJ. 2015. PHENIX: an R package to estimate a size-controlled phenotypic integration index. *Applications in Plant Sciences* **3:** 1400104. https://doi.org/10.3732/apps.1400104.

Uyeda JC, Harmon LJ. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic Biology* **63:** 902–918. https://doi.org/10.1093/sysbio/syu057.

Van Valen L. 1974. Multivariate structural statistics in natural history. *Journal of Theoretical Biology* **45:** 235–247. https://doi.org/10.1016/0022-5193(74)90053-8.

Van Valen L. 2005. The statistics of variation. In: Hallgrímsson B, Hall BK, eds. *Variation*. Amsterdam: Elsevier, 29–47. https://doi.org/10.1016/B978-012088777-4/50005-3.

de Waal DJ, Nel DG. 1973. On some expectations with respect to Wishart matrices. *South African Statistical Journal* **7:** 61–67.

Wagner GP. 1984. On the eigenvalue distribution of genetic and phenotypic dispersion matrices: evidence for a nonrandom organization of quantitative character variation. *Journal of Mathematical Biology* **21:** 77–95. https://doi.org/10.1007/BF00275224.

Walker JA. 2000. Ability of geometric morphometric methods to estimate a known covariance matrix. *Systematic Biology* **49:** 686–696. https://doi.org/10.1080/106351500750049770.

Waller NG. 2020. Generating correlation matrices with specified eigenvalues using the method of alternating projections. *American Statistician* **74:** 21–28. https://doi.org/10.1080/00031305.2017.1401960.

Watanabe J. 2018. Clade-specific evolutionary diversification along ontogenetic major axes in avian limb skeleton. *Evolution* **72:** 2632–2652. https://doi.org/10.1111/evo.13627.

Appendix A

In this part, relationships between eigenvalue dispersion indices and individual eigenvalues are derived under certain restrictive conditions, in order to facilitate interpretation and to clarify algorithms used in simulations. For simplicity, it is assumed $\bar{\lambda} = 1$ in the following discussions; general cases easily follow by scaling.

Let us first consider the simple conditions where the first q ($< p$) population eigenvalues are equally large and the rest $p - q$ eigenvalues are equally small: $\lambda_1 = \cdots =$

$\lambda_q \geq \lambda_{q+1} = \cdots = \lambda_p$ ("q-large λ conditions" in simulations). By noting $\sum \lambda_i = q\lambda_1 + (p-q)\lambda_p = p$, it is seen that

$$V_{\text{rel}}(\mathbf{\Sigma}) = \frac{\sum_{i=1}^{p}(\lambda_i - 1)^2}{p(p-1)} = \frac{(p-q)}{q(p-1)}\left(1-\lambda_p\right)^2, \quad \text{(A1)}$$

and hence

$$\lambda_1 = 1 + \sqrt{\frac{(p-1)(p-q)}{q}V_{\text{rel}}},$$

$$\lambda_p = 1 - \sqrt{\frac{q(p-1)}{p-q}V_{\text{rel}}}. \quad \text{(A2)}$$

By noting the constraint $0 \leq \lambda_p \leq 1$, an upper limit of V_{rel} can be seen from equation A1:

$$V_{\text{rel}}(\mathbf{\Sigma}) \leq \frac{(p-q)}{q(p-1)} = \frac{1}{q} - \frac{q-1}{q(p-1)}. \quad \text{(A3)}$$

It is then obvious that, under these constraints, a value of V_{rel} greater than 0.5 cannot happen when $q > 1$; that is, such a large value implies the dominance of a single component of variance. The same arguments equally apply to correlation matrices.

When $q = 1$ for the correlation matrix, $V_{\text{rel}}(\mathbf{P})$ completely specifies the magnitude of correlation in every pair of variables. This point can be seen from the definition of eigendecomposition:

$$\rho_{ij} = \sum_{k=1}^{p} \lambda_k \upsilon_{ik}\upsilon_{jk}$$

$$= (\lambda_1 - \lambda_p)\upsilon_{i1}\upsilon_{j1} + \lambda_p \sum_{k=1}^{p} \upsilon_{ik}\upsilon_{jk}, \quad \text{(A4)}$$

where the (i,j)-th element of the eigenvector matrix denoted as υ_{ij}. Remember that $\sum_{k=1}^{p} \upsilon_{ik}\upsilon_{jk} = \delta_{ij}$, the Kronecker delta. Then, by noting equation A2 with $q = 1$,

$$1 = \rho_{ii} = (\lambda_1 - \lambda_p)\upsilon_{i1}^2 + \lambda_p = (p\upsilon_{i1}^2 - 1)\sqrt{V_{\mathrm{rel}}} + 1, \tag{A5}$$

therefore $\upsilon_{i1}^2 = p^{-1/2}$ for any i (that is, the coefficients of the first eigenvector are equal in magnitude). Finally, we have

$$\rho_{ij}^2 = (\lambda_1 - \lambda_p)^2 \upsilon_{i1}^2 \upsilon_{j1}^2 = V_{\mathrm{rel}} \tag{A6}$$

for any combination of i and j ($i \neq j$); the magnitude of correlation is identical across all pairs. Taken differently, $\lambda_2 = \cdots = \lambda_p = |\rho|$. These relationships have previously been noted by Anderson (1963) and Pavlicev et al. (2009).

The population eigenvalues of the linearly and quadratically decreasing λ conditions used in simulations are defined as $\lambda_i = (p - i + 1)\lambda_p$ and $\lambda_i = (p - i + 1)^2\lambda_p$ ($i = 1, 2, \dots, p$) for linearly and quadratically decreasing conditions, respectively. Under the assumption of a constant average eigenvalue, it is a simple matter of algebra to obtain the actual values of λ_p and $V_{\mathrm{rel}}(\mathbf{\Sigma})$, which are simple functions of p. The latter equals $1/3(p + 1)$ and $(8p + 11)/5(p + 1)(2p + 1)$ for the linearly and quadratically decreasing λ conditions, respectively.

Appendix B

In this part, the first two moments of $V(\mathbf{S})$ and $V_{\mathrm{rel}}(\mathbf{S})$ under the arbitrary $\mathbf{\Sigma}$ are derived, assuming multivariate normality. Derivation of the moments of the latter requires evaluation of moments of the ratio $\sum l_i^2 / (\sum l_i)^2 = \mathrm{tr}(\mathbf{A}^2)/(\mathrm{tr}\,\mathbf{A})^2$, which are not guaranteed to coincide with the ratio of moments except under the null hypothesis. Here we utilize the approximation based on the Taylor series expansion given in equation 32. In turn, we need $\mathrm{E}[\mathrm{tr}(\mathbf{A}^2)]$, $\mathrm{E}[(\mathrm{tr}\,\mathbf{A})^2]$,$\mathrm{Var}[\mathrm{tr}(\mathbf{A}^2)]$, $\mathrm{Var}[(\mathrm{tr}\,\mathbf{A})^2]$, and $\mathrm{Cov}[\mathrm{tr}(\mathbf{A}^2), (\mathrm{tr}\,\mathbf{A})^2]$.

We will follow Srivastava & Yanagihara's (2010) approach to obtain these moments. As in the text, let the $n \times p$ matrix $\mathbf{Z}$ be $(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)^T$, where $\mathbf{z}_i \sim N_p(\mathbf{0}_p, \mathbf{\Sigma})$ for $i =$

$1, 2, \dots, n$. Consider the cross-product matrix

$$\mathbf{A} = \mathbf{Z}^T\mathbf{Z}, \tag{B1}$$

such that $\mathbf{A} \sim W_p(\boldsymbol{\Sigma}, n)$. Let the spectral decomposition of $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \boldsymbol{\Upsilon}\boldsymbol{\Lambda}\boldsymbol{\Upsilon}^T, \tag{B2}$$

with the orthogonal matrix of eigenvectors $\boldsymbol{\Upsilon}$ and the diagonal matrix of eigenvalues $\boldsymbol{\Lambda}$. Let the $n \times p$ matrix $\mathbf{J}$ be $(\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_n)^T$, where $\mathbf{j}_i$ are i.i.d. $N_p(\mathbf{0}, \mathbf{I}_p)$, such that $\mathbf{Z} = \mathbf{J}\boldsymbol{\Sigma}^{1/2}$ with $\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Upsilon}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Upsilon}^T$. Then, it is possible to write

$$\mathbf{A} = \boldsymbol{\Sigma}^{1/2}\mathbf{J}^T\mathbf{J}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Upsilon}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Upsilon}^T\mathbf{J}^T\mathbf{J}\boldsymbol{\Upsilon}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Upsilon}^T = \boldsymbol{\Upsilon}\boldsymbol{\Lambda}^{1/2}\mathbf{V}^T\mathbf{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Upsilon}^T, \tag{B3}$$

where $\mathbf{V} = \mathbf{J}\boldsymbol{\Upsilon} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p)$ with $\mathbf{v}_i$ being i.i.d. $N_n(\mathbf{0}, \mathbf{I}_n)$. Furthermore, let $w_{ij} = \mathbf{v}_i^T\mathbf{v}_j$, such that w_{ii} are i.i.d. chi-square variables with n degrees of freedom χ_n^2. Obviously $w_{ij} = w_{ji}$. Note that

$$\operatorname{tr}\mathbf{A} = \operatorname{tr}\left(\boldsymbol{\Upsilon}\boldsymbol{\Lambda}^{1/2}\mathbf{V}^T\mathbf{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Upsilon}^T\right) = \operatorname{tr}(\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{V}) = \textstyle\sum_{i=1}^{n}\lambda_i w_{ii}. \tag{B4}$$

From well-known results on normal and chi-square moments, we have the following:

$$\mathrm{E}(w_{ii}^r) = n(n+2)\dots(n+2r-2), \qquad r = 1, 2, \dots,$$

$$\begin{aligned}
\mathrm{E}\left(w_{ii}^r w_{ij}^2\right) &= \mathrm{E}\left[\operatorname{tr}\left(w_{ii}^r\mathbf{v}_i\mathbf{v}_i^T\mathbf{v}_j\mathbf{v}_j^T\right)\right] \\
&= \operatorname{tr}\left[\mathrm{E}\left(w_{ii}^r\mathbf{v}_i\mathbf{v}_i^T\right)\mathrm{E}\left(\mathbf{v}_j\mathbf{v}_j^T\right)\right] \\
&= \operatorname{tr}\left[\mathrm{E}\left(w_{ii}^r\mathbf{v}_i\mathbf{v}_i^T\right)\mathbf{I}_n\right] \\
&= \mathrm{E}\left(w_{ii}^{r+1}\right) \\
&= n(n+2)\dots(n+2r), \qquad i \neq j, \qquad r = 0, 1, \dots,
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left(w_{ii}w_{jj}w_{ij}^2\right) &= \mathrm{E}\left[\operatorname{tr}\left(w_{ii}\mathbf{v}_i\mathbf{v}_i^T w_{jj}\mathbf{v}_j\mathbf{v}_j^T\right)\right] \\
&= \sum_{\alpha,\beta,\gamma,\delta}^{n}\mathrm{E}\left(v_{i\alpha}^2 v_{i\gamma}v_{i\delta}\right)\mathrm{E}\left(v_{j\beta}^2 v_{j\delta}v_{j\gamma}\right) \\
&= 9n + 6n(n-1) + n(n-1)^2 \\
&= n(n+2)^2, \qquad i \neq j,
\end{aligned}$$

$$\mathrm{E}\left(w_{ij}^2 w_{ik}^2\right) = \sum_{\alpha,\beta}^{n} \mathrm{E}\left(v_{i\alpha}^2 v_{i\beta}^2\right) \mathrm{E}\left(v_{j\alpha}^2\right) \mathrm{E}\left(v_{k\beta}^2\right)$$

$$= 3n + n(n-1)$$

$$= n(n+2), \qquad i \neq j \neq k,$$

$$\mathrm{E}\left(w_{ij}^4\right) = \sum_{\alpha,\beta,\gamma,\delta}^{n} \mathrm{E}\left(v_{i\alpha} v_{i\beta} v_{i\gamma} v_{i\delta}\right) \mathrm{E}\left(v_{j\alpha} v_{j\beta} v_{j\delta} v_{j\gamma}\right)$$

$$= 9n + 3n(n-1)$$

$$= 3n(n+2), \qquad i \neq j,$$

(B5)

where some intervening equations result from direct enumeration of the nonzero moments.

From the above expectations, one can evaluate the desired moments as follows:

$$\mathrm{E}[\mathrm{tr}(\mathbf{A}^2)] = \mathrm{E}[\mathrm{tr}(\mathbf{\Lambda} \mathbf{V}^T \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \mathbf{V})]$$

$$= \mathrm{E}\left(\sum_{i,j}^{n} \lambda_i \lambda_j w_{ij}^2\right)$$

$$= \sum_{i}^{n} \lambda_i^2 \,\mathrm{E}(w_{ii}^2) + \sum_{i \neq j}^{n} \lambda_i \lambda_j \,\mathrm{E}\left(w_{ij}^2\right)$$

$$= n(n+2) \sum_{i}^{n} \lambda_i^2 + n \sum_{i \neq j}^{n} \lambda_i \lambda_j\,,$$

$$\mathrm{E}[(\mathrm{tr}\,\mathbf{A})^2] = \mathrm{E}[[\mathrm{tr}(\mathbf{\Lambda} \mathbf{V}^T \mathbf{V})]^2]$$

$$= \mathrm{E}\left(\sum_{i,j}^{n} \lambda_i \lambda_j w_{ii} w_{jj}\right)$$

$$= \sum_{i}^{n} \lambda_i^2 \,\mathrm{E}(w_{ii}^2) + \sum_{i \neq j}^{n} \lambda_i \lambda_j \,\mathrm{E}(w_{ii}) \,\mathrm{E}\left(w_{jj}\right)$$

$$= n(n+2) \sum_{i}^{n} \lambda_i^2 + n^2 \sum_{i \neq j}^{n} \lambda_i \lambda_j\,,$$

$$\mathrm{E}[[\mathrm{tr}(\mathbf{A}^2)]^2] = \mathrm{E}[[\mathrm{tr}(\mathbf{\Lambda}\mathbf{V}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{V})]^2]$$

$$= \mathrm{E}\left(\sum_{i,j,k,l}^{n} \lambda_i\lambda_j\lambda_k\lambda_l w_{ij}^2 w_{kl}^2\right)$$

$$= \sum_{i}^{n} \lambda_i^4 \,\mathrm{E}(w_{ii}^4) + 4\sum_{i\neq j}^{n} \lambda_i^3\lambda_j \,\mathrm{E}\left(w_{ii}^2 w_{ij}^2\right)$$

$$+ \sum_{i\neq j}^{n} \lambda_i^2\lambda_j^2\left[\mathrm{E}(w_{ii}^2)\,\mathrm{E}\left(w_{jj}^2\right) + 2\,\mathrm{E}\left(w_{ij}^4\right)\right]$$

$$+ \sum_{i\neq j\neq k}^{n} \lambda_i^2\lambda_j\lambda_k\left[2\,\mathrm{E}(w_{ii}^2)\,\mathrm{E}\left(w_{jk}^2\right) + 4\,\mathrm{E}\left(w_{ij}^2 w_{ik}^2\right)\right]$$

$$+ \sum_{i\neq j\neq k\neq l}^{n} \lambda_i\lambda_j\lambda_k\lambda_l \,\mathrm{E}\left(w_{ij}^2\right)\mathrm{E}(w_{kl}^2)$$

$$= n(n+2)(n+4)(n+6)\sum_{i}^{n} \lambda_i^4 + 4n(n+2)(n+4)\sum_{i\neq j}^{n} \lambda_i^3\lambda_j$$

$$+ n(n+2)(n^2+2n+6)\sum_{i\neq j}^{n} \lambda_i^2\lambda_j^2 + 2n(n+2)\sum_{i\neq j\neq k}^{n} \lambda_i^2\lambda_j\lambda_k$$

$$+ n^2 \sum_{i\neq j\neq k\neq l}^{n} \lambda_i\lambda_j\lambda_k\lambda_l\,,$$

$$\mathrm{E}[(\mathrm{tr}\,\mathbf{A})^4] = \mathrm{E}[[\mathrm{tr}(\mathbf{\Lambda}\mathbf{V}^T\mathbf{V})]^4]$$

$$= \mathrm{E}\left(\sum_{i,j,k,l}^{n} \lambda_i\lambda_j\lambda_k\lambda_l w_{ii} w_{jj} w_{kk} w_{ll}\right)$$

$$= \sum_{i}^{n} \lambda_i^4 \,\mathrm{E}(w_{ii}^4) + 4 \sum_{i \neq j}^{n} \lambda_i^3 \lambda_j \,\mathrm{E}(w_{ii}^3)\,\mathrm{E}(w_{jj}) + 3 \sum_{i \neq j}^{n} \lambda_i^2 \lambda_j^2 \,\mathrm{E}(w_{ii}^2)\,\mathrm{E}(w_{jj}^2)$$

$$+ 6 \sum_{i \neq j \neq k}^{n} \lambda_i^2 \lambda_j \lambda_k \,\mathrm{E}(w_{ii}^2)\,\mathrm{E}(w_{jj})\,\mathrm{E}(w_{kk})$$

$$+ \sum_{i \neq j \neq k \neq l}^{n} \lambda_i \lambda_j \lambda_k \lambda_l \,\mathrm{E}(w_{ii})\,\mathrm{E}(w_{jj})\,\mathrm{E}(w_{kk})\,\mathrm{E}(w_{ll})$$

$$= n(n+2)(n+4)(n+6) \sum_{i}^{n} \lambda_i^4 + 4n^2(n+2)(n+4) \sum_{i \neq j}^{n} \lambda_i^3 \lambda_j$$

$$+ 3n^2(n+2)^2 \sum_{i \neq j}^{n} \lambda_i^2 \lambda_j^2 + 6n^3(n+2) \sum_{i \neq j \neq k}^{n} \lambda_i^2 \lambda_j \lambda_k$$

$$+ n^4 \sum_{i \neq j \neq k \neq l}^{n} \lambda_i \lambda_j \lambda_k \lambda_l\,,$$

$$\mathrm{E}[\mathrm{tr}(\mathbf{A}^2) \cdot (\mathrm{tr}\,\mathbf{A})^2] = \mathrm{E}[\mathrm{tr}(\mathbf{\Lambda V}^T \mathbf{V \Lambda V}^T \mathbf{V}) \cdot [\mathrm{tr}(\mathbf{\Lambda V}^T \mathbf{V})]^2]$$

$$= \mathrm{E}\left(\sum_{i,j,k,l}^{n} \lambda_i \lambda_j \lambda_k \lambda_l w_{ij}^2 w_{kk} w_{ll} \right)$$

$$= \sum_{i}^{n} \lambda_i^4 \,\mathrm{E}(w_{ii}^4) + \sum_{i \neq j}^{n} \lambda_i^3 \lambda_j [2\,\mathrm{E}(w_{ii}^3)\,\mathrm{E}(w_{jj}) + 2\,\mathrm{E}(w_{ii}^2 w_{ij}^2)]$$

$$+ \sum_{i \neq j}^{n} \lambda_i^2 \lambda_j^2 [\mathrm{E}(w_{ii}^2)\,\mathrm{E}(w_{jj}^2) + 2\,\mathrm{E}(w_{ii} w_{jj} w_{ij}^2)]$$

$$+ \sum_{i \neq j \neq k}^{n} \lambda_i^2 \lambda_j \lambda_k [\mathrm{E}(w_{ii}^2)\,\mathrm{E}(w_{jj})\,\mathrm{E}(w_{kk}) + 4\,\mathrm{E}(w_{ii} w_{ij}^2)\,\mathrm{E}(w_{kk})$$

$$+ \mathrm{E}(w_{ii}^2)\,\mathrm{E}(w_{jk}^2)] + \sum_{i \neq j \neq k \neq l}^{n} \lambda_i \lambda_j \lambda_k \lambda_l \,\mathrm{E}(w_{ij}^2)\,\mathrm{E}(w_{kk})\,\mathrm{E}(w_{ll})$$

$$= n(n+2)(n+4)(n+6)\sum_{i}^{n}\lambda_i^4 + 2n(n+1)(n+2)(n+4)\sum_{i\neq j}^{n}\lambda_i^3\lambda_j$$

$$+ n(n+2)^3\sum_{i\neq j}^{n}\lambda_i^2\lambda_j^2 + n^2(n+2)(n+5)\sum_{i\neq j\neq k}^{n}\lambda_i^2\lambda_j\lambda_k$$

$$+ n^3\sum_{i\neq j\neq k\neq l}^{n}\lambda_i\lambda_j\lambda_k\lambda_l\,, \tag{B6}$$

where notations of the form $i \neq j \neq k \neq l$ represent inequality of every pairwise combination of the subscripts concerned.

Although equations B6 can be evaluated for any $\mathbf{\Sigma}$, calculating the product of all possible combinations of eigenvalues is rather cumbersome especially when p is large. For this practical reason, it would be preferable to simplify these expressions by noting

$$\sum_{i}^{n}\lambda_i^r = \operatorname{tr}(\mathbf{\Lambda}^r)\,, \qquad r = 1, 2, \ldots,$$

$$\sum_{i\neq j}^{n}\lambda_i\lambda_j = \left(\sum_{i=1}^{n}\lambda_i\right)^2 - \sum_{i=1}^{n}\lambda_i^2 = (\operatorname{tr}\mathbf{\Lambda})^2 - \operatorname{tr}(\mathbf{\Lambda}^2)\,,$$

$$\sum_{i\neq j}^{n}\lambda_i^3\lambda_j = \operatorname{tr}\mathbf{\Lambda}\operatorname{tr}(\mathbf{\Lambda}^3) - \operatorname{tr}(\mathbf{\Lambda}^4)\,,$$

$$\sum_{i\neq j}^{n}\lambda_i^2\lambda_j^2 = 3[\operatorname{tr}(\mathbf{\Lambda}^2)]^2 - \operatorname{tr}(\mathbf{\Lambda}^4)\,,$$

$$\sum_{i\neq j\neq k}^{n}\lambda_i^2\lambda_j\lambda_k = (\operatorname{tr}\mathbf{\Lambda})^2\operatorname{tr}(\mathbf{\Lambda}^2) - 2\operatorname{tr}\mathbf{\Lambda}\operatorname{tr}(\mathbf{\Lambda}^3) - [\operatorname{tr}(\mathbf{\Lambda}^2)]^2 + 2\operatorname{tr}(\mathbf{\Lambda}^4)\,,$$

$$\sum_{i\neq j\neq k\neq l}^{n}\lambda_i\lambda_j\lambda_k\lambda_l = (\operatorname{tr}\mathbf{\Lambda})^4 - 6(\operatorname{tr}\mathbf{\Lambda})^2\operatorname{tr}(\mathbf{\Lambda}^2) + 8\operatorname{tr}\mathbf{\Lambda}\operatorname{tr}(\mathbf{\Lambda}^3) + 3[\operatorname{tr}(\mathbf{\Lambda}^2)]^2 - 6\operatorname{tr}(\mathbf{\Lambda}^4)\,. \tag{B7}$$

Then, equations B6 can be written as follows:

$$
\begin{aligned}
\mathrm{E}[\mathrm{tr}(\mathbf{A}^2)] &= n(\mathrm{tr}\,\boldsymbol{\Lambda})^2 + n(n+1)\,\mathrm{tr}(\boldsymbol{\Lambda}^2)\,, \\
\mathrm{E}[(\mathrm{tr}\,\mathbf{A})^2] &= n^2(\mathrm{tr}\,\boldsymbol{\Lambda})^2 + 2n\,\mathrm{tr}(\boldsymbol{\Lambda}^2)\,, \\
\mathrm{E}[[\mathrm{tr}(\mathbf{A}^2)]^2] &= n^2(\mathrm{tr}\,\boldsymbol{\Lambda})^4 + 2n(n^2+n+4)(\mathrm{tr}\,\boldsymbol{\Lambda})^2\,\mathrm{tr}(\boldsymbol{\Lambda}^2) + 16n(n+1)\,\mathrm{tr}\,\boldsymbol{\Lambda}\,\mathrm{tr}(\boldsymbol{\Lambda}^3) \\
&\quad + n(n^3+2n^2+5n+4)[\mathrm{tr}(\boldsymbol{\Lambda}^2)]^2 + 4n(2n^2+5n+5)\,\mathrm{tr}(\boldsymbol{\Lambda}^4)\,, \\
\mathrm{E}[(\mathrm{tr}\,\mathbf{A})^4] &= n^4(\mathrm{tr}\,\boldsymbol{\Lambda})^4 + 12n^3(\mathrm{tr}\,\boldsymbol{\Lambda})^2\,\mathrm{tr}(\boldsymbol{\Lambda}^2) + 12n^2\,\mathrm{tr}\,\boldsymbol{\Lambda}\,\mathrm{tr}(\boldsymbol{\Lambda}^3) \\
&\quad + 32n^2[\mathrm{tr}(\boldsymbol{\Lambda}^2)]^2 + 48n\,\mathrm{tr}(\boldsymbol{\Lambda}^4)\,. \\
\mathrm{E}[\mathrm{tr}(\mathbf{A}^2)\cdot(\mathrm{tr}\,\mathbf{A})^2] &= n^3(\mathrm{tr}\,\boldsymbol{\Lambda})^4 + n^2(n^2+n+10)(\mathrm{tr}\,\boldsymbol{\Lambda})^2\,\mathrm{tr}(\boldsymbol{\Lambda}^2) \\
&\quad + 8n(n^2+n+2)\,\mathrm{tr}\,\boldsymbol{\Lambda}\,\mathrm{tr}(\boldsymbol{\Lambda}^3) + 2n(n^2+n+4)[\mathrm{tr}(\boldsymbol{\Lambda}^2)]^2 \\
&\quad + 24n(n+1)\,\mathrm{tr}(\boldsymbol{\Lambda}^4)\,,
\end{aligned}
\tag{B8}
$$

Finally,

$$
\begin{aligned}
\mathrm{Var}[\mathrm{tr}(\mathbf{A}^2)] &= \mathrm{E}[[\mathrm{tr}(\mathbf{A}^2)]^2] - \mathrm{E}[\mathrm{tr}(\mathbf{A}^2)]^2 \\
&= 8n(\mathrm{tr}\,\boldsymbol{\Lambda})^2\,\mathrm{tr}(\boldsymbol{\Lambda}^2) + 16n(n+1)\,\mathrm{tr}\,\boldsymbol{\Lambda}\,\mathrm{tr}(\boldsymbol{\Lambda}^3) \\
&\quad + 4n(n+1)[\mathrm{tr}(\boldsymbol{\Lambda}^2)]^2 + 4n(2n^2+5n+5)\,\mathrm{tr}(\boldsymbol{\Lambda}^4)\,, \\
\mathrm{Var}[(\mathrm{tr}\,\mathbf{A})^2] &= \mathrm{E}[(\mathrm{tr}\,\mathbf{A})^4] - \mathrm{E}[(\mathrm{tr}\,\mathbf{A})^2]^2 \\
&= 8n^3(\mathrm{tr}\,\boldsymbol{\Lambda})^2\,\mathrm{tr}(\boldsymbol{\Lambda}^2) + 32n^2\,\mathrm{tr}\,\boldsymbol{\Lambda}\,\mathrm{tr}(\boldsymbol{\Lambda}^3) + 8n^2[\mathrm{tr}(\boldsymbol{\Lambda}^2)]^2 \\
&\quad + 48n\,\mathrm{tr}(\boldsymbol{\Lambda}^4)\,, \\
\mathrm{Cov}[\mathrm{tr}(\mathbf{A}^2), (\mathrm{tr}\,\mathbf{A})^2] &= \mathrm{E}[\mathrm{tr}(\mathbf{A}^2)\cdot(\mathrm{tr}\,\mathbf{A})^2] - \mathrm{E}[\mathrm{tr}(\mathbf{A}^2)]\,\mathrm{E}[(\mathrm{tr}\,\mathbf{A})^2] \\
&= 8n^2(\mathrm{tr}\,\boldsymbol{\Lambda})^2\,\mathrm{tr}(\boldsymbol{\Lambda}^2) + 8n(n^2+n+2)\,\mathrm{tr}\,\boldsymbol{\Lambda}\,\mathrm{tr}(\boldsymbol{\Lambda}^3) + 8n[\mathrm{tr}(\boldsymbol{\Lambda}^2)]^2 \\
&\quad + 24n(n+1)\,\mathrm{tr}(\boldsymbol{\Lambda}^4)\,.
\end{aligned}
\tag{B9}
$$

Inserting equations B8 and B9 into equations 12, 19, and 32 yields the desired results.

Identical results can be derived from del Waal & Nel's (1973) results on the expectations of elementary symmetric functions of eigenvalues and their products for a

Wishart matrix. However, these results appear to have been proved only under the condition $n > p - 1$ (see also Constantine, 1963; Muirhead, 1982: chapter 7). The above derivation is valid for any combination of p and n.

Appendix C

This part demonstrates that $\mathrm{Cov}\left(r_{ij}^2, r_{kl}^2\right) = 0$ for $(i, j) \neq (k, l)$ under the condition $\mathbf{P} = \mathbf{I}_p$, as cursorily mentioned by Schott (2005). Under this condition, a sample covariance can be written as $s_{ij} = n_*^{-1}\left(\sigma_{ii}\sigma_{jj}\right)^{1/2}\mathbf{v}_i^T\mathbf{v}_j$, with $\mathbf{v}_i$ and $\mathbf{v}_j$ being i.i.d. $N_n(\mathbf{0}, \mathbf{I}_n)$. Therefore, a sample correlation coefficient can be written as $r_{ij} = s_{ij}\left(s_{ii}s_{jj}\right)^{-1/2} = \mathbf{u}_i^T\mathbf{u}_j$, where $\mathbf{u}_i = (\mathbf{v}_i^T\mathbf{v}_i)^{-1/2}\mathbf{v}_i$ are uniformly distributed on the surface of the unit hypersphere in the n-dimensional space. By noting $\mathbf{u}_i^T\mathbf{u}_i = 1$, it is possible to see $\mathrm{E}(\mathbf{u}_i\mathbf{u}_i^T) = n^{-1}\mathbf{I}_n$ for any i, because the elements of $\mathbf{u}_i$ are symmetric and uncorrelated with one another (a formal demonstration requires introduction of the density function; see Anderson, 2003: p. 49). With these preliminaries, it is easily seen, for $i \neq j \neq k$,

$$
\begin{aligned}
\mathrm{E}\left(r_{ij}^2 r_{ik}^2\right) &= \mathrm{E}\left(\mathbf{u}_i^T\mathbf{u}_j\mathbf{u}_j^T\mathbf{u}_i\mathbf{u}_i^T\mathbf{u}_k\mathbf{u}_k^T\mathbf{u}_i\right) \\
&= \mathrm{E}\left[\mathbf{u}_i^T\mathrm{E}\left(\mathbf{u}_j\mathbf{u}_j^T\right)\mathbf{u}_i\mathbf{u}_i^T\mathrm{E}(\mathbf{u}_k\mathbf{u}_k^T)\mathbf{u}_i\right] \\
&= n^{-2}\mathrm{E}[\mathbf{u}_i^T\mathbf{u}_i\mathbf{u}_i^T\mathbf{u}_i] \\
&= n^{-2} = \mathrm{E}\left(r_{ij}^2\right)\mathrm{E}(r_{ik}^2).
\end{aligned}
\tag{C1}
$$

The second equation is valid because $\mathbf{u}_i$, $\mathbf{u}_j$, and $\mathbf{u}_k$ are stochastically independent from one another. Therefore, $\mathrm{Cov}\left(r_{ij}^2, r_{ik}^2\right) = 0$ for partly overlapping subscripts. Similarly, $\mathrm{Cov}\left(r_{ij}^2, r_{kl}^2\right) = 0$ for non-overlapping subscripts, although this could also be seen as a direct consequence of the independence between r_{ij} and r_{kl} in this case.

Appendix D

In this part, an asymptotic expression for the variance of $V_{\mathrm{rel}}(\mathbf{R})$ is derived, somewhat heuristically, for arbitrary non-null conditions with $p > 2$. Konishi (1979) gave an asymptotic theory for the distribution of an arbitrary function of eigenvalues of a sample correlation matrix $f(l_1, \ldots, l_p)$ under multivariate normality. In particular, when $n \to \infty$, $\sqrt{n}[f(l_1, \ldots, l_p) - f(\lambda_1, \ldots, \lambda_p)]$ was shown to be normally distributed with mean 0 and variance

$$\tau^2 = 2 \sum_{\alpha,\beta=1}^{p} \lambda_\alpha \lambda_\beta \left[\delta_{\alpha\beta} - (\lambda_\alpha + \lambda_\beta) \sum_{i=1}^{p} \upsilon_{i\alpha}^2 \upsilon_{i\beta}^2 + \sum_{i,j=1}^{p} \rho_{ij}^2 \upsilon_{i\alpha}^2 \upsilon_{j\beta}^2 \right] f_\alpha f_\beta , \tag{D1}$$

where the summations are over all combinations of subscripts, $\delta_{\alpha\beta}$ is the Kronecker delta, $\upsilon_{i\alpha}$ is the (i, α)-th element of the population eigenvector matrix $\mathbf{Y}$, and $f_\alpha = \partial f / \partial l_\alpha |_{(l_1,\ldots,l_p)=(\lambda_1,\ldots,\lambda_p)}$, the partial derivative of f with respect to l_α evaluated at $(l_1, \ldots, l_p) = (\lambda_1, \ldots, \lambda_p)$. Note that Konishi's (1979; corollary 2.2) original notation also concerned potential multiplicity of population eigenvalues, which is ignored here for simplicity; the population eigenvectors corresponding to multiplicated eigenvalues can in practice be chosen arbitrarily as a suite of orthogonal vectors in the appropriate subspace, as is done in numerical determination of eigenvectors. The derivative of $V_{\mathrm{rel}}(\mathbf{R})$ is simply

$$f_\alpha = \frac{\partial}{\partial l_\alpha} V_{\mathrm{rel}}(\mathbf{R}) \bigg|_{(l_1,\ldots,l_p)=(\lambda_1,\ldots,\lambda_p)} = \frac{2}{p(p-1)} \lambda_\alpha . \tag{D2}$$

Inserting equation D2 into equation D1, we obtain τ^2/n as an asymptotic expression of the variance of $V_{\mathrm{rel}}(\mathbf{R})$ (eq. 37).

An empirically equivalent result can be obtained from the alternative expression of $V_{\mathrm{rel}}(\mathbf{R})$ as average squared correlation coefficients (eq. 11), from a similar theory for functions of a sample correlation matrix by Konishi (1979: theorem 6.2). However, that alternative expression does not seem to bear much practical advantage, for it typically takes substantially more computational time to evaluate as p grows.

Table 1. Summary statistics of selected simulation results for eigenvalue variance of covariance matrix $V(\mathbf{S})$. Theoretical expectation ($\mathrm{E}[V(\mathbf{S})]$) and standard deviation ($\mathrm{SD}[V(\mathbf{S})]$), as well as empirical median, mean, standard deviation (ESD), and bias of mean in standard error unit ($T = \sqrt{5000}\{\mathrm{Mean} - \mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})]\}/\mathrm{ESD}$, which should roughly follow t distribution with 4999 degrees of freedom if the expectation is exact) from 5000 simulation runs are shown for selected conditions. See Table S1 for full results.

	$\mathrm{E}[V(\mathbf{S})]$	$\mathrm{SD}[V(\mathbf{S})]$	Median	Mean	ESD	T
$p = 2$, $V(\mathbf{\Sigma}) = 0$						
$N = 8$	0.2857	0.3582	0.1731	0.2876	0.3464	0.3845
$N = 16$	0.1333	0.1501	0.0876	0.1330	0.1466	−0.1703
$N = 32$	0.0645	0.0686	0.0438	0.0646	0.0689	0.0489
$N = 64$	0.0317	0.0327	0.0221	0.0315	0.0315	−0.6336
$p = 4$, $V(\mathbf{\Sigma}) = 0$						
$N = 8$	0.2143	0.1551	0.1734	0.2131	0.1534	−0.5437
$N = 16$	0.1000	0.0602	0.0861	0.0996	0.0603	−0.4826
$N = 32$	0.0484	0.0261	0.0429	0.0481	0.0261	−0.8676
$N = 64$	0.0238	0.0120	0.0216	0.0239	0.0119	0.5321
$p = 16$, $V(\mathbf{\Sigma}) = 0$						
$N = 8$	0.1518	0.0304	0.1482	0.1510	0.0307	−1.8948
$N = 16$	0.0708	0.0102	0.0702	0.0708	0.0102	−0.0003
$N = 32$	0.0343	0.0038	0.0341	0.0342	0.0037	−0.4806
$N = 64$	0.0169	0.0015	0.0168	0.0169	0.0015	−0.6023
$p = 64$, $V(\mathbf{\Sigma}) = 0$						
$N = 8$	0.1473	0.0203	0.1464	0.1477	0.0204	1.2947

Table 1 (continued)

	$\mathrm{E}[V(\mathbf{S})]$	$\mathrm{SD}[V(\mathbf{S})]$	Median	Mean	ESD	T
$N = 16$	0.0688	0.0067	0.0688	0.0689	0.0066	1.7306
$N = 32$	0.0333	0.0024	0.0332	0.0333	0.0024	0.6721
$N = 64$	0.0164	0.0009	0.0163	0.0164	0.0009	−0.4407
$p = 256, V(\mathbf{\Sigma}) = 0$						
$N = 8$	0.1440	0.0097	0.1436	0.1440	0.0099	0.1215
$N = 16$	0.0672	0.0031	0.0672	0.0672	0.0031	0.9739
$N = 32$	0.0325	0.0011	0.0325	0.0325	0.0011	−0.8132
$N = 64$	0.0160	0.0004	0.0160	0.0160	0.0004	0.0842
$p = 1024, V(\mathbf{\Sigma}) = 0$						
$N = 8$	0.1431	0.0048	0.1430	0.1431	0.0048	−0.6874
$N = 16$	0.0668	0.0015	0.0668	0.0668	0.0015	1.7324
$N = 32$	0.0323	0.0005	0.0323	0.0323	0.0005	0.3227
$N = 64$	0.0159	0.0002	0.0159	0.0159	0.0002	0.3522
$p = 2, q = 1, V(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.6857	0.8717	0.3967	0.6901	0.8663	0.3572
$N = 16$	0.5333	0.4852	0.3972	0.5426	0.4981	1.3183
$N = 32$	0.4645	0.3023	0.3936	0.4616	0.3011	−0.6854
$N = 64$	0.4317	0.2002	0.4044	0.4337	0.2019	0.6886
$p = 4, q = 1, V(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.6429	0.7343	0.3949	0.6372	0.7364	−0.5426
$N = 16$	0.5133	0.4137	0.3968	0.5132	0.4183	−0.0225
$N = 32$	0.4548	0.2598	0.3937	0.4544	0.2661	−0.1087
$N = 64$	0.4270	0.1728	0.4025	0.4292	0.1756	0.8791
$p = 16, q = 1, V(\mathbf{\Sigma}) = 0.4$						

Table 1 (continued)

	$\mathrm{E}[V(\mathbf{S})]$	$\mathrm{SD}[V(\mathbf{S})]$	Median	Mean	ESD	T
$N = 8$	0.6107	0.6427	0.4130	0.6165	0.6509	0.6282
$N = 16$	0.4983	0.3668	0.4060	0.5036	0.3748	0.9922
$N = 32$	0.4476	0.2322	0.4057	0.4492	0.2314	0.5001
$N = 64$	0.4234	0.1552	0.4005	0.4227	0.1546	−0.3205
$p = 64, q = 1, V(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.6027	0.6206	0.4227	0.6052	0.5988	0.2997
$N = 16$	0.4946	0.3554	0.4061	0.5021	0.3631	1.4613
$N = 32$	0.4458	0.2255	0.4003	0.4445	0.2244	−0.4124
$N = 64$	0.4225	0.1509	0.3986	0.4237	0.1546	0.5322
$p = 256, q = 1, V(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.6007	0.6151	0.4140	0.6076	0.6137	0.7995
$N = 16$	0.4936	0.3526	0.4000	0.4873	0.3492	−1.2751
$N = 32$	0.4453	0.2239	0.4068	0.4465	0.2213	0.3813
$N = 64$	0.4223	0.1499	0.4003	0.4234	0.1505	0.5054
$p = 1024, q = 1, V(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.6002	0.6138	0.3961	0.5940	0.6361	−0.6814
$N = 16$	0.4934	0.3519	0.4015	0.4977	0.3638	0.8432
$N = 32$	0.4452	0.2235	0.4037	0.4489	0.2304	1.1498
$N = 64$	0.4222	0.1496	0.4020	0.4240	0.1526	0.8140
$p = 2, q = 1, V(\mathbf{\Sigma}) = 0.8$						
$N = 8$	1.0857	1.2651	0.6587	1.0792	1.3030	−0.3520
$N = 16$	0.9333	0.7343	0.7440	0.9364	0.7250	0.3003
$N = 32$	0.8645	0.4698	0.7732	0.8675	0.4675	0.4571
$N = 64$	0.8317	0.3157	0.7840	0.8333	0.3172	0.3574

Table 1 (continued)

	$E[V(\mathbf{S})]$	$SD[V(\mathbf{S})]$	Median	Mean	ESD	T
$p = 4, q = 1, V(\mathbf{\Sigma}) = 0.8$						
$N = 8$	1.0714	1.2182	0.6903	1.0730	1.2081	0.0917
$N = 16$	0.9267	0.7096	0.7454	0.9328	0.7139	0.6117
$N = 32$	0.8613	0.4549	0.7682	0.8585	0.4565	−0.4368
$N = 64$	0.8302	0.3061	0.7903	0.8350	0.3069	1.1137
$p = 16, q = 1, V(\mathbf{\Sigma}) = 0.8$						
$N = 8$	1.0607	1.1840	0.6778	1.0687	1.1804	0.4765
$N = 16$	0.9217	0.6917	0.7517	0.9147	0.6842	−0.7221
$N = 32$	0.8589	0.4442	0.7847	0.8683	0.4387	1.5248
$N = 64$	0.8290	0.2992	0.7925	0.8358	0.3067	1.5855
$p = 64, q = 1, V(\mathbf{\Sigma}) = 0.8$						
$N = 8$	1.0580	1.1756	0.6979	1.0755	1.2081	1.0231
$N = 16$	0.9204	0.6873	0.7463	0.9304	0.7024	1.0083
$N = 32$	0.8583	0.4415	0.7643	0.8495	0.4352	−1.4315
$N = 64$	0.8287	0.2975	0.7873	0.8292	0.2977	0.1296
$p = 256, q = 1, V(\mathbf{\Sigma}) = 0.8$						
$N = 8$	1.0574	1.1735	0.6982	1.0818	1.1935	1.4492
$N = 16$	0.9201	0.6862	0.7373	0.9122	0.6746	−0.8321
$N = 32$	0.8581	0.4409	0.7749	0.8663	0.4540	1.2689
$N = 64$	0.8286	0.2971	0.7824	0.8219	0.2873	−1.6582
$p = 1024, q = 1, V(\mathbf{\Sigma}) = 0.8$						
$N = 8$	1.0572	1.1729	0.6881	1.0573	1.1723	0.0072
$N = 16$	0.9200	0.6859	0.7655	0.9344	0.6943	1.4624
$N = 32$	0.8581	0.4407	0.7709	0.8531	0.4414	−0.7945

Table 1 (continued)

	$\mathrm{E}[V(\mathbf{S})]$	$\mathrm{SD}[V(\mathbf{S})]$	Median	Mean	ESD	T
$N = 64$	0.8286	0.2969	0.7840	0.8292	0.2987	0.1364

1506

Table 2. Summary statistics of selected simulation results for relative eigenvalue variance of covariance matrix $V_{\mathrm{rel}}(\mathbf{S})$. (Approximate) theoretical expectation ($\mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})]$) and standard deviation ($\mathrm{SD}[V_{\mathrm{rel}}(\mathbf{S})]$), as well as empirical median, mean, standard deviation (ESD), and bias of mean in standard error unit (T) from 5000 simulation runs are shown for selected conditions. See Table 1 for further information and Table S2 for full results.

	$\approx \mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})]$	$\approx \mathrm{SD}[V_{\mathrm{rel}}(\mathbf{S})]$	Median	Mean	ESD	T
$p = 2$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0$						
$N = 8$	0.2500	0.1936	0.2079	0.2514	0.1924	0.4961
$N = 16$	0.1250	0.1102	0.0953	0.1259	0.1098	0.6000
$N = 32$	0.0625	0.0587	0.0448	0.0628	0.0597	0.3234
$N = 64$	0.0313	0.0303	0.0224	0.0309	0.0289	−0.8301
$p = 4$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0$						
$N = 8$	0.2000	0.0840	0.1856	0.2002	0.0858	0.1644
$N = 16$	0.0968	0.0433	0.0907	0.0968	0.0434	0.0620
$N = 32$	0.0476	0.0219	0.0438	0.0474	0.0221	−0.6146
$N = 64$	0.0236	0.0110	0.0218	0.0237	0.0110	0.6544
$p = 16$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0$						
$N = 8$	0.1579	0.0193	0.1562	0.1580	0.0197	0.3764
$N = 16$	0.0744	0.0091	0.0738	0.0746	0.0094	1.4158
$N = 32$	0.0361	0.0044	0.0357	0.0361	0.0044	−1.2604
$N = 64$	0.0178	0.0022	0.0177	0.0178	0.0022	0.4338
$p = 64$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0$						
$N = 8$	0.1467	0.0047	0.1462	0.1466	0.0046	−0.6909
$N = 16$	0.0686	0.0022	0.0685	0.0686	0.0022	0.5618

Table 2 (continued)

	$\approx \mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})]$	$\approx \mathrm{SD}[V_{\mathrm{rel}}(\mathbf{S})]$	Median	Mean	ESD	T
$N = 32$	0.0332	0.0010	0.0332	0.0332	0.0011	0.0446
$N = 64$	0.0164	0.0005	0.0163	0.0164	0.0005	−0.5869
$p = 256$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0$						
$N = 8$	0.1438	0.0012	0.1437	0.1438	0.0012	−0.0173
$N = 16$	0.0672	0.0005	0.0671	0.0672	0.0005	1.4545
$N = 32$	0.0325	0.0003	0.0325	0.0325	0.0003	−0.7481
$N = 64$	0.0160	0.0001	0.0160	0.0160	0.0001	0.3608
$p = 1024$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0$						
$N = 8$	0.1431	0.0003	0.1431	0.1431	0.0003	−0.5105
$N = 16$	0.0668	0.0001	0.0668	0.0668	0.0001	0.0231
$N = 32$	0.0323	0.0001	0.0323	0.0323	0.0001	−1.3055
$N = 64$	0.0159	0.0000	0.0159	0.0159	0.0000	−0.0022
$p = 2$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.4377	0.2825	0.4939	0.4804	0.2334	12.9356
$N = 16$	0.4232	0.1982	0.4457	0.4384	0.1764	6.0670
$N = 32$	0.4132	0.1378	0.4169	0.4152	0.1297	1.1224
$N = 64$	0.4070	0.0963	0.4108	0.4083	0.0945	0.9214
$p = 4$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.4319	0.2065	0.4730	0.4675	0.1658	15.1734
$N = 16$	0.4200	0.1444	0.4316	0.4288	0.1272	4.8902
$N = 32$	0.4113	0.1001	0.4147	0.4125	0.0944	0.9058
$N = 64$	0.4060	0.0699	0.4073	0.4067	0.0682	0.7503
$p = 16$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.4275	0.1691	0.4657	0.4626	0.1346	18.4219

Table 2 (continued)

	$\approx \mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})]$	$\approx \mathrm{SD}[V_{\mathrm{rel}}(\mathbf{S})]$	Median	Mean	ESD	T
$N = 16$	0.4174	0.1175	0.4302	0.4270	0.1049	6.4710
$N = 32$	0.4098	0.0813	0.4161	0.4129	0.0762	2.8682
$N = 64$	0.4052	0.0566	0.4070	0.4058	0.0548	0.7118
$p = 64, q = 1, V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.4264	0.1613	0.4670	0.4599	0.1273	18.6017
$N = 16$	0.4168	0.1119	0.4286	0.4266	0.0996	6.9651
$N = 32$	0.4095	0.0773	0.4125	0.4109	0.0726	1.3497
$N = 64$	0.4050	0.0538	0.4063	0.4056	0.0532	0.7287
$p = 256, q = 1, V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.4261	0.1594	0.4637	0.4606	0.1249	19.5222
$N = 16$	0.4166	0.1105	0.4262	0.4228	0.0977	4.4222
$N = 32$	0.4094	0.0763	0.4148	0.4120	0.0718	2.5791
$N = 64$	0.4050	0.0531	0.4066	0.4058	0.0520	1.1405
$p = 1024, q = 1, V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.4$						
$N = 8$	0.4261	0.1590	0.4566	0.4548	0.1261	16.1041
$N = 16$	0.4166	0.1102	0.4272	0.4246	0.0990	5.7314
$N = 32$	0.4094	0.0761	0.4140	0.4123	0.0718	2.8690
$N = 64$	0.4050	0.0530	0.4068	0.4058	0.0521	1.1494
$p = 2, q = 1, V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.8$						
$N = 8$	0.7847	0.1153	0.8314	0.7947	0.1473	4.8016
$N = 16$	0.7929	0.0866	0.8164	0.7963	0.1004	2.4327
$N = 32$	0.7969	0.0625	0.8090	0.7990	0.0672	2.2207
$N = 64$	0.7986	0.0445	0.8030	0.7983	0.0470	−0.3966
$p = 4, q = 1, V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.8$						

Table 2 (continued)

	$\approx \mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})]$	$\approx \mathrm{SD}[V_{\mathrm{rel}}(\mathbf{S})]$	Median	Mean	ESD	T
$N = 8$	0.7841	0.0920	0.8207	0.7927	0.1170	5.1526
$N = 16$	0.7927	0.0692	0.8086	0.7945	0.0793	1.6175
$N = 32$	0.7968	0.0498	0.8039	0.7967	0.0537	−0.1966
$N = 64$	0.7985	0.0354	0.8026	0.7993	0.0368	1.5352
$p = 16$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.8$						
$N = 8$	0.7838	0.0810	0.8153	0.7913	0.1049	5.0688
$N = 16$	0.7926	0.0609	0.8067	0.7943	0.0696	1.7415
$N = 32$	0.7968	0.0437	0.8044	0.7989	0.0452	3.3474
$N = 64$	0.7985	0.0310	0.8019	0.7990	0.0323	1.0675
$p = 64$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.8$						
$N = 8$	0.7837	0.0787	0.8157	0.7915	0.1022	5.4258
$N = 16$	0.7926	0.0592	0.8072	0.7956	0.0669	3.2349
$N = 32$	0.7967	0.0425	0.8021	0.7967	0.0447	−0.0106
$N = 64$	0.7985	0.0301	0.8018	0.7989	0.0308	0.8231
$p = 256$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.8$						
$N = 8$	0.7836	0.0781	0.8150	0.7930	0.0994	6.6508
$N = 16$	0.7926	0.0588	0.8058	0.7944	0.0663	1.9392
$N = 32$	0.7967	0.0422	0.8033	0.7976	0.0455	1.3138
$N = 64$	0.7985	0.0299	0.8010	0.7983	0.0303	−0.5863
$p = 1024$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{\Sigma}) = 0.8$						
$N = 8$	0.7836	0.0780	0.8140	0.7904	0.1011	4.7652
$N = 16$	0.7926	0.0587	0.8081	0.7959	0.0665	3.5530
$N = 32$	0.7967	0.0421	0.8026	0.7967	0.0448	−0.0218
$N = 64$	0.7985	0.0299	0.8012	0.7987	0.0307	0.4338

Table 3. Summary statistics of selected simulation results for relative eigenvalue variance of correlation matrix $V_{\text{rel}}(\mathbf{R})$. Theoretical expectation ($\text{E}[V_{\text{rel}}(\mathbf{R})]$) and (approximate) standard deviation ($\text{SD}[V_{\text{rel}}(\mathbf{R})]$), as well as empirical median, mean, standard deviation (ESD), and bias of mean in standard error unit (T) from 5000 simulation runs are shown for selected conditions. See Table 1 for further information and Table S3 for full results.

	$\text{E}[V_{\text{rel}}(\mathbf{R})]$	$\approx \text{SD}[V_{\text{rel}}(\mathbf{R})]$	Median	Mean	ESD	T
$p = 2$, $V_{\text{rel}}(\mathbf{P}) = 0$						
$N = 8$	0.1429	0.1650	0.0792	0.1429	0.1635	0.0363
$N = 16$	0.0667	0.0856	0.0336	0.0679	0.0862	1.0235
$N = 32$	0.0323	0.0435	0.0163	0.0326	0.0427	0.6113
$N = 64$	0.0159	0.0219	0.0075	0.0156	0.0205	−1.0380
$p = 4$, $V_{\text{rel}}(\mathbf{P}) = 0$						
$N = 8$	0.1429	0.0673	0.1332	0.1432	0.0673	0.3148
$N = 16$	0.0667	0.0349	0.0608	0.0661	0.0344	−1.0696
$N = 32$	0.0323	0.0178	0.0292	0.0321	0.0175	−0.6906
$N = 64$	0.0159	0.0090	0.0142	0.0157	0.0087	−1.0749
$p = 16$, $V_{\text{rel}}(\mathbf{P}) = 0$						
$N = 8$	0.1429	0.0151	0.1411	0.1428	0.0151	−0.0917
$N = 16$	0.0667	0.0078	0.0663	0.0669	0.0080	1.7743
$N = 32$	0.0323	0.0040	0.0319	0.0322	0.0039	−1.7517
$N = 64$	0.0159	0.0020	0.0158	0.0159	0.0020	0.7127
$p = 64$, $V_{\text{rel}}(\mathbf{P}) = 0$						
$N = 8$	0.1429	0.0037	0.1425	0.1428	0.0036	−0.4617
$N = 16$	0.0667	0.0019	0.0666	0.0667	0.0019	0.4407

	$\mathrm{E}[V_{\mathrm{rel}}(\mathbf{R})]$	$\approx \mathrm{SD}[V_{\mathrm{rel}}(\mathbf{R})]$	Median	Mean	ESD	T
$N = 32$	0.0323	0.0010	0.0322	0.0323	0.0010	−0.0725
$N = 64$	0.0159	0.0005	0.0159	0.0159	0.0005	−0.6525
$p = 256$, $V_{\mathrm{rel}}(\mathbf{P}) = 0$						
$N = 8$	0.1429	0.0009	0.1428	0.1429	0.0009	0.0105
$N = 16$	0.0667	0.0005	0.0667	0.0667	0.0005	1.8676
$N = 32$	0.0323	0.0002	0.0323	0.0323	0.0002	−0.9701
$N = 64$	0.0159	0.0001	0.0159	0.0159	0.0001	0.3244
$p = 1024$, $V_{\mathrm{rel}}(\mathbf{P}) = 0$						
$N = 8$	0.1429	0.0002	0.1428	0.1429	0.0002	−0.9639
$N = 16$	0.0667	0.0001	0.0667	0.0667	0.0001	0.5220
$N = 32$	0.0323	0.0001	0.0323	0.0323	0.0001	−1.8393
$N = 64$	0.0159	0.0000	0.0159	0.0159	0.0000	−0.0357
$p = 2$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{P}) = 0.4$						
$N = 8$	0.4318	0.2495	0.4362	0.4294	0.2516	−0.6571
$N = 16$	0.4111	0.1844	0.4230	0.4147	0.1832	1.3640
$N = 32$	0.4046	0.1326	0.4056	0.4041	0.1322	−0.2806
$N = 64$	0.4021	0.0944	0.4058	0.4025	0.0954	0.3294
$p = 4$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{P}) = 0.4$						
$N = 8$	0.4318	0.2078	0.4314	0.4287	0.1750	−1.2271
$N = 16$	0.4111	0.1420	0.4141	0.4093	0.1317	−0.9769
$N = 32$	0.4046	0.0988	0.4054	0.4034	0.0960	−0.8743
$N = 64$	0.4021	0.0693	0.4033	0.4022	0.0687	0.1270
$p = 16$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{P}) = 0.4$						
$N = 8$	0.4318	0.1682	0.4321	0.4329	0.1391	0.5780

	$\mathrm{E}[V_{\mathrm{rel}}(\mathbf{R})]$	$\approx \mathrm{SD}[V_{\mathrm{rel}}(\mathbf{R})]$	Median	Mean	ESD	T
$N = 16$	0.4111	0.1149	0.4149	0.4120	0.1073	0.5980
$N = 32$	0.4046	0.0799	0.4084	0.4055	0.0773	0.7996
$N = 64$	0.4021	0.0561	0.4030	0.4021	0.0552	−0.0283
$p = 64$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{P}) = 0.4$						
$N = 8$	0.4318	0.1598	0.4366	0.4332	0.1304	0.7802
$N = 16$	0.4111	0.1092	0.4144	0.4129	0.1015	1.2079
$N = 32$	0.4046	0.0760	0.4052	0.4039	0.0733	−0.6457
$N = 64$	0.4021	0.0533	0.4029	0.4021	0.0535	0.0493
$p = 256$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{P}) = 0.4$						
$N = 8$	0.4318	0.1578	0.4348	0.4343	0.1278	1.3865
$N = 16$	0.4111	0.1078	0.4118	0.4091	0.0995	−1.4526
$N = 32$	0.4046	0.0750	0.4078	0.4052	0.0726	0.5664
$N = 64$	0.4021	0.0526	0.4033	0.4024	0.0522	0.4500
$p = 1024$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{P}) = 0.4$						
$N = 8$	0.4318	0.1573	0.4266	0.4286	0.1287	−1.7550
$N = 16$	0.4111	0.1075	0.4127	0.4111	0.1008	−0.0454
$N = 32$	0.4046	0.0748	0.4072	0.4055	0.0726	0.8769
$N = 64$	0.4021	0.0524	0.4035	0.4024	0.0524	0.4578
$p = 2$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{P}) = 0.8$						
$N = 8$	0.7831	0.1596	0.8250	0.7853	0.1571	0.9790
$N = 16$	0.7918	0.1015	0.8141	0.7928	0.1033	0.6253
$N = 32$	0.7961	0.0674	0.8076	0.7976	0.0680	1.5536
$N = 64$	0.7981	0.0462	0.8024	0.7976	0.0473	−0.6617
$p = 4$, $q = 1$, $V_{\mathrm{rel}}(\mathbf{P}) = 0.8$						

	$\mathrm{E}[V_{\mathrm{rel}}(\mathbf{R})]$	$\approx \mathrm{SD}[V_{\mathrm{rel}}(\mathbf{R})]$	Median	Mean	ESD	T
$N = 8$	0.7831	0.1073	0.8149	0.7840	0.1249	0.4605
$N = 16$	0.7918	0.0733	0.8059	0.7913	0.0815	−0.4682
$N = 32$	0.7961	0.0510	0.8027	0.7953	0.0543	−1.0173
$N = 64$	0.7981	0.0358	0.8019	0.7987	0.0370	1.2077
$p = 16, q = 1, V_{\mathrm{rel}}(\mathbf{P}) = 0.8$						
$N = 8$	0.7831	0.0939	0.8102	0.7835	0.1116	0.2370
$N = 16$	0.7918	0.0642	0.8045	0.7913	0.0716	−0.5326
$N = 32$	0.7961	0.0446	0.8035	0.7976	0.0457	2.4043
$N = 64$	0.7981	0.0313	0.8013	0.7984	0.0325	0.7108
$p = 64, q = 1, V_{\mathrm{rel}}(\mathbf{P}) = 0.8$						
$N = 8$	0.7831	0.0912	0.8098	0.7841	0.1083	0.6205
$N = 16$	0.7918	0.0623	0.8046	0.7928	0.0687	0.9579
$N = 32$	0.7961	0.0433	0.8009	0.7955	0.0452	−0.9741
$N = 64$	0.7981	0.0304	0.8012	0.7983	0.0309	0.4419
$p = 256, q = 1, V_{\mathrm{rel}}(\mathbf{P}) = 0.8$						
$N = 8$	0.7831	0.0905	0.8098	0.7857	0.1056	1.7052
$N = 16$	0.7918	0.0618	0.8031	0.7915	0.0681	−0.3408
$N = 32$	0.7961	0.0430	0.8022	0.7963	0.0460	0.3738
$N = 64$	0.7981	0.0302	0.8004	0.7977	0.0305	−0.9565
$p = 1024, q = 1, V_{\mathrm{rel}}(\mathbf{P}) = 0.8$						
$N = 8$	0.7831	0.0903	0.8088	0.7830	0.1074	−0.0831
$N = 16$	0.7918	0.0617	0.8057	0.7930	0.0683	1.2443
$N = 32$	0.7961	0.0429	0.8014	0.7954	0.0454	−0.9758
$N = 64$	0.7981	0.0301	0.8006	0.7981	0.0308	0.0710

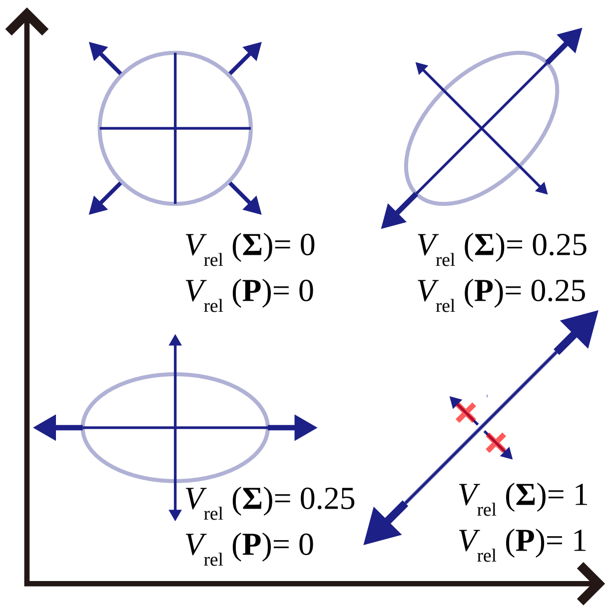

Figure 1. Schematic illustration of eigenvalue dispersion indices in bivariate cases. Ellipses representing equiprobability contours are shown on the Cartesian space of two hypothetical variables for four conditions, as well as the relative eigenvalue variance of the corresponding covariance and correlation matrices ($V_{rel}(\mathbf{\Sigma})$ and $V_{rel}(\mathbf{P})$, respectively). The scale is arbitrary but identical for the two axes. The axes of each ellipse are proportional to square roots of the two eigenvalues of the respective covariance matrix. $V_{rel}(\mathbf{\Sigma})$ represents eccentricity of variation and is sensitive to differing scale changes between axes but not to rotation (change of eigenvectors), whereas $V_{rel}(\mathbf{P})$ represents magnitude of correlation and is insensitive to scale changes. Arrows schematically represent variation along major axes (whose directions are arbitrary when $V_{rel}(\mathbf{\Sigma}) = 0$).

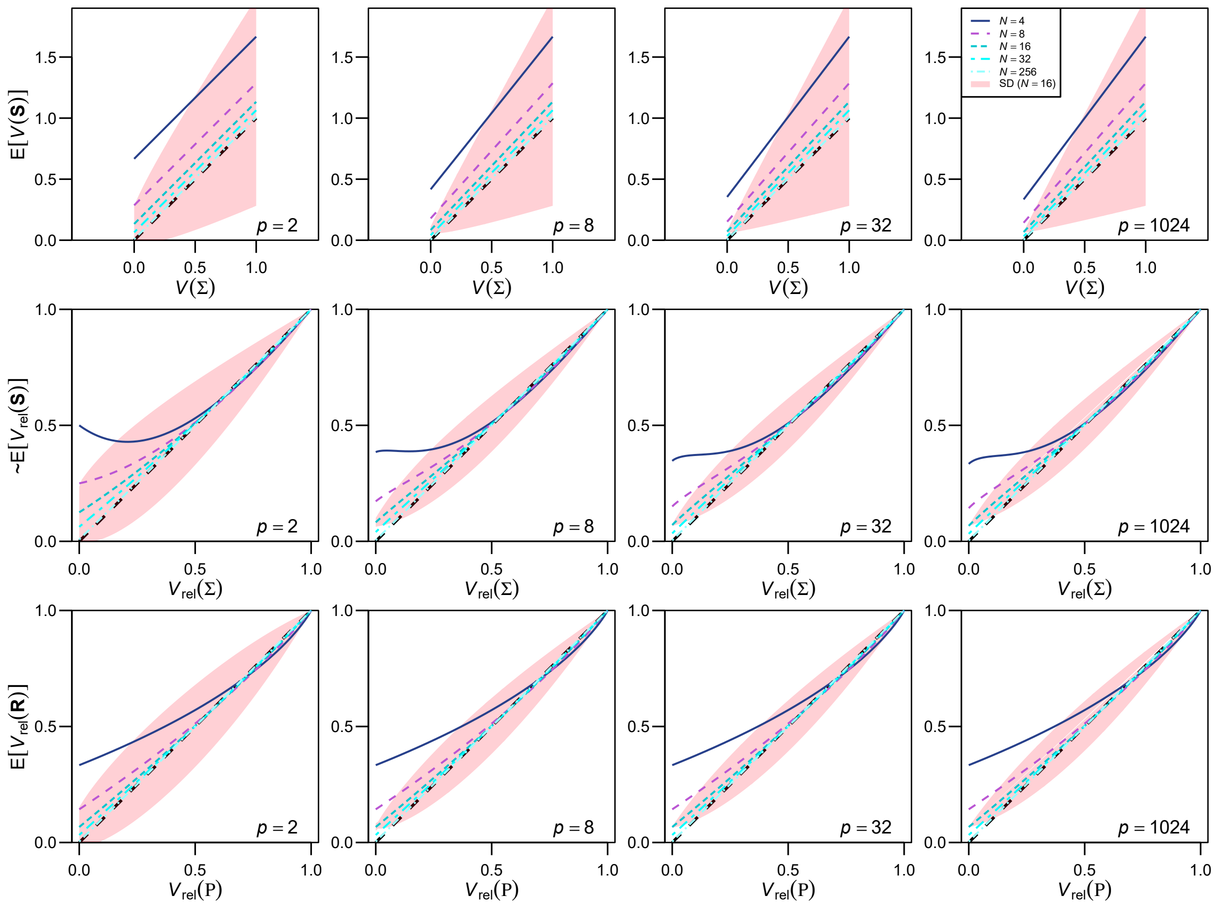

Figure 2. Profiles of the expectations of eigenvalue dispersion measures in selected conditions. The expectations of $V(\mathbf{S})$ (top row), $V_{\mathrm{rel}}(\mathbf{S})$ (approximate; middle row), and $V_{\mathrm{rel}}(\mathbf{R})$ (bottom row) are drawn with solid lines, for $p = 2$, 8, 32, and 1024 (from left to right) and for $N = 4$, 8, 16, 32, and 256 (from top to bottom on the left end of each box). In all cases, $n = N - 1$. The breadth of one standard deviation at $N = 16$ is also shown around the mean profiles with pink fills; these are approximations for $V_{\mathrm{rel}}(\mathbf{S})$ and for $V_{\mathrm{rel}}(\mathbf{R})$ with $p > 2$ (exact for $V_{\mathrm{rel}}(\mathbf{R})$ with $p = 2$). Note that actual distributions might be skewed unlike these fills. There are generally many suites of eigenvalues corresponding to a single value of V_{rel}, and $\mathrm{E}[V_{\mathrm{rel}}(\mathbf{R})]$ can also depend on eigenvector configurations; the profiles shown here are from such eigenvalue configurations that there is one large eigenvalue, with the rest being equally small, in which case $\mathrm{E}[V_{\mathrm{rel}}(\mathbf{R})]$ does not depend on eigenvector configurations. The population covariance matrix $\mathbf{\Sigma}$ is scaled so that $\mathrm{tr}(\mathbf{\Sigma}) = p(p - 1)^{-1/2}$. The initial decrease of the $\mathrm{E}[V_{\mathrm{rel}}(\mathbf{S})]$ profiles in some cases seems to be an artifact of approximation. See text for further technical details.

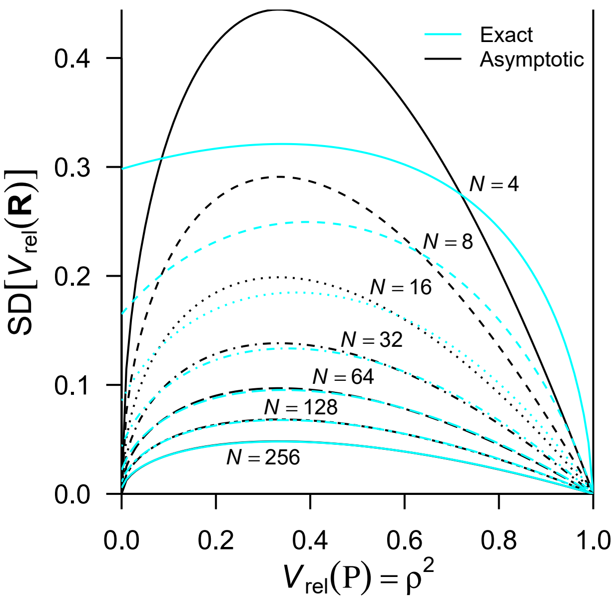

Figure 3. Comparison of exact and asymptotic standard deviations of $V_{\text{rel}}(\mathbf{R})$. Profiles of the exact (cyan lines) and asymptotic (black lines) standard deviations for $p = 2$ are shown across the entire range of the population value $V_{\text{rel}}(\mathbf{P})$, for $N = 4$, 8, 16, 32, 64, 128, and 256 (from top to bottom as labeled; shown with different line styles). Note that the asymptotic profiles converge to 0 when $V_{\text{rel}}(\mathbf{P}) = 0$.

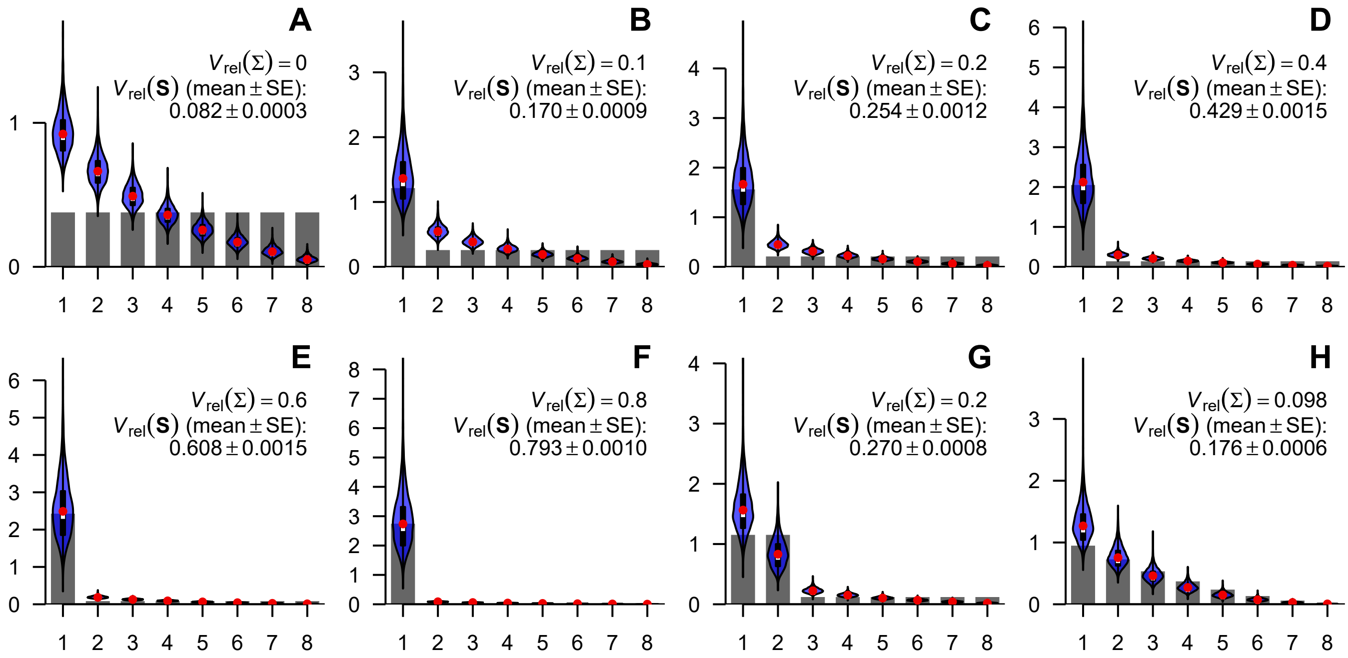

Figure 4. Selected population eigenvalue structures used in simulations and distributions of sample eigenvalues, examples for $p = 8$. The eigenvalues of population covariance matrix are shown as scree plots, and distributions of sample eigenvalues with $N = 16$ are shown as violin plots. **A**, null condition; **B–G**, q-large λ conditions, $q = 1$ (**B–F**) or 2 (**G**), with $V_{rel}(\mathbf{\Sigma}) = 0.1, 0.2, 0.4, 0.6, 0.8$, and 0.2, respectively; **H**, quadratically decreasing λ condition. Red dots denote empirical means of sample eigenvalues, whereas white bars (mostly overlapping with red dots) denote medians. Thick black bars within violins denote interquartile ranges. Note different scales of vertical axes.

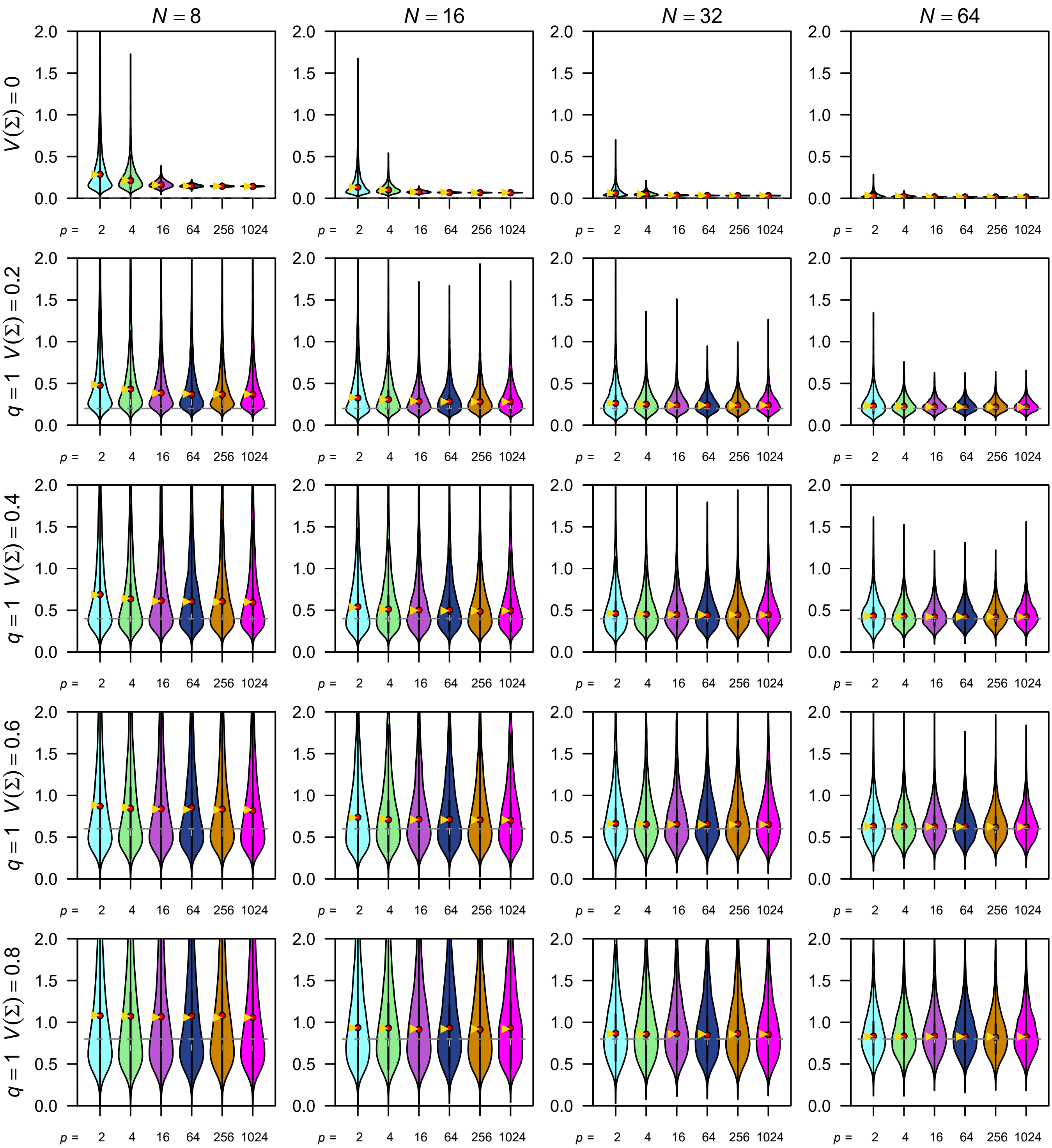

Figure 5. Selected results of simulation for the eigenvalue variance of covariance matrix $V(\mathbf{S})$. Empirical distributions of simulated $V(\mathbf{S})$ values are shown as violin plots, whose tails extend to the extreme values. Red dots denote empirical means, whereas yellow triangles denote expectations (which are exact). Thick black bars within violins denote interquartile ranges, with white bars near the center (in most cases overlapping with red dots) denote medians. Rows of panels correspond to varying population values of $V(\mathbf{\Sigma})$ (under 1-large λ conditions), whereas columns correspond to varying sample size N. Columns within each panel correspond to varying number of variables p. Note that extreme values in some panels are cropped for visual clarity. See Figure S2–S4 for full results.

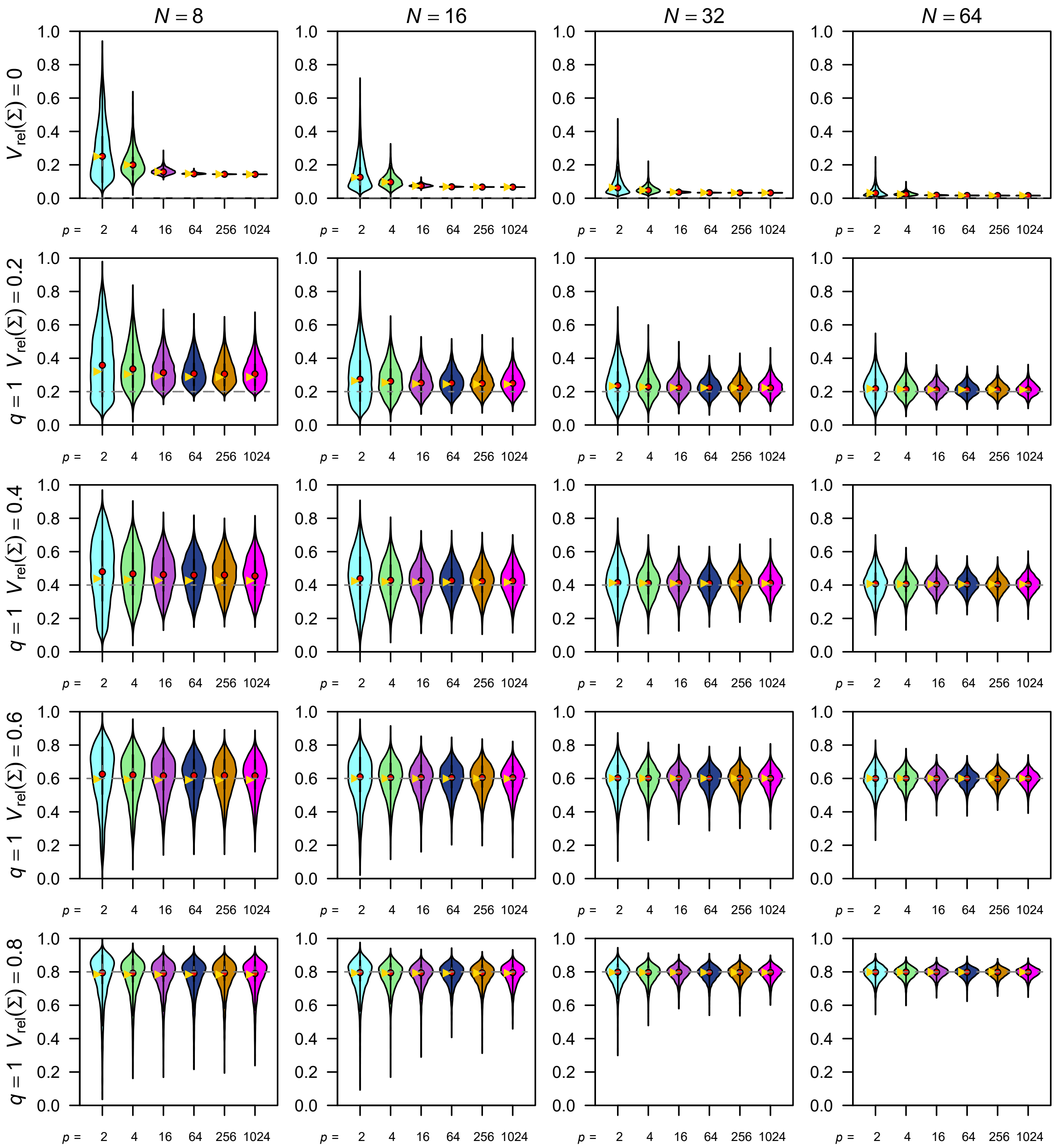

Figure 6. Selected results of simulation for the relative eigenvalue variance of covariance matrix $V_{\text{rel}}(\mathbf{S})$. Empirical distributions of simulated $V_{\text{rel}}(\mathbf{S})$ values are shown as violin plots. Yellow triangles denote expectations (which are approximate except under the null condition). Rows of panels correspond to varying population values of $V_{\text{rel}}(\mathbf{\Sigma})$ (under 1-large λ conditions). Other legends are as in Fig. 5. See Figure S4–S6 for full results.

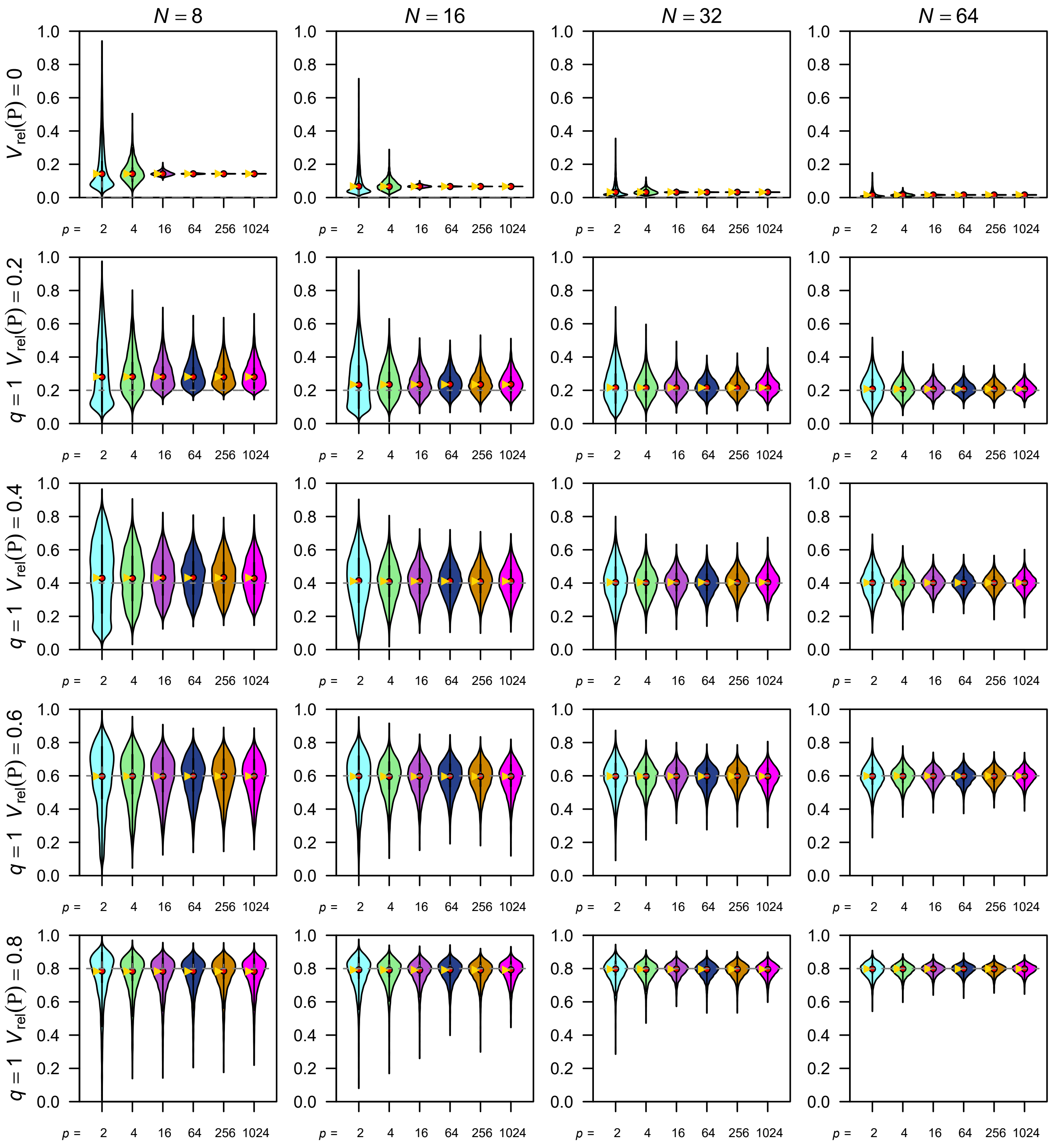

Figure 7. Selected results of simulation for the relative eigenvalue variance of correlation matrix $V_{rel}(\mathbf{R})$. Empirical distributions of simulated $V_{rel}(\mathbf{R})$ values are shown as violin plots. Yellow triangles denote expectations (which are exact). Rows of panels correspond to varying population values of $V_{rel}(\mathbf{P})$ (under 1-large λ conditions). Other legends are as in Fig. 5. See Figure S4, S7, and S8 for full results.