Examining Modularity via Partial Correlations: A Rejoinder to a Comment by Paul Magwene

PHILIPP MITTEROECKER\(^1\),\(^2\)∗ AND FRED L. BOOKSTEIN\(^3\),\(^4\)

\(^1\)Department of Theoretical Biology, University of Vienna, Althanstrasse 14, A-1091 Vienna, Austria;
\(^2\)Konrad Lorenz Institute for Evolution and Cognition Research, Adolf Lorenz Gasse 2, A-3422 Altenberg, Austria;
\(^3\)Department of Anthropology, University of Vienna, Althanstrasse 14, A-1091 Vienna, Austria;
\(^4\)Department of Statistics, University of Washington, Padréford B-207, Seattle, WA 98195-4322, USA;

∗Correspondence to be sent to: Department of Theoretical Biology, University of Vienna, Althanstrasse 14, A-1091 Vienna, Austria:
E-mail: philipp.mitteroecker@univie.ac.at.

We work, as Paul Magwene does, within the general tradition of path models set out nearly a century ago by Sewall Wright. In this approach, when 2 variables each depend linearly on the same cause, the part of their covariance that explanation accounts for is simply the product of the 2 path coefficients (regressions on the common cause) times the variance of that cause. Any additional covariance observed is attributed to additional common causes and is conveniently quantified by the partial correlation of the 2 outcome variables conditional on the first common cause (which is the correlation between the two residuals from the regression on this cause). A partial correlation thus combines a test for the adequacy of one particularly simple explanation of the covariances among observed variables together with an assessment of the maximum covariance possible that any additional factor, however hypothesized, might contribute.

The notion of correlations among residuals extend easily from the case of 3 variables to many. Pearl (2000) is a fine summary of the way in which this sort of logic can be ramified to incorporate fairly sophisticated statistical analyses about alternate causal interpretations of a data set. The first appearance of this elaboration in the context of morphological integration and modularity was apparently due to Paul Terentjev in the astoundingly early year of 1933. He investigated partial correlations among measurements on frogs conditioned on 1 measure of overall size together with 1 estimated more local morphogenetic factor. Similarly, in Mitteroecker and Bookstein (2007, 2008), we demonstrated interpretation of a pattern of covariances within and between modules after partialling out 1, 2, or more common factor estimates.

But it is a typical misapplication of partial correlation methods to partial out the wrong variables, or too many variables, thereby “throwing out the baby with the bathwater,” getting rid of the evidence for more than one causal interpretation at the same time. The extreme case consists of partialling every other variable of a measured data set out of the relation between any two of them. These quantities cannot all be meaningful, but there is no algorithmic impediment to computing them. As is well-known, they are all computed at once as the negatives of the off-diagonal cells of the scaled inverse of the observed covariance matrix among the original variables. It is this matrix that Magwene (2001) originally submitted to our attention as a conveyance for information about modularity. It does not serve in that capacity, as we noted in our comment of 2007. It overpartials, with one exception: when a module consists of only 2 variables.

When invoking a path model of local and common factors as suggested in Mitteroecker and Bookstein (2007, 2008), a model also adopted by Magwene in his reply, one coherent application of partial correlations would be the following. If common factors are estimated (e.g., by partial least squares or Wright-style factor analysis), the covariances or correlations conditional on these common factor scores reflect the statistical associations induced by local factors (i.e., modular factors). According to the classic theory, these correlations are supposed to be high within the modules but close to 0 between different modules. When the common factors are not estimated explicitly, the covariances between variables from the same module may be conditioned on all variables of the other modules as they are affected by the same common factors as well. The common factors—as far as they are reflected by these variables—are therefore “controlled for” in the analysis of within-module covariances, which should then be due to the local factors only. Of course, we would not condition the covariances on the local factors affecting this particular module, as it is the differential effect of these local factors that we want to study. Conditioning the covariances on variables of the same module (and thus also on the local factors) would actually control for the signal that we are interested in.

In his 2001 paper, Magwene instructs the reader to inspect the partial correlations conditional on all other variables, but, as we argued, those numbers do not correspond to any valid causal interpretation. Whenever the size of modules is greater than 2, the partial correlations are conditioned on all factors, whether common or
local, except those that affect only the particular pair of variables whose partial correlation is under inspection. Whenever there are more than 2 variables per module, it is not clear to which biological question this would supply the answer.

In his reply to our original argument, Magwene, after summarizing the classic path model arguments on which we agree, presents the skeleton of a path model with two modules of three variables each and some corresponding covariances (Figures 1a and 1b in Magwene 2009). He argues that a simulation using such a path model is consistent with his claims about the inverse covariance matrix. As he did not specify the actual path coefficients involved, however, we are free to assign them ourselves. In that event, we do not need to simulate anything—we can compute the results of Magwene’s (2001) original method by exact algebra, as follows. We assign the simplest possible range of numerical values to the necessary quantities in the path model: to wit, we set all the factor loadings to 1.0, all the factor variances to 1.0, and all the unique variances (variances of measured variables uncorrelated with factors and with all other variables) to 1.0. Crucially, we reserve the right to vary the number of variables per module. Let this count be \( k \), so that the total number of variables is \( 2k \).

The modeled covariance matrix \( \Sigma \) for 2 modules of \( k \) measurements each is then exactly

\[
\begin{pmatrix}
3 & 2 & \ldots & 2 & 2 & 1 & \ldots & 1 \\
2 & 3 & \ldots & 2 & 2 & 1 & \ldots & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
2 & 2 & \ldots & 3 & 2 & 1 & \ldots & 1 \\
2 & 2 & \ldots & 2 & 3 & 1 & \ldots & 1 \\
1 & 1 & \ldots & 1 & 1 & 3 & \ldots & 2 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 1 & \ldots & 1 & 1 & 2 & \ldots & 3
\end{pmatrix}
\]

of which the exact inverse is

\[
(3k^2 + 4k + 1)^{-1} \times 
\begin{pmatrix}
3k^2 + k - 1 & -(3k + 2) & \ldots & -(3k + 2) & -1 & \ldots & -1 \\
-(3k + 2) & 3k^2 + k - 1 & \ldots & -(3k + 2) & -1 & \ldots & -1 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
-(3k + 2) & -(3k + 2) & \ldots & 3k^2 + k - 1 & -1 & \ldots & -1 \\
-(3k + 2) & -(3k + 2) & \ldots & -(3k + 2) & 3k^2 + k - 1 & \ldots & -1 \\
-1 & -1 & \ldots & -1 & -1 & \ldots & -1 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
-1 & -1 & \ldots & -1 & -1 & \ldots & -1
\end{pmatrix}
\]

as can be verified by multiplying them in either order—their product is the identity. To terms of highest order, the exact (not simulated) partial correlations within block are of the order of \( 1/k \), and those between modules, \( 1/3k^2 \).

This exact algebra is not particularly helpful to Magwene’s claim. Even though the between-module correlations tend to be lower than the within-module correlations, to be statistically significant, a correlation of about \( 1/k \) requires a sample size of about \( 4k^2 \). For reasonable levels of effort at morphometric measurement, \( k \) will be of the order of tens, and the necessary sample size thus in the hundreds or thousands. This is not practicable for our more common applications, such as vertebrate anatomy. More paradoxically, as the number of variables rises, the strength of the evidence for modularity falls, ultimately to 0, regardless of the fact of modularity. This is a wholly undesirable property of a bio-metric method, as it punishes conscientiousness of measurement rather than rewarding it. In short, Magwene’s 3-variable model is not far enough from the 2-variable model to elucidate the properties of his 2001 method in the general case; those properties are highly undesirable in all realistic morphometric applications. From the matrix prototype above, readers can easily determine for themselves how fast the entries of the inverse covariance matrix drop to 0 when the number of variables increases.

In contrast to Magwene’s methodology, our approach demands prior knowledge of the local factors, or at least a heuristic hypothesis, and is not suited for the identification of modules. As we demonstrated in our 2007 paper, under most realistic assumptions, attempts at morphometric identification of modules are unreliable even when the correlations under study are all positive. But it is ultimately for biological rather than primarily mathematical reasons that no algebraic approach is suited for that identification task. Given that spatial autocorrelation among the original measurements is ubiquitous, it is always questionable whether a “successful” decomposition of the phenotypic covariance structure reflects the dissociation of local genetic and developmental factors or instead mere spatial contiguity or axial ordering. (Spatially adjacent variables tend to have higher correlations than more distant ones, even in the absence of dissociated developmental control, so that many arbitrary notions of spatially coherent modules would exhibit the same expected pattern of correlations.)

But there is an additional problem as well. In geometric morphometrics, which is our principal domain of application here, there is no direct connection between the existence of modules and the magnitude of correlations. In any shape-coordinate scheme, some covariances must be 0 by the nature of Procrustes registration, and likewise within the range of possible shape variables some measures must be uncorrelated with exogenous causes of form even within a single module. In general, then, no examination of off-diagonal parts of an inverse correlation matrix, nor any other geometry-free approach to partial correlations, can substitute for close inspection of the actual pattern of shape coordinates that is being decomposed.

In the end, empirical evidence about developmental factors and assessments of causal relationships provided by embryology and developmental biology is far more persuasive about modularity than merely morphometric identification could possibly be. In our view, it is the task of the morphometrician to quantify and calibrate general notions of modularity based on the
available data from outside morphometrics and to relate evolutionary or functional differences to these signals. Modules come from the organism, not from the algebra.

REFERENCES


Received 27 June 2008; reviews returned 22 February 2009; accepted 25 February 2009
Associate Editor: Norman MacLeod