Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Multivariate Analysis 97 (2006) 79–102

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

Latent models for cross-covariance

Jacob A. Wegelin[a,∗,1], Asa Packer[b], Thomas S. Richardson[c,2]

[a]Division of Biostatistics, Department of Public Health Sciences, School of Medicine, University of California, Davis, One Shields Avenue, MS 1-C, Davis, CA 95616, USA
[b]BioSonics, Inc., 4027 Leary Way NW, Seattle, WA 98107, USA
[c]University of Washington, Department of Statistics, Box 354322, Seattle, WA 98195, USA

Received 18 December 2003
Available online 20 January 2005

Abstract

We consider models for the covariance between two blocks of variables. Such models are often used in situations where latent variables are believed to present. In this paper we characterize exactly the set of distributions given by a class of models with one-dimensional latent variables. These models relate two blocks of observed variables, modeling only the cross-covariance matrix. We describe the relation of this model to the singular value decomposition of the cross-covariance matrix. We show that, although the model is underidentified, useful information may be extracted. We further consider an alternative parameterization in which one latent variable is associated with each block, and we extend the result to models with r-dimensional latent variables.

AMS 1991 subject classification: Primary 62H20, 62H25; Secondary 62M45, 62P10

Keywords: Canonical correlation; Latent variables; Partial least squares; Reduced-rank regression; Singular value decomposition

---

∗ Corresponding author. Tel./fax: +1 206 903 9540.
E-mail address: jawegelin@ucdavis.edu (J.A. Wegelin)
URL: http://wegelin.ucdavis.edu.

[1] Research supported in part by National Science Foundation Grant No. DMS-0071818.
[2] Research supported in part by National Science Foundation Grant No. DMS-9972008.

doi:10.1016/j.jmva.2004.11.009

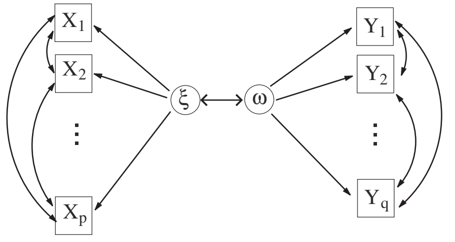

Fig. 1. A symmetric latent correlation model with a single pair of latent variables.

1. Introduction

Cross-covariance problems arise in the analysis of multivariate data that can be divided naturally into two blocks of variables, **X** and **Y**, observed on the same units. In a cross-covariance problem we are interested, not in the within-block covariances, but in the way the **Y**s vary with the **X**s.

The field of behavioral teratology furnishes examples of cross-covariance problems. In a study of the relationship between fetal alcohol exposure and neurobehavioral deficits reported by Sampson et al. [17] and by Streissguth et al. [23] there are 13 **X** variables, each corresponding to a different measure of the mother's reported alcohol consumption during pregnancy, and 11 **Y** variables, each corresponding to a different IQ subtest. In a study of the relationship between corpus callosum shape and neuropsychological deficits related to prenatal exposure to alcohol, there are 80 shape coordinates and 260 neuropsychological measures [5]. The researchers are not primarily interested in the relationships between the different measures of the mother's alcohol intake, between the shape coordinates, or between the different neuropsychological measures. In the first study the researchers are interested in the relationship between alcohol intake and IQ. In the second study they are interested in the relationship between callosal shape and behavioral impairment. None of these phenomena can be measured directly.

A natural model to associate with the cross-covariance problem is the symmetric *paired latent correlation model*.[3] A path diagram (corresponding to a *semi-Markovian system of equations*, [15, pp. 30, 141] is seen in Fig. 1. Path diagrams and their notation are defined in Appendix A. The paired latent correlation model family is formally specified in Section 2.2. The model depicted in Fig. 1 belongs to this family. With each block of observed variables is associated a scalar latent variable with unit variance, $\boldsymbol{\xi}$ for the **X** block and $\boldsymbol{\omega}$ for the **Y** block. The observed variables are linear functions of their parents, the latent variables, plus error. Correlated errors are indicated by bidirected edges. Under this model **X** and **Y** are conditionally independent given either or both of the latent variables.

A common approach in the factor analysis literature is to assume that within-block errors are uncorrelated. This approach is incompatible, however, with our wish to model only

[3] The term "symmetric" refers to the fact that both the **X**s and **Y**s are children of their respective latent variables, and that the latents have correlated errors. Asymmetric models will not be considered until Section 5.

between-block covariance. Thus in the current model the correlations of the within-block errors are unconstrained.

One problem with the latent variable model depicted in Fig. 1 is that it is underidentified. That is, there are parameter values that cannot be distinguished on the basis of data. Furthermore it is not clear to which set of distributions over the observed variables the model corresponds. In this paper we overcome these problems by showing that the latent model with k pairs of latent variables corresponds to the set of all distributions over the observed variables in which the cross-covariance, $\mathbf{\Sigma}_{XY}$, is of rank k. Consequently the latent model is appropriate for the setting we described, where we do not seek to model, and place no constraints on, the within-block covariance. Furthermore, in the case of a single pair of latent variables, or rank $(\mathbf{\Sigma}_{XY}) = 1$, this solution furnishes a precise answer to the question of identifiability. In estimating these models we are able to exploit well-developed methods: moment-based approaches using the singular value decomposition, and likelihood-based approaches used for reduced-rank regression.

As a corollary we prove covariance equivalence (see [15, p. 145]) of two latent-variable models containing different numbers of latent variables. To our knowledge this is the first result of this kind. Note furthermore that we have specified this model to be symmetric in $\mathbf{X}$ and $\mathbf{Y}$. In fact it will turn out that there are asymmetric variants that are equivalent to the symmetric model. Finally we contrast model equivalence results in the case where the errors are unrestricted with the situation where they are assumed to be diagonal.

For related work that attempts to characterize the set of distributions over the observed variables induced by a latent model see [6,18,19]. For work that attempts to solve the problem of identification of a single-factor model, see [22,24]. For other approaches to identifying relationships between variables, in situations where the response is unobserved, see Kuroki et al. [12].

2. Models

All models in the current work are specified only by the covariance matrix, also known as the variance-covariance or dispersion matrix. If we assume an underlying Gaussian density, then this specifies the distribution; however none of the results that shall be derived depend on any particular probability density function. Furthermore, it shall be seen in Section 6 that, in the examples cited from behavioral teratology, scientific findings were arrived at without assuming a particular density for the data.

2.1. Reduced-rank-regression models

A rank-r *reduced-rank-regression model* is the set of $(p+q)\times(p+q)$ positive semidefinite matrices satisfying a rank constraint on the cross-covariance matrix:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{YX} & \mathbf{\Sigma}_{YY} \end{bmatrix}, \ \text{rank}(\mathbf{\Sigma}_{XY}) = r, \ \text{where } \mathbf{\Sigma}_{XY} \ \text{is} \ p \times q. \tag{1}$$

Anderson discusses inference for this model [2].

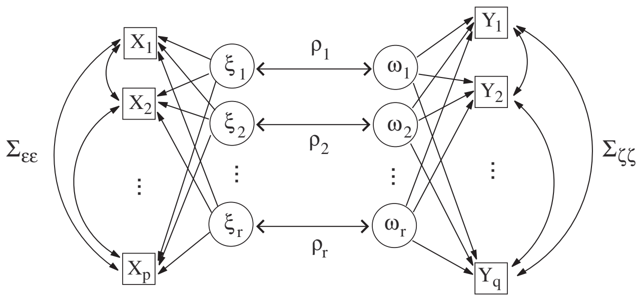

Fig. 2. Path diagram representing a symmetric paired latent correlation model, as specified in Section 2.2. Not depicted are edge coefficients $a_{11}, \ldots, a_{pr}$ and $b_{11}, \ldots, b_{qr}$, associated with the directed edges from $\boldsymbol{\xi}$ to the X variables and from $\boldsymbol{\omega}$ to the Y variables respectively.

2.2. *Paired latent correlation models*

The r-variate symmetric *paired latent correlation model* is a generalization of the model introduced in Section 1. It is the set of distributions over the latent r-vectors $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$, the observed variables $\mathbf{X}$ and $\mathbf{Y}$, and the errors $\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}$, specified as follows.

$$\left.\begin{array}{l}\mathbf{x} = \mathbf{A}\boldsymbol{\xi} + \boldsymbol{\epsilon} \text{ and } \mathbf{y} = \mathbf{B}\boldsymbol{\omega} + \boldsymbol{\zeta}, \\ \text{where} \\ \mathbf{Var}\left(\boldsymbol{\xi}^T, \boldsymbol{\omega}^T\right)^T = \begin{bmatrix} \mathbf{I}_r & \mathbf{R} \\ \mathbf{R} & \mathbf{I}_r \end{bmatrix}, \ \mathbf{R} = \operatorname{diag}\left(\rho_1, \ldots, \rho_r\right), \\ \mathbf{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}_{\varepsilon\varepsilon} \in I\!R^{(p\times p)}, \ \mathbf{Var}(\boldsymbol{\zeta}) = \boldsymbol{\Sigma}_{\zeta\zeta} \in I\!R^{(q\times q)}, \\ \boldsymbol{\epsilon}, \ \left(\boldsymbol{\xi}^T, \boldsymbol{\omega}^T\right)^T, \text{ and } \boldsymbol{\zeta} \text{ are mutually independent,} \\ \mathbf{A} \in I\!R^{(p\times r)}, \text{ and } \mathbf{B} \in I\!R^{(q\times r)}. \end{array}\right\} \tag{2}$$

Thus the parameters are $\rho_1, \ldots, \rho_r$, $\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, and $\boldsymbol{\Sigma}_{\zeta\zeta}$, where $|\rho_k| \leqslant 1$ for all k and where $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ and $\boldsymbol{\Sigma}_{\zeta\zeta}$ must be positive semidefinite. The columns of $\mathbf{A}$ and $\mathbf{B}$ are called loadings. A path diagram for a symmetric paired latent correlation model may be seen in Fig. 2. The reader will observe that this model is underidentified. For the case where there is a single pair of latent variables, however ($r = 1$), we shall precisely characterize the degree of non-identifiability, and suggest a natural convention that makes the model identifiable.

2.3. *Diagonal latent models*

The r-variate symmetric *diagonal latent model* is the set of distributions over the latent r-vector $\boldsymbol{\eta}$, the observed p-vector $\mathbf{X}$, the p-vector of errors $\boldsymbol{\epsilon}$, the observed q-vector $\mathbf{Y}$, and the q-vector of errors $\boldsymbol{\zeta}$, specified as follows.

$$\left.\begin{array}{l}\mathbf{x} = \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\epsilon} \text{ and } \mathbf{y} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta}, \\ \text{where} \\ \mathbf{Var}\left(\boldsymbol{\eta}\right) = \mathbf{I}_r, \ \mathbf{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}_{\varepsilon\varepsilon} \in I\!R^{(p\times p)}, \ \mathbf{Var}(\boldsymbol{\zeta}) = \boldsymbol{\Sigma}_{\zeta\zeta} \in I\!R^{(q\times q)}, \\ \boldsymbol{\epsilon}, \boldsymbol{\eta}, \text{ and } \boldsymbol{\zeta} \text{ are mutually independent,} \\ \mathbf{A} \in I\!R^{(p\times r)}, \text{ and } \mathbf{B} \in I\!R^{(q\times r)}. \end{array}\right\} \tag{3}$$

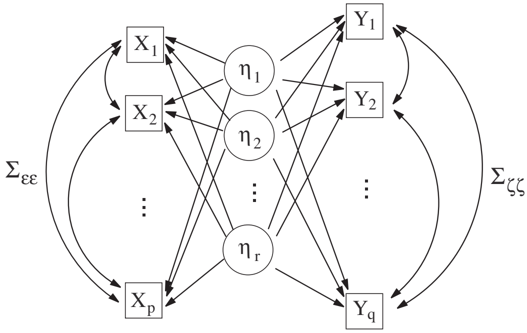

Fig. 3. Path diagram representing a symmetric diagonal latent model, as specified in Section 2.3. As in Fig. 2, the edge coefficients $a_{11}, \ldots, a_{pr}$ associated with the directed edges from $\boldsymbol{\xi}$ to the X variables, and $b_{11}, \ldots, b_{qr}$, associated with the directed edges from $\boldsymbol{\omega}$ to the Y variables, are not depicted.

Thus the parameters are the matrices $\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, and $\boldsymbol{\Sigma}_{\zeta\zeta}$, subject to the constraint that $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ and $\boldsymbol{\Sigma}_{\zeta\zeta}$ must both be positive semidefinite. The reader will observe that the r-variate diagonal latent model is a special case of the r-variate paired latent correlation model where $\boldsymbol{\xi} \equiv \boldsymbol{\omega}$. A path diagram for a symmetric diagonal latent model may be seen in Fig. 3.

3. Existence of latent parameterizations

Every set of parameter values for the symmetric paired latent correlation model induces a covariance matrix (1) over the observed variables as follows.

$$\left.\begin{aligned} \boldsymbol{\Sigma}_{XX} &= \mathbf{A}\mathbf{A}^T + \boldsymbol{\Sigma}_{\varepsilon\varepsilon}, \\ \boldsymbol{\Sigma}_{YY} &= \mathbf{B}\mathbf{B}^T + \boldsymbol{\Sigma}_{\zeta\zeta}, \\ \boldsymbol{\Sigma}_{XY} &= \mathbf{A}\mathbf{R}\mathbf{B}^T. \end{aligned}\right\} \tag{4}$$

A set of parameter values for the diagonal latent model induces a special case of this covariance matrix, with $\mathbf{R}$ equal to the identity.

The Eq. (4) define a map from the space of r-variate symmetric paired latent correlation models into the space of rank-r reduced-rank-regression models. The existence of such a map immediately raises the question whether every covariance in the rank-r reduced-rank-regression model can be obtained by a set of parameter values in the r-variate paired latent correlation model—i.e., is the map onto. The answer is yes. We say that the parameter values for the latent model *parameterize* or *are a parameterization of* the reduced-rank-regression covariance matrix that they induce. This result is stated as follows.

Theorem 1. *For each covariance matrix* (1) *in the rank-r reduced-rank-regression model there is at least one set of parameter values in the r-variate symmetric diagonal latent model that induces it.*

To prove this theorem we require the following lemma.

Lemma 2. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be real-valued matrices, each with n rows. Let* r_x *be the rank of* $\mathbf{X}$, r_y *the rank of* $\mathbf{Y}$, *and r the rank of* $\mathbf{X}^T\mathbf{Y}$. *Without loss of generality suppose* $r_x \leqslant r_y$. *Then there are matrices* $\mathbf{U}$ *and* $\mathbf{V}$ *such that* $\mathbf{U}$ *is a basis for the range (the column space) of* $\mathbf{X}$, $\mathbf{V}$ *is a basis for the range of* $\mathbf{Y}$, $\mathbf{U}^T\mathbf{U} = \mathbf{I}_{r_x}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{r_y}$, *and*

$$\mathbf{U}^T\mathbf{V} = [\mathbf{D}|\,\mathbf{0}], \tag{5}$$

where $\mathbf{0}$ *is an* $r_x \times (r_y - r_x)$ *matrix of zeroes, absent if* $r_x = r_y$, $\mathbf{D}$ *is an* $r_x \times r_x$ *diagonal matrix satisfying*

$$\mathbf{D} = \operatorname{diag}(\cos(\theta_1), \ldots, \cos(\theta_r), 0, \ldots, 0), \tag{6}$$

the last $(r_x - r)$ *diagonal entries of* (6) *are zero if* $r_x > r$, *and*

$$\cos(\theta_1) \geqslant \cdots \geqslant \cos(\theta_r) > 0.$$

Proof. Omitted. This is a restatement of Corollary 9.12 in Afriat [1]. □

Proof of Theorem 1. Suppose $\mathbf{\Sigma}$ is a covariance matrix satisfying (1). Let $\mathbf{X}$ and $\mathbf{Y}$ be matrices such that

$$\mathbf{X}^T\mathbf{X} = \mathbf{\Sigma}_{XX},\ \mathbf{X}^T\mathbf{Y} = \mathbf{\Sigma}_{XY}, \text{and } \mathbf{Y}^T\mathbf{Y} = \mathbf{\Sigma}_{YY}.$$

Such matrices are guaranteed to exist. For instance they may be obtained by partitioning the symmetric positive semidefinite square root of $\mathbf{\Sigma}$ ([8, p. 543]). Let n be the number of rows in $\mathbf{X}$, and let $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{D}$ be as in Lemma 2.

Then $\mathbf{U}$ is $n \times r_x$ and $\mathbf{V}$ is $n \times r_y$. Let $\mathbf{E}$ be an $r_x \times p$ matrix and $\mathbf{F}$ an $r_y \times q$ matrix such that

$$\mathbf{X} = \mathbf{UE},\ \mathbf{Y} = \mathbf{VF}. \tag{7}$$

Define the $p \times r_x$ matrix $\mathbf{A}$ and the $q \times r_x$ matrix $\mathbf{B}$ by

$$\begin{aligned} \mathbf{A} &= \mathbf{E}^T\sqrt{\mathbf{D}}, \\ \mathbf{B} &= \mathbf{F}^T\begin{bmatrix} \sqrt{\mathbf{D}} \\ \mathbf{0}^T \end{bmatrix}, \end{aligned}$$

where $\mathbf{D}$ and $\mathbf{0}$ have the same value as in (5). Then by (5)

$$\begin{aligned} \mathbf{AB}^T &= \mathbf{E}^T\mathbf{U}^T\mathbf{VF} \\ &= \mathbf{X}^T\mathbf{Y}. \end{aligned}$$

Then

$$\begin{aligned} \mathbf{\Sigma}_{XX} - \mathbf{AA}^T &= \mathbf{X}^T\mathbf{X} - \mathbf{E}^T\mathbf{DE} \\ &= \mathbf{E}^T\mathbf{U}^T\mathbf{UE} - \mathbf{E}^T\mathbf{DE} \\ &= \mathbf{E}^T\left(\mathbf{I}_{r_x} - \mathbf{D}\right)\mathbf{E} \\ &= \mathbf{E}^T\operatorname{diag}(1 - \cos(\theta_1), \ldots, 1 - \cos(\theta_r), 1, \ldots, 1)\,\mathbf{E}, \end{aligned} \tag{8}$$

a positive semidefinite matrix. By a similar argument

$$\begin{aligned} \mathbf{\Sigma}_{YY} - \mathbf{BB}^T &= \mathbf{F}^T\left(\mathbf{I}_q - \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right)\mathbf{F} \\ &= \mathbf{F}^T\operatorname{diag}(1 - \cos(\theta_1), \ldots, 1 - \cos(\theta_r), 1, \ldots, 1)\,\mathbf{F}, \end{aligned} \tag{9}$$

also positive semidefinite. Define the $p \times p$ matrix $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ and the $q \times q$ matrix $\boldsymbol{\Sigma}_{\zeta\zeta}$ by

$$\boldsymbol{\Sigma}_{\varepsilon\varepsilon} = \boldsymbol{\Sigma}_{XX} - \mathbf{A}\mathbf{A}^T,$$
$$\boldsymbol{\Sigma}_{\zeta\zeta} = \boldsymbol{\Sigma}_{YY} - \mathbf{B}\mathbf{B}^T.$$

The values of $\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, and $\boldsymbol{\Sigma}_{\zeta\zeta}$ satisfy the definition of an r-variate diagonal latent model, stated in (3), and they induce $\boldsymbol{\Sigma}$. □

Remark on the proof of Theorem 1. The columns of $\mathbf{U}$ and $\mathbf{V}$ are *principal vectors*, and the θ_k are *principal angles*. Golub and van Loan report an algorithm for computing $\mathbf{U}$ and $\mathbf{V}$ in the case when $\mathbf{X}$ and $\mathbf{Y}$ are of full rank [7, pp. 603f]. Björck and Golub [3] discuss numerical methods, including the case where $\mathbf{X}$ and $\mathbf{Y}$ are rank-deficient. In a statistical context the $\cos(\theta_k)$ are known as *canonical correlations* and the principal vectors as *canonical correlation variables* or *canonical variates*. Mardia, Kent and Bibby develop these concepts within a statistical context for the case where $\boldsymbol{\Sigma}$ has full rank [13, Chapter 10, pp. 281–299], as does Anderson [2]. The `cancor()` function in S-PLUS [14] may be used to compute canonical correlations. S-PLUS also computes two matrices, respectively $r_x \times r_x$ and $r_y \times r_y$, which may be used to compute $\mathbf{U}$ and $\mathbf{V}$ from $\mathbf{X}$ and $\mathbf{Y}$ provided $\mathbf{X}$ and $\mathbf{Y}$ have full rank.

Corollary 3. *The values of both $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ and $\boldsymbol{\Sigma}_{\zeta\zeta}$, derived in the proof of Theorem 1, are strictly positive definite if and only if $\boldsymbol{\Sigma}$ is strictly positive definite.*

Proof. $\boldsymbol{\Sigma}$ is strictly positive definite if and only if the columns of the combined matrix $[\mathbf{X}|\mathbf{Y}]$ are linearly independent. This condition holds if and only if the following three conditions hold.

(1) The columns of $\mathbf{X}$ are linearly independent of the columns of $\mathbf{Y}$, so that the first principal angle satisfies $\cos(\theta_1) < 1$ (the first canonical correlation is less than one in absolute value). Note that this is the only way that

$$\text{diag}\,(1 - \cos(\theta_1), \ldots, 1 - \cos(\theta_r), 1, \ldots, 1)$$

can have full rank.

(2) The following equivalent conditions hold.
- The columns of $\mathbf{X}$ are linearly independent.
- $r_x = p$.
- $\text{rank}(\mathbf{E}) = p$.

(i) The following equivalent conditions hold.
- The columns of $\mathbf{Y}$ are linearly independent.
- $r_y = q$.
- $\text{rank}(\mathbf{F}) = q$.

Thus if $\boldsymbol{\Sigma}$ is strictly positive definite, both (8) and (9) are of full rank, that is, strictly positive definite. Suppose on the other hand that (8) and (9) are of full rank. The matrix at (9) is $p \times p$, the product of a $p \times r_x$ matrix, an $r_x \times r_x$ matrix, and an $r_x \times p$ matrix. Since $r_x \leqslant p$, this matrix can be of full rank only if $r_x = p$. Furthermore it can be of full rank only if the middle matrix is of full rank, which requires $\rho_1 < 0$. Similarly if (8) is of full rank it follows that $r_y = q$ and $\rho_1 < 0$. Thus $\boldsymbol{\Sigma}$ is of full rank. □

Remark on Corollary 3. Corollary 3 notwithstanding, for a given strictly positive definite covariance matrix in the reduced-rank-regression model there may be parameterizations, different from those derived in the proof of Theorem 1, with singular within-block covariance. For instance, we shall see in Theorem 6 that, if we hypothesize a single pair of latent variables, we can always parameterize the model such that the within-block error covariance matrices are singular for both blocks.

Corollary 4. *Each rank-r reduced-rank-regression model can be parameterized by at least one r-variate paired latent correlation model.*

Proof. Recall that the rank-r diagonal latent model is the special case of the rank-r paired latent correlation model where the paired latent variables $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ are not only perfectly correlated, but equal. Thus, given a covariance matrix (1) in the rank-r reduced-rank-regression model, to construct a paired latent correlation parameterization, we need only let $\boldsymbol{\eta}$ be the latent variable of the diagonal latent model guaranteed by Theorem 1, and let $\boldsymbol{\xi} \equiv \boldsymbol{\omega} \equiv \boldsymbol{\eta}$. Then the parameters of the paired latent correlation model are obtained by setting $\mathbf{R} = \mathbf{I}_r$, and letting $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, $\boldsymbol{\Sigma}_{\zeta\zeta}$, $\mathbf{A}$, and $\mathbf{B}$ equal the values obtained for the diagonal latent model. □

Remark on Corollary 4. The correlations between $\boldsymbol{\xi}_k$ and $\boldsymbol{\omega}_k$, written ρ_k, are distinct from the canonical correlations that appear in the proof of Theorem 1. The correlation between latents within pairs, guaranteed by Corollary 4, is unity for each of the r pairs of latent variables. The canonical correlation that appears in the proof of Theorem 1, on the other hand, is only unity if $\boldsymbol{\Sigma}$ is singular.

3.1. Equivalence of model spaces

Consider the following three spaces of covariance matrices over the observed variables $\mathbf{X}$ and $\mathbf{Y}$.

(1) Those corresponding to the reduced-rank-regression model.
(2) Those induced by the symmetric paired latent correlation model.
(3) Those induced by the symmetric diagonal latent model.

It follows from definitions and from Eqs. (4) that Set 3 $\subset$ Set 2 $\subset$ Set 1. By Theorem 1, however, we see that any rank-r reduced-rank-regression model can be parameterized by a symmetric diagonal latent model. Thus Set 1 $\subset$ Set 3. Hence Set 1 = Set 2 =Set 3. We state this as the following corollary.

Corollary 5. *The sets of covariance matrices over the observed variables induced by the r-variate symmetric paired latent correlation model and the r-variate symmetric diagonal latent model are equal to the set of covariance matrices belonging to the rank-r reduced-rank-regression model.*

4. Characterization of univariate latent parameterizations

We now restrict our attention to the case where $\boldsymbol{\Sigma}_{XY}$ has unit rank. In this case, for each distribution $\boldsymbol{\Sigma}$ in the reduced-rank-regression model (1) there correspond precise bounds on the values of the parameters in the diagonal latent and paired latent correlation models. In particular, the loadings, which specify the linear relationship between the latent variable for a block and the observed variables of that block, are identified up to sign and scale.

In the case of a single pair of latent variables, we replace the notation $\mathbf{A}$ and $\mathbf{B}$ in (2) and (3) by $\mathbf{a}$ and $\mathbf{b}$, to emphasize that each matrix of loadings is a column vector. A set of parameter values for the diagonal latent model that parameterize (1) will satisfy the following conditions:

$$\left.\begin{array}{c} \mathbf{\Sigma}_{XX} = \mathbf{a}\mathbf{a}^T + \mathbf{\Sigma}_{\varepsilon\varepsilon}, \\ \mathbf{\Sigma}_{YY} = \mathbf{b}\mathbf{b}^T + \mathbf{\Sigma}_{\zeta\zeta}, \\ \mathbf{\Sigma}_{\varepsilon\varepsilon} \text{ and } \mathbf{\Sigma}_{\zeta\zeta} \text{ positive semidefinite} \end{array}\right\} \tag{10}$$

and

$$\mathbf{\Sigma}_{XY} = \mathbf{a}\mathbf{b}^T. \tag{11}$$

We summarize our result as follows.

Theorem 6. *Let $\mathbf{\Sigma}$ be as in (1), with rank $r = 1$. Then*

(1) *There are positive constants $\alpha_{\min}$ and $\alpha_{\max}$, depending on $\mathbf{\Sigma}$,* with $\alpha_{\min} \leqslant \alpha_{\max}$, *such that each element in the equivalence class of diagonal latent parameterizations of $\mathbf{\Sigma}$ is determined by a point $\alpha \in [\alpha_{\min}, \alpha_{\max}]$. For $\alpha = \alpha_{\min}$, the dispersion matrix of the within-block errors of the $\mathbf{Y}$ block, $\mathbf{\Sigma}_{\zeta\zeta}$, is singular. For $\alpha = \alpha_{\max}$, the dispersion matrix of the within-block errors of the $\mathbf{X}$ block, $\mathbf{\Sigma}_{\varepsilon\varepsilon}$, is singular.*

(2) *Each element in the equivalence class of paired latent correlation parameterizations of $\mathbf{\Sigma}$ is determined by a point in the following set:*

$$\left\{(\rho, \alpha) : \ |\rho| \leqslant 1 \ \textit{and} \ \frac{\alpha_{\min}}{|\rho|} \leqslant \alpha \leqslant \alpha_{\max}\right\}. \tag{12}$$

(3) *There are vectors $\mathbf{u}$ and $\mathbf{v}$, and a scalar d, all depending on $\mathbf{\Sigma}_{XY}$, such that the loadings for any diagonal latent or paired latent correlation parameterization of $\mathbf{\Sigma}$ can be expressed as*

$$\mathbf{a} = \alpha\mathbf{u} \quad \text{and} \quad \mathbf{b} = \frac{\mathbf{v}}{\alpha\rho} d.$$

(4) *In any paired latent correlation parameterization of $\mathbf{\Sigma}$,*

$$|\rho| \geqslant \frac{\alpha_{\min}}{\alpha_{\max}}$$

and when the bound is attained, both $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ and $\mathbf{\Sigma}_{\zeta\zeta}$ are singular.

(5) *If the set (12) defined in Part 2 contains only one point, then $\mathbf{\Sigma}$ is singular.*

To prove Theorem 6 we require the following lemma.

Lemma 7. *Let $\mathbf{A}$ and $\mathbf{C}$ be symmetric matrices of the same dimension, $\mathbf{C}$ positive semidefinite. Let $h : [0, \infty) \mapsto I\!R$ be defined by*

$$h(\alpha) = \textit{the smallest eigenvalue of } (\mathbf{A} - \alpha\mathbf{C}).$$

Then

(1) *The function h is monotone nonincreasing. If* $\mathbf{C}$ *is strictly positive definite, the function is strictly monotone decreasing.*

(2) $\lim_{\alpha\downarrow 0} h(\alpha) = h(0)$.

(3) *If* $\mathbf{C}$ *has at least one positive eigenvalue*, $\lim_{\alpha\uparrow\infty} h(\alpha) = -\infty$.

Proof. Let $\mathbf{z}(\alpha)$ be the eigenvector belonging to the smallest eigenvalue of $(\mathbf{A} - \alpha\mathbf{C})$, without loss of generality let $||\mathbf{z}(\alpha)|| = 1$, and recall that, with this convention, $\mathbf{z}(\alpha)^T\,(\mathbf{A} - \alpha\mathbf{C})\,\mathbf{z}(\alpha)$ equals the smallest eigenvalue.

Part 1: Let $\beta > \alpha$.

$$\begin{aligned} h(\alpha) &= \mathbf{z}(\alpha)^T\,(\mathbf{A} - \alpha\mathbf{C})\,\mathbf{z}(\alpha) \\ &= \mathbf{z}(\alpha)^T\mathbf{A}\mathbf{z}(\alpha) - \alpha\mathbf{z}(\alpha)^T\mathbf{C}\mathbf{z}(\alpha) \\ &\geqslant \mathbf{z}(\alpha)^T\mathbf{A}\mathbf{z}(\alpha) - \beta\mathbf{z}(\alpha)^T\mathbf{C}\mathbf{z}(\alpha) \qquad (13) \\ &= \mathbf{z}(\alpha)^T\,(\mathbf{A} - \beta\mathbf{C})\,\mathbf{z}(\alpha) \\ &\geqslant \mathbf{z}(\beta)^T\,(\mathbf{A} - \beta\mathbf{C})\,\mathbf{z}(\beta) \qquad (14) \\ &= h(\beta). \end{aligned}$$

At line (13) the inequality is not strict because possibly $\mathbf{z}(\alpha)^T\mathbf{C}\mathbf{z}(\alpha) = 0$. If $\mathbf{C}$ is strictly positive definite, the inequality at this line is strict and hence h is strictly decreasing. The inequality at line (14) occurs because $\mathbf{z}(\beta)$, by definition, minimizes the quadratic form.

Part 2: This is a consequence of a well-known theorem regarding the eigenvalues of a diagonalizable matrix under perturbation. See, for example, [10, Theorem 6.3.2, p. 365].

Part 3: Let $\mathbf{y}$ be an eigenvector corresponding to a positive eigenvalue of $\mathbf{C}$.

$$\min_{||\mathbf{x}||=1} \mathbf{x}^T\,(\mathbf{A} - \alpha\mathbf{C})\,\mathbf{x} \;\leqslant\; \mathbf{y}^T\mathbf{A}\mathbf{y} - \alpha\mathbf{y}^T\mathbf{C}\mathbf{y}.$$

The first term is constant. The second term approaches negative infinity as α approaches infinity. □

Proof of Theorem 6. Let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{\Sigma}_{\varepsilon\varepsilon}$, and $\mathbf{\Sigma}_{\zeta\zeta}$ be a diagonal latent parameterization of $\mathbf{\Sigma}$, guaranteed by Theorem 1, and accordingly decompose $\mathbf{\Sigma}$ as follows:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{a}\mathbf{a}^T & \mathbf{a}\mathbf{b}^T \\ \mathbf{b}\mathbf{a}^T & \mathbf{b}\mathbf{b}^T \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} \mathbf{\Sigma}_{\varepsilon\varepsilon} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\zeta\zeta} \end{bmatrix},$$

so that

$$\mathbf{\Sigma} = \mathbf{Q} + \mathbf{E}.$$

Following the convention of the singular value decomposition, let us define vectors $\mathbf{u}$, and $\mathbf{v}$, and scalar d, such that

$$\mathbf{\Sigma}_{XY} = \mathbf{u}\mathbf{v}^T d, \;\; ||\mathbf{u}|| = ||\mathbf{v}|| = 1,$$

where $||\cdot||$ represents the Euclidean norm. We may do this because $\mathbf{\Sigma}_{XY}$ has unit rank. Furthermore, let us assume that a sign convention has been adopted, so that $\mathbf{u}$ and $\mathbf{v}$ are determined uniquely. Now, for $\alpha > 0$, let

$$\mathbf{a}(\alpha) \equiv \alpha\mathbf{u}, \;\; \mathbf{b}(\alpha) \equiv \frac{\mathbf{v}d}{\alpha}.$$

Then $\mathbf{a}(\alpha)$ and $\mathbf{b}(\alpha)$ satisfy $\mathbf{\Sigma}_{XY} = \mathbf{a}(\alpha)\,[\mathbf{b}(\alpha)]^T$. Consider the set of positive real numbers α such that the matrices defined by

$$\begin{aligned}\mathbf{\Sigma}_{\varepsilon\varepsilon}(\alpha) &\equiv \mathbf{\Sigma}_{XX} - \mathbf{a}(\alpha)\,[\mathbf{a}(\alpha)]^T \text{ and}\\ \mathbf{\Sigma}_{\zeta\zeta}(\alpha) &\equiv \mathbf{\Sigma}_{YY} - \mathbf{b}(\alpha)\,[\mathbf{b}(\alpha)]^T\end{aligned} \tag{15}$$

are positive semidefinite, satisfying the model definition in (3). By Theorem 1 this set contains at least one point. We call this the *feasible set for the diagonal latent model*.

Part 1: It remains to show that this set is a closed bounded interval, and to characterize the dispersion matrices when α is on the boundary. Define $f : (0,\infty) \mapsto \mathbb{R}$ and $g : (0,\infty) \mapsto \mathbb{R}$ by

$$\begin{aligned}f(\alpha) &= \min\{\text{eigenvalues of } \mathbf{\Sigma}_{\varepsilon\varepsilon}(\alpha)\},\\ g(\alpha) &= \min\left\{\text{eigenvalues of } \mathbf{\Sigma}_{\zeta\zeta}(\alpha)\right\}.\end{aligned} \tag{16}$$

It may be shown that these functions are continuous ([10]). Furthermore, $\mathbf{\Sigma}_{XY}$ is not identically zero, since it has unit rank, and thus neither $\mathbf{a}$ or $\mathbf{b}$ is identically zero. It follows from Lemma 7 that

- f is monotone nonincreasing and goes to $-\infty$ as $\alpha \to \infty$;
- g is monotone nondecreasing and goes to $-\infty$ as $\alpha \downarrow 0$.

Let

$$\mathcal{F} = \{\alpha : f(\alpha) < 0\} \text{ and } \mathcal{G} = \{\alpha : g(\alpha) < 0\}.$$

By the continuity of f and g these sets are open, but by monotonicity they are intervals. Let α_1 and α_2 be the finite endpoints of $\mathcal{F}$ and $\mathcal{G}$ respectively, so that we have

$$\mathcal{F} = (\alpha_1, \infty) \text{ and } \mathcal{G} = (0, \alpha_2).$$

The set of feasible α values is $\mathbb{R} \setminus (\mathcal{F} \cup \mathcal{G}) = [\alpha_2, \alpha_1]$, where $\alpha_2 \leqslant \alpha_1$ follows from the fact that this set is nonvoid. Let us therefore call these values respectively $\alpha_{\min}$ and $\alpha_{\max}$. Following the definitions at (15) and (16), $\mathbf{\Sigma}_{\varepsilon\varepsilon}(\alpha_{\min})$ and $\mathbf{\Sigma}_{\zeta\zeta}(\alpha_{\max})$ are singular.

Part 2: For each point (ρ, α) in the set defined at (12), we specify a complete set of parameter values for the symmetric paired latent correlation model as follows:

$$\mathbf{a} = \alpha\mathbf{u},\;\; \mathbf{b} = \frac{\mathbf{v}}{\alpha\rho}d,\;\; \mathbf{\Sigma}_{\varepsilon\varepsilon} = \mathbf{\Sigma}_{XX} - \mathbf{a}\mathbf{a}^T, \text{ and } \mathbf{\Sigma}_{\zeta\zeta} = \mathbf{\Sigma}_{YY} - \mathbf{b}\mathbf{b}^T. \tag{17}$$

It is straightforward to check that these parameter values satisfy the model definition in (2). If any of the constraints defining the set (12) is violated, on the other hand, at least one of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ and $\mathbf{\Sigma}_{\zeta\zeta}$ will fail to be positive semidefinite, or we will have $|\rho| > 1$, and the parameterization will fail. Thus the set defined at (12) characterizes all paired latent correlation parameterizations.

Part 3: Since the diagonal latent model is a constrained version of the paired latent correlation model with $\rho = 1$, this has been proved.

Part 4: The constraints in (12) imply that $|\rho| \geqslant \dfrac{\alpha_{\min}}{\alpha_{\max}}$. If we set ρ equal to this bound, it also follows from (12) that $\alpha \geqslant \alpha_{\max}$, i.e., $\alpha = \alpha_{\max}$, so that $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ is singular. But by plugging this value of ρ into (17) we obtain

$$\mathbf{\Sigma}_{\zeta\zeta} = \mathbf{\Sigma}_{YY} - \frac{d^2}{\alpha^2_{\min}}\mathbf{v}\mathbf{v}^T$$

and by the definition of $\alpha_{\min}$ it follows that $\mathbf{\Sigma}_{\zeta\zeta}$ is singular.

Part 5: If the set (12) contains only one point, then the latent model is identified, and by Part 4 the unique values of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ and $\mathbf{\Sigma}_{\zeta\zeta}$ are both singular. Thus rank $(\mathbf{\Sigma}_{\varepsilon\varepsilon}) < p$ and rank $\left(\mathbf{\Sigma}_{\zeta\zeta}\right) < q$. By definition, rank $(Q) = 1$. By a theorem regarding the rank of the sum of two matrices (see, for instance, [10, p. 13]), rank $(\mathbf{E}) \leqslant p+q-2$, so that rank $(\mathbf{\Sigma}) \leqslant p+q-1$ and the matrix is singular. □

Remarks on Theorem 6. Theorem 6 defines a map from $\mathbf{\Sigma}$, the population covariance matrix for the observed variables in a reduced-rank-regression model, to the set of all symmetric diagonal latent and paired latent correlation parameterizations. Since maximum-likelihood estimation procedures are available for reduced-rank-regression [2], the problem of maximum-likelihood estimation for these latent models is solved.

Let us call the lower bound on correlation $\rho_{\min}$. Given a distribution $\mathbf{\Sigma}$, the choice of the quantity ρ within the feasible interval $\left[\rho_{\min}, 1\right]$ entails a tradeoff. When $\rho = \rho_{\min}$, the covariance matrices $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ and $\mathbf{\Sigma}_{\zeta\zeta}$ are singular. When $\rho = 1$, on the other hand, at least one of the error variances may be nonsingular. Thus we may choose to have the latent variables either perfectly correlated ($\rho = 1$) but poorly measured, or measured with minimal error but poorly correlated with each other ($\rho = \rho_{\min}$). This latter choice guarantees that the parameters are identified up to the sign and scale of the loadings.

For any fixed $\rho > \rho_{\min}$, and in particular for the diagonal latent model when $\rho_{\min} < 1$, there are many feasible values for α, and the choice of α entails another tradeoff. By choosing a greater value of α we choose to diminish the least eigenvalue of the error covariance for the **X** block and to increase the corresponding value for the **Y** block.

A schematic diagram of a feasible set for positive values of ρ may be seen in Fig. 4. A numeric example may be found in [27].

Although the singularity of $\mathbf{\Sigma}$ is a necessary condition for the parameterization to be unique, it is not sufficient. This may be seen by the following example. The matrix at (18) has eigenvalues 4.56, 1, 0.44, and 0.

$$\mathbf{\Sigma} = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \tag{18}$$

Let $p = q = 2$. The matrix (18) represents a degenerate distribution, since the two **Y** variables are perfectly correlated. The fact that there are infinitely many feasible parameterizations follows from the fact that $\mathbf{v}\mathbf{v}^T$ is proportional to $\mathbf{\Sigma}_{YY}$. The feasible set is $\left[\sqrt{2}, \sqrt{3}\right]$. The value $\alpha = \sqrt{2}$ entails zero error for the **Y**-block, so that each **Y** variable measures the latent $\boldsymbol{\eta}$ exactly. All feasible values of α, however, entail a singular error covariance for the **Y**-block. For all values of α, whether feasible or not, $\mathbf{\Sigma}_{\zeta\zeta}(\alpha) = \begin{bmatrix} \delta & \delta \\ \delta & \delta \end{bmatrix}$ for some $\delta \in I\!R$.

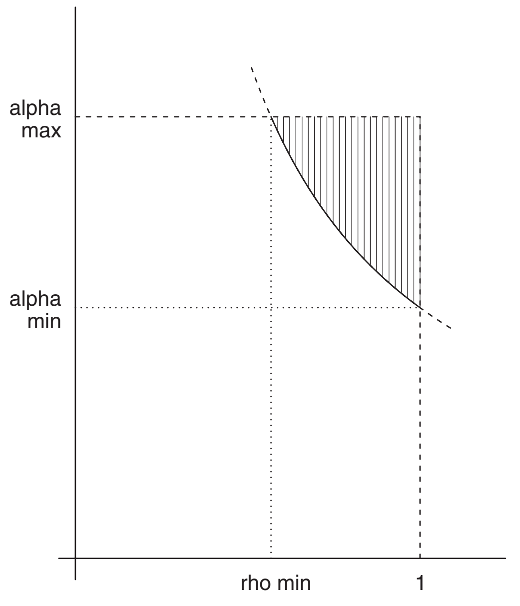

Fig. 4. Schematic of the feasible values for ρ and α in the paired latent correlation model, as developed in Theorem 6. Feasible values are in the shaded region in the upper right-hand corner of the rectangle. The right-hand boundary of the feasible set corresponds to the diagonal latent model. The feasible value guaranteed by Theorem 1 lies on this boundary.

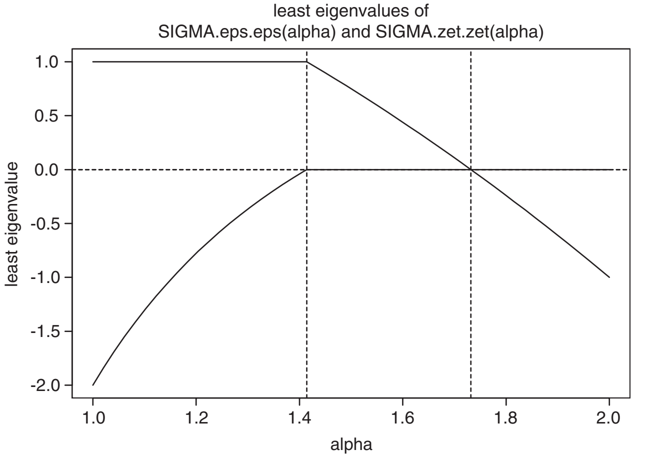

Fig. 5. The least eigenvalues of $\mathbf{\Sigma}_{\varepsilon\varepsilon}(\alpha)$ (the nonincreasing function) and $\mathbf{\Sigma}_{\zeta\zeta}(\alpha)$ for the matrix (18), p. 16.

When $\sqrt{2} \leqslant \alpha$, so that $\delta \geqslant 0$, the least eigenvalue is 0. For $\alpha < \sqrt{2}$, hence $\delta < 0$, $\mathbf{\Sigma}_{\zeta\zeta}(\alpha)$ is not a covariance and the least eigenvalue is strictly increasing in α. The least eigenvalues are plotted against α in Fig. 5, p. 17.

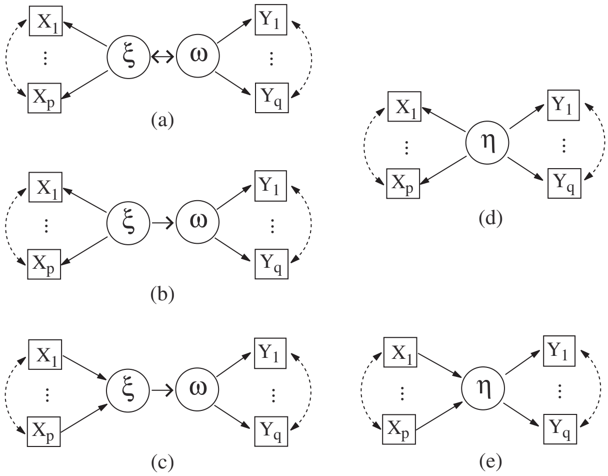

Fig. 6. Path diagrams corresponding to two-block latent variable models. Under (I) the dashed edges are present; under (II) they are absent. Under (I) all models are covariance equivalent over **X** and **Y**.

5. Related equivalence results

All the models we have considered so far place no restrictions on the within-block covariance matrices; we have allowed the error terms in each block to have an arbitrary covariance matrix. This is in keeping with motivation for *cross*-covariance modeling: the goal is only to model the relations between blocks of variables.

In this section, we present a number of model equivalence results concerning cross covariance models, which are important for the purposes of interpretation. [4] We then contrast these results with the case in which the errors within each block are assumed to be uncorrelated, as is common in the factor analysis and structural equation modeling literature.

The models with uncorrelated errors typically are identified after fixing the variance of the latent variable. Thus these models could be fitted to the behavioral teratology data. However, to do so would be artificial since the researchers in this study did not believe that it was reasonable to suppose that errors within blocks were uncorrelated. Our primary purpose in considering models with uncorrelated errors is theoretical: we wish to examine the implications that this additional assumption has for model equivalence.

All the graphs we consider are shown in Fig. 6. Graphs (a) and (b) represent two path diagrams in which the latent variables $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ are parents of the observed variables. The only difference between the models is that (a) specifies that $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ are correlated, while in (b) $\boldsymbol{\xi}$ is a parent of $\boldsymbol{\omega}$. The graph shown in Fig. 6 (c) differs from that shown in (b) in that the **X** variables are parents of $\boldsymbol{\xi}$. The graph in (d) is analogous to (a) and (b) but the pair of latent variables $\boldsymbol{\xi}$, $\boldsymbol{\omega}$ are replaced by a single variable. Likewise (e) represents the analogue

[4] An earlier version of these results, in the context of the multivariate normal distribution has appeared previously in conference proceedings [26].

to (c) with a single latent variable. Graphs (c) and (e) are instances of MIMIC models, so called because they contain observed variables that are Multiple Indicators and MultIple Causes of latent variables (see [4, pp. 331, 397] for examples).

We consider the five models corresponding to these graphs under two sets of conditions on the error terms:

(I) $\mathbf{Cov}\left(\epsilon_i, \zeta_j\right) = 0$, but $\mathbf{Cov}\,(\epsilon_i, \epsilon_k)$ and $\mathbf{Cov}\left(\zeta_j, \zeta_\ell\right)$ are unrestricted.

(II) $\mathbf{Cov}\left(\epsilon_i, \zeta_j\right) = 0$, and
$\mathbf{Cov}\,(\epsilon_i, \epsilon_k) = \mathbf{Cov}\left(\zeta_j, \zeta_\ell\right) = 0$ for $i \neq k, j \neq \ell$.

Let $\mathcal{M}_{\mathrm{a}}^{\mathrm{I}}$ denote the set of covariance matrices over $\mathbf{X}$ and $\mathbf{Y}$ given by graph (a) in Fig. 6 under condition (I) on the errors, likewise for $\mathcal{M}_{\mathrm{a}}^{\mathrm{II}}$, $\mathcal{M}_{\mathrm{b}}^{\mathrm{I}}$, $\mathcal{M}_{\mathrm{b}}^{\mathrm{II}}$ and so on. Corollary 5 of the previous section thus shows that $\mathcal{M}_{\mathrm{a}}^{\mathrm{I}} = \mathcal{M}_{\mathrm{d}}^{\mathrm{I}}$. We extend these results further in the next theorem.

Theorem 8. *The following relations hold*:

$$\mathcal{M}_{\mathrm{a}}^{\mathrm{I}} = \mathcal{M}_{\mathrm{b}}^{\mathrm{I}} = \mathcal{M}_{\mathrm{c}}^{\mathrm{I}} = \mathcal{M}_{\mathrm{d}}^{\mathrm{I}} = \mathcal{M}_{\mathrm{e}}^{\mathrm{I}},$$
$$\mathcal{M}_{\mathrm{a}}^{\mathrm{II}} = \mathcal{M}_{\mathrm{b}}^{\mathrm{II}} \neq \mathcal{M}_{\mathrm{c}}^{\mathrm{II}} = \mathcal{M}_{\mathrm{e}}^{\mathrm{II}} \neq \mathcal{M}_{\mathrm{d}}^{\mathrm{II}} \neq \mathcal{M}_{\mathrm{a}}^{\mathrm{II}}.$$

(*The first and third inequalities require* $p > 1$. *The second also requires* $q > 1$.)

In words: When the within-block errors are not restricted, all of the latent structures in Fig. 6 parametrize the same set of covariance matrices, and hence are indistinguishable. When the errors are uncorrelated, on the other hand, the following conditions hold:

- We can distinguish structures in which ξ is a parent of the $\mathbf{X}$'s from those in which the $\mathbf{X}$'s are parents of ξ.
- When the $\mathbf{X}$'s are parents of ξ we cannot distinguish between models with one and two latent variables.
- When ξ is a parent of the $\mathbf{X}$'s we can distinguish models with two latent variables from those containing only one.

The existence of equivalent models containing different numbers of hidden variables is important for the purpose of interpretation. It highlights the danger of postulating a second latent variable within a given structure, when the data—even a very large set of data—could not furnish evidence either for or against its existence unless the researchers make additional assumptions.

5.1. *Proofs of equivalence results*

In order to prove the results in Theorem 8 we need several definitions. (See Section A for basic graphical definitions.) Following Richardson and Spirtes [16] we say that a path diagram, which may contain directed edges ($\rightarrow$) and bi-directed edges ($\leftrightarrow$), is *ancestral* if:

(a) there are no directed cycles;

(b) if there is an edge $x \leftrightarrow y$ then x is not an ancestor of y, (and vice versa).

Conditions (a) and (b) may be summarized by saying that if x and y are joined by an edge and there is an arrowhead at x then x is *not* an ancestor of y; this is the motivation for the term 'ancestral'. [5]

A non-endpoint vertex v on a path is said to be a *collider* if two arrowheads meet at v, i.e. $\rightarrow v \leftarrow$, $\leftrightarrow v \leftrightarrow$, $\leftrightarrow v \leftarrow$ or $\rightarrow v \leftrightarrow$; all other non-endpoint vertices on a path are *non-colliders*. A path $\boldsymbol{\pi}$ between α and β is said to be *m-connecting given Z* if the following hold:

(i) no non-collider on $\boldsymbol{\pi}$ is on Z;
(ii) every collider on $\boldsymbol{\pi}$ is an ancestor of a vertex in Z.

Two vertices α and β are said to be *m-separated given Z* if there is no path m-connecting α and β given Z. Disjoint sets of vertices A and B are said to be *m-separated given Z* if there is no pair α, β with $\alpha \in A$ and $\beta \in B$ such that α and β are m-connected given Z. Two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are said to be *m-separation equivalent* if for all disjoint sets A, B, Z (where Z may be empty), A and B are m-separated given Z in $\mathcal{G}_1$ if and only if A and B are m-separated given Z in $\mathcal{G}_2$. A covariance matrix Σ is said to obey the zero partial covariance restrictions implied by $\mathcal{G}$ if

$$\Sigma_{AB.Z} \equiv \Sigma_{AB} - \Sigma_{AZ}\Sigma_{ZZ}^{-1}\Sigma_{ZB} = 0$$

whenever A is m-separated from B given Z in $\mathcal{G}$.

An ancestral graph is said to be *maximal* if for every pair of non-adjacent vertices α, β there exists some set Z such that α and β are m-separated given Z. For any non-maximal ancestral graph $\mathcal{G}$ there exists a unique maximal ancestral graph $\bar{\mathcal{G}}$, such that $\mathcal{G}$ and $\bar{\mathcal{G}}$ are m-separation equivalent and $\mathcal{G}$ is a subgraph of $\bar{\mathcal{G}}$. (See [16, Theorem 5.1].) In fact, any edge present in $\bar{\mathcal{G}}$ but not $\mathcal{G}$ will be bi-directed.

5.1.1. M-separation contrasted with d-separation

The concept of m-separation, which may be applied to graphs containing both directed and bi-directed edges, is an extension of Pearl's d-separation criterion, which applies to graphs that have only directed edges [15].

Richardson and Spirtes show that the set of covariance matrices obtained by parameterizing the maximal ancestral path diagram $\mathcal{G}$ is exactly the set of Gaussian distributions that obey the zero partial covariance restrictions implied by $\mathcal{G}$ [16]. More formally, we have: [6]

Theorem 9. *If $\mathcal{G}$ is a maximal ancestral graph then the following equality holds regarding Gaussian covariance matrices*:

$$\begin{aligned}&\{\Sigma \mid \Sigma \text{ results from some assignment of}\\ &\quad \text{parameter values to the path coefficients and error covariance associated with } \mathcal{G}\}\\ &= \{\Sigma \mid \Sigma \text{ satisfies the zero partial covariance restrictions implied by } \mathcal{G}\}.\end{aligned}$$

---

[5] In [16], ancestral graphs are actually defined to include undirected edges as well as directed and bidirected edges. The definition considered here corresponds to the subclass without this third kind of edge.

[6] Theorem 8.14 and Corollary 8.19 in [16] are stated in terms of conditional independence in Gaussian distributions. However, the statements given here follow directly using the fact that X is conditionally independent of Y given Z in a multivariate Gaussian distribution if and only if $\Sigma_{XY.Z} = 0$. See e.g. Whittaker [28], p.158 and Corollary 6.3.4, p.164.

See Theorem 8.14 in [16].

As an immediate Corollary we have:

Corollary 10. *If $\mathcal{G}_1$ and $\mathcal{G}_2$ are two m-separation equivalent maximal ancestral graphs then they parameterize the same sets of covariance matrices.*

See Corollary 8.19 in [16]. These results do not generally hold for path diagrams that are not both maximal and ancestral.

The sets of covariance matrices given by the models under (I) correspond to the path diagrams shown in Fig. 6 in which there are bi-directed edges between all variables within the same block, thus $\mathbf{X}_i \leftrightarrow \mathbf{X}_k$ $(i \neq k)$ and $\mathbf{Y}_j \leftrightarrow \mathbf{Y}_\ell$ $(j \neq \ell)$.

5.2. *Proof of Theorem 8*

We first show $\mathcal{M}_\mathrm{a}^\mathrm{I} = \mathcal{M}_\mathrm{b}^\mathrm{I} = \mathcal{M}_\mathrm{c}^\mathrm{I}$. Observe that in each of the graphs in Fig. 6(a)–(c), the following m-separation relations hold:

(i) $\mathbf{X}_i$ is m-separated from $\mathbf{Y}_j$ by any non-empty subset of $\{\boldsymbol{\xi}, \boldsymbol{\omega}\}$,
(ii) $\mathbf{X}_i$ is m-separated from $\boldsymbol{\omega}$ by $\boldsymbol{\xi}$,
(iii) $\mathbf{Y}_j$ is m-separated from $\boldsymbol{\xi}$ by $\boldsymbol{\omega}$.

Further, when bi-directed edges are present between vertices within each block all other pairs of vertices are adjacent so there are no other m-separation relations. Consequently these graphs are m-separation equivalent and maximal since there is a separating set for each pair of non-adjacent vertices. It then follows directly by Corollary 10 that these graphs parameterize the same sets of covariance matrices over the set $\{\mathbf{X}, \mathbf{Y}, \boldsymbol{\omega}, \boldsymbol{\xi}\}$, hence they induce the same sets of covariance matrices on the margin over the observed variables $\{\mathbf{X}, \mathbf{Y}\}$.

The proof that $\mathcal{M}_\mathrm{d}^\mathrm{I} = \mathcal{M}_\mathrm{e}^\mathrm{I}$ is very similar. When bi-directed edges are present within each block the only pairs of non-adjacent vertices are $\mathbf{X}_i$ and $\mathbf{Y}_j$ which are m-separated by $\boldsymbol{\xi}$. It then follows as before that these graphs are m-separation equivalent and maximal and hence by Corollary 10 they parameterize the same sets of covariance matrices over $\{\mathbf{X}, \mathbf{Y}, \boldsymbol{\xi}\}$, and consequently over $\{\mathbf{X}, \mathbf{Y}\}$.

Since we have already shown $\mathcal{M}_\mathrm{a}^\mathrm{I} = \mathcal{M}_\mathrm{d}^\mathrm{I}$ in Corollary 5, the proof of equivalences concerning models with error structure given by (I) is complete. It remains to prove the results concerning models of type (II). These correspond to the path diagrams in Fig. 6, without the bi-directed edges between vertices within the same block. Subsequent references to graphs in this figure will be to the graphs without these within-block edges.

First note that the m-separation relations given by (i), (ii), (iii) above continue to hold when there are no edges between vertices within each block. In graphs (a) and (b) we also have:

(iv) $\mathbf{X}_i$ and $\mathbf{X}_j$ are m-separated given $\boldsymbol{\xi}$,
(v) $\mathbf{Y}_i$ and $\mathbf{Y}_j$ are m-separated given $\boldsymbol{\omega}$.

Consequently these graphs are m-separation equivalent and maximal. Hence

$\mathcal{M}_\mathrm{a}^\mathrm{II} = \mathcal{M}_\mathrm{b}^\mathrm{II}$ by Corollary 10. In the path diagrams corresponding to (c) and (e), we have

(vi) $\mathbf{X}_i$ and $\mathbf{X}_j$ are m-separated by the empty set.

Consequently the variables in the **X** block are uncorrelated in $\mathcal{M}_c^{II} = \mathcal{M}_e^{II}$, while this is not so under $\mathcal{M}_a^{II}, \mathcal{M}_b^{II}, \mathcal{M}_d^{II}$. This establishes two of the inequalities. By direct calculation it may be seen that for any covariance matrix in $\mathcal{M}_d^{II}$, the following *tetrad* constraint holds

$$\mathbf{Cov}\,(\mathbf{X}_i, \mathbf{X}_k)\,\mathbf{Cov}\left(\mathbf{Y}_j, \mathbf{Y}_\ell\right) = \mathbf{Cov}\left(\mathbf{X}_i, \mathbf{Y}_j\right)\mathbf{Cov}\,(\mathbf{X}_k, \mathbf{Y}_\ell)$$

but this does not in general hold for covariance matrices in $\mathcal{M}_a^{II} = \mathcal{M}_b^{II}$. This establishes the third inequality. It only remains to show that $\mathcal{M}_c^{II} = \mathcal{M}_e^{II}$.

First observe that the set of m-separation relations which hold among $\{\mathbf{X}, \mathbf{Y}, \boldsymbol{\omega}\}$ in the graph in (c), i.e. (i), (ii), (iii) and (v), is identical to the set of relations holding among $\{\mathbf{X}, \mathbf{Y}, \boldsymbol{\eta}\}$ in (e), i.e. (i), (ii), (iii) and

(vii) $\mathbf{Y}_j$ and $\mathbf{Y}_\ell$ are m-separated by $\boldsymbol{\eta}$,

where $\boldsymbol{\eta}$ is substituted for $\boldsymbol{\omega}$. Consequently, by Theorem 9, applied to the graph in (e), any covariance matrix over $\{\mathbf{X}, \mathbf{Y}, \boldsymbol{\omega}\}$ that is obtained from the graph in (c) may also be parameterized by the graph in (e) after substituting $\boldsymbol{\eta}$ for $\boldsymbol{\omega}$. It then follows that $\mathcal{N}_c^{II} \subseteq \mathcal{N}_e^{II}$. Hence any covariance matrix over $\{\mathbf{X}, \mathbf{Y}\}$ parameterized by (c) may be parameterized by (e). To prove the opposite inclusion it is sufficient to observe that any covariance matrix over $\{\mathbf{X}, \mathbf{Y}, \boldsymbol{\eta}\}$ that is parameterized by the graph in (e) may be parameterized by the graph in (c) by letting the error variance for $\boldsymbol{\xi}$ equal the error variance for $\boldsymbol{\eta}$ and setting $\boldsymbol{\omega} \equiv \boldsymbol{\xi}$. This completes the proof. □

6. Discussion

6.1. Partition of parameters

Recall that in a cross-covariance problem we wish to model the relationship between the **X** and **Y** blocks rather than the within-block covariances. With this in mind, it may be observed that the parameters of the latent-variable models discussed in this paper fall naturally into three disjoint sets. These differ in terms of whether their values matter in a cross-covariance problem and whether they are identified.

- The within-block errors, $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ and $\boldsymbol{\Sigma}_{\zeta\zeta}$, govern only the within-block covariance, and hence are nuisance parameters in a cross-covariance problem. They are not identified.
- The parameters governing correlation between the latent variables in the paired latent correlation model, $\rho_1, \ldots, \rho_r$, govern only cross-covariance, as may be seen in Eq. (4). These parameters are not identified, but we see in Theorem 1 that they may, in every case, be assigned the value of one without loss of generality.
- The loadings **A** and **B** govern the covariance between the latent and observed variables. These possess a straightforward interpretation, since they are proportional to covariances between the latent variables for one block and the observed variables for the other block. This follows from the definition of the parameters of the paired latent correlation model, Eq. (2), for

  $$\mathbf{Cov}\,(\mathbf{X}, \omega) = \mathbf{AR} \text{ and } \mathbf{Cov}\,(\mathbf{Y}, \xi) = \mathbf{BR}.$$

  Furthermore, in the rank $r = 1$ case, we have seen in part (3) of Theorem 6 that the vectors of loadings are equal, up to sign and scale, to the singular vectors of the cross-covariance. Thus they are identified in the rank-one case.

To summarize: The parameters that are not identified relate to the within-block covariances, while all parameters relating to the cross-covariance are identified.

6.2. Interpretation and estimation of loadings

In the first example from behavioral teratology in Section 1, it is the relative magnitude and direction of the loadings that matter scientifically. The researchers hypothesized a single pair of latent variables for modeling the relationship between the 13 **X** variables, measuring maternal drinking during pregnancy, and the 11 **Y** variables, which were IQ subtests administered to the children at age seven. The latent variable for the **X**s, called $\boldsymbol{\xi}$ in the current work, was conceptualized as "net alcohol exposure," and the latent variable for the **Y**s, which we call $\boldsymbol{\omega}$, as "net intelligence deficit" [23, p. 60]. Thus ξ was a function of **X** and $\boldsymbol{\omega}$ was a function of ξ, and the model is that depicted in Fig. 6(c), under condition (I), that is, with unconstrained within-block covariance. By Theorem 8, we know that this asymmetric model is an alternate parameterization of the paired latent correlation model.

Of interest to Sampson and Streissguth were two sets of parameters:

- The relative magnitudes and directions of the 13 different covariances between the latent variable for anatomic change and the measures of maternal drinking during pregnancy, and
- The relative magnitudes and directions of the 11 different covariances between the latent variable for alcohol exposure and the IQ subtests.

These quantities follow immediately from the loadings, for when there is only one pair of latent variables we have

$$\mathbf{Cov}\,(\mathbf{X}, \omega) = \rho\mathbf{a} \text{ and } \mathbf{Cov}\,(\mathbf{Y}, \xi) = \rho\mathbf{b}.$$

The loadings for **X** are proportional to the direct (and total) effect of $\boldsymbol{\xi}$ on the X_i's; likewise the loadings for **Y** are proportional to the direct (and total) effect of $\boldsymbol{\omega}$ on the Y_i's.

If we assume a model in which there is an edge $\xi \to \boldsymbol{\omega}$ then the loadings for **Y** are proportional to the total effect of ξ on Y_i.

We see in the proof of Theorem 6 that the loadings **a** and **b** are equal, up to sign and scale, to the left and right singular vectors of the cross-covariance $\mathbf{\Sigma}_{XY}$. In fact, the researchers used the singular value decomposition to estimate loadings from the sample cross-covariance, and drew a path diagram to justify this approach without formal specification of a statistical model [23, p. 64]. This application of the singular value decomposition is known as partial least squares (PLS) [23]. [7] Theorem 1 links this use of the singular value decomposition to a formally specified model.

[7] The term "PLS" includes a large class of other methods, most of them not linked to formally specified models, many of them used for prediction rather than modeling. For a survey of two-block PLS methods the reader is referred to [25].

6.3. Lack of identifiability of correlation between latent variables

It is customary in linear structural equation models to constrain the variances of all latent variables (except for error variances) to equal one. This custom has been followed in the current work (Section 2.2). In spite of this constraint, the paired latent correlation model with $r = 1$ yields only a lower bound on the correlation. This is in marked contrast to the empirical correlations obtained by the canonical correlation and PLS methods. This fact is an unavoidable consequence of the model, and it should thus be borne in mind by researchers who postulate a particular latent structure that it can only be identified by making further assumptions.

Notice that in the latent model (c) in Fig. 6, if we standardize the variances of $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ to 1, then the correlation is equal to the path coefficient for the edge $\boldsymbol{\xi} \to \boldsymbol{\omega}$, which is the total and direct effect of $\boldsymbol{\xi}$ on $\boldsymbol{\omega}$.

Kuroki et al. present conditions under which the square of the total effect may be identified, in the case where the response is unobserved [12].

6.4. Comparison of correlation measures

Canonical analysis yields scores for two derived variables, linear combinations respectively of the **X** and **Y** blocks. The canonical correlation is maximal over all linear combinations of the **X** and **Y** blocks [11]. Sampson and Streissguth, on the other hand, used the singular value decomposition to compute PLS scores. These scores are also linear combinations of the **X** and **Y** blocks, chosen to maximize the absolute value of the covariance between the scores, subject to the constraint that the vectors of coefficients each have Euclidean norm one. In both of these cases, the correlations between the derived variables are subject to attenuation [20]. The minimal correlation developed in Theorem 6, on the other hand, is not subject to attenuation, since it is a correlation between latent variables, rather than between linear combinations of the observed variables. In absolute value, this minimal correlation may be substantially greater than the canonical correlation, which in turn is greater than or equal to the correlation between PLS scores.

6.5. Possible implications for behavioral teratology

If we choose to believe that the data in the first example in Section 1 were generated by a paired latent correlation model with $r = 1$, and in particular that the covariance between the two blocks originates in a pair of latent variables, then part (4) of Theorem 6 establishes that there is no way, using data, to determine whether the latent variables have the minimal correlation, are perfectly correlated, or have some correlation between these extremes. This shortcoming of the model cannot be overcome by a larger sample size, since an increase in sample size would have no effect on the size of the interval of feasible correlations.

The lack of identifiability of the correlation draws our attention to the question of what these latent variables actually are. One tempting a priori interpretation is that differences in the mothers' drinking patterns during pregnancy caused differences in the children's brain anatomy, which are summarized by the latent variable for the alcohol exposure block; that these anatomical differences caused a pattern of differences in the children's cognitive ability, summarized by the latent variable for IQ deficit; and that these differences in cognitive ability caused differences in scores on the IQ sub-tests. If this model is correct, there is a lower bound on the correlation between anatomy and IQ that is reasonable in the context of the data. This value could be substantially greater in absolute value than the correlation of -0.202 between PLS scores seen in [23, p. 64]. (If the lower bound is attained, the within-block error covariances must be singular.)

6.6. On numerical stability and estimation

The proof of Theorem 1 depends on canonical vectors. These vectors are not identified when the number of variables exceeds the number of observations. Furthermore, when either $\mathbf{X}^T\mathbf{X}$ or $\mathbf{Y}^T\mathbf{Y}$ is ill-conditioned, algorithms for computation of canonical vectors are unstable. These facts do not affect the proof, however, because it is a proof of existence and is not concerned with numerical computation.

We have not sought to develop a method for estimation of the parameters of the latent variable models presented here. Theorem 6, however, by linking the singular value decomposition method (PLS) to a rank-one paired latent correlation model, establishes that PLS is an estimator for the loadings in the rank-one case. Since it is based on the singular value decomposition, it is numerically stable, regardless of whether there are more variables than observations. Because the singular value decomposition is continuous, the estimator is consistent.

6.7. Underidentification vs. identification via assumptions

A more standard approach, adopted in SEM modelling, is to assume that observed "indicator" variables are uncorrelated given the latent variables. While one might prefer models in which the parameters are identified, these additional assumptions are not, in general, testable via standard goodness-of-fit tests—for instance, the case where there are only 3 indicators per latent, and the model is just-identified. (Even in cases where such an assumption is testable via tetrad constraints, as in [21], failure to reject the null hypothesis does not imply that the assumption holds.) Our goal here is to see how much information may be extracted from the data, without resorting to additional assumptions of the type required for identification. As we have illustrated with the study of Sampson and Streissguth, bounds on underidentified parameters can sometimes be informative.

6.8. Future work

Although Theorem 1 establishes that a latent parameterization of a reduced-rank-regression model always exists, questions remain regarding the case where the rank exceeds one. First, the feasible set of correlations between latent variables has not been characterized when $r > 1$. Furthermore, it is not known whether the loadings in the higher-rank model are determined, up to sign and scale, by the cross-covariance matrix. The second example in Section 1 lends urgency to these questions, for the researchers computed and interpreted two pairs of PLS latent scores.

Appendix A. Graphical terminology

A *path diagram* is a graph with vertex set $\mathbf{V}$ containing *directed* ($\to$) and *bi-directed* ($\leftrightarrow$) edges. If there is an edge from x to y, $x \to y$, then x is said to be a *parent* of y, and conversely, y is said to be a *child* of x. If there is an edge $x \to y$, or $x \leftrightarrow y$ then we say there is *an arrowhead at* y on this edge. The set of parents of a vertex is defined as

$$\mathrm{pa}(v) = \{w \mid w \to v\}.$$

A *sequence of edges* between α and β in a path diagram is an ordered (multi)set of edges $\langle \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n \rangle$, such that there exists a sequence of vertices (not necessarily distinct) $\langle \alpha \equiv \omega_1, \ldots, \omega_{n+1} \equiv \beta \rangle (n \geqslant 0)$, where edge $\boldsymbol{e}_i$ has endpoints ω_i, ω_{i+1}. A sequence of edges for which the corresponding sequence of vertices contains no repetitions is called a *path*. The vertices α and β are the *endpoints*; all other vertices in the path are *non-endpoint vertices* with respect to that path. A path of the form $v \to \cdots \to w$ on which every edge is of the form $\to$, with the arrowhead pointing towards w, is said to form a *directed path* from v to w. If there is such a directed path from v to $w \neq v$, or if $v = w$, then v is an *ancestor* of w. A directed path from v to w together with an edge $w \to v$ is said to form a *directed cycle*.

A path diagram with m vertices parameterizes a set of m-by-m covariance matrices as follows. First, to each vertex in $\mathbf{V}$ corresponds a variable Y_i. Then each variable Y_i is a linear function of its parents in the graph together with an error term, ε_i.

$$Y_i = \sum_{Y_j \in \mathrm{pa}(Y_i)} \beta_{ij} Y_j + \varepsilon_i$$

this may be written in vector form as

$$\mathbf{V} = (\mathbf{I} - \mathbf{\Gamma})^{-1} \varepsilon,$$

where $(\Gamma)_{ij} = \beta_{ij}$, $\mathbf{V} = (Y_1, \ldots, Y_p)^T$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_p)^T$. The parameter β_{ij} is said to be the *path coefficient* associated with the $Y_j \to Y_i$ edge.

The bi-directed edges place restrictions on the covariance matrix over the error terms, $\Sigma_{\varepsilon\varepsilon}$, as follows:

$$\text{there is no edge } Y_i \leftrightarrow Y_j \text{ in the path diagram} \quad \Rightarrow \quad (\Sigma_{\varepsilon\varepsilon})_{ij} = 0 \tag{A.1}$$

subject to the restriction that $\Sigma_{\varepsilon\varepsilon}$ be positive semi-definite. The path coefficients β_{ij}, together with the covariance over the errors $\Sigma_{\varepsilon\varepsilon}$, specify a covariance matrix over Y given by

$$\Sigma_{\mathbf{VV}} = (\mathbf{I} - \mathbf{\Gamma})^{-1} \Sigma_{\varepsilon\varepsilon} (\mathbf{I} - \mathbf{\Gamma})^{-T}.$$

Note that when we depict path diagrams (e.g. Figs. 1–3), error variables are not included. A distinction is made between observed variables, in squares, and unobserved variables, in circles.

We associate with a given path diagram the set of covariance matrices which may be obtained by choosing arbitrary values for the path coefficients β_{ij} and the covariance matrix $\Sigma_{\varepsilon\varepsilon}$, subject to the restriction given by (A.1). When all variables are observed this set of covariance matrices is said to be the *model* specified by the path diagram.

When unobserved variables are present, the model consists of the set of covariances matrices over the observed variables that are the submatrices of some covariance matrix over all the variables, for some choice of values for the path coefficients, β_{ij}, and the covariance matrix $\Sigma_{\varepsilon\varepsilon}$.

References

[1] S.N. Afriat, Orthogonal and oblique projectors and the characteristics of pairs of vector spaces, Proc. Cambridge Philos. Soc. (Math. Phys. Sci.) 53 (4) (1957) 800–816.

[2] T.W. Anderson, Asymptotic distribution of the reduced rank regression estimator under general conditions, Ann. Statist. 27 (4) (1999) 1141–1154.

[3] A. Björck, G.H. Golub, Numerical methods for computing angles between linear subspaces, Math. Comput. 27 (123) (1973) 579–594.

[4] K.A. Bollen, Structural Equations With Latent Variables, Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1989.

[5] F.L. Bookstein, A.P. Streissguth, P.D. Sampson, P.D. Connor, H.M. Barr, Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure, NeuroImage 15 (2002) 233–251.

[6] D. Geiger, D. Heckerman, H. King, C. Meek, On the geometry of DAG models with hidden variables, in: D. Heckerman, J. Whittaker (Eds.), Preliminary Papers of the Seventh International Workshop on AI and Statistics, January 4–7, Fort Lauderdale, FL, 1999, pp. 76–85.

[7] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., Johns Hopkins, Baltimore, MD, 1996.

[8] D.A. Harville, Matrix Algebra from a Statistician's Perspective, Springer, Berlin, 1997.

[10] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge, Cambridge, 1985.

[11] H. Hotelling, Relations between two sets of variables, Biometrika 28 (1936) 321–377.

[12] M. Kuroki, M. Miyakawa, Graphical identifiability criteria for total effects in studies with an unobserved response variable, Behaviormetrika 31 (1) (2004) 13–28.

[13] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate analysis, Probability and Mathematical Statistics, A Series of Monographs and Textbooks, Academic Press, New York, 1979.

[14] MathSoft, Inc. S-PLUS Version 3.4 Release 1 for Silicon Graphics Iris, IRIX 5.3, 1996. Software.

[15] J. Pearl, Causality: Models, reasoning, and inference, Cambridge University Press, Cambridge, 2000.

[16] T. Richardson, P. Spirtes, Ancestral graph markov models, Ann. Statist. 30 (4) (2002) 962–1030.

[17] P.D. Sampson, A.P. Streissguth, H.M. Barr, F.L. Bookstein, Neurobehavioral effects of prenatal alcohol: part II, Partial Least Squares analysis, Neurotoxicol. Teratol. 11 (5) (1989) 477–491.

[18] R. Settimi, J.Q. Smith, On the geometry of Bayesian graphical models with hidden variables, in: G. Cooper, S. Moral (Eds.), Uncertainty in Artificial Intelligence: Proceedings of the 14th Conference, San Francisco, Morgan Kaufmann, Los Altos, CA, 1998, pp. 472–279.

[19] R. Settimi, J.Q. Smith, Geometry, moments and Bayesian networks with hidden variables, in: D. Heckerman, J. Whittaker (Eds.), pp. 293–298.

[20] C. Spearman, The proof and measurement of association between two things, Amer. J. Psychol. 15 (1904) 72–101.

[21] P. Spirtes, C.N. Glymour, R. Scheines, Causation Prediction and Search, MIT Press, Cambridge, MA, 2000.

[22] E. Stanghellini, Identification of a single-factor model using graphical gaussian rules, Biometrika 84 (1) (1997) 241–244.

[23] A.P. Streissguth, F.L. Bookstein, P.D. Sampson, H.M. Barr, The enduring effects of prenatal alcohol exposure on child development: birth through seven years, a partial least squares solution, International Academy for Research in Learning Disabilities Monograph Series, vol. 10, University of Michigan Press, Ann Arbor, 1993.

[24] P. Vicard, On the identification of a single-factor model with correlated residuals, Biometrika 87 (1) (2000) 199–205.

[25] J.A. Wegelin, A survey of partial least squares (PLS) methods, with emphasis on the two-block case, Technical Report 371, Department of Statistics, University of Washington, seattle, 2000. http://www.stat.washington.edu/www/research/reports/2000/tr371.ps

[26] J.A. Wegelin, T.S. Richardson, Cross-covariance modelling via DAGs with hidden variables, in: J. Breese, D. Koller (Eds.), Uncertainty in Artificial Intelligence: Proceedings of the Seventeenth Conference (UAI-2001), August 2–5, 2001, University of Washington, Seattle, Washington, Morgan Kaufmann, Las Altos, 2001, pp. 546–553.

[27] J.A. Wegelin, T.S. Richardson, D.L. Ragozin, Rank-one latent models for cross-covariance, Technical Report 391, Department of Statistics, University of Washington, Seattle, 2001, http://www.stat.washington.edu/www/research/reports/2001/tr391.pdf.

[28] J. Whittaker, Graphical models in Applied Multivariate Statistics. Wiley, 1990.