STATISTICA, anno LXXIV, n. 4, 2014

A MATRIX-VARIATE REGRESSION MODEL WITH CANONICAL STATES: AN APPLICATION TO ELDERLY DANISH TWINS

Laura Anderlucci
Dipartimento di Scienze Statistiche, Università di Bologna, Bologna, Italia
Angela Montanari
Dipartimento di Scienze Statistiche, Università di Bologna, Bologna, Italia
Cinzia Viroli[1]
Dipartimento di Scienze Statistiche, Università di Bologna, Bologna, Italia.

1. Introduction

Many situations and many modern scientific applications are frequently producing data sets with a matrix-variate instead of a vector structure. These are contexts where p variables are observed on n subjects in r different situations, such as multivariate panel data or spatio-temporal data (see, for example, Basford & McLachlan , 1985; Anderlucci & Viroli , 2015; Genton , 2007). The three dimensions are referred to multivariate or spatial responses, measured over time on the same individuals. Another example of matrix-variate data comes from the population-based twin studies, where variability between and within couples of twins is of interest. The trivariate structure is reflected by considering a sample of twin couples (i.e. the sampling units), where several responses are collected, separately on each twin. The aim is to investigate whether observations are function of some covariates.

There are several approaches to study matrix-variate data as function of some covariates in the statistical literature. In particular, Mitchell et al. (2006) developed a likelihood ratio test for testing the correct specification of a regression model with time-independent covariates and time-varying and variable-varying regression coefficients. Brien & Fitzmaurice (2005) focused on the comparison between random effect models and a regression model where covariates have the same dimensionality of the responses and regression coefficients do not vary between and within variables and replicates. It represents a particular and constrained *multivariate* regression analysis adapted for dealing with three-way data. Viroli (2012) proposed a more general approach that allows for q predictors that can differently affect the p observed measurements through a matrix of parameters having dimension $p \times q$. More recently, Zhou & Li (2014) proposed a class of regularized matrix regression methods based on spectral regularization to analyze electroencephalography data set of alcoholism, where a sample of subjects was exposed to a stimulus.

In this work we define a novel matrix-variate regression model that decomposes the contribution of the different covariates on the responses in a canonical representation thus

[1] Corresponding Author. E-mail: cinzia.viroli@unibo.it

allowing for a better interpretation. It can be viewed as an extension of the general matrix-variate regression analysis proposed by Viroli (2012), but it resembles in some aspects the Haseman-Elston regression (see Haseman & Elston , 1972). The extension has been motivated by population twin-studies where one aspect of interest is to evaluate how the components of the trait variability of the twins are related to selected covariates.

The paper is organized as follows. In the next section we review the matrix-variate normal distribution and the general matrix-variate regression model proposed in Viroli (2012). Then, the extension of the matrix-variate regression model with canonical states and the inferential properties of the estimators are presented. In Section 3 the real application is illustrated and the model interpretation is discussed. We conclude this work with a discussion on the proposed model and on the obtained results.

2. The model

2.1. Preliminaries: the general matrix-variate regression model

Let Y_i denote the $p \times r$ observed matrix containing the numerical p responses in r states for each unit (with $i = 1, \ldots, n$). Suppose we observe a set of covariates for each state, so that X_i is the design matrix that includes covariates and intercept of dimension $q \times r$.

A general matrix-variate regression model for Y_i takes the form

$$Y_i = \Theta X_i + U_i, \quad (i = 1, \ldots, n), \tag{1}$$

where Θ is a matrix of dimension $p \times q$ of unknown parameters and U_i is the error term. Since U_i is a matrix of size $p \times r$, we assume that it is distributed according to a matrix-variate normal distribution (Gupta & Nagar , 2000), $U_i \sim \phi^{(p \times r)}(0, \Phi, \Omega)$, that is

$$f(U_i) = (2\pi)^{-\frac{rp}{2}} |\Phi|^{-\frac{p}{2}} |\Omega|^{-\frac{r}{2}} \exp\left\{ -\frac{1}{2} tr\left(\Omega^{-1} U_i \Phi^{-1} U_i^\top\right) \right\}, \tag{2}$$

where Φ is a $r \times r$ covariance matrix containing the variances and covariances 'within' the states and Ω is a $p \times p$ covariance matrix containing the variance and covariances 'between' the p variables.

The matrix normal distribution plays a pivotal role in the family of matrix-variate distributions (see Gupta & Nagar , 2000) thanks to its mathematical properties and to its capability to be a reference model for most multiway phenomena according to the central limit theorem. For these reasons, the matrix-variate Gaussian distribution is receiving a growing interest in the statistical literature as demonstrated by some recent applications that include spatio-temporal analysis (Mitchell et al. , 2005), Bayesian graphical models (Wang & West , 2009) and model-based clustering (Viroli , 2011).

The matrix-variate regression analysis can be viewed as an extended multivariate regression analysis that deals with matrix-variate responses instead of vectors. It is always possible to stack the columns of the response matrix into a single vector - thus obtaining a vector response of dimension pr - and to fit a multivariate regression model. However, such approach would lead to a solution that is different from the one yielded by a matrix-variate regression analysis, since the identification of the two sources of variability (i.e. 'within' and 'between' dispersions) would not be possible. In fact, in the matrix-variate regression model the total variability, say Σ, is separable in the form $\Sigma = \Omega \otimes \Phi$ (where $\otimes$ is the Kronecker product). This *separability condition* leads to a structured covariance matrix,

which is convenient for our purposes because it combines parsimony and flexibility. In fact, it allows to estimate the the correlation of the items between and within states by resorting to a smaller number of parameters, which are $r(r+1)/2+p(p+1)/2$ instead of $pr(pr+1)/2$.

2.2. A Matrix-Variate Regression Model with Canonical States

Let $\mathbf{x}_i$ be a vector of dimension $q \times 1$ that includes the set of stacked covariates and an additional element equal to 1 with the aim of incorporating the intercept in the model. Now, we decompose the systematic part of the general model (1) by introducing a set of canonical vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_r$ defined as $\mathbf{e}_1 = (1, 0, \ldots, 0)^\top$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)^\top, \ldots, \mathbf{e}_r = (0, \ldots, 0, 1)^\top$. We define the following regression model for Y_i

$$Y_i = \Gamma_1 \mathbf{x}_i \mathbf{e}_1^\top + \Gamma_2 \mathbf{x}_i \mathbf{e}_2^\top + \ldots + \Gamma_r \mathbf{x}_i \mathbf{e}_r^\top + U_i \qquad (i = 1, \ldots, n) \tag{3}$$

where Γ_1, $\Gamma_2, \ldots, \Gamma_r$ are r unknown $p \times q$ matrices of parameters. From one hand, the inclusion of the canonical vectors helps the interpretability of the model parameters, since it allows to decompose the effect of the covariates on the different states. On the other hand, is also simplifies the estimation problem as it will be explained in the next subsection.

2.3. Inference

The maximum likelihood estimation for the model parameters is not generally achievable in closed-form if no constraint is imposed on the matrices Φ and Ω of the matrix-normal distribution. In fact, suppose $\hat{\Theta}$ is the maximum likelihood estimate of Θ in model (1). It is easy to show that the maximum likelihood estimators of Φ and Ω can be obtained by evaluating and differentiating the residual log-likelihood

$$\begin{aligned} \log L(\Phi, \Omega | Y_1, \ldots, Y_n) = & - \frac{rpn}{2} \log(2\pi) - \frac{pn}{2} \log|\Phi| - \frac{rn}{2} \log|\Omega| \\ & - \frac{1}{2} \sum_{i=1}^{n} tr\left(\Omega^{-1} \tilde{Y}_i \Phi^{-1} \tilde{Y}_i^\top \right) \end{aligned}$$

where $\tilde{Y}_i = Y_i - \hat{\Theta} X_i$. The maximization problem leads to a system of coupled equations

$$\begin{cases} \hat{\Phi} = \frac{\sum_{i=1}^{n} \tilde{Y}_i^\top \hat{\Omega}^{-1} \tilde{Y}_i}{pn} \\ \hat{\Omega} = \frac{\sum_{i=1}^{n} \tilde{Y}_i \hat{\Phi}^{-1} \tilde{Y}_i^\top}{rn} \end{cases},$$

that implies that there is not a closed-form analytic solution for estimating the two covariance matrices. Their values must be computed in an iterative way (see, for instance, Dutilleul , 1999) and the solution is unique up to a multiplicative constant, say $a \neq 0$, since $\Phi \otimes \Omega = a\Phi \otimes \frac{1}{a}\Omega$.

A way to solve this problem could be to assume that the covariates describe all the correlation between the states, so that Φ is diagonal. Moreover, without loss of generality, let us suppose that the state variances in Φ are equal to one. The advantage of these choices is twofold. First, it solves the identifiability problem $\Phi \otimes \Omega = a\Phi \otimes \frac{1}{a}\Omega$ for $a \neq 0$ and it allows to develop finite sample inference on the model estimators; secondly, as we will show in our illustrative example, in some cases it may help the interpretation of the

results. Under these assumptions, $U_i \sim \phi^{(p\times r)}(0, I_r, \Omega)$ and $Y_i|M_i \sim \phi^{(p\times r)}(M_i, I_r, \Omega)$ where $M_i = \Gamma_1 \mathbf{x}_i \mathbf{e}_1^\top + \ldots + \Gamma_r \mathbf{x}_i \mathbf{e}_r^\top$.

Parameter estimators and related properties can be obtained by rephrasing the model (3) in a vectorized form. Introducing the vec operator such that $\mathbf{y}_i = \text{vec}(Y_i)$ and $\mathbf{u}_i = \text{vec}(U_i)$ are vectors of length rp, we can express (3) as

$$\mathbf{y}_i = \left(\mathbf{e}_1 \mathbf{x}_i^\top \otimes I_p\right) \boldsymbol{\gamma}_1 + \left(\mathbf{e}_2 \mathbf{x}_i^\top \otimes I_p\right) \boldsymbol{\gamma}_2 + \cdots + \left(\mathbf{e}_r \mathbf{x}_i^\top \otimes I_p\right) \boldsymbol{\gamma}_r + \mathbf{u}_i, \quad (i = 1, \ldots, n), \tag{4}$$

where I_p is the identity matrix of order p and $\boldsymbol{\gamma}_r = \text{vec}(\Gamma_r)$. Setting $\mathbf{y}^\top = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)$, $\mathbf{u}^\top = (\mathbf{u}_1^\top, \ldots, \mathbf{u}_n^\top)$ and $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, the model (4) can be formulated as

$$\mathbf{y} = \left\{(I_n \otimes \mathbf{e}_1) X^\top \otimes I_p\right\} \boldsymbol{\gamma}_1 + \left\{(I_n \otimes \mathbf{e}_2) X^\top \otimes I_p\right\} \boldsymbol{\gamma}_2 + \cdots + \left\{(I_n \otimes \mathbf{e}_r) X^\top \otimes I_p\right\} \boldsymbol{\gamma}_r + \mathbf{u}. \tag{5}$$

Under the above assumptions, the maximum likelihood estimator of the generic $\boldsymbol{\gamma}_r$ coincides with the generalized least squares estimator of $\boldsymbol{\gamma}_r$ given by

$$\hat{\boldsymbol{\gamma}}_r = \left\{\left(XX^\top\right)^{-1} X (I_n \otimes \mathbf{e}_r^\top) \otimes I_p\right\} \mathbf{y} \tag{6}$$

with

$$\hat{\Gamma}_r = Y \left\{X^\top \left(XX^\top\right)^{-1} \otimes \mathbf{e}_r\right\} \tag{7}$$

being the maximum likelihood estimator of Γ_r, where $Y = (Y_1, \ldots, Y_n)$ is a matrix of dimension $p \times rn$. To derive expression (6) we can define the known matrix of dimension $rpn \times rpq$, $W = \left\{(I_n \otimes \mathbf{e}_1) X^\top \otimes I_p, \ldots, (I_n \otimes \mathbf{e}_r) X^\top \otimes I_p\right\}$ and the unknown vector $\boldsymbol{\gamma}$ of length rpq that contains $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_r$. Then model (5) can be rephrased as $\mathbf{y} = W\boldsymbol{\gamma} + \mathbf{u}$ from which $\hat{\boldsymbol{\gamma}} = \left(W^\top W\right)^{-1} W^\top \mathbf{y}$. Result follows by observing that $W^\top W$ is a diagonal block matrix because the canonical vectors are orthogonal. Note that the least square estimator is also equal to the maximum likelihood estimator as a consequence of the matrix-normal properties. More specifically, if $U_i \sim \phi^{(p\times r)}(0, I_r, \Omega)$ then $\mathbf{u}_i = \text{vec}(U_i)$ is distributed according to a multivariate normal, that is $\mathbf{u}_i \sim \phi^{(pr)}(0, I_r \otimes \Omega)$. Therefore $\mathbf{u} \sim \phi^{(rpn)}(0, I_n \otimes I_r \otimes \Omega)$. This also implies that, the regression coefficient estimates are always the same regardless the structure imposed on the covariance matrix of the errors.

Maximum likelihood estimator of Ω is given by S_Ω/rn where S_Ω is the residual sum of squares matrix

$$S_\Omega = \sum_{i=1}^{n} \left(Y_i - \hat{M}_i\right) \left(Y_i - \hat{M}_i\right)^\top. \tag{8}$$

In the next theorems we derive the distributional form of the two estimators for Ω and Γ. Proofs are given in the Appendix.

THEOREM 1. *Under the constraint $p < r(n-q)$, the residual sum of squares matrix S_Ω is distributed as a Wishart with parameter Ω and degrees of freedom $r(n-q)$, $S_\Omega \sim \mathcal{W}_p(r(n-q), \Omega)$.*

COROLLARY 2. *The unbiased estimator of Ω is $\hat{\Omega} = \frac{S_\Omega}{r(n-q)}$.*

The next theorem establishes some inferential properties of the estimator $\hat{\Gamma} = [\hat{\Gamma}_1, \ldots, \hat{\Gamma}_r]$ for $\Gamma = [\Gamma_1, \ldots, \Gamma_r]$ where $\hat{\Gamma}$ and Γ have dimension $p \times rq$.

LEMMA 3. *Let $\hat{\Gamma} = [\hat{\Gamma}_1, , \ldots, \hat{\Gamma}_r]$ be the estimator defined in (7). $\hat{\Gamma}$ is a linear and unbiased estimator of Γ with covariance matrix*

$$var\left(\hat{\Gamma}\right) = \left(XX^\top\right)^{-1} \otimes I_r \otimes \Omega.$$

THEOREM 4. *Let $\hat{\Gamma} = [\hat{\Gamma}_1, \ldots, \hat{\Gamma}_r]$ be the estimator defined in (7) and S_Ω the residual sum of squares matrix in (8). Then $S_\Omega^{-\frac{1}{2}}\hat{\Gamma}$ has a matrix-variate $\mathcal{T}$ distribution*

$$S_\Omega^{-\frac{1}{2}}\hat{\Gamma} \sim \mathcal{T}_{rq,p}\left(r(n-q) - p - 1, S_\Omega^{-\frac{1}{2}}\Gamma, (XX^\top)^{-1} \otimes I_r, I_p\right)$$

where $r(n-q) - p - 1$ are the degrees of freedom.

Theorem 2 allows to test the null hypothesis of linear independence of the responses from the covariates, $H_0 : \Gamma = 0$ through the statistics

$$T = S_\Omega^{-\frac{1}{2}}\hat{\Gamma} \sim \mathcal{T}_{rq,p}\left(r(n-q) - p - 1, 0, (XX^\top)^{-1} \otimes I_r, I_p\right).$$

Since the matrix-variate $\mathcal{T}$ distribution is closed under linear transformation and partitions, we could also test the null hypothesis about the single regression coefficient, $H_0 : \Gamma_{jh} = 0$, with $j = 1, \ldots, p$ and $h = 1, \ldots, rq$ via

$$T_{jh} \sim t_{r(n-q)-p-1}\left(\{(XX^\top)^{-1} \otimes I_r\}_{hh}\right).$$

This would employ a battery of tests, with the possible consequence of an inflation of the true significance level. A generalized strategy for simultaneous testing has been considered in Viroli (2012) and can be adapted here as follows. Consider the general linear hypothesis in the form $H_0 : M\Gamma C^\top = 0$, where M is a matrix of dimension $c \times p$ with rank $c \leq p$, and C is a $g \times rq$ matrix with rank $g \leq rq$. Now, under the null hypothesis, the transformation $\hat{\delta} = \text{vec}(M\hat{\Gamma}C^\top) = (C \otimes M)\hat{\boldsymbol{\gamma}}$ has distribution

$$\hat{\delta} \sim \phi^{(cg)}\left(0, C(XX^\top \otimes I_r)^{-1}C^\top \otimes M\Omega M^\top\right).$$

Moreover, the matrix $H = M\hat{\Gamma}C^\top\left(C(XX^\top \otimes I_r)^{-1}C^\top\right)^{-1}C\hat{\Gamma}M^\top$ is distributed as Wishart, $H \sim \mathcal{W}_c(g, M\Omega M^\top)$ independently from $G = MS_\Omega M^\top$, which is $G \sim \mathcal{W}_c(r(n-q), M\Omega M^\top)$. Tests on $H_0 : M\Gamma C^\top = 0$ can be carried out based on the characteristic roots of $G(H+G)^{-1}$. More specifically,

$$\tilde{\lambda} = |G|/|H+G| = \prod_{i=1}^{c}(1+\lambda_i)^{-1}$$

has a Wilks' lambda distribution with parameters c, $r(n-q)$ and g (see Mardia et al. , 2003), where λ_i $(i = 1, \ldots, c)$ are the eigenvalues of HG^{-1}.

REMARK 5. *From this theory, it is clear that the constraint $\Phi = I_r$ is necessary to obtain the finite sample properties of the estimators. This constrain can be relaxed by allowing a full matrix Φ in order to better describe the variability between and within states. In this scenario, theorems 1 and 2 can be easily extended, but the distributional results only hold asymptotically.*

3. Illustration

3.1. Data description

We illustrate the model properties on data coming from the Longitudinal Study of Aging Danish Twins (LSADT) that has been conducted every two years between 1995 and 2005 (see Christensen & Vaupel , 2009) with the aim of investigating the causes of variation in survival, health, diseases, loss of abilities, and cognitive functions among the elderly Danish twins aged 75 years and older. The information collected during the interviews cover health, physical functioning, cognitive functioning, depression symptomatology, social factors, lifestyle characteristics, and quality of life. Previous studies on these data (e.g. McGue & Christensen , 2007; Christensen et al. , 1999) investigated the relative influence of genetic and environmental factors on the overall level of physical and cognitive functioning and found that some determinants, like the social activity, were significantly correlated with the mental health status.

In this study we have considered information on $n = 362$ couples of respondents of the first wave in 1995 (the remaining 18 alive twins have been discarded because of missing data). We considered $p = 3$ responses measuring the cognitive functioning. The first variable is given by the total score obtained in the battery of tests of orientation, registration, attention and calculation. It summarizes the 'mental status'. The second variable is a measure of the 'episodic memory' because it summarizes the scores obtained in the word recall, delayed word recall and language tests. The third variable is the total score obtained in the Lawton's IADL questionnaire that measures the capability of self-maintaining and performing the main activities of daily living, like shopping, housekeeping and food preparation (see, for major details, Lawton & Brody , 1969). Since the cognitive functioning may be measured by the range of all the described variables, it is important to consider all the three responses simultaneously. The cognitive functioning may be affected by some determinants and the interest is in investigating the association of the overall cognitive functioning with some covariates, including age, gender, education and indicators of prior or existing major health conditions such as cancer, heart disease, stroke, diabetes or psychiatric disorders (see Table 1).

3.2. The analysis

The direct application of model (1) to multivariate twin data may have a drawback related to the exchangeability of the twin members, that is, for each couple i we do not know which subject should come first (i.e. information about the birth order is not always available, in this sense twin members can be arbitrarily exchanged). Thus the model is not fully identified because the diagonal elements of Φ are not distinguishable and interpretation of Θ may be not clear.

In order to overcome this limitation, we introduce an invertible transformation on the original measurements. In some sense, the idea is similar to the Haseman-Elston regression analysis (Haseman & Elston , 1972) and its subsequent revisited models (Elston et al. , 2000; Zhang et al. , 2008), developed to detect genetic linkage between quantitative traits and genetic markers. Basically, the Haseman-Elston regression analysis considers pairs of siblings and regresses the squared differences in their trait values on the proportion of alleles shared at a given locus with the aim of detecting if the locus is linked to the trait. Alternatively, the squared trait sum or the mean-corrected trait product have been used

TABLE 1
Description of covariates

Question	Type
Do you think that your health is generally?	1. Excellent; ...; 5. Very poor
Are you happy and satisfied with your life at present?	1. Yes, always; ...; 5. No, never
Did a doctor ever tell you that you had diabetes?	Dummy: 1. Yes; 0. No
Did a doctor ever tell you that you had rheumatoid arthritis?	Dummy: 1. Yes; 0. No
Did a doctor ever tell you that you had Parkinson's disease?	Dummy: 1. Yes; 0. No
Did a doctor ever tell you that you had epilepsy?	Dummy: 1. Yes; 0. No
Did a doctor ever tell you that you had cancer?	Dummy: 1. Yes; 0. No
Did a doctor ever tell you that you had stroke?	Dummy: 1. Yes; 0. No
Did a doctor ever tell you that you had heart attack?	Dummy: 1. Yes; 0. No
Have you ever suffered from a depression?	Dummy: 1. Yes; 0. No
What is your body weight?	Kg
How tall are you?	Cm
For how many years did you and your twin live together?	Years
Do you live alone?	Dummy: 1. Yes; 0. No
What type of school education did you receive?	Years of school
Are you or have you been married? (reference: Married)	Dummy: 1. Never married; 0. Other
Are you or have you been married? (reference: Married)	Dummy: 1. Divorced/separated; 0. Other
Are you or have you been married? (reference: Married)	Dummy: 1. Widower/widow; 0. Other
Do you smoke?	Dummy: 1. Yes; 0. No
How much do you drink alcohol per week?	Number of drinks
Age	Years
Sex	Dummy: 1. Male; 0. Female
Zygosity	1. Monozygotic; 0; Dizygotic

(see, for example, Wright , 1997; Drigalenko , 1998).

Let y_{ij1} and y_{ij2} be the two observed values of the jth variable in the ith couple of twins so that $r = 2$. Here, for each variable j ($j = 1, \ldots, p$), we consider the midpoint and the range of the two measurements of each couple of twins, instead of their observed values:

$$\begin{cases} y_{ij1} = \frac{y_{ij1}+y_{ij2}}{2} \\ y_{ij2} = |y_{ij1} - y_{ij2}| \end{cases}$$

Therefore the transformed Y_i is still a matrix of dimension $p \times 2$, but the order of its two columns is univocally determined.

A similar transformation is applied on the covariates. More precisely, if a covariate assumes individual-varying numerical values within each couple, midpoint and range are computed so that it is split into two new variables. Please note that, by considering the range, we lose the individual twin-specific connection between covariates and outcomes. But, what is really important in this application is to measure the effect of a big or a small covariate range on the observed range of an outcome, as a measure of the genetic link between the couple of twins. The positive or negative direction of the effect of a covariate on the responses is captured by regressing the corresponding midpoints.

If a covariate is numerical but it assumes the same value within the couples of twins (for instance, the age) the range is steadily zero and it may be dropped from the analysis. Qualitative predictors can be treated in a different manner. An individual-varying dummy (for instance, the individual smokes or does not smoke) is split into two dummies that are: 'both individuals in the couple have that modality' (both smoke) and 'only one has that modality' (only one smokes). The case 'none has that modality' (i.e. none smokes) is incorporated in the intercept. A non-individual-varying dummy (for instance, the twins are homozygous) remains unaltered. In case of categorical variables, each level can be split in the same way.

All the obtained and transformed covariates are stacked in the q-dimensional vector $\mathbf{x}_i$ and model (3) reduces to

$$Y_i = \Gamma_1 \mathbf{x}_i \mathbf{e}_1^\top + \Gamma_2 \mathbf{x}_i \mathbf{e}_2^\top + U_i \qquad (i = 1, \ldots, n). \tag{9}$$

In this manner, we measure the effect of the covariates on the response midpoints and response ranges separately. This is evident from the two sets of parameters in Γ_1 and Γ_2 that also imply that the model is characterized by two vectors of intercepts of length p.

The interpretation of the first set of parameters, Γ_1, is straightforward. A significant covariate coefficient for a given response midpoint measures the change in the expected value of the response midpoint of the individuals (without considering the genetic link between them) when the corresponding covariate is varied, while all other independent variables are held constant.

Of major interest is the interpretation of the second set of parameters, Γ_2. Suppose a regression coefficient of a covariate for a given response range is significant and positive. The covariate can be expressed as midpoint or range, and this yields two different scenarios. In the first situation we could say that, as the midpoint of the covariate increases, the range of the outcome increases as well because of the positive coefficient. This implies that the correlation between the twin members for a given response decreases, being the range inversely related to the correlation between twin members. Therefore the covariate could partly explain the variation due to the genetic link. On the contrary, a significant

TABLE 2

Estimation results of different models. The proposed matrix-variate regression analysis is compared with multivariate regression analysis, the structural equation models by Mx with ACE components and multivariate mixed effect model with random slopes. The statistics are: log-likelihood, the Akaike's Information Criterion and the Bayesian Information Criterion.

Model	logL	AIC	BIC
Matrix-variate regression	-6358.31	13040.61	13131.12
Multivariate regression	-6452.97	13259.93	13358.83
Mx SEM with ACE	-6736.37	13646.74	13695.35
Multivariate mixed effect model	-6750.39	13668.78	13715.72

and negative coefficient means that, if the corresponding covariate midpoint increases, the response range decreases, and therefore the twins will tend to assume more similar values and their correlation will be stronger. In the second scenario, for a positive coefficient we have that as the covariate range increases the response range increases as well and viceversa. Or, in other terms, big differences between covariate values imply big differences between twin responses, regardless their order. This means that the covariate has a relevant effect in explaining the variability between the twin members. In this perspective, a test of the null hypothesis of the second set of model intercepts becomes particularly important: an intercept not significantly different from zero implies that there is no expected variation between twins, given the covariate effect. On the contrary, if the null hypothesis is rejected we can conclude that there is an expected variability between the twins responses, not fully explained by the covariates.

3.3. Model estimation

The proposed matrix-variate regression model has been estimated on the re-coded responses and covariates, according to the transformation previously outlined. We started from the complete model with all the covariates. On the basis of the methodology presented in the last section, we have checked the association between the covariates on all the response variables. By choosing $M = I_p$ and $C = [1, 0, 0, \ldots, 0]$, $C = [0, 1, 0, \ldots, 0]$, ..., $C = [0, 0, 0, \ldots, 1]$ all the null hypotheses have been checked. In a backward strategy, only those covariates that were significant at each fit have been retained in the model.

The estimation details of the fitted model, including log-likelihood and information criteria, are reported in Table 2. For comparison purposes, we have also fitted a structural equation model using the `Mx` software (Neale et al. , 2006; Neale , 2003) with the so called ACE components, that are random effects describing the additive genetic factor (A), the shared environment (C), and the specific environment (E). We also estimated a multivariate mixed effect model using all the covariates previously described. Models with random intercepts and a multivariate mixed model with random intercepts and slopes have been fitted using the function `lme` in `R`. The estimation of the model with random effects both on intercepts and slopes does not converge and it crashes after some hours. This is mainly due to the complex estimation task deriving by a high number of covariates (i.e. 25) and to the multivariate nature of the problem (p=3).

Among all the considered approaches, the matrix-variate regression model is preferable, according to the information criteria AIC and BIC.

Table 3 reports the estimated coefficients of the final matrix-variate regression model. In

brackets the p-values associated to the significance test of each single coefficient are reported. These values have been obtained by testing the general hypothesis $H_0 : M\Gamma C^\top = 0$ with C set as before, and M a row vector of dimension p with value one in correspondence of the covariate position (and zero elsewhere).

The obtained values indicate that, as the perceived health increases (the correspondent covariate decreases), the mid-value of the three responses increases. Therefore the perceived health status and the cognitive functioning are positively strictly associated. The same positive effect can be observed for the level of education, as expectable. In general, females obtain higher scores in the episodic memory and Lawton's IADL tests than males, and as age increases, the cognitive functioning worsens. Surprisingly enough, the dummies 'one had cancer' and 'one had depression' have positive significant effects on Mental Status and Episodic Memory with respect to the status 'none had cancer (or depression)' incorporated in the intercept.

In regard to the second set of coefficients, it is important to observe that the intercepts are not significantly different from zero, meaning that the range within twins is fully explained by the set of covariates. The midpoint and the range of the perceived health status have a significant impact on the correlations between the mental status scores of the twins. More precisely, as the health status worsens and its variability increases, the correlation decreases. The mental status correlation is also positively affected by the level of happiness and satisfaction, that is, happier individuals tend to have more similar values of mental status of their twin. Other predictors are significantly associated with the responses, as the civil status with the modality 'never married', the sex and the age. Of particular importance is the role of the zygosity, that is significant for all the three responses, denoting that it explains a large part of correlation (the coefficients of the monozygotic dummy are all negative).

4. Conclusions

In this work a matrix-variate regression model with canonical states has been proposed and discussed. The proposed model has the merit to offer an interesting tool for evaluating the role of the covariates in explaining the variation of a set of quantitative responses. As a result of the particular parameterization of model (3) we can separate the effect of the covariates on the two sources of variability within and between the twins. Moreover, the separability of the total covariance matrix into the two covariance matrices Ω and Φ can be thought of as a separability condition on the two sources of correlations typical of twin studies that arise mostly by genetic factors (the within variability between the couples) and mostly by non-familiar factors (the midpoint of the couples) thus improving the model interpretation.

Compared with classical genetic multivariate models for twin and family studies (such as the structural equation modeling or the mixed effect models), it can aid the understanding of the contribution of covariates to the observed variation. On the contrary, if the interest is mainly focused on the genetic decomposition of the observed variation into genetic and environmental components, classical approaches should be preferable. For this reason, the presented approach should not be considered as an alternative to the classical approaches but as a complementary tool that could provide further interesting insight.

The model could be extended in several ways. Random effects could be incorporated in a matrix-variate perspective in order to reflect the genetic theory decomposition. In this

TABLE 3

Estimated regression coefficients of the matrix-variate regression model. In brackets the p-values associated to the null hypothesis of each single coefficient are reported. In the table Y_1 refers to the 'mental status', Y_2 is the 'episodic memory' and Y_3 refers to the 'Lawton's IADL'.

	Γ_1			Γ_2		
	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
Intercept	**23.57** (0.00)	**74.53** (0.00)	**65.51** (0.00)	-2.33 (0.49)	15.88 (0.18)	-7.62 (0.09)
Mid.Rate.health	**-0.72** (0.00)	**-4.09** (0.00)	**-1.79** (0.00)	**0.62** (0.02)	0.43 (0.64)	**1.02** (0.00)
Range.Rate.health	-0.11 (0.55)	-0.25 (0.71)	0.37 (0.15)	**0.43** (0.02)	0.42 (0.53)	0.36 (0.16)
Mid.Happy	-0.64 (0.07)	-2.27 (0.07)	-0.57 (0.23)	**0.82** (0.02)	-1.45 (0.24)	0.04 (0.94)
Range.Happy	-0.14 (0.54)	0.61 (0.46)	-0.31 (0.32)	0.28 (0.24)	1.45 (0.08)	**0.81** (0.01)
Both.Cancer	-1.19 (0.44)	-7.60 (0.16)	0.81 (0.69)	0.22 (0.89)	-1.76 (0.74)	-1.86 (0.37)
One.Cancer	**0.96** (0.03)	**3.99** (0.01)	0.18 (0.77)	-0.74 (0.10)	-1.71 (0.28)	-0.49 (0.42)
One.Stroke	-0.77 (0.12)	-0.35 (0.84)	**-1.46** (0.03)	**1.11** (0.02)	-0.29 (0.87)	**2.19** (0.00)
Both.Live.alone	0.72 (0.35)	3.95 (0.15)	**4.73** (0.00)	-1.15 (0.14)	4.89 (0.07)	**-2.30** (0.03)
One.Live.alone	0.49 (0.41)	2.12 (0.30)	**2.38** (0.00)	-0.22 (0.71)	**6.18** (0.00)	-0.24 (0.76)
Mid.Education	**0.31** (0.00)	**0.62** (0.04)	0.07 (0.55)	-0.09 (0.29)	0.45 (0.15)	0.20 (0.09)
Range.Education	0.10 (0.29)	**0.97** (0.00)	**0.25** (0.04)	-0.03 (0.76)	0.14 (0.65)	**-0.44** (0.00)
Both.Never.married	-1.57 (0.10)	-5.73 (0.09)	-0.11 (0.93)	**1.97** (0.04)	-5.62 (0.10)	-0.33 (0.80)
One.Never.married	-0.69 (0.27)	-3.89 (0.07)	-0.51 (0.54)	**1.28** (0.04)	-1.92 (0.38)	-0.18 (0.83)
Both.Divorced	-1.01 (0.65)	-3.64 (0.64)	1.61 (0.59)	-0.75 (0.74)	7.63 (0.33)	-3.08 (0.30)
One.Divorced	-0.06 (0.93)	1.47 (0.55)	-0.07 (0.94)	0.29 (0.68)	0.55 (0.82)	1.19 (0.20)
Both.Widow	-0.76 (0.39)	-5.72 (0.06)	-1.41 (0.23)	0.97 (0.27)	-3.53 (0.25)	1.00 (0.40)
One.Widow	-0.94 (0.12)	**-6.18** (0.00)	-0.49 (0.55)	0.47 (0.44)	-3.02 (0.16)	0.03 (0.97)
Both.Depression	0.77 (0.40)	-1.74 (0.59)	-0.01 (0.99)	**-2.59** (0.00)	**-6.47** (0.05)	-0.07 (0.96)
One.Depression	**1.04** (0.01)	**3.65** (0.01)	0.01 (0.98)	**-1.51** (0.00)	-2.16 (0.15)	0.15 (0.79)
Mid.Alcohol	0.07 (0.78)	-0.58 (0.49)	0.10 (0.75)	0.35 (0.14)	0.02 (0.98)	-0.37 (0.26)
Range.Alcohol	0.01 (0.97)	0.36 (0.62)	0.38 (0.17)	0.01 (0.98)	0.11 (0.88)	**-0.82** (0.00)
Age	-0.07 (0.10)	**-0.44** (0.00)	**-0.36** (0.00)	0.04 (0.38)	-0.08 (0.60)	**0.12** (0.03)
Both.Male	0.24 (0.58)	**-5.60** (0.00)	**-4.37** (0.00)	**-0.90** (0.04)	-0.18 (0.91)	**1.78** (0.00)
One.Male	-1.00 (0.22)	**-6.70** (0.02)	**-2.36** (0.03)	0.26 (0.75)	0.02 (0.99)	0.26 (0.81)
Monozygotic	0.00 (0.99)	-0.72 (0.54)	**1.32** (0.00)	**-0.61** (0.04)	**-3.16** (0.01)	**-1.35** (0.00)

way, a matrix-variate variant of the classical variance component model could be obtained. A further generalization of the method could consider mixture models (McLachlan & Peel , 2000; Ng & McLachlan , 2014) to deal with potential unobserved heterogeneity among couples of twins.

Acknowledgements

Data used for this research were provided by LSADT, which is supported in part by funds from Duke University under an award from the U.S. National Institutes of Health, and by the Danish National Research Foundation. The findings, opinions and recommendations expressed therein are those of the authors and are not necessarily those of LSADT, Duke University, the National Institutes of Health or the Danish National Research Foundation.

A. Proofs of Theorems

Without loss of generality, let us assume that $r = 2$. Extension to $r > 2$ is straightforward.

Proof of Theorem 1

Model (5) may be rewritten in a matrix-variate form as

$$\begin{aligned} Y &= \Gamma_1 \left(X \otimes \mathbf{e}_1^\top\right) + \Gamma_2 \left(X \otimes \mathbf{e}_2^\top\right) + U \\ &= [\Gamma_1, \Gamma_2] \begin{bmatrix} X \otimes \mathbf{e}_1^\top \\ X \otimes \mathbf{e}_2^\top \end{bmatrix} + U = [\Gamma_1, \Gamma_2] \left(X \otimes I_2\right) + U \end{aligned}$$

from which $\hat{Y} = [\hat{\Gamma}_1, \hat{\Gamma}_2] \left(X \otimes I_2\right)$. Substituting (7) in the previous expression we obtain

$$\hat{Y} = Y \left\{ X^\top \left(XX^\top\right)^{-1} X \otimes I_2 \right\}. \tag{10}$$

Now, we have $S_\Omega = \tilde{Y}\tilde{Y}^\top$ with $\tilde{Y} = Y - \hat{Y}$ a matrix of dimension $p \times 2n$. Putting (10) into the previous expression we obtain

$$\tilde{Y} = Y \left\{ I_{2n} - X^\top \left(XX^\top\right)^{-1} X \otimes I_2 \right\}$$

where $H = I_{2n} - X^\top \left(XX^\top\right)^{-1} X \otimes I_2$ is the hat matrix, that is symmetric idempotent with rank $2(n-q)$. Then

$$S_\Omega = YHY^\top = UHU^\top$$

because $(X \otimes I_2)H = 0$. Now we use Theorem 3.2.5 in Gupta & Nagar (2000) that establishes that if $U \sim \phi^{(p \times rn)}(0, I_{rn}, \Omega)$ and H $(rn \times rn)$ is a symmetric idempotent matrix with rank $r(n-q) > p$ then $UHU^\top \sim \mathcal{W}_p(r(n-q), \Omega)$.

Proof of Lemma 1

From (7) we have

$$\hat{\Gamma} = Y \left\{ X^\top \left(XX^\top\right)^{-1} \otimes I_2 \right\}.$$

Similarly, model (5) may be rewritten in a matrix-variate form as

$$\begin{aligned} Y &= \Gamma_1 \left(X \otimes \mathbf{e}_1^\top\right) + \Gamma_2 \left(X \otimes \mathbf{e}_2^\top\right) + U \\ &= [\Gamma_1, \Gamma_2] \begin{bmatrix} X \otimes \mathbf{e}_1^\top \\ X \otimes \mathbf{e}_2^\top \end{bmatrix} + U = \Gamma \left(X \otimes I_2\right) + U \end{aligned} \tag{11}$$

Therefore

$$\begin{aligned} E(\hat{\Gamma}) &= E(Y)\left\{X^\top \left(XX^\top\right)^{-1} \otimes I_2\right\} \\ &= \Gamma\left(X \otimes I_2\right)\left\{X^\top \left(XX^\top\right)^{-1} \otimes I_2\right\} = \Gamma \end{aligned}$$

and

$$var(\hat{\Gamma}) = \left\{\left(XX^\top\right)^{-1} X \otimes I_2\right\} var(Y) \left\{X^\top \left(XX^\top\right)^{-1} \otimes I_2\right\} = \left(XX^\top\right)^{-1} \otimes I_2 \otimes \Omega$$

since $var(Y) = I_{2n} \otimes \Omega$.

Proof of Theorem 2

From Lemma 1 we have

$$\hat{\Gamma} \sim \phi^{(p\times 2q)}\left(\Gamma, (XX^\top)^{-1} \otimes I_2, \Omega\right)$$

since $\hat{\Gamma}$ is a linear combination of Y. This also implies that

$$\Omega^{-\frac{1}{2}}\hat{\Gamma} \sim \phi^{(p\times 2q)}\left(\Omega^{-\frac{1}{2}}\Gamma, (XX^\top)^{-1}, I_p\right)$$

since matrix normal distributions are closed under linear transformations (see, for major details, Theorem 2.3.10 in Gupta & Nagar , 2000). Note that $\hat{\Gamma}$ and S_Ω are independent. This is true because $\hat{\Gamma}$ is independent from $\tilde{Y}$ where $\hat{\Gamma} = YA$ with $A = X^\top \left(XX^\top\right)^{-1} \otimes I_2$, $\tilde{Y} = YH$ and $A^\top H = 0$. Then observe

$$\Omega^{-\frac{1}{2}} S_\Omega \left(\Omega^{-\frac{1}{2}}\right)^\top \sim \mathcal{W}_p\left(2(n-q), I_p\right).$$

Now let $X_1 = \Omega^{-\frac{1}{2}}\hat{\Gamma}$ and $X_2 = \Omega^{-\frac{1}{2}} S_\Omega \left(\Omega^{-\frac{1}{2}}\right)^\top$. Then using result given in Dickey (1967) the transformation

$$\left(X_2^{-\frac{1}{2}}\right)^\top X_1 = S_\Omega^{-\frac{1}{2}}\hat{\Gamma} \sim \mathcal{T}_{2q,p}\left(2(n-q)-p-1, S_\Omega^{-\frac{1}{2}}\Gamma, (XX^\top)^{-1} \otimes I_2, I_p\right).$$

REFERENCES

L. Anderlucci, C. Viroli (2015). *Covariance Pattern Mixture Models for the Analysis of Multivariate Heterogeneous Longitudinal Data.* The Annals of Applied Statistics, 9, pp. 777-800.

K. E. Basford, G. J. McLachlan (1985). *The Mixture Method of Clustering applied to three-way data.* Journal of Classification, 2, pp. 109-125.

L. Brien, G. Fitzmaurice (2005). *Regression models for the analysis of longitudinal Gaussian data from multiple sources.* Statistics in Medicine, 24, pp. 1725-1744.

K. Christensen, J. W. Vaupel (2009). *Longitudinal Study of Aging Danish Twins, 1995.* ICPSR21041-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

K. Christensen, N. V Holm, M. McGue, L. Corder, J. W. Vaupel (1999). *A Danish population-based twin study on general health in the elderly.* Journal of Aging and Health, 11, pp. 49-64.

J. M. Dickey (1967). *Matrix variate generalizations of the multivariate* t *distribution and the inverted multivariate* t *distribution.* The Annals of Mathematical Statistics. 2, pp. 511-518.

E. Drigalenko (1998). *How sib pairs reveal linkage.* The American Journal of Human Genetics, 63, pp. 1242–1245.

P. Dutilleul (1999). *The MLE algorithm for the matrix normal distribution.* Journal of Statistical Computation and Simulation, 64, pp. 105-123.

R. C. Elston, S. Buxbaum, K. B. Jacobs, J. M. Olson (2000). *Haseman and Elston revisited.* Genetics Epidemiology, 19, pp. 1-17.

M. G. Genton (2007). *Separable approximations of space-time covariance matrices.* Environmetrics, Special Issue for METMA3, 18, pp. 681-695.

A. K. Gupta, D. K. Nagar (2000). *Matrix Variate Distributions.* Chapman & Hall/CRC.

J. K. Haseman, R. C. Elston (1972). *The investigation of linkage between a quantitative trait and a marker locus.* Behavior Genetics, 2, pp. 3-19.

M. P. Lawton, E. M. Brody (1969). *Assessment of older people: Self-maintaining and instrumental activities of daily living.* Gerontologist, 9, pp. 179-186.

K. V. Mardia, J. T. Kent, J, M. Bibby (2003). *Multivariate Analysis.* Academic Press.

M. McGue, K. Christensen (2007). *Social activity and healthy aging: A study of aging Danish twins.* Twin Research and Human Genetics, 10, pp. 255-265.

G. J. McLachlan, D. Peel (2000). *Finite Mixture Models.* Wiley, New York.

M. W. Mitchell, M. G. Genton, M. L. Gumpertz (2005). *Testing for separability of space-time covariances.* Envinronmetrics, 16, pp. 819-831.

M. W. Mitchell, M. G. Genton, M. L. Gumpertz (2006). *A likelihood ratio test for separability of covariances.* Journal of Multivariate Analysis, 97, pp. 1025-1043.

M. C. Neale, S. M. Boker, G. Xie, H. H. Maes (2006). Mx: Statistical Modeling (Seventh Edition), Virginia Commonwealth University, Department of Psychiatry, Richmond, VA.

M. C. Neale (2003). *A finite mixture distribution model for data collected from twins.* Twin Research, 6, pp. 235-239.

S. K. Ng, G. J McLachlan (2014). *Mixture of random effects models for clustering multilevel growth trajectories.* Computational Statistics & Data Analysis, 71, pp. 43-51.

C. Viroli (2011). *Finite mixtures of matrix normal distributions for classifying three-way data.* Statistics and Computing, 21, pp. 511-522.

C. Viroli (2012). *On matrix-variate regression analysis.* Journal of Multivariate Analysis, 111, pp. 296-309.

H. Wang, M. West (2009). *Bayesian Analysis of Matrix Normal Graphical Models.* Biometrika, 96, pp. 821-834.

F. A. Wright (1997). *The phenotypic difference discards sib-pair QTL linkage information.* The American Journal of Human Genetics, 60, pp. 740-742.

Y. M. Zhang, H. Y Lu, L. L. Yao (2008). *Multiple quantitative trait loci Haseman-Elston regression using all markers on the entire genome*, Theoretical and Applied Genetics, 117, pp. 683-690.

H. Zhou, L. Li (2014). *Regularized matrix regression.* Journal of the Royal Statistical Society - Series B, 76, pp. 463-483.

SUMMARY

A Matrix-Variate Regression Model with Canonical States: An Application to Elderly Danish Twins

In many situations we observe a set of variables in different states (e.g. times, replicates, locations) and the interest can be to regress the matrix-variate observed data on a set of covariates. We define a novel matrix-variate regression model characterized by canonical components with the aim of analyzing the effect of covariates in describing the variability within and between the different states. Despite the seeming complexity, inference can be easily performed through maximum likelihood. We derive the inferential properties of the model estimators and a general approach for hypothesis testing. Finally, the proposed method is applied to data coming from the Longitudinal Study of Aging Danish Twins (LSADT), so to investigate the causes of variation in cognitive functioning.

Keywords: Linear Regression; Matrix-variate normal distribution; Maximum Likelihood; Structural equation modeling; Twin data.