On matrix-variate regression analysis

Cinzia Viroli
Department of Statistics, University of Bologna, via Belle Arti, 41, Bologna, Italy

ARTICLE INFO

Article history:
Received 29 November 2011
Available online 3 May 2012

AMS subject classifications:
primary 62H20
secondary 62H12

Keywords:
Linear regression
Maximum likelihood
Random matrix
Three-way data

ABSTRACT

Three-way data arise in different application domains when multiple responses are measured at different time points or locations. A new regression model for analyzing three-way data is proposed. By assuming the matrix normal distribution for the error term, we will show that the proposed model represents the natural generalization of multiple and multivariate regression analysis. Inferential properties of the model estimators are derived. The model fit is illustrated on a real application.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Usually when referring to regression analysis the adjective multiple indicates that a set of covariates are being considered instead of only one and the adjective multivariate refers to the extension of multiple regression analysis by simultaneously considering a set of \( p \) (with \( p > 1 \)) correlated responses rather than a single response (see, for instance, [20]). The key step that allows this extension is the passage from the assumption of univariate Gaussian distribution to the multivariate Gaussian distribution of the model errors. The aim of this work is to extend multivariate regression analysis to deal with a set of variables simultaneously observed on several occasions. Considering the set of \( n \) observations the responses may be arranged in a three-way structure. Three-way data can occur in different application domains which include: spatial multivariate data, longitudinal data on multiple response variables, measurements at different time points and locations (thus leading to the so-called spatio-temporal data) or objects rated on multiple attributes by multiple experts. In all these situations, the data take the form of a three-dimensional array (instead of a matrix) characterized by three entities or modes: units (indexed by rows), variables (indexed by columns) and occasions (the layers). By denoting the set of multivariate responses observed in \( r \) occasions as \( p \), we have a \( p \times r \) observed matrix for each statistical unit.

In this work the problem of (linearly) regressing three-dimensional data to a set of covariates will be addressed. Let \( Y_i \) be the \( p \times r \) observed matrix for each statistical unit, with \( i = 1, \ldots, n \). Given a set of predictors \( X_i \) of dimension \( q \times r \) we define the general linear model

\[
Y_i = \Theta X_i + U_i, \quad (i = 1, \ldots, n).
\]

(1)

This regression model will be called matrix-variate regression analysis.

In this perspective the model errors, \( U_i \) of dimension \( p \times r \), should be thought of as realizations of a random matrix instead of a \( p \)-dimensional vector of random variables. On the probability theory side, the objective can be achieved by resorting to the so-called matrix-variate distributions. In the 1980s some authors began to study this family of distributions (see, among others, [4,5]), but their potential applications have been partially unexplored until the computational advances of
the recent years have made it possible. Recent applications include spatio-temporal analysis [21], Procrustes analysis [35], Bayesian graphical models [39] and model-based clustering [36,37]. Among the matrix-variate distributions, the matrix normal distribution [24,25,7] plays the same pivotal role which the multivariate normal distribution does in the family of multivariate distributions. The reasons are its mathematical tractability, which still holds in the matrix-variate context, its various properties and its role as reference model for most multivariate phenomena, which is guaranteed by the central limit theorem.

A random matrix \( Y \) of dimension \( p \times r \) has a matrix normal distribution with parameters \( M, \Omega \) and \( \Phi \) if

\[
f(Y|M, \Omega, \Phi) = (2\pi)^{-\frac{rp}{2}}|\Phi|^{-\frac{r}{2}}|\Omega|^{-\frac{p}{2}} \exp \left\{-\frac{1}{2} \text{tr} \left( \Omega^{-1}(Y - M)\Phi^{-1}(Y - M)^\top \right) \right\}
\]

where \( M \) is a matrix of dimension \( p \times r \) representing the expected value of \( Y \), \( \Omega \) is a covariance matrix of dimension \( p \times p \) containing the variances and covariances of the \( p \) variables or locations (with respect to all occasions/times) and \( \Phi \) is a covariance matrix of dimension \( r \times r \) containing the variances and covariances between the \( r \) occasions or times (with respect to all variables/locations). An equivalent definition specifies the \((p \times r)\)-matrix normal distribution as a special case of the \(pr\)-dimensional normal distribution when its covariance matrix, \( \Sigma \), is separable in the form \( \Sigma = \Omega \otimes \Phi \) (where \( \otimes \) is the Kronecker product). Thus a matrix normal distribution has the desirable feature of being able to estimate both the between and the within variable variation by resorting to a smaller set of parameters, which are \( r(r + 1)/2 \) instead of \( pr(pr + 1)/2 \) instead of \( pr(pr + 1)/2 \).

The matrix normal distribution is chosen to model the error term and consequently the response conditional means in matrix-variate regression analysis. This choice implies we are assuming the separability of the total variability into two sources originated by the within and between variable variation, expressed via \( \Sigma = \Omega \otimes \Phi \). Thanks to the matrix-normal distribution assumption we can derive asymptotic and finite-sample inferential properties of the parameter estimators, thus making possible to test hypotheses and construct confidence intervals. Furthermore, this assumption makes the matrix-variate regression analysis a very general framework in which both multivariate regression and multiple regression analysis can be obtained as special cases when \( r = 1 \) (or \( p = 1 \)), or \( r = 1 \) and \( p = 1 \), respectively.

The paper is organized as follows. The next section presents a review of the literature solutions proposed for regressing three-way data to a set of predictors. In Sections 2–4 the model is presented starting by the simplest scheme of only time-independent covariates in order to gradually come to a general model which include time-dependent and independent covariates and a time-dependent mean response. The inferential properties of the estimators are derived. Section 5 covers the problem of dealing with general linear hypotheses for the model parameters. In Section 6 the goodness of fit of the model is addressed. A real application is presented in Section 7. We conclude this work with a discussion.

1.1. A background

The problem of regressing three-way data to a set of covariates is not new and has been variously addressed by many authors.

An important class of regression models for dealing with three-way data is given by the random effects linear models. [28,29] considered the analysis of repeated-measurements where several characteristics are measured on each individual. In this formulation, a multivariate regression model with varying individual parameters and random-effects is defined and estimated. [16] discussed a general family of random effect models, which includes both growth models and repeated-measures models as special cases. [31] applied the linear mixed effects model to analyze three-way data with one covariate by taking compound symmetry as well as AR(1) correlation structures on \( \Phi \). An increasing interest in these models has been developed within the framework of hierarchical data where observations are nested within groups. These methods and further extensions have resulted in the established statistical framework called multilevel data analysis [10], which includes a variety of statistical methods for analyzing repeated measures or longitudinal data such as generalized latent variable models, multiple membership models and multivariate growth models (see, among others, [26,3,32]). Typically the role of random effects is to account for the component-specific covariances or correlations, which may be or not may be structured. In the proposed matrix-variate regression we incorporate only fixed effects (the covariates) since, for assumption, the separability of the variability within and between variables and occasions holds through \( \Sigma = \Omega \otimes \Phi \). Moreover, no individual-varying parameters are assumed, so that the proposed approach represents the direct generalization of multivariate regression analysis to matrix of observations (instead of vectors of observations).

In a different perspective, three recent works consider a multivariate regression model for three-way data where the separability condition, \( \Sigma = \Omega \otimes \Phi \), is explicitly assumed.

[23] solved the problem of performing a MANOVA test for multivariate repeated measures data assumed to be divided in \( g \) groups. For this purpose they assumed a multivariate regression model obtained by stacking all the repeated measurements on \( p \) variables along the first mode of observations in the form

\[
y_j = B^\top x_j + u_j, \quad (j = 1, \ldots, nr),
\]

where \( y_j \) is a \( p \)-dimensional vector, \( B \) is a matrix of parameters of dimension \( q \times p \), \( x_j \) is a design vector of length \( q \) and \( u_j \) is the \( p \)-dimensional vector of errors. In this model they assumed a Kronecker product covariance matrix to analyze separately
the within and between covariation of the variables. The aim of this paper is to propose a more general matrix-variate formulation for different types of three-way data (including multivariate repeated measures) which explicitly consider matrix of observations as in (1) without unfolding one of the modes of the three-way structure.

In the work of [22], a likelihood ratio test for the separability of the covariance in a particular multivariate regression model is considered. More precisely, the multivariate regression analysis is adapted to deal with three-way data by assuming the model

\[ y_i = B^\top x_i + u_i, \quad (i = 1, \ldots, n), \]

where \( y_i \) is a \( pr \)-dimensional vector, \( B \) is a matrix of parameters of dimension \( q \times pr \), \( x_i \) is a vector of \( q \) time-independent predictors and \( u_i \) is the \( pr \)-dimensional vector of errors. In this model \( u_i \) is assumed to be distributed according to a multivariate Gaussian with a separable covariance matrix. Our proposal differentiates by this approach because we will consider both time-dependent and independent covariates and time-dependent mean responses. Moreover, switching the focus from a multivariate perspective to a matrix-variate one allows us to construct a probabilistic framework for the inferential properties of the model parameters.

[1] focused on the comparison between random effect models and a regression model based on the separability condition where the covariates have the same dimensionality of the responses and regression coefficients do not vary between and within variables and occasions. More specifically, they considered the multivariate regression model

\[ y_i = X_i\theta + u_i, \quad (i = 1, \ldots, n), \]

where \( y_i \) is a \( pr \)-dimensional vector, \( X_i \) is a matrix of covariates of dimension \( pr \times q \) (which are supposed to be both time-varying and variable-varying) and \( \theta \) is a vector of parameters of length \( q \). The error term \( u_i \) is a \( pr \)-dimensional vector with multivariate Gaussian distribution. By comparing this model with the matrix-variate regression model in (1) two main differences emerge. First, we model the more realistic situation in which predictors can differently affect the \( p \) observed measurements through a matrix of parameters having dimension \( p \times q \) and then, again, we embrace a new matrix-variate perspective that allows to cast the problem in a more general inferential framework.

A parallel line of research has been devoted to the so-called multiway regression models mainly motivated by chemical applications [34,12]. Multiway regression models may be classified into three methods. The unfold Partial Least Squares (PLS) [40] consists in unfolding the modes of the three-way data so as to convert them to the conventional matrix data, and thereby to apply a two-way PLS. In the second approach, the two-way PLS algorithm is generalized to data of higher order using a multilinear Parallel Factor Analysis (PARAFAC) structure [2]. The third method is the multiway covariates regression [33] in which the principal covariates regression model is generalized to data of higher order. These methodologies are based on least squares optimization strategies to estimate the parameters of the model which do not require explicit distributional assumptions. On the contrary, we will adopt a likelihood-based approach. We will show that advantages are many and they include the ability to demonstrate some convergence properties, the possibility to perform hypothesis testing and the capability to compare different models. In the next section a matrix-variate regression model with time-independent covariates is introduced.

2. Model with time-independent covariates

We consider the case where several characteristics are measured on each individual at each occasion or, alternatively, a univariate variable is measured in different locations and experimental conditions (or times). We assume that we have \( n \) individuals, \( p \) characteristics or locations and \( r \) occasions or conditions. This situation yields a \( p \times r \) observed matrix, \( Y_i \), for each statistical unit, with \( i = 1, \ldots, n \). We assume a model for \( Y_i \) to be of the form

\[ Y_i = A z_j^\top + U_i, \quad (i = 1, \ldots, n), \tag{2} \]

where \( A \) is a matrix of dimension \( p \times q \) of unknown parameters, \( z_j \) is a vector of \( q \) covariates, which vary among individuals but are supposed to be equal across the different occasions, and \( j \) is an \( r \times 1 \) vector of ones. The error term matrix, \( U_i \), is supposed to be distributed according to a matrix normal distribution, \( U_i \sim \mathcal{N}_{p \times r}(0, \Phi, \Omega) \), that is

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

where \( \Phi \) is a \( r \times r \) covariance matrix containing the variances and covariances between the \( r \) occasions (or times) and \( \Omega \) is a \( p \times p \) covariance matrix containing the variance and covariances of the \( p \) variables.

Introducing the vec operator which stack the columns of a matrix, such that \( y_i = \text{vec}(Y_i) \) and \( u_i = \text{vec}(U_i) \) are vectors of dimension \( pr \times 1 \), we can express (2) as

\[ y_i = (j \otimes I_p) \text{vec}(A z_j^\top) + u_i = (j \otimes I_p) (z_j^\top \otimes I_p) \text{vec}(A) + u_i \]

\[ = (j, z_j^\top \otimes I_p) \gamma + u_i, \quad (i = 1, \ldots, n), \tag{3} \]
which lead to a system of coupled equations.

\[ \hat{\gamma} = \left( (Z^T)^{-1} Z \otimes \frac{I}{r} \otimes I_p \right) y. \]

From the properties of the matrix normal distribution it follows that if \( U_i \sim \mathcal{N}_{p r n}(0, \Phi \otimes \Omega) \) then \( u_i = \text{vec}(U_i) \) is distributed according to the multivariate normal \( \mathcal{N}_{p r n}(0, \Phi \otimes \Omega) \) (see, for major details, [8,11]). Therefore \( u \sim \mathcal{N}_{p r n}(0, I_n \otimes \Phi \otimes \Omega) \) and \( yZ \sim \mathcal{N}_{p r n}(Z^T \otimes j, I_n \otimes \Phi \otimes \Omega) \). Note that if \( \Omega \) and \( \Phi \) are diagonal matrices (this means that there is no correlation between the responses and across all occasions) the matrix-regression analysis in (4) is equivalent to perform independently several \( p \times r \) multiple regression analyses.

### 2.1. Inference on model parameters

Thanks to the representation of the matrix-variate regression model as a particular multivariate regression model in expression (4), the generalized least square of \( y \) is the same as the least square estimator. It also coincides with the maximum likelihood estimator given by

\[ \hat{\gamma} = \left( (Z^T)^{-1} Z \otimes \frac{I}{r} \otimes I_p \right) y. \]

By using the vec and Kronecker properties, we can derive the estimator for \( \Gamma' \) as

\[ \hat{\gamma} = \left( Z^T (Z^T)^{-1} \otimes \frac{I}{r} \right)^{-1} Z^T \]

the maximum likelihood estimator of \( \Gamma' \), where \( Y = (Y_1, \ldots, Y_n) \) is a matrix of dimension \( p \times r n \). By setting \( \hat{M}_i = \hat{\gamma} \; z_i j_i^T \), and \( \hat{U}_i = Y_i - \hat{M}_i \), estimation for \( \Phi \) and \( \Omega \) can be obtained by evaluating and differentiating the log-likelihood

\[ \log L(\Phi, \Omega | Y_1, \ldots, Y_n) = \frac{p n}{2} \log(2\pi) - \frac{p n}{2} \log |\Phi| - \frac{m}{2} \log |\Omega| - \frac{1}{2} \sum_{i=1}^{n} \text{tr} \left( \Omega^{-1} \hat{U}_i \Phi^{-1} \hat{U}_i^T \right), \]

which leads to a system of coupled equations

\[
\begin{align*}
\hat{\Phi} &= \frac{1}{n p} \sum_{i=1}^{n} \hat{U}_i \Delta \hat{U}_i^T \\
\hat{\Omega} &= \frac{1}{n r} \sum_{i=1}^{n} \hat{U}_i \hat{\Phi}^{-1} \hat{U}_i^T.
\end{align*}
\]

This means that there is not a closed form analytic solution for estimating the two covariance matrices. Their values must be computed in an iterative fashion (see, for instance, the strategies proposed in [7, 18, 30]). The solution is unique up to a multiplicative constant, say \( a \neq 0 \), since \( \Phi \otimes \Omega = a \Phi \otimes \frac{1}{a} \Omega \). In practice, a way to obtain a unique solution is to impose the identifiability constraint \( \text{tr} \Phi = r \) or alternatively \( \sum_{h,c} \Phi_i = r^2 \), where \( h \) and \( c \) index rows and columns of \( \Phi \) and \( \phi \) is the single element in \( \Phi \).

In the next theorem we state some inferential properties of the parameter estimator \( \hat{\gamma} \).

**Theorem 1.** \( \hat{\gamma} \) is a linear and unbiased estimator of \( \gamma \) with covariance matrix

\[ \Sigma_{\gamma | \phi, \omega} = \left( (Z^T)^{-1} \otimes \Omega \right) \sum_{h,c} \phi_{h,c} / r^2. \]

**Proof.** In order to verify the estimator unbiasedness we consider that model (2) can be rephrased by including all the observations as

\[ Y = \Gamma'Z (I_n \otimes j,^T) + U, \]

where \( Y \) and \( U \) are matrices of dimension \( p \times r n \). By substituting the previous equation into (5) and evaluating the expected value we obtain

\[ \mathbb{E}(\hat{\gamma}) = \Gamma'Z (I_n \otimes j,^T) \left( Z^T (Z^T)^{-1} \otimes \frac{j,}{r} \right) = \Gamma \left( I_q \otimes \frac{1}{r} \right) = \Gamma. \]
This implies that $E(\hat{y}) = y$. The covariance matrix of $\hat{y}$ is

$$
\text{Var}(\hat{y}) = \left\{ (Z\hat{Z})^{-1}Z \otimes \frac{j^T}{r} \otimes I_p \right\} \text{Var}(y) \left\{ (Z\hat{Z})^{-1}Z \otimes \frac{j^T}{r} \otimes I_p \right\}^T
$$

$$
= \left\{ (Z\hat{Z})^{-1}Z \otimes \frac{j^T}{r} \otimes I_p \right\} (I_n \otimes \Phi \otimes \Omega) \left\{ Z^T(Z\hat{Z})^{-1} \otimes \frac{j}{r} \otimes I_p \right\}
$$

$$
= (Z\hat{Z})^{-1} \otimes \frac{j^T \Phi j}{r^2} \otimes \Omega = (Z\hat{Z})^{-1} \otimes \Omega \sum_{h,c} \phi_{h,c}/r^2. \quad \Box
$$

**Corollary 2.1.** The finite-sample distribution of $(\hat{y} - y)$ is the multivariate Gaussian

$$(\hat{y} - y) \sim \mathcal{N}_{pq}(0, \Sigma_{y|\Phi, \Omega})$$

with $\Sigma_{y|\Phi, \Omega}$ defined in (7).

**Corollary 2.1** is true since $\hat{y}$ is a linear combination of $y$.

**Corollary 2.2.** Under the identifiability constraint $\sum_{h,c} \phi_{h,c} = r^2$

$$(\hat{y} - y) \rightarrow \mathcal{N}_{pq}(0, (Z\hat{Z})^{-1} \otimes \hat{\Omega})$$

in probability as $n \rightarrow \infty$.

**Corollary 2.2** follows by considering that $\hat{y}$ is the maximum likelihood estimator of $y$ and $\hat{\Omega}$ is a consistent estimator of $\Omega$.

Theorem 1 and its corollaries establish some interesting properties of the parameter estimator, $\hat{y}$, but the finite-sample distribution of $\hat{y}$ cannot be used for hypothesis testing, since it depends on the unknown population matrices $\Phi$ and $\Omega$. This is due to the system of coupled equations (6) which does not allow to obtain closed-form estimators for $\Phi$ and $\Omega$. A possible way to overcome this drawback is to assume $\Phi = I$. However this constraint is too much restrictive and not realistic in this case, unless the observed $Y_i$ are assumed to be conditionally time-independent given a design matrix to model time-interactions, as shown in the next section.

### 3. Model with time-dependent mean response

We consider the model for $Y_i$

$$
Y_i = AW + U_i, \quad (i = 1, \ldots, n), \quad (8)
$$

where $A$ is a matrix of unknown parameters of dimension $p \times m$ and $W$ is an $m \times r$ design matrix to model time-relations via a polynomial or a non-linear function of order $m - 1$. For convenience we impose that the matrix $W$ has as its first row a row of ones to incorporate the intercept in the model. We assume that the mean response describes all the cross-time variability of $Y_i$ so that $Y_i | W \sim \mathcal{N}_{p	imes r}(AW, I_r, \Omega)$ and $U_i \sim \mathcal{N}_{p	imes r}(0, I_r, \Omega)$.

Now (8) can also be expressed as

$$
y_i = (W^T \otimes I_p)\lambda + u_i, \quad (i = 1, \ldots, n),
$$

where $\lambda = \text{vec}(A)$, so that we have

$$
y = (j_n \otimes W^T \otimes I_p)\lambda + u. \quad (9)
$$

#### 3.1. Inference on model parameters

Starting from the model formulation in (9) the least square and maximum likelihood estimator of $\lambda$ is

$$
\hat{\lambda} = \left\{ J_n \otimes (WW^T)^{-1}W \otimes I_p \right\} y. \quad (10)
$$

from which

$$
\hat{\Lambda} = Y \left\{ \frac{J_n}{n} \otimes W^T(WW^T)^{-1} \right\}. \quad (11)
$$
The maximum likelihood estimator of $\Omega$ is given by $S_\Omega / \text{nr}$ where $S_\Omega$ is the residual sum of squares matrix

$$S_\Omega = \sum_{i=1}^{n} (Y_i - \hat{A}W)(Y_i - \hat{A}W)^T.$$  \hfill (12)

In the next theorems we derive the distributional form of the two estimators, $S_\Omega$ and $\hat{A}$.

**Theorem 2.** Under the constraint $p < \text{nr} - m$, the residual sum of squares matrix $S_\Omega$ is distributed as a Wishart with parameter $\Omega$ and degrees of freedom $\text{nr} - m$, $S_\Omega \sim \mathcal{W}_p(\text{nr} - m, \Omega)$.

**Proof.** We have $S_\Omega = \hat{U}\hat{U}^T$ with $\hat{U} = Y - \hat{Y}$ a matrix of dimension $p \times \text{rn}$ and $\hat{Y} = \hat{A}(j_n^\top \otimes W)$. Substituting the value of $\hat{A}$ from (11) into $\hat{U}$ we obtain

$$\hat{U} = Y - Y \left\{ \frac{j_n}{n} \otimes W^T(WW^T)^{-1} \right\} (j_n^\top \otimes W) = Y \left\{ I_m - \frac{j_n}{n} \otimes W^T(WW^T)^{-1}W \right\},$$

where $j_n$ is a matrix of ones of order $n$ and $H = I_m - \frac{j_n}{n} \otimes W^T(WW^T)^{-1}W$ is the hat matrix, that is symmetric idempotent with rank $\text{nr} - m$. Now

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

because $(j_n^\top \otimes W)H = 0$. Now we use Theorem 3.2.5 in [11] that establishes that if $U \sim \phi(p \times \text{rn})(0, I_m, \Omega)$ and $H (\text{rn} \times \text{rn})$ is a symmetric and idempotent matrix with rank $\text{nr} - m > p$ then $UHU^T \sim \mathcal{W}_p(\text{nr} - m, \Omega)$. \hfill $\Box$

**Corollary 3.1.** The unbiased estimator of $\Omega$ is

$$\hat{\Omega} = \frac{S_\Omega}{\text{nr} - m}.$$  

The next theorem represents the direct generalization of the Theorem of Gauss–Markov to matrix-variate regression analysis. Its proof is relatively straightforward with respect to multivariate regression analysis, but it will be outlined here for the sake of completeness.

**Theorem 3 (Gauss–Markov).** Let $\hat{\lambda}$ be the estimator defined in (10), $\hat{\lambda}$ is a linear unbiased estimator with covariance matrix

$$\text{Var}(\hat{\lambda}) = \frac{(WW^T)^{-1}}{n} \otimes \Omega$$

which is smaller than that of any other linear unbiased estimator.

**Proof.** Substituting $Y = A(j_n^\top \otimes W) + U$ into (11) it is easy to check that $\hat{\lambda}$ is an unbiased estimator for $\lambda$. This implies that $\hat{\lambda}$ is unbiased, as well. The covariance matrix of $\hat{\lambda}$ can be easily obtained from (11) as

$$\text{Var}(\hat{\lambda}) = \left\{ \frac{j_n}{n} \otimes (WW^T)^{-1}W \otimes I_p \right\} \text{Var}(Y) \left\{ \frac{j_n}{n} \otimes W^T(WW^T)^{-1} \otimes I_p \right\}$$

$$= \left\{ \frac{j_n}{n} \otimes (WW^T)^{-1}W \otimes I_p \right\} (I_n \otimes I_r \otimes \Omega) \left\{ \frac{j_n}{n} \otimes W^T(WW^T)^{-1} \otimes I_p \right\},$$

from which the covariance matrix (13) follows. Now consider another linear unbiased estimator of $\lambda$ in the form $\hat{\lambda}^* = Cy$, where $C$ is a matrix of dimension $pm \times pm$. From (9) it follows that $\hat{\lambda}^*$ is unbiased if and only if $C(j_n \otimes W^T \otimes I_p) = I_m$. Without loss of generality we consider $C = \left\{ \frac{k_i}{n} \otimes (WW^T)^{-1}W \otimes I_p \right\} + D$. Then $C(j_n \otimes W^T \otimes I_p) = I_m$ implies $D(j_n \otimes W^T \otimes I_p) = 0$. Now it is easy to show that $\hat{\lambda}^* = \lambda + \left\{ \frac{k_i}{n} \otimes (WW^T)^{-1}W \otimes I_p \right\} + D$ and $\text{Var}(\hat{\lambda}^*) = \text{Var}(\hat{\lambda}) + D(I_n \otimes I_r \otimes \Omega)D^T$. The theorem follows by observing that $D(I_n \otimes I_r \otimes \Omega)D^T$ is positive semi-definite. \hfill $\Box$

The next theorem establishes the distributional form of $S_{\Omega}^{-\frac{1}{2}} \hat{\lambda}$.

**Theorem 4.** Let $\hat{\Lambda}$ be the estimator defined in (11) and $S_{\Omega}$ the residual sum of squares matrix in (12). Then $S_{\Omega}^{-\frac{1}{2}} \hat{\Lambda}$ has a matrix-variate $T$ distribution

$$S_{\Omega}^{-\frac{1}{2}} \hat{\Lambda} \sim T_{m \times p} \left( \text{nr} - m - p - 1, S_{\Omega}^{-\frac{1}{2}} \Lambda, \frac{(WW^T)^{-1}}{n}, I_p \right),$$

where $\text{nr} - m - p - 1$ are the degrees of freedom.
Proof. Since \( \hat{\Lambda} \) is a linear estimator of \( Y \) which has a matrix normal distribution, it is distributed as

\[
\hat{\Lambda} \sim \mathcal{N}_{p \times m} \left( \Lambda, \frac{(WW^\top)^{-1}}{n}, \Omega \right).
\]

because matrix normal distributions are closed under linear transformations (see, for major details, [11]). This also implies that

\[
\Omega^{-\frac{1}{2}} \hat{\Lambda} \sim \mathcal{N}_{p \times m} \left( (WW^\top)^{-1} \frac{1}{n}, \Omega \right).
\]

Note that \( \hat{\Lambda} \) and \( S_{\Omega} \) are independent. This is true because \( \hat{\Lambda} \) is independent of \( \hat{U} \) where \( \hat{\Lambda} = YA \) with \( A = \frac{k_1}{n} \otimes W^\top (WW^\top)^{-1}, \hat{U} = YH \) and \( A^\top H = 0 \). Then observe

\[
\Omega^{-\frac{1}{2}} S_{\Omega} \left( \Omega^{-\frac{1}{2}} \right)^\top \sim W_p (nr - m, I_p).
\]

Now let \( X_1 = \Omega^{-\frac{1}{2}} \hat{\Lambda} \) and \( X_2 = \Omega^{-\frac{1}{2}} S_{\Omega} \left( \Omega^{-\frac{1}{2}} \right)^\top \). Then using result given in [6] the transformation

\[
\begin{align*}
(X_2^{-\frac{1}{2}})^\top X_1 &= S_{\Omega}^{-\frac{1}{2}} \hat{\Lambda} \sim T_{mxp} \left( nr - m - p - 1, S_{\Omega}^{-\frac{1}{2}} A, \frac{(WW^\top)^{-1}}{n}, I_p \right). \quad \Box
\end{align*}
\]

Theorem 4 allows to test the null hypothesis of absence of time structure in the matrix-variate data, \( H_0 : \Lambda = 0 \) through the statistics

\[
T = S_{\Omega}^{-\frac{1}{2}} \hat{\Lambda} \sim T_{mxp} \left( nr - m - p - 1, 0, \frac{(WW^\top)^{-1}}{n}, I_p \right).
\]

Since the matrix-variate \( T \) distribution is closed under linear transformation and partitions we could also test the null hypothesis on a single element of \( \Lambda \), \( H_0 : \Lambda_{jh} = 0 \) with \( j = 1, \ldots, p \) and \( h = 1, \ldots, r \) via

\[
T_{jh} \sim t_{nr - m - p - 1} \left( \frac{(WW^\top)^{-1}}{n} \right).
\]

This approach could be employed to test different time structures for the \( p \) variables, by applying a battery of tests. However, it is well known that it may cause the inflation of the true significance level. A better and generalized strategy for simultaneous testing will be presented in Section 5.

4. A generalization

We now consider a general model which includes covariates and a time-dependent mean response. Let \( Z_i \) be a \( q \times r \) matrix consisting of a set of time-dependent covariates, say \( Z_i^{(1)} \), and an eventual set of time-independent predictors as in Section 2, so that \( Z_i^\top = \left( Z_i^{(1)^\top}, \mathbf{j}_i Z_i^{(2)^\top} \right)^\top \).

We assume the model for \( Y_i \)

\[
Y_i = \Gamma Z_i + \Lambda W + U_i, \quad (i = 1, \ldots, n), \tag{14}
\]

where \( \Gamma \) and \( \Lambda \) are matrices of unknown parameters of dimension \( p \times q \) and \( p \times m \), respectively, and \( W \) is the \( m \times r \) design matrix to model time-relations defined in Section 3. The error term matrix, \( U_i \), is assumed to be distributed as a matrix normal, that is \( U_i \sim \mathcal{N}_{p \times r}(0, I, \Omega) \).

Expressing model (14) in vector notation, we have

\[
\mathbf{y}_i = (Z_i^\top \otimes I_p) \mathbf{y} + (W^\top \otimes I_p) \lambda + U_i, \quad (i = 1, \ldots, n), \tag{15}
\]

and considering all observations

\[
\mathbf{y} = (Z^\top \otimes I_p) \mathbf{y} + (\mathbf{j}_n \otimes W^\top \otimes I_p) \lambda + \mathbf{u},
\]

where \( Z \) is the matrix of predictors of dimension \( q \times m \). Substituting \( X^\top = (Z^\top, \mathbf{j}_n \otimes W^\top) \) and \( \theta = \text{vec}(\theta) = \text{vec} (\Gamma, \Lambda) \), this model has the compact form

\[
\mathbf{y} = (X^\top \otimes I_p) \theta + \mathbf{u}, \tag{16}
\]

and model (14) can be expressed as

\[
Y_i = \Theta X_i + U_i, \quad (i = 1, \ldots, n), \tag{17}
\]

where \( \Theta \) is the matrix containing the compact set of parameters to be estimated of dimension \( p \times (q + m) \) and \( X_i \) is the predictor matrix of dimension \((q + m) \times r\).
4.1. Inference on model parameters

Under parametrization in (16) the maximum likelihood estimator of \( \theta \) takes the simple form

\[
\hat{\theta} = \{XX^\top\}^{-1}X \otimes I_p \} y,
\]
from which it is easy to derive the maximum likelihood estimators of \( \Theta \)

\[
\hat{\Theta} = YX^\top (XX^\top)^{-1}.
\]

The maximum likelihood estimator of \( \Omega \) has a similar expression of the analogue estimator derived in the previous section, and it is given by \( S_{\Omega}/n_r \) where \( S_{\Omega} \) is the residual sum of squares matrix

\[
S_{\Omega} = \sum_{i=1}^n \left(Y_i - \hat{\Theta}X_i\right) \left(Y_i - \hat{\Theta}X_i\right)^\top.
\]

In the next theorems the inferential properties of \( S_{\Omega} \) and \( \hat{\Theta} \) are derived. Proofs of theorems follow the same steps of those of previous section and are omitted for easy of presentation.

**Theorem 5.** Under the constraint \( p < nr - q - m \), the residual sum of squares matrix \( S_{\Omega} \) is distributed as a Wishart with parameter \( \Omega \) and degrees of freedom \( nr - q - m, S_{\Omega} \sim W_p(nr - q - m, \Omega) \).

The proof is similar to the proof of Theorem 2 and follows by expressing \( S_{\Omega} = \hat{U}\hat{U}^\top \) with \( \hat{U} = Y - \hat{\Theta}X \).

**Corollary 4.1.** The unbiased estimator of \( \Omega \) is

\[
\hat{\Omega} = \frac{S_{\Omega}}{nr - q - m}.
\]

**Theorem 6** (Gauss–Markov). Let \( \hat{\theta} \) be the estimator defined in (18). \( \hat{\theta} \) is a linear unbiased estimator with covariance matrix

\[
\text{Var}(\hat{\theta}) = (XX^\top)^{-1} \otimes \Omega
\]

which is smaller than that of any other linear unbiased estimator.

**Theorem 7.** Let \( \hat{\Theta} \) be the estimator defined in (19) and \( S_{\Omega} \) the residual sum of squares matrix in (20). Then \( S_{\Omega}^{-\frac{1}{2}} \hat{\Theta} \) has a matrix-variate \( T \) distribution

\[
S_{\Omega}^{-\frac{1}{2}} \hat{\Theta} \sim T_{(q+m)\times p} \left(nr - q - m - p - 1, S_{\Omega}^{-\frac{1}{2}} \Theta, (XX^\top)^{-1}, I_p\right),
\]

where \( nr - q - m - p - 1 \) are the degrees of freedom.

**Theorem 7** states that \( \hat{\Theta} \) is a linear and unbiased estimator of \( \Theta \) with covariance matrices \( I_p \) and \( (XX^\top)^{-1} \). The theorem also allows to perform a battery of test to check hypotheses on the single parameters. In the next section we consider simultaneous linear hypotheses for the model parameters \( \Theta \).

4.2. Relation with multivariate growth curve models

Growth curve models have been studied extensively in the literature because of their flexibility to analyze those experiments which consider the response of an individual over a period of time, or over different replications. Traditional analyses of growth data, following the model suggested by [27], are based on a single response; see for example [38] or [9] and references therein. Extensions to situations where several responses are observed have been considered in [28,19,26,17] among others. Using the notation of this section, a multivariate growth model can be formulated as

\[
y_i = (z_i^\top \otimes I_p \otimes W^\top) \beta + (I_p \otimes W_c^\top) s_i + u_i, \quad (i = 1, \ldots, n),
\]

where \( y_i \) is the \( pr \)-vector of responses, the first addendum is the fixed part of the model and the second term represents the random part; \( z_i \) is a \( q \times 1 \) set of covariates (including the intercept), \( W \) is a time-matrix of dimension \( m \times r \), \( \beta \) is a vector of \( pqm \) parameters and \( W_c \) is a design matrix of dimension \( r \times c \). The random effect \( s_i \) is distributed as a Gaussian with zero vector mean and covariance matrix \( D \) of dimension \( pc \times pc \) and it is assumed to be independent of the error term \( u_i \), which is a Gaussian with zero vector mean and covariance matrix \( I_r \otimes \Omega \) of dimension \( pr \times pr \). Now using the Kronecker properties the previous model may be rewritten in the form

\[
Y_i = (z_i^\top \otimes I_p) BW + S_i W_c + U_i, \quad (i = 1, \ldots, n).
\]
This formulation is unusual in the growth model framework but it resembles dimensionality of expression (14). Indeed, in (21), \( Y \) has dimension \( p \times r \) and \( B \) is a matrix of unknown parameters of dimension \( (pq) \times m \). Now comparing the previous model with (14) it is clear that, without the random part, the multivariate growth curve model is a particular matrix-normal regression model with \( U_i \sim \mathcal{N}_{pq} (0, I_r, \Omega) \). However, in (14), we have a form of 'separability' between covariates and design matrix \( W \) because they are treated as two different additive terms. This leads to several desirable properties. First of all, from a computational point of view, model (14) involves less parameters to be estimated, which are \( p(q+m) \) instead of \( pqm \) for the fixed part, having the error term, \( U_i \), the same distribution between the two configurations. In model (21) the additional free elements in \( D \) have also to be estimated. Second, the separability of the regression coefficients allows to better interpret the model with regards to the diverse components that affect the responses. Moreover, this also allows to perform hypothesis testing separately for the covariates and the polynomial components of the time structure in \( W \).

4.3. A comparison on simulated data

Performance of the matrix-variate regression analysis and of the multivariate growth curve model have been evaluated on simulated data. We have considered \( p = 5 \) observed variables in \( r = 6 \) different times. The temporal trend of the data is supposed to be governed by the quadratic function \( 6 - 1.5t + t^2 \). The observed data are related to a covariate with values randomly generated by a Uniform distribution in \((-5, 5)\) and with regression coefficients randomly generated by a Uniform distribution in \((-1, 1)\). The error term has been randomly generated by a matrix-variate distribution with zero mean and covariance matrices \( \Phi = I_r \) and \( \Omega = I_p \). From this simulation design, we have generated 100 different samples in three scenarios with \( n = 50, n = 100 \) and \( n = 200 \), respectively. On the simulated data the root mean squared error (RMSE) and the mean absolute error (MAE) for prediction have been computed for the two models. Table 1 reports the obtained results. From the table it is clear that the matrix-variate regression model fits better the data in all the situations \( n = 50, n = 100 \) and \( n = 200 \).

5. Testing hypothesis

Without loss of generality we consider here testing hypothesis about the regression coefficients and the covariance matrix in the general model described in Section 4.

5.1. General linear hypothesis about the regression coefficients

The issue of testing general linear hypothesis for the regression coefficients in multivariate regression analysis has been addressed using various methods in the statistical literature (please see [20] for some of the most used procedures and references therein). Here we wish to develop the procedure for the matrix-variate regression model. Consider testing linear hypotheses of the form \( H_0 : M \hat{\Theta} C^\top = 0 \), where \( M \) is a matrix of dimension \( c \times p \) with rank \( c \leq p \), and \( C \) is a \( g \times (q+m) \) matrix with rank \( g \leq (q+m) \). This represents a general formulation for hypothesis on the regression coefficients because it includes all the interesting situations. For example, by setting \( M = I_p \) and \( C = [1, 0, \ldots, 0] \) the null hypothesis about the relevance of the first regressor versus all the responses can be tested. Alternatively, if \( M = [1, 0, \ldots, 0] \) and \( C = [1, 0, \ldots, 0] \) we test the single hypothesis about the significance of the first regressor versus the first response (as an alternative to procedure resulting from Theorem 7). Let \( \hat{\Delta} \) be \( \hat{\Delta} = M \hat{\Theta} C^\top \) and consequently \( \hat{\delta} = \text{vec}(\hat{\Delta}) = (C \otimes M)\hat{\theta} \). Now from (18) it is evident that \( \hat{\theta} \) is a linear combination of \( \delta \). Hence under the null hypothesis

\[
\hat{\delta} \sim \mathcal{N}_{cg} \left( 0, C (XX^\top)^{-1} C^\top \otimes M\Omega M^\top \right).
\]

Thus we can define a hypothesis testing matrix as

\[
H = M \hat{\Theta} C^\top \left( C (XX^\top)^{-1} C^\top \right)^{-1} C \hat{\Theta} M^\top
\]
which is distributed as Wishart, \( H \sim W_{c}(g, M \Omega M^{\top}) \). \( H \) is independent of \( E = MS_{g}M^{\top} \), which is \( E \sim W_{c}(nr - q - m, M \Omega M^{\top}) \). Tests of \( H_{0} : M \Theta C^{\top} = 0 \) can be carried out based on the characteristic roots of \( E(H + E)^{-1} \). More specifically,
\[
\hat{\lambda} = |E|/|H + E| = \prod_{i=1}^{c}(1 + \lambda_{i})^{-1}
\]
has a Wilks' lambda distribution with parameters \( c, nr - q - m \) and \( g \) (see [20]), where \( \lambda_{i} (i = 1, \ldots, c) \) are the eigenvalues of \( HE^{-1} \).

5.2. Hypothesis about the covariance matrix

We wish to test that a covariance matrix is equal to a specified positive definite matrix \( \Omega_{0} \). The null hypothesis is
\[
H_{0} : \Omega = \Omega_{0}.
\]
This test has considered in [15] under the multivariate normal assumption. By Theorem 5, \( \frac{S_{g} - \Omega_{0}}{n - q - m} \) is an unbiased estimator of \( \Omega \). Then \( \hat{\Sigma} = \frac{1}{n-r-q-m}(S_{g} \otimes I_{r}) \) is an unbiased estimator of \( \Sigma = \Omega \otimes I_{r} \). Moreover, the \( y_{i} \), in expression (15) is distributed according to a multivariate normal with mean \((Z_{i}^{\top} \otimes I_{p})y + (W^{\top} \otimes I_{p})\lambda\) and covariance matrix \( S_{g} \otimes I_{r} \). The null hypothesis \( H_{0} : \Omega = \Omega_{0} \) is equivalent to \( H_{0} : \Sigma = \Sigma_{0} \) with \( \Sigma_{0} = \Omega_{0} \otimes I_{r} \). The test statistic to test the null hypothesis is
\[
W = -2 \log \lambda = (nr - q - m) \left[ \log |\Sigma_{0}| - \log |\hat{\Sigma}| + \text{tr} \left( \hat{\Sigma} \Sigma_{0}^{-1} \right) - pr \right]
\]
where \( \lambda \) is the standard likelihood ratio criterion. [14] developed approximations to \( W \) using a \( \chi^{2} \) approximation. Multiplying \( W \) by \( 1 - \rho \) where \( \rho = (2p^{2}r^{2} + 3pr - 1)/(6(nr - q - m)(pr + 1)) \) the quantity
\[
(1 - \rho)W \overset{d}{\rightarrow} \chi^{2}(pr(pr + 1)/2).
\]

6. Measures of goodness of fit

In multivariate regression analysis, the simple squared multiple correlation coefficient \( R^{2} \) cannot be computed because \( p > 1 \). The index is generalized to a matrix of correlation indexes of dimension \( p \times p \) in order to measure the multivariate correlation between responses and covariates (see for major details, [20,13] and references therein). In our matrix-variate perspective two matrices related to the squared multiple correlation coefficients exist, of dimension \( p \times p \) and \( r \times r \) respectively. They measure the proportion of variance explained by the model with respect to the two sources of variability. The first one is computed along the first mode of each observed matrix, \( Y_{i} \), i.e. the set of dependent variables; and the second one is computed along the second mode which contains the \( r \) occasions. This means that the first indicator, say \( R^{2}_{(p)} \), measures how well the linear predictors explain the variability of the \( p \) characteristics in the whole observed interval. On the contrary, the second squared correlation index, \( R^{2}_{(r)} \), indicates if the time-relationships of all responses are captured by the design matrix and the time-dependent covariates. More precisely, let \( Y \) be the \( p \times nr \) matrix defined in (5) and define \( Y' \) as the matrix of dimension \( np \times r \) in which the observations are stacked along the first mode, that is \( Y'^{\top} = (Y_{1}', \ldots, Y_{n}'^{\top})^{\top} \). For convenience we assume that the \( p \) rows of \( Y \) and the \( r \) columns of \( Y' \) have zero mean. Analogously let \( \hat{Y} \) and \( \hat{Y}' \) be the fitted matrices which lie on the regression hyperplane. Then, given \( YY^{\top} \) the matrix of dimension \( p \times p \) measuring the variability of the \( p \) responses and the residual sum of squares matrix \( S_{g} = (Y - \hat{Y})(Y - \hat{Y})^{\top} \) the product \( D_{(p)} = \left( YY^{\top} \right)^{-1}(Y - \hat{Y})(Y - \hat{Y})^{\top} \) is a first generalization of \( 1 - R^{2} \) in the univariate case. \( D_{(p)} \) is a \( p \times p \) matrix that varies between the identity matrix (when no part of the variation of \( Y \) is explained by the model) and the zero matrix (when all the variation of \( Y \) is explained by the model). [13] proposed the trace-based coefficient or the determinant-based coefficient as measure of multivariate correlation. Here, we define \( R^{2}_{(p)} \) as the trace based transform of \( I_{p} - D_{(p)} \):
\[
R^{2}_{(p)} = \frac{\text{tr}(I_{p} - D_{(p)})}{p}.
\]
In a similar manner, \( Y'^{\top}(I_{r} \otimes \hat{\Sigma})^{-1}Y' \) is the matrix of dimension \( r \times r \) measuring the variability of the \( r \) occasions and \( (Y' - \hat{Y}')^{\top}(I_{r} \otimes \hat{\Sigma})^{-1}(Y' - \hat{Y}') \) is the corresponding residual sum of squares matrix. Then \( D_{(r)} = \left( Y'^{\top}(I_{r} \otimes \hat{\Sigma})^{-1}Y' \right)^{-1}(Y' - \hat{Y}')^{\top}(I_{r} \otimes \hat{\Sigma})^{-1}(Y' - \hat{Y}') \) varies between the identity matrix and the zero matrix. A possible squared correlation index is
\[
R^{2}_{(r)} = \frac{\text{tr}(I_{r} - D_{(r)})}{r}.
\]

7. Real example

In this section we analyze data about quality of papermaking tested by the Finnish Pulp and Paper Research Institute. The dataset consists of \( n = 48 \) batches of pine sulfate pulp. An essential operation in the papermaking process is beating the pulp fibers before paper manufacturing. The length and kind of beating affect almost all the physical properties of the paper. Here we study \( p = 4 \) handsheets quality variables as a function of \( r = 5 \) beating times 5, 15, 30, 45 and 60 min. The
Table 2
Estimated regression coefficients of the matrix-variate regression model with the single covariate of electric conductivity. In the first column $\lambda_i$ with $i = 0, 1, 2$ are referred to the intercept, linear and log-linear terms of the curve. The coefficient $\gamma$ is referred to the model covariate. In brackets the $p$-values associated to the null hypothesis of each single coefficient are reported.

<table>
<thead>
<tr>
<th></th>
<th>Tensile index</th>
<th>Burst index</th>
<th>Tear index</th>
<th>Drainability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0$</td>
<td>3.1270 (0.000)</td>
<td>0.2503 (0.000)</td>
<td>3.5548 (0.000)</td>
<td>2.7803 (0.000)</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$-0.0061 (0.000)$</td>
<td>$-0.0086 (0.000)$</td>
<td>$0.0044 (0.000)$</td>
<td>$0.0283 (0.000)$</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.4669 (0.000)</td>
<td>0.5668 (0.000)</td>
<td>$-0.3817 (0.000)$</td>
<td>$-0.0897 (0.000)$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$-0.0214 (0.001)$</td>
<td>$-0.0033 (0.993)$</td>
<td>$-0.0458 (0.000)$</td>
<td>$-0.0013 (0.999)$</td>
</tr>
</tbody>
</table>

Fig. 1. Paper quality variables versus beating time.

measured variables are the tensile index (Ng/g), the burst index (kPa m²/g), the tear index (nN m²/g) and the drainability of pulp (**SR number**). Because the quality of the manufacturing process can be measured by the range of all the physical characteristics of the paper, it is important to consider all the quality variables simultaneously. The final quality of the paper may be also affected by some characteristics of the pulp, like its dryness of pH, which can be included in the model as covariates. Here we consider $q = 4$ characteristics which are the pulp viscosity (dm³/kg), the brightness (ISO%), the values of electric conductivity (mS/m) and pH. Fig. 1 shows the plots of the logarithm of the four paper quality variables against beating times.

The data have been previously analyzed in [26] with a multivariate growth curve model with $q = 1$ covariate (the electric conductivity). From the previous analysis the non-linear function $y = \lambda_0 + \lambda_1 t + \lambda_2 \log(t)$ appeared to fit well the temporal trend in the data.

We now describe results based on the matrix-variate regression model where $W$ has been chosen according to the previous non-linear function, so that $m = 3$. For illustrative purposes, we have first fitted the matrix-variate regression model (17) with the single covariate of electric conductivity ($q = 1$). Table 2 reports the estimated coefficients of the matrix-variate regression model the single covariate of electric conductivity. On the basis of methodology presented in Section 5 we have checked the association of the time components and of the covariates on all the observed variables. By choosing $M = I_p$ and $C = [1, 0, 0, 0], C = [0, 1, 0, 0], C = [0, 0, 1, 0], C = [0, 0, 0, 1]$ the four null hypotheses

$$H_0 : \theta_j = 0, \quad \text{with} \quad j = 1, \ldots, (q + m)$$

were checked (where $\theta_j$ are the $p \times 1$ columns of $\Theta$). The attained value of the test statistics indicated that all the hypotheses were rejected. In order to test single null hypothesis on each time component and on the covariate versus each observed response, we have set $M = [1, 0, 0, 0], M = [0, 1, 0, 0], M = [0, 0, 1, 0], M = [0, 0, 0, 1]$ and $C$ as above. The obtained $p$-values are reported in brackets in Table 2. They indicate that the covariate does not affect the burst index and the level of
Table 3
Estimated regression coefficients of the matrix-variate regression model with all covariates. In the first column \( \lambda_i \) with \( i = 0, 1, 2 \) are referred to the intercept, linear and log-linear terms of the curve. The coefficient \( \gamma_i \) with \( i = 1, 2, 3, 4 \) are referred to the model covariates. In brackets the \( p \)-values associated to the null hypothesis of each single coefficient are reported.

<table>
<thead>
<tr>
<th></th>
<th>Tensile index</th>
<th>Burst index</th>
<th>Tear index</th>
<th>Drainability</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_0 )</td>
<td>4.1620 (0.000)</td>
<td>3.6922 (0.000)</td>
<td>8.1458 (0.000)</td>
<td>2.1086 (0.000)</td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>-0.0061 (0.000)</td>
<td>-0.0086 (0.000)</td>
<td>0.0044 (0.000)</td>
<td>0.0283 (0.000)</td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>0.4669 (0.000)</td>
<td>0.5668 (0.000)</td>
<td>-0.3817 (0.000)</td>
<td>-0.0897 (0.000)</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>0.0004 (0.000)</td>
<td>0.0004 (0.001)</td>
<td>-0.0000 (1.000)</td>
<td>0.0008 (0.000)</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-0.0169 (0.000)</td>
<td>-0.0470 (0.000)</td>
<td>-0.0529 (0.000)</td>
<td>-0.0022 (0.971)</td>
</tr>
<tr>
<td>( \gamma_3 )</td>
<td>-0.0074 (0.425)</td>
<td>0.0249 (0.030)</td>
<td>-0.0342 (0.000)</td>
<td>0.0019 (0.266)</td>
</tr>
<tr>
<td>( \gamma_4 )</td>
<td>0.0374 (0.000)</td>
<td>0.0937 (0.000)</td>
<td>0.0443 (0.000)</td>
<td>0.0429 (0.078)</td>
</tr>
</tbody>
</table>

Fig. 2. Model residuals versus beating time.

drainability although it is significant in the whole matrix-variate regression model. The goodness of fit of the model can be measured separately for responses on the whole period of time and for the time curve of all responses taken together. The determinant-based coefficients are \( R^2_p = 0.549 \) and \( R^2_r = 0.551 \), respectively.

By including all the covariates in the matrix-variate regression model we have obtained the estimated regression coefficients reported in Table 3 (with \( p \)-values in brackets). Here \( \gamma_1, \ldots, \gamma_4 \) are referred to viscosity, brightness, electric conductivity and pH, respectively. With reference to all responses, the terms of the time non-linear function and the covariates resulted to be significant by applying joint test hypotheses.

In Fig. 2 the residuals of the matrix-variate model have been fitted against the beating times. The residuals are clearly independent of time thus indicating the matrix-variate model has captured the nonlinear trend of the original variables depicted in Fig. 1.

A further proof of that is offered by the goodness of fit coefficients developed in Section 6. With the additional covariates the determinant-based coefficients are now \( R^2_p = 0.698 \) and \( R^2_r = 0.551 \). The diagonal elements of the matrix \( (I_p - D(p)) \) give an idea of well variability of each response is described by the model. They are 0.729 for the tensile index, 0.371 for the burst index, 0.751 for the tear index and 0.943 for drainability.

The estimated \( \Omega \) matrices in the matrix-variate models with one and four covariates respectively are

\[
\Omega^{(1)} = 10^{-3} \begin{bmatrix}
1.23 & 1.85 & 0.78 & 0.59 \\
1.85 & 5.07 & 2.88 & 0.92 \\
0.78 & 2.88 & 4.22 & 0.80 \\
0.59 & 0.92 & 0.80 & 4.36
\end{bmatrix} \quad \Omega^{(4)} = 10^{-3} \begin{bmatrix}
0.75 & 0.79 & -0.03 & 0.08 \\
0.79 & 2.59 & 0.73 & 0.08 \\
-0.03 & 0.73 & 2.02 & 0.52 \\
0.08 & 0.08 & 0.52 & 3.54
\end{bmatrix}
\]
with associated correlation matrices

\[ \rho(1) = \begin{bmatrix} 1.00 & 0.74 & 0.34 & 0.25 \\ 0.74 & 1.00 & 0.62 & 0.20 \\ 0.34 & 0.62 & 1.00 & 0.19 \\ 0.25 & 0.20 & 0.19 & 1.00 \end{bmatrix} \quad \rho(4) = \begin{bmatrix} 1.00 & 0.57 & -0.02 & 0.05 \\ 0.57 & 1.00 & 0.32 & 0.03 \\ -0.02 & 0.32 & 1.00 & 0.19 \\ 0.05 & 0.03 & 0.19 & 1.00 \end{bmatrix}. \]

The difference between the values of the correlations between the two models is not surprising since the two matrices contain the correlations among the model residuals. It indicates that the additional covariates describe an additional part of the correlations between the responses.

8. Conclusions

The key idea of this work was to represent the three-way data analysis problem in a matrix-variate perspective, by assuming the data are realizations from random matrices instead of the conventional random univariate or multivariate variables. On the probability theory side, the objective has been achieved by resorting to the so-called matrix-variate normal distribution. Switching the focus from a multivariate perspective to a matrix-variate one has inspired a new class of regression models for investigating the temporal/spatial trend in three-way data or the linear relation with some predictors. We have referred to them as matrix-variate regression models. Matrix-variate regression models can be viewed as a generalization of multivariate and multiple regression analysis for data coming from the simultaneous observation of several variables (or locations) at different times (or situations).

The model presentation has followed three steps. First, a simple model with time-independent covariates and with unconstrained covariance matrices \( \Phi \) and \( \Omega \) has been introduced. The model is quite easily estimable via maximum likelihood and asymptotic results for testing the model parameters have been derived. However when \( n \) is not large, like in the example presented in Section 7, we need some finite-sample tools for performing hypothesis testing on the model parameters. This has suggested to consider the restriction \( \Phi = I_p \) and to introduce a design matrix to model the time behavior of the responses in the second step of our presentation. Then, in Section 4, the model has been extended to include time-dependent and independent covariates. It is important to note that an alternative way to afford the same problem could have been to impose \( \Omega = I_p \) and leaving \( \Phi \) unconstrained. This could be useful when, in a certain real application, the time behavior of responses does not seem to follow a particular linear or non-linear function and there are significant covariates that make \( \Omega = I_p \) plausible. This alternative strategy could also allow the possibility to model the error temporal matrix, \( \Phi \), with some typical structures, like the AR(1), the compound symmetry or the Toeplitz ones.

We believe that the matrix-variate regression model introduced and discussed in this paper provides an interesting new tool for regressing three-way data. The model has several advantages compared to the proposed solutions of the literature. First, it represents the natural extension of multiple and multivariate regression analysis. Second, it can model separately the variability between and within the repeated responses. It also allows to incorporate separately the effects of covariates from the time-dependent mean response terms in \( W \), thus improving and simplifying both testing hypothesis and model interpretation. The main differences with the class of mixed models or growth curve models, have been outlined in Section 4. Here, we would also highlight that the introduction of the matrix normal assumption for the error terms makes possible to capture all the component-specific correlations (by modeling random matrices of dimension \( p \times r \) instead of multivariate variables) without the need of including random components. The most recent multiway regression models accomplish the same regression task but they lack of an inferential apparatus and they involve iterative estimation problems against the closed-form solutions of the proposed matrix-variate regression.

Many future research directions can arise out of this work. For instance, one could consider the generalization to deal with not numerical responses, like binary, count or compositional data or the introduction of penalty terms to keep under control the estimation process when dimensionality increases.

Acknowledgment

We thank Tapio Nummi for providing the dataset on paper quality.

References