



Communications in Statistics - Theory and Methods

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/lsta20>

General multivariate linear models for longitudinal studies

Gwown Shieh^a

^a Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan, 30050, ROC

Published online: 27 Jun 2007.

To cite this article: Gwown Shieh (2000) General multivariate linear models for longitudinal studies, Communications in Statistics - Theory and Methods, 29:4, 735-753, DOI: [10.1080/03610920008832512](https://doi.org/10.1080/03610920008832512)

To link to this article: <http://dx.doi.org/10.1080/03610920008832512>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

GENERAL MULTIVARIATE LINEAR MODELS FOR LONGITUDINAL
STUDIES

Gwown Shieh

Department of Management Science
National Chiao Tung University
Hsinchu, Taiwan 30050, ROC

Key words and phrases: Doubly multivariate linear models; GMANOVA; Growth curve models; MANOVA; Pooled time series and cross-sectional data; Repeated measures design; Seemingly unrelated regression models; Time-varying covariates.

ABSTRACT

Longitudinal studies occur frequently in many different disciplines. To fully utilize the potential value of the information contained in a longitudinal data, various multivariate linear models have been proposed. The methodology and analysis are somewhat unique in their own ways and their relationships are not well understood and presented. This article describes a general multivariate linear model for longitudinal data and attempts to provide a constructive formulation of the components in the mean response profile. The objective is to point out the extension and connections of some well-known models that have been obscured by different areas of application. More importantly, the model is expressed in a unified regression form from the subject matter considerations. Such an approach is simpler and more intuitive than other ways to modeling and parameter estimation. As a consequence the analyses of the general class of models for longitudinal data can be easily implemented with standard software.

1. INTRODUCTION

The defining characteristic of longitudinal designs or repeated measures designs is that each individual or subject is observed on several occasions. The distinction between these two types of designs is that measurements are recorded at several different time periods or under different experimental conditions. Here repeated measures designs are considered as a subset of longitudinal designs and the contents apply directly to repeated measures analysis. For a general discussion of linear models for the analysis of longitudinal studies, the reader is referred to Ware (1985).

In many problems, the primary goal is to develop a profound linear model for the mean response profile. This article attempts to provide a constructive introduction of multivariate linear models for longitudinal studies through the formulation and derivation of the mean response profile. The extension and connections of some well-known models of different disciplines are the major emphasis. According to the intrinsic feature that makes longitudinal designs so different from others, that is, the “time” and “subject” dimensions, the components of the mean response profile are characterized specifically by the nature of the covariates and design matrices involved. First, the covariates are classified as fixed or time-varying. Second, it is often makes sense for longitudinal studies to impose some restrictions relating the sequence of observations of each subject at different time periods. In this case, the mean response profile is assumed to include a time-related within-subject design matrix. Overall an incidence matrix is needed for situations of unbalanced and incomplete structure.

Table 1. The classification scheme for components of the mean response profile in a longitudinal study.

covariate design matrix	fixed	fixed and time-varying
incidence matrix	MANOVA model*	seemingly unrelated regression model*
time-related within-subject design matrix	GMANOVA model*	pooled time series and cross-sectional model*

*These are special cases.

Based on this classification scheme, there are four combinations or components. It turns out that they are either by themselves well-known models or include well-known models as special cases, see Table 1. It permits new insights into the specification of the mean response profile, especially the connections with the well-known models that have been obscured by different areas of application. In the vein of modeling the mean response profile, some of the components and models in Table 1 have been compared or combined. For example, Chinchilli and Elswick (1985), Hand and Crowder (1996), Park and Woolson (1992), Patel (1986), Stanek and Koch (1985), Timm and Mieczkowski (1997) and Verbyla (1988), among others. However none of them has made such distinction.

2. SPECIFICATION OF THE MODEL

Consider the following situation. Let y_{ij} , $j = 1, \dots, T_i$, be the sequence of observed measurements on the i^{th} of N subjects, and t_{ij} , $j = 1, \dots, T_i$, be the corresponding time periods at which the measurements are taken on the i^{th} subject.

Associated with each y_{ij} are the values, x_{ijk} , $k = 1, \dots, p$, of p explanatory variables or covariates. Let $y_i = (y_{i1}, \dots, y_{iT_i})'$ be the $T_i \times 1$ vector of observed sequence of measurements on the i^{th} subject. We can model this vector by $y_i = X_i\beta + \varepsilon_i$. Here β is a $p \times 1$ vector of unknown parameters and common for all subjects. It includes the parameters associated with fixed and time-varying covariates describing both differences between and within subjects. X_i is a $T_i \times p$ design matrix of covariates for the i^{th} subject with j^{th} row $(x_{ij1}, \dots, x_{ijp})$. ε_i is a vector of random error with covariance matrix Σ_i . It is taken to be an appropriate $T_i \times T_i$ submatrix of a $T \times T$ covariance matrix Σ for all possible T time periods. By this representation, the mean response profile $X_i\beta$ of y_i can be considered separately from the model for the covariance matrix $\Sigma_i = \text{Var}(\varepsilon_i)$. We will focus on the mean response profile in this article. For the discussion of alternative models of covariance, see Diggle, Liang and Zeger (1994, Chapter 5), Hand and Crowder (1996, Chapter 6), Jennrich and Schluchter (1986), Ware (1985) and Wolfinger (1993).

Now we shall discuss the components for the mean response profile in Table 1.

2.1. Component for fixed covariates and incidence matrix

Let f_{i1} be a $r_1 \times 1$ between-subject vector which accommodates r_1 values of fixed covariates for subject i for $i = 1, \dots, N$ and let ξ_i be a $T \times r_1$ matrix of parameters with j^{th} row $(\xi_{i(j1)}, \dots, \xi_{i(jr_1)})$, the parameters corresponding to r_1 baseline covariates at time j , $j = 1, \dots, T$. For unbalanced and incomplete longitudinal data, the mean response profile is of the form

$$K_i \xi_i f_{ii} = (K_i \otimes f_{ii}') \text{vec}(\xi_i') = X_{ii} \beta_i, \quad (1)$$

where $X_{ii} = (K_i \otimes f_{ii}')$ and $\beta_i = \text{vec}(\xi_i')$, and K_i is a $T_i \times T$ incidence matrix of 0's and 1's, and vec is the matrix operator by stacking the columns of the matrix into a column vector.

Example 2.1. It is easy to see that the classical MANOVA model is a special case of (1) with $K_i = I_T$ and f_{ii} is the vector of 0's and 1's which indicates the treatment assignment. Moreover, the MANCOVA model is another special case with $K_i = I_T$ and f_{ii} includes values of categorical and continuous covariates.

2.2. Component for fixed covariates and time-related within-subject design matrix

For most longitudinal designs, the relationship within the vector y_i is of primary interest. It is often more appropriate to consider an extension of the mean response profile in (1) as

$$K_i P \xi_{2i} f_{2i} = (P_i \otimes f_{2i}') \text{vec}(\xi_{2i}') = X_{2i} \beta_{2i}, \quad (2)$$

where $X_{2i} = (P_i \otimes f_{2i}')$, $\beta_{2i} = \text{vec}(\xi_{2i}')$, P is the $T \times m_2$ within-subject design matrix which describes the change pattern for the model over T time periods through $P \xi_{2i}$. The most obvious candidates as attributes for a subject's response profile are the linear, quadratic, cubic, etc., terms of the polynomial regression on time. In this case, the j^{th} row of P is $(t_j^0, t_j^1, \dots, t_j^{m_2})$, $m_2 \leq T$. ξ_{2i} is the $m_2 \times r_2$ parameter matrix, f_{2i} is the $r_2 \times 1$ between-subject design vector composed of r_2 values of fixed covariates, $P_i = K_i P$ has dimensions $T_i \times m_2$ and is the time-related within-subject design matrix for i^{th} subject, $i = 1, \dots, N$.

Example 2.2. For the completely balanced case, $K_i = I_T$ and $P_i = P$ for all i , it leads to the general MANOVA model or growth curve model (Potthoff and Roy, 1964) where $(y_1, \dots, y_N) = P\xi_2 F_2 + (\epsilon_1, \dots, \epsilon_N)$ with $F_2 = (f_{21}, \dots, f_{2N})$.

2.3. Component for both types of covariates and incidence matrix

Let $(i_1, \dots, i_{T_i})$ be the subset of $(1, \dots, T)$ corresponding to the time periods that y_i being observed. Let $C_{1(i_j')}$ be a $c_{1ij'} \times 1$ vector of values of $c_{1ij'}$ time-varying covariates at time period $i_j' = i_1, \dots, i_{T_i}$ and be a $c_{1ij} \times 1$ vector of 0's otherwise for subject i . It is not required here that the same group of covariates is recorded at each time period. Define f_{3i} as the $r_3 \times 1$ between-subject design vector composed of r_3 values of fixed covariates, and ξ_{3j} is the $1 \times (c_{1j}r_3)$ parameter vector associated with both time-varying and fixed covariates at time j . Hence one gets

$$K_i \text{diag}(\xi_{31}, \dots, \xi_{3T}) M_{1i} = (C_{1i}' \otimes f_{3i}') K_i^* \xi_3^* = X_{3i} \beta_3, \quad (3)$$

$$\text{where } X_{3i} = (C_{1i}' \otimes f_{3i}') K_i^*, \beta_3 = \xi_3^*, \text{diag}(\xi_{31}, \dots, \xi_{3T}) = \begin{bmatrix} \xi_{31} & & 0 \\ & \xi_{32} & \\ 0 & & \xi_{3T} \end{bmatrix},$$

$$M_{1i} = \begin{bmatrix} C_{1(i_1)'} \\ C_{1(i_2)'} \\ \vdots \\ C_{1(i_{T_i})'} \end{bmatrix} \otimes f_{3i}, C_{1i} = \begin{bmatrix} C_{1(i_1)} & & 0 \\ & C_{1(i_2)} & \\ & & \ddots \\ 0 & & & C_{1(i_{T_i})} \end{bmatrix} \text{ is } \sum_{j'=1}^{T_i} c_{1ij'} \times T_i, \xi_3^* = \begin{bmatrix} \xi_{31}' \\ \xi_{32}' \\ \vdots \\ \xi_{3T}' \end{bmatrix} \text{ is}$$

$r_3 \sum_{j=1}^T c_{1j} \times 1$, and K_i^* is a $\sum_{j'=1}^{T_i} c_{1ij'} \times r_3 \sum_{j=1}^T c_{1j}$ matrix with $T_i \times T$ blocks and the (j', j)

block = (j', j) element of K_i times $I_{r_3} c_{1j}$ for $j' = 1, \dots, T_i$ and $j = 1, \dots, T$.

Example 2.3. If $c_{1j} = 1$ for $j = 1, \dots, T$ and $C_{1(i_j')} = 1$ for $j' = 1, \dots, T_i$, then (3)

reduces to (1) which includes only fixed covariates. If $f_{3i} = 1$ for all i , then (3) reduces to a model with only time-varying covariates and the parameter ξ_{3j} is common for all subjects, $j = 1, \dots, T$. Furthermore, if each subject has its own set of covariates recorded across all possible periods, then (3) has a close relationship with the seemingly unrelated regression model which will be discussed in Section 3.2.

Although y_i 's may be unbalanced or incomplete, a typical case is that they are designed to have the same set of covariates evaluated for all subjects across all possible time periods. Hence a structural pattern may exist for the parameters of time-varying covariates. This is the motivation of next component.

2.4. Component for both types of covariates and time-related within-subject design matrix

We show that (2) is an extension of (1) by imposing some restriction on the parameters associated with the same covariates across different time periods. We now apply the same idea to the set of covariates being recorded for all possible time periods in (3). Let $C_{2(i_j)}$ be a $c_2 \times 1$ vector of values of c_2 time-varying covariates at time period $i_j' = i_1, \dots, i_{T_1}$ and be a $c_2 \times 1$ vector of 0's otherwise for subject i . Let f_{4i} be the $r_4 \times 1$ between-subject design vector composed of r_4 values of fixed covariates. Define ξ_4 as the $m_4 \times (c_2 r_4)$ parameter matrix associated with both time-varying and fixed covariates. Let Q be the $T \times m_4$ within-subject design matrix with j^{th} row $Q_j = (t_j^0, t_j^1, \dots, t_j^{m_4})$, $m_4 \leq T$, $j = 1, \dots, T$. Now we have the most general component

$$K_i \text{diag}(Q_1 \xi_4, \dots, Q_T \xi_4) M_{2i} = (C_{2i}' \otimes f_{4i}') Q_i^* \xi_4^* = X_{4i} \beta_4, \quad (4)$$

where $X_{4i} = (C_{2i}' \otimes f_{4i}') Q_i^*$, $\beta_4 = \xi_{4i}^*$, $\text{diag}(Q_1 \xi_{41}, \dots, Q_T \xi_{4T}) = \begin{bmatrix} Q_1 \xi_{41} & & 0 \\ & Q_2 \xi_{42} & \\ 0 & & Q_T \xi_{4T} \end{bmatrix}$.

$$M_{2i} = \begin{bmatrix} C_{2(i_1)} \\ C_{2(i_2)} \\ \vdots \\ C_{2(i_T)} \end{bmatrix} \otimes f_{4i}, \quad C_{2i} = \begin{bmatrix} C_{2(i_1)} & & 0 \\ & C_{2(i_2)} & \\ & & \ddots \\ 0 & & & C_{2(i_T)} \end{bmatrix} \text{ is } c_2 T_i \times T_i, \quad \xi_{4i}^* = \text{vec}(\xi_{4i}'), \quad Q_i^* =$$

$(K_i Q) \otimes I_{c_2 r_4}$ is $T_i c_2 r_4 \times m_4 c_2 r_4$.

Example 2.4. For $Q = I_T$ and $C_{2(i_j')} = 1$ for $j' = 1, \dots, T_i$, then (4) reduces to (3) with $c_{ij} = c_2$ for all j and (2), respectively.

It follows from preceding discussion that a general multivariate linear model for longitudinal studies may include all (or some) of the components (1)-(4). One can write

$$y_i = X_i \beta + \varepsilon_i, \quad (5)$$

where $X_i = (X_{1i}, X_{2i}, X_{3i}, X_{4i})$ and $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$ with $\text{Var}(\varepsilon_i) = K_i \Sigma K_i'$ for $i = 1, \dots, N$.

If one is interested in completely balanced situation where $K_i = I_T$, then a useful expression is

$$\begin{aligned} y &= (y_1, \dots, y_N) \\ &= \xi_1 F_1 + P \xi_2 F_2 + \text{diag}(\xi_{31}, \dots, \xi_{3T}) M_1 + \text{diag}(Q_1 \xi_{41}, \dots, Q_T \xi_{4T}) M_2 \\ &\quad + (\varepsilon_1, \dots, \varepsilon_N), \end{aligned} \quad (6)$$

where $F_1 = (F_{11}, \dots, F_{1N})$, $M_1 = (M_{11}, \dots, M_{1N})$ and $M_2 = (M_{21}, \dots, M_{2N})$. In case only some of the four components are necessary, one can modify the model and it leads to a simpler or reduced model of (5) or (6).

3. APPLICATIONS

In this section we show that several well-known and widely applied models are special cases of (5).

3.1 A mixture of MANOVA and GMANOVA models

We already show that the mean response profiles of classical MANOVA and GMANOVA models are special cases of components (1) and (2) in Examples 2.1 and 2.2, respectively. Chinchilli and Elswick (1985) proposed a mixture of the two models and it might arise in a one-way MANOVA design in which it is convenient to express treatment effects as a polynomial growth curve over time, whereas it might not be desirable to express covariate effects in the same manner. This mixed model essentially equivalent to (6) with only the first two components of mean response profile retained and consequently is a special case of (5).

3.2 Seemingly unrelated regression models

The seemingly unrelated regression (SUR) models proposed in Zellner (1962) is defined as

$$y_i^*_{T \times 1} = X_i^*_{T \times t_i} \beta_i^*_{t_i \times 1} + \varepsilon_i^*_{T \times 1} \text{ for } i = 1, \dots, N. \quad (7)$$

It describes possibly different regression equations for N correlated vectors of dependent variables y_i^* with respect to corresponding specification matrices X_i^*

and unknown parameters β_i^* . In many applications, y_i^* and X_i^* contain respective observations and values of covariates for T different time periods. The subscript i will correspond to a particular economic or geographical unit such as a household, a firm or a state. It seems that (7) agrees with (5). However the common application of SUR models rely on the important assumption of contemporaneous correlation between disturbances from different equations at a given time period, that is, $E(\epsilon_i^*) = 0$, $\text{Cov}(\epsilon_i^*, \epsilon_{i'}^*) = \sigma_{ii'} I_T$ and $\sigma_{ii'}$ is the (i, i') element of Σ ($N \times N$) for i and $i' = 1, \dots, N$. When contemporaneous correlation exists, it may be more efficient to estimate all equations jointly, rather than to estimate each one separately. In other words, a gain in efficiency can be achieved most of the time by combing a number of equations that, at first glance, seem unrelated. But in the meanwhile the contemporaneous correlation may be responsible for the somehow limited connection and recognition in application outside economic and social science. Afterward Stanek and Koch (1985) tried to recognize the connection between the SUR model and growth curve model by showing the equivalence of parameter estimates for the two models with a proper choice (more exactly, a clever transformation) of SUR model. Here we provide a simpler and more intuitive way of viewing a SUR model. First define y^* , β^* and ϵ^* as the column concatenation of y_i^* , β_i^* and ϵ_i^* , respectively. So that $\text{Var}(\epsilon^*) = \Sigma \otimes I_T$. The SUR model can be written as $y^* = \text{diag}(X_1^*, \dots, X_N^*) \beta^* + \epsilon^*$, where $\text{diag}(X_1^*, \dots, X_N^*)$ is block-diagonal matrix with diagonal elements $X_1^*, \dots, X_N^*$. For our purpose of relating a SUR model to a more natural setting with covariance matrix $I_T \otimes \Sigma$, we

rearrange y^* and ε^* into $y^{**} = (y_{11}^*, y_{21}^*, \dots, y_{N1}^*, \dots, y_{1T}^*, y_{2T}^*, \dots, y_{NT}^*)$ and $\varepsilon^{**} = (\varepsilon_{11}^*, \varepsilon_{21}^*, \dots, \varepsilon_{N1}^*, \dots, \varepsilon_{1T}^*, \varepsilon_{2T}^*, \dots, \varepsilon_{NT}^*)$ where y_{ij}^* and ε_{ij}^* are j^{th} element of y_i^* and ε_i^* ,

respectively. Then a new expression is $y^{**} = X^{**}\beta^* + \varepsilon^{**}$, where $X^{**} = \begin{bmatrix} X_1^{**} \\ \vdots \\ X_T^{**} \end{bmatrix}$, X_j^{**}

$= \text{diag}(X_{1j}^*, \dots, X_{Nj}^*)$, X_{ij}^* is the j^{th} row of X_i^* , and $\text{Var}(\varepsilon^{**}) = I_T \otimes \Sigma$. For each section of y^{**} and ε^{**} with length N ,

$$y_j^{**} = (y_{1j}^*, \dots, y_{Nj}^*)' = X_j^{**} \beta^* + \varepsilon_j^{**}, \quad (8)$$

where $\text{Var}(\varepsilon_j^{**}) = \Sigma$ and ε_j^{**} are independent for all $j = 1, \dots, T$. Now it is easy to see that the $X_j^{**} \beta^*$ in (8) is agree with the component (3) by setting $K_i = I_N$, $f_{3i} = 1$, $C_{ii}' = X_j^{**}$ and the subscript i now from 1 to T . Besides the possible baseline differences, a typical SUR model will involve only time-varying covariates and each subject has its own set of (t_i) covariates recorded across all time periods. Hence for every SUR model with N units and T time periods, there exists a corresponding multivariate linear model with time-varying covariates in the context of longitudinal designs with T subjects and N time periods. We believe such connection is more appropriate and congruent with the attribute of the time-varying covariates involved in a SUR model. As mentioned above, the connection in Stanek and Koch (1985) is through some proper transformation of SUR model with the restriction $t_j = t$ for all $j = 1, \dots, T$. Furthermore, a growth curve model usually does not involve any time-varying covariates other than age or time elapsed as in Example 2.2 of Section 2.2. We now conclude that SUR models are

in the class of models (5). It will be shown in later sections that such phenomenon remains to exist in other extensions of SUR models. For example, SUR models with baseline differences will be discussed in Section 3.4.

3.3 A multivariate model for repeated measures designs with time-varying covariates

Patel (1986) introduced a multivariate linear model that accommodates both fixed and time-varying covariates for the analysis of complete repeated measures designs. The model is $y = \xi A + \sum_{q=1}^c \Gamma_q X_q + \varepsilon$, where y is a $T \times N$ matrix of observations, ξ is a $T \times r$ matrix of parameters, A is a $r \times N$ design matrix of fixed covariates, Γ_q is a $T \times T$ diagonal parameter matrix with diagonal elements $\Gamma_{q1}, \dots, \Gamma_{qT}$ and X_q is a $T \times N$ matrix of values of time-varying covariates. The model assumes that the columns of ε are independently distributed with a common multivariate normal distribution with mean 0 and a $T \times T$ dispersion matrix Σ . The first term ξA is exactly the same as the first component $\xi_1 F_1$ in (6) with $\xi_1 = \xi$ and $F_1 = A$. The second term $\sum_{q=1}^c \Gamma_q X_q$ can be rewritten in the form of the third component of (6) with $\xi_{3j} = (\Gamma_{1j}, \dots, \Gamma_{cj})$, $C_{1(j)} = (X_{1(j)}, \dots, X_{c(j)})$, $X_{q(j)}$ is the (j, i) element of X_q , and $f_{3i} = 1$ for $i = 1, \dots, N$ and $j = 1, \dots, T$. Without the general expression of (5), Patel (1986) proposed an iterative scheme for the estimation of parameters. The estimation procedure is somehow cumbersome and it motivates Verbyla (1988) to rewriting it as a SUR model and estimation can be carried out simply. Here we show it belongs to the class of models (5). The estimation procedure is standard and well known (Jennrich and Schluchter, 1986).

Although the form of the model is algebraically equivalent to a SUR model, it is more natural to keep the between-subject design matrix of fixed covariates apart from the classical SUR model where all covariates are usually time-varying.

3.4 An extension of the SUR model for unbalanced and incomplete longitudinal data with fixed covariates

Park and Woolson (1992) attempted to view the SUR model in the context of longitudinal design and extended it to a general class of multivariate models to handle unbalanced and incomplete longitudinal data. According to their definition of the model, the expression of the dimensions for subject and time period contradicts the common contemporaneous correlation structure in a SUR model. The reason is obviously due to the equivalence of a SUR model (with N subjects and T time periods) and a multivariate linear model in the context of longitudinal designs (with T subjects and N time periods) as discussed in Section 3.2. Following the same idea in Section 3.2, this phenomenon can be shown to hold for the unbalanced and incomplete case with varying T_i in (7). Therefore an unbalanced and incomplete SUR model with N subjects and varying $(T_i, i = 1, \dots, N)$ time periods is equivalent to a multivariate linear model in the context of longitudinal designs with $T (\geq T_i, i = 1, \dots, N)$ subjects and varying $(N_j \leq N, j = 1, \dots, T)$ time periods.

Besides the component of the fixed covariates for each subject is defined to be identical for all time periods in Park and Woolson (1992). Hence it only includes intercept terms and can not reflect the possible systematic change pattern associated with the fixed covariates across all time periods. This can be found in their expressions (in Sections 4.2 and 5) where the time-alike covariate (age) is

treated as time-varying covariate. A careful examination will reveal that the fixed covariates are also included in the term and it leads to an ambiguity of the model.

For the purpose of extending a SUR model for unbalanced and incomplete longitudinal data with fixed covariates, we can construct a model with component (3) combined with (1), (2) or both whenever necessary.

3.5 Models for pooled time series and cross-sectional data

Estimation of relationships that combine time series and cross-sectional data is a problem frequently encountered in economics. The typical problem is to specify a model that will adequately allow for differences in behavior over cross-sectional units as well as any differences in behavior over time for a given cross-sectional unit. Judge, et al. (1985, Chapter 13) and Johnston (1991, Section 10.3) consider a variety of models for pooled time series and cross-sectional data. In general the models can be written as $y_{ij} = \beta_{1ij} + \sum_{k=1}^K x_{kij} \beta_{kij} + \varepsilon_{ij}$, where y_{ij} is the value of dependent variable, x_{kij} is the value of k^{th} explanatory variable and ε_{ij} is the error term for cross-sectional unit i and time period j . The parameters β_{kij} , as the subscripts suggest, for the most general case they can be different for different subjects and in different time periods. However, in most cases more restrictive assumptions will be made. For example, they may be common for all i and j or may be varying only over i or only over j . We will restrict our attention on the most general case since the others can be treated as special cases and the model is

$$y_{ij} = (\beta_1 + \mu_{1i} + \lambda_{1j}) + \sum_{k=2}^K (\beta_k + \mu_{ki} + \lambda_{kj}) x_{kij} + \varepsilon_{ij}, \quad (9)$$

$i = 1, \dots, N$ refers to a cross-sectional unit and $j = 1, \dots, T$ refers to a given time

period. As previous models, (9) can be organized by cross-sectional unit as $y_i = (y_{i1}, \dots, y_{iT})'$ and it can be written in the form of (6) with the first and fourth components retained. The corresponding $\xi_i = \beta_i J_{T \times N} + 1_{T \times 1} \otimes (\mu_{i1}, \dots, \mu_{iN}) + (1_{N \times 1} \otimes (\lambda_{i1}, \dots, \lambda_{iT}))'$, $J_{T \times N}$ is a $T \times N$ matrix of all 1's, $1_{T \times 1}$ is a $T \times 1$ vector of all 1's, $F_1 = I_N$, $Q = I_T$, the j^{th} row of ξ_4 is $(\beta_2, \dots, \beta_K) \otimes (1_{N \times 1})' + (\mu_{21}, \dots, \mu_{2N}, \dots, \mu_{K1}, \dots, \mu_{KN}) + (\lambda_{2j}, \dots, \lambda_{Kj}) \otimes (1_{N \times 1})'$, $C_{2(ij)} = (x_{2ij}, \dots, x_{Kij})'$, $c_2 = K - 1$ and f_{4i} is the i^{th} column of I_N .

The other models for pooled time series and cross-sectional data with intercept β_{1ij} and/or slope $(\beta_{2ij}, \dots, \beta_{Kij})$ coefficients varying only over i , only over j , or common for all i and j are special cases of (9) with respective parameters λ 's, μ 's, or both omitted from (9). The error term $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})'$ are assumed to be independent for $i = 1, \dots, N$ with $E(\varepsilon_i) = 0$ and $\text{Var}(\varepsilon_i) = \Sigma$. However, it is possible to assume ε_i is contemporaneously correlated disturbance as in the SUR model. Under this assumption we only need to organize the observations y_{ij} with respect to time period instead of subject.

3.6 Doubly multivariate linear models

In this section we extend the discussion to the doubly multivariate linear models which allow one to analyze vector-valued repeated measurements or longitudinal data by obtaining d -variates on N subjects at each of T time periods (Timm and Mieczkowski, 1997, Chapter 6).

The model is $Y = BF + E$, where Y is the $dT \times n$ matrix of observations with i^{th} column Y_i is the dT responses of subject i , B is a $dT \times r$ parameter matrix, F is the $r \times N$ between-subject design matrix of full rank r , and E is a $dT \times N$

matrix of random errors. The columns E_i are independent for $i = 1, \dots, N$ with mean 0 and $\text{Var}(E_i) = \Sigma^*$, a $dT \times dT$ positive definite matrix. The mean response profile BF in this model is nothing more than component (1) with T being extended to dT for there are d -variates measured at each time period. The covariance Σ^* may have a kronecker product structure $\Sigma^* = \Sigma_1^* \otimes \Sigma_2^*$, where Σ_1^* and Σ_2^* are $d \times d$ and $T \times T$ matrices, respectively (Galecki, 1994).

Next we focus on a combination of SUR model and double multivariate linear model. Liu (1993) proposed a seemingly unrelated multivariate regression (SUMR) model with number of subjects limited to 2. In general the model is of the form

$$Y_i^*_{T \times d_i} = X_i^*_{T \times t_i} B_i^*_{t_i \times d_i} + E_i^*_{T \times d_i}, i = 1, \dots, N,$$

Which is a direct extension of model (7) with d_i measurements, instead of just 1, on i th subject at each of T time periods. The rows of $E^* = (E_1^*, \dots, E_N^*)$ are independently distributed with mean 0 and a common covariance matrix. Let Y_{ij}^* be the j^{th} row of Y_i^* , $i = 1, \dots, N$. By stacking $Y_{1j}^*, \dots, Y_{Nj}^*$ together into one column. We obtain an expression similar to (8)

$$Y_j^{**} = (Y_{1j}^*, \dots, Y_{Nj}^*)' = X_j^{**} B^* + E_j^{**}, \quad (10)$$

where Y_j^{**} is a $d \times 1$ vector of $d = \sum_{i=1}^N d_i$ observations for all subjects at time period j ,

$X_j^{**} = \text{diag}(I_{d_1} \otimes X_{1j}^*, \dots, I_{d_N} \otimes X_{Nj}^*)$, a $d \times \sum_{i=1}^N d_i t_i$ diagonal matrix, X_{ij}^* is the j^{th} row of

X_i^* , $B^* = (\text{vec}(B_1^*), \dots, \text{vec}(B_N^*))'$ is a $\sum_{i=1}^N d_i t_i \times 1$ vector of parameters, $E_j^{**} = (E_{1j}^*, \dots,$

E_{Nij}^* , E_{ij}^* is the j^{th} row of E_i^* . In terms of a general multivariate linear model for longitudinal studies, Y_j^{**} can be viewed as the d responses observed at all N time periods on j^{th} subject, X_{ij}^* is the vector of time-varying covariates for j^{th} subject at i^{th} time period. The mean response profile of model (10) is still in the form of component (3). Hence a SUMR model with d_i measurements for $i = 1, \dots, N$ subjects at all T time periods can be transformed into a general multivariate linear model with time-varying covariates for d time periods and T subjects. This reduces to the regular case studied in Section 3.2 when $d_i = 1$ for $i = 1, \dots, N$.

4. CONCLUDING REMARKS

We propose a class of general multivariate linear models for longitudinal studies with the mean response profile composed of four components. The classification scheme of the components is based on the nature of covariates and design matrices involved in a longitudinal design. We show that several well-known models are special cases of the general multivariate linear model. Furthermore their estimation and connections are unified, clarified and extended through the general regression expression. It can be easily adapted to standard software for implementation. However we do not cover the covariance modeling. For selection of structural covariance matrices, see Jennrich and Schluchter (1986) and Wolfinger (1993) for further details.

Although we have confined the discussion to the fixed effects model, the general multivariate linear model can be easily extended into a mixed effects model by introducing additional term for random effects and adjusting

accordingly the covariance matrix for the random effects and errors. Finally the generalized linear models and nonlinear models for longitudinal or repeated measurement data, for example, Diggle, Liang and Zeger (1994, Chapter 7), Davidian and Giltinan (1995) and Vonesh and Chinchilli (1997), still involve the same process of modeling the mean response profile. The results presented here may shed some light on the modeling of more complicated situations. Hopefully the analysis then will be simplified to some extent and efforts can be focused on other important issues.

BIBLIOGRAPHY

- Chinchilli, V. M. and Elswick, R. K. (1985), A mixture of the MANOVA and GMANOVA models, *Communications in Statistics - Theory and Methods*, 14, 3075-3085.
- Davidian, M. and Giltinan, D. M. (1995), *Nonlinear Models for Repeated Measurement Data*, London: Chapman and Hall.
- Diggle, P. J., Liang K.-Y. and Zeger, S. L. (1994), *Analysis of Longitudinal Data*, Oxford: Oxford University Press.
- Galeki, A. T. (1994), General class of covariance structures for two or more repeated factors in longitudinal data analysis, *Communications in Statistics - Theory and Methods*, 23, 3105-3119.
- Hand, D. J. and Crowder, M. J. (1996), *Practical Longitudinal Data Analysis*, London: Chapman and Hall.
- Hecker, H. (1987), A generalization of GMANOVA-model, *Biometrical Journal*, 7, 763-770.
- Jennrich, R. I. and Schluchter, M. D. (1986), Unbalanced repeated-measures models with structured covariance matrices, *Biometrics*, 42, 805-820.
- Johnston, J. (1991), *Econometric Methods*, New York: McGraw-Hill.
- Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H. and Lee, T. C. (1985), *The Theory and Practice of Econometrics*, New York: Wiley.
- Liu, A. (1993), An efficient estimation of seemingly unrelated multivariate regression models with application to growth curves analysis. *Statistica Sinica*, 3, 421-434.

- Park, T. and Woolson, R. F. (1992), Generalized multivariate models for longitudinal data, *Communications in Statistics – Simulations*, 21, 925-946.
- Patel, H. I. (1986), Analysis of repeated measures designs with changing covariates in clinical trials, *Biometrika*, 73, 707-715.
- Potthoff, R. F. and Roy, S. N. (1964), A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, 51, 313-326.
- Stanek, E. J. and Koch, G. G. (1985), The equivalence of parameter estimates from growth curve models and seemingly unrelated regression models, *The American Statistician*, 39, 149-152.
- Timm, N. H. and Mieczkowski, T. A. (1997), *Univariate and Multivariate General Linear Models: Theory and Applications Using SAS Software*, Cary, NC: SAS Institute INC.
- Verbyla, A. P. (1988), Analysis of repeated measures designs with changing covariates, *Biometrika*, 75, 172-174.
- Verbyla, A. P. and Venables, W. N. (1988), An extension of the growth curve model, *Biometrika*, 75, 129-138.
- Vonesh, E. F. and Chinchilli, V. M. (1997), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, New York: Marcel Dekker.
- Ware, J. H. (1985), Linear models for the analysis of longitudinal studies, *The American Statistician*, 39, 95-101.
- Wolfinger, R. (1993), Covariance structure selection in general mixed models, *Communications in Statistics – Simulation*, 22, 1079-1106.
- Zellner, A. (1962), An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of American Statistical Association*, 57, 348-368.

Received March, 1999; Revised November, 1999.