***Open Journal of Statistics*， 2013， 3， 27-35**

<http://dx.doi.org/10.4236/ojs.2013.31004>Published Online February 2013 [(http://www.scirp.org/journal/ojs)](http://www.scirp.org/journal/ojs))

**Joint Variable Selection of Mean-Covariance Model for Longitudinal Data**

### Dengke Xu1， Zhongzhan Zhang1， Liucang Wu1，2

1College of Applied Sciences， Beijing University of Technology， Beijing， China 2Faculty of Science， Kunming University of Science and Technology， Kunming， China [Email: zzhang@bjut.edu.cn](mailto:zzhang@bjut.edu.cn)

Received October 23， 2012; revised November 24， 2012; accepted December 9， 2012

## 摘要

在本文中，我们重新参数化纵向数据分析中的协方差结构，通过修正Cholesky分解，即分解协方差阵自身。在这样的修正Cholesky分解的基础上， 个体内的协方差矩阵被分解为一个包括移动平均系数的下三角矩阵，以及一个含有变换方差的对角矩阵， 用协变量的线性函数建模得到。然后， 我们提出一个惩罚极大似然方法，用于在基于这个分解的联合均值-协方差模型中进行变量选择。在一定的正则条件下，我们建立模型中的惩罚极大似然参数估计的一致性以及渐进正态性。为了评估所提出的变量选择程序在有限样本中的表现，最后还进行了模拟数据研究。

**关键词:** 联合均值方差模型; 变量选择; Cholesky 分解; 纵向数据分析; 惩罚极大似然估计

## 介绍

近些年来，在广义线性模型和多元正态误差的前提下， 联合均值协方差建模的方法被Pourahmadi [1，2]试探性地提出. 此类模型的主要优点包括方便对参数估计进行统计解释和计算，这一点会在第二节进行详细描述. 另一方面，协方差矩阵的估计是纵向研究中的重要问题. 一个优质的协方差估计可以提高回归系数的有效性. 此外，协方差估计本身也是非常有用的[3]. 许多学者都对协方差估计问题进行了研究. Pourahmadi [1，2]考虑了协方差矩阵修正Cholesky 分解分量对应的广义线性模型*.*Fan[4] and Wu [5]提出了利用协方差函数的半参数模型. 最近，在没有额外的计算成本的条件下， Rothman *et al.* [6] 提出了一个新的协方差矩阵的Cholesky 因子的回归解释 ，通过参数化其自身以及保证估计协方差的正定性，基于这个分解[6]， Zhang and Leng [7]提出了联合均值协方差分析的有效极大似然估计。

众所周知，作为建模策略的其中一部分，变量选择是大多数统计分析中的一个重要课题，并且在过去的三十年中被人广泛探讨过了. 在传统线性回归的设定下，许多变量选择标准(AIC和BIC等等) 已经在实践中得到了广泛应用. 然而，这些选择方法会造成巨大的计算成本.由于在许多情况下我们需要理想的计算效率，人们开发了许多收缩方法，包括但不仅限于: 非负garrotte方法 [8]，LASSO [9]方法，岭回归方法[10]，SCAD方法 [11]，以及一步稀疏估计方法[12]. 最近，Zhang and Wang [13] 提出了一个新的标准， 名为PIC方法， to simultaneously select explanatory variables in the mean model and variance model in heteroscedastic linear models based on the model struc- ture. Zhao and Xue [14] presented a variable selection procedure by using basis function approximations and a partial group SCAD penalty for semiparametric varying coefficient partially linear models with longitudinal data.

In this paper we show that the modified Cholesky de- composition of the covariance matrix， rather than its in- verse， also has a natural regression interpretation， and therefore all Cholesky-based regularization methods can be applied to the covariance matrix itself instead of its inverse to obtain a sparse estimator with guaranteed posi- tive definiteness. Furthermore， we aim to develop an

28 D. K. XU *ET AL*.

efficient penalized likelihood based method to select important explanatory variables that make a significant

T

*i*  ** ，， ** 

*i*1 *imi*

is an *mi* 1 vector and *i* is an

contribution to the joint modelling of mean and covari-

*mi*  *mi* 

positive definite matrix *i*  1，， *n* . As a

ance structures for longitudinal data. With proper choices of the penalty functions and the tuning parameters， we

tool for regularizing the inverse covariance matrix，

Pourahmadi [1] suggested using the modified Cholesky

*i*

establish the consistency and asymptotic normality of the

factorization of  1 . To parametrize

*i* ， Pourahmadi

resulting estimator. Simulation studies are used to illus-

[1] first proposed to decompose it as

*Ti* *i Ti*  *Di* . The

trate the proposed methodologies. Compared with exist-

T

lower triangular matrix

*Ti* is unique with 1’s on its di-

ing methods， our procedure offers the following differ-

agonal and the below diagonal entries of *Ti*

are the

ences and improvements. Firstly， Zhang and Leng [7] discussed efficient maximum likelihood estimates and

negative autoregressive parameters

*j* 1

*ijk*

in the model

model selection for joint mean-covariance analysis based BIC. As is well known， BIC selection method would

*yij*  *ij*  *ijk*  *yik*  *ik*   *ij* .

*k* 1

suffer from expensive computational costs. However， our

The diagonal entries of *Di*

are the innovation vari-

method can select significant variables and obtain the

ances as ** 2  Var **  .

parameter estimators simultaneously in the joint model- According to the idea of the proposed decomposition

*ij ij*

*i i*

ling of mean and covariance structures for longitudinal

in Rothman *et al.* [6]， we let

*L*  *T* 1 ， a lower triangular

data， that implies that our method can avoid the heavy computational burden. Secondly， in this paper we assume the covariates may be of high dimension， which become increasingly common in many health studies， and our

matrix with 1’s on its diagonal， we can write

  *TDT* T . We actually use a new statistically mean- ingful representation that reparameterizes the covariance matrices by the modified Cholesky decomposition advo-

*i i i i*

method also can select the important subsets of the cova-

cated by Rothman *et al*. [6]. The entries

*lijk* in

*Li* can

tiates. Thirdly， we reparameterize covariance structures in longitudinal data analysis through the modified Cho- lesky decomposition of itself， which is brought closer to time series analysis， for which the moving average model

be interpreted as the moving average coefficients in

*j* 1

*yij*  *ij*  *lijk ik*  *ij* ， *j*  2，， *mi* ，

*k* 1

may provide an alternative， equally powerful and parsi-

monious representation.

where *i*1  *yi*1  *i*1 and *i*  *N* 0， *Di*  for

 T

*i*

The rest of this paper is organized as follows. In Sec-

*i*  *i*1 ，，*im* . Note that the parameters *lijk*

and

tion 2 we first describe a reparameterization of covari-

log ** 2  are unconstrained.

ance matrix itself through the modified Cholesky de- composition and introduce the joint mean and covariance

*ij*

Based on the modified Cholesky decomposition and

motivated by [1，2] and Ye and Pan [15]， the uncon-

models for longitudinal data. We then propose a variable

strained parameters

*ij* ， *lijk*

and

 *ij* 

log ** 2

are modeled

selection method for the joint models via penalized like- lihood function. Asymptotic properties of the resulting

in terms of the generalized linear regression models (JMVGLRM)

estimators are considered in Section 3. In Section 4 we give the computation of the penalized likelihood estima-

*g* **   *x*T ** ， *l*  *z*T ** ， log** 2   *h*T** . (1)

*ij ij ijk ijk ij ij*

tor as well as the choice of the tuning parameters. In Sec-

Here

*g* 

is a monotone and differentiable known

tion 5 we carry out simulation studies to assess the finite

link function， and

*xij* ，

*zijk* and

*hij* are the *p* × 1， *q* × 1

sample performance of the method. and *d* × 1 vectors of covariates， respectively. The covari-

ate

*xij*

and

*hij* are the usual covariates used in regres-

## Variable Selection for Joint

sion analysis， while *zijk* is usually taken as a polyno-

## Mean-Covariance Model

### Modified Cholesky Decomposition of the

mial of time difference

 T

*tik*  *tij* . In addition， denote

 T

*X*  *x* 1 ，， *x*

and

*H*  *h* ，， *h*

. We further

*i i imi*

*i i*1

*imi*

### Covariance Matrix

Suppose that there are *n* independent subjects and the *i*th

refer to ** as moving average coefficients and ** as innovation coefficients. In this paper we assume that the

subject has *mi* repeated measurements. Specifically， de-

covariates

*xij* ，

*zijk* and

*hij* may be of high dimension

note the response vector

*y*   *y*

，， *y*

T for the *i*th

and we would select the important subsets of the covari-

*i i*1

*imi*

ates *xij* ， *zijk* and *hij* ， simultaneously. We first assume

subject，

*i*  1，， *n* ， which are observed at time

all the explanatory variables of interest， and perhaps their

*t*  *t* ，， *t* T . We assume that the response vector is

1

*i i imi*

normally distributed as *yi* ~ *N*  *i* ， *i*  ， where

interactions as well， are already included into the initial models. Then， we aim to remove the unnecessary ex- planatory variables from the models.

D. K. XU *ET AL*. 29

* 1. **Penalized Maximum Likelihood for** penalized likelihood estimate. We first introduce some

### JMVGLRM

Many traditional variable selection criteria can be con-

notations. Let **0

thermore， let

denote the true values of ** . Fur-

T T T

sidered as a penalized likelihood which balances model-

ling biases and estimation variances [11]. Let ** 

**  ** ，，** T  ** 1 

0 01 0*s*  0

， ** 2   .

0 

denote the log-likelihood function. For the JMVGLRM， we propose the penalized likelihood function

 

For ease of presentation and without loss of generality， it is assumed that ** 1 consists of all nonzero compo-

*p* 0 2

*L* **   **    *p* 1  *i* 

**

nents of **0

1

and that **0

 0 . Denote the dimension of

*i* 1

*q d*

(2)

**0 by

*s*1 . Let

  *p* 2  * j*    *p* 3  *k* 

** **

*n*

*an*  max  *p*

 **0

*j*  :**0 *j*

 0

*j* 1 *k* 1

where **  ** ，，** T  ** ，， ** ;** ，，** ; ** ，， ** T

1 *s*

1 *p* 1 *q* 1 *d*

and

1 *j*  *s*

with *s*  *p*  *q*  *d* and

*p* *l*  

is a given penalty

*bn*  max *p* *n*  **0 *j*  : **0 *j*  0 .

function with the tuning



**

parameter

** *l*  *l*  1， 2， 3 . The

1 *j*  *s*

tuning parameters can be chosen by a data-driven crite-

Here we denote ** *l*  as ** to emphasize its de-

rion such as cross validation (CV)， generalized cross-

*n*

pendence on sample size *n* . * n*

is equal to either ** 1 ，

validation (GCV) [9]， or the BIC-type tuning parameter

** 2 or

** 3 ， depending on whether

**0 *j* is a component

selector [16] which is described in Section 4. Here we

of **0 ，

** 0 or **0 1  *j*  *s*  .

use the same penalty function

*p*  for all the regres-

To obtain the asymptotic properties in the paper， we

sion coefficients but with different tuning parameters require the following regularity conditions:

** 1 ，

** 2 and

** 3 for the mean parameters， moving

(C1): The covariate vectors

*xij* ， *zijk*

and

*hij* are fixed.

average parameters and log-innovation variances， respec- Also， for each subject the number of repeated measure-

tively. Note that the penalty functions and tuning pa- rameters are not necessarily the same for all the parame- ters. For example， we wish to keep some important vari-

ments，

*mi* ， is fixed

*i*  1，， *n*， *j*  1，， *mi* ， *k*  1，， *j* 1 .

ables in the final model and therefore do not want to pe- (C2): The parameter space is compact and the true

nalize their coefficients. In this paper， we use the

value

**0 is in the interior of the parameter space.

smoothly clipped absolute deviation (SCAD) penalty

(C3): The design matrices *Xi*

and *Hi*

in the joint

whose first derivative satisfies

models are all bounded， meaning that all the elements of

the matrices are bounded by a single finite real number.

 *a*  *t*  

*p*   * I* *t*  **    *I* *t*  ** 

**Theorem 1** Assume

1.  *O*

*n*1 2  ，

1.  0

and

**  *a* 1 ** 

*n p n*

 

* n*  0

as *n*  . Under the conditions (C1)-(C3)，

for some *a*  2 [11]. Following the convention in [11]， wewe setset *aa*  3.73.7 in our work. The SCAD penalty is a

with probability tending to 1 there must exist a local maximizer ˆ of the penalized likelihood function

spline function on an interval near zero and constant out-

**

*n*

*L* **  in (2) such that **ˆ is a -consistent estimator

side， so that it can shrink small value of an estimate to zero while having no impact on a large one.



*n*

*n*

The penalized maximum likelihood estimator of ** ，

of **0 .

The following theorem gives the asymptotic normality property of **ˆ . Let

*n*

** 

** 



denoted by **ˆ ， maximizes the function *L* ** 

in (2).

With appropriate penalty functions， maximizing

*n*

*L* ** 

*An*  diag 

*p*

*n*

**

1

01

，， *p* 

1

0*s*1

with respect to ** leads to certain parameter estimators

vanishing from the initial models so that the correspond-

  

1 

 1 

  1 

 1 T

ing explanatory variables are automatically removed.

1.  *p*

*n*

**

*n*

**01

sgn **01

，， *p*

*n*

**

1

**0*s*

sgn **0 *s* ，

Hence， through maximizing

1

*L* ** 

we achieve the goal

where

** 1 is the *j*th component of

** 1 1  *j*  *s*  .

of selecting important variables and obtaining the pa- 0 *j* 0 1

Denote the Fisher information matrix of ** by

*In* **  .

rameter estimators， simultaneously. In Section 4， we pro-

**

*n*

vide the technical details and an algorithm for calculating

**Theorem 2** Assume that the penalty function

*p* *t* 

the penalized maximum likelihood estimator

**ˆ .

satisfies

*p* *t* 

lim inf lim inf *n*  0

1. **Asymptotic Properties** *n t* 0 * n*

  

We next study the asymptotic properties of the resulting

and *In*  *In* **0  *n*

converges to a finite and positive

30 D. K. XU *ET AL*.

definite matrix *I* **0  as *n*  . Under the same mild

 *p*1  **01 

*p*1  **0 *p*  *p*2  ** 01 

conditions as these given in Theorem 1， if

* n*  0

and

** **0

  diag  **1

** **

，， ， 1 ，

** 01

*p*

*n n*  as



*n*  ， then the

-consistent esti-

 **0 *p*

T T T 



*n*

**01

*n*





*n*   *n*

**



**ˆ 2  

*p*   **

 *p*   ** 

*p*   **0*d*  

isfy

1)

**ˆ2  0

with probability tending to 1.

1 2

where

1 *s*

1 *p* 1 *q* 1 *d*

2 0*q*

， *q*

**

| ** 0*q* |

*n*

3 01

， 1

**

| **01 |

3 

，， *d* ，

**0*d*







2) 



*n*

1  

1

 ˆ1 1 

 1

1 

*In*

*L* *N*

*s*

1

*In*

0， *Is*

1

* *An*

 ，

*n* **0

 *In*

* *An cn*

** 

and

T

** ，，** T

 ** ，， ** ，** ，，** ， ** ，， ** T

where “ *L* ” stands for the convergence in distribu-

**0  **

01 ，，**0*s* 

tion;

*I* 1

is the

 *s*  *s* 

submatrix of

*I* corre-

*n* 1 1 *n* T

0

sponding to the nonzero components

** 1 and

*I* is

1

*s*

 **01，， **0 *p* ，** 01，，** 0*q* ， **01 ，， **0*d*  .

the  *s*1  *s*1  identity matrix.

Accordingly， the quadratic maximization problem for

Remark: The proofs of the Theorems 1 and 2 are simi-

lar to [11]. To save space， the proofs are omitted.

*L* **  leads to a solution iterated by

2

1

 

 

## Computation

**  **

**0 

   *n*  **  *n*  ** **

**0 

 .

### Algorithm

1 0  **** T

** 0

 

** 0 0

** 

Because *L* **  is irregular at the origin， the commonly used gradient method is not applicable. Now， we develop

Secondly， as the data are normally distributed the log-likelihood function **  can be written as

an iterative algorithm based on the local quadratic ap-

** 1 log 1

** T 1 **

*n n*



proximation of the penalty function

*p*  as in [11].

  

2

 *i*    *yi* 

2

*i*  *i*

 *yi*  *i* 

*i*1 *i*1

Firstly， note the first two derivatives of the log-likeli- *n n*

*i i i*

1

hood function **  are continuous. Around a given



 log  *Di*  

1 ** T *D*1** .

point **0 ， the log-likelihood function can be approxi-

2 *i*1

2 *i*1

mated by

T

Therefore， the resulting score functions are

**   **

  **0  

** ** 

** 

*U *   *U* T

** ，*U* T **

，*U* T ** ，

0  **  0

T

  

**

1  

2  

3  

 

 1  

T   **0    .

where

####  0

2

2

 **** T  ** **0

 

*n*

*U*  **   **  

*X* T 1  *y*  **  *X * ;

Also， given an initial value

*t*0 we can approximate the

1 **



*i*1

*i i i i i i*

penalty function

*p* *t*  by a quadratic function [11] *n* T

*U* **   **    *i*

*i i*

2

*D*1** ;

*p*  *t*   *p*  *t*

*n*

  1 *p*  *t*0  *t* 2  *t*2 ，

** *i*1 **

2 *t*0

0 0

*i i i mi*

3

*U* **   **   1  *H* T *D*1 *f* 1 .

for

*t*  *t*0 .

** 2 *i*1

Therefore， the penalized likelihood function (2) can be

local approximated， apart from a constant term， by

Here

  

 *X * 

 diag

*g*

1 

*x*T **

，， *g*

1  *x*T

**  ，

*L* **   **   **0   ** ** 

1

T

*i i i i imi*

*g* 1  is the derivative of the inverse of the link func-

0  **  0

tion *g* 1  ， and

**  **

，，**

T with

  *i i*1

*imi*

 1 ** ** T  **0   ** 

2

  *n* T

  ，

*j* 1 T

2 0  **** T 

 **

2

**0 **

2

* *0 **

*ij*  *rij*  *lijk ik* ; *fi*   *fi*1 ，， *fim* 

where

 

with

*fij ij*

*k* 1

and 1 is a vector of

*m*

*i*

*i*

1 ’s. Denote

*In* ** 

2** 

 *E*

**** T

D. K. XU *ET AL*. 31

models， and proved its model selection consistency prop- erty， *i.e.*， the optimal parameter chosen by BIC can iden- tify the true model with probability tending to one. Hence，

 2** 

 *E*

**** T

2** 

*E*

**** T

2**  

*E* 

**** T

we use their suggestion throughout this paper. So the BIC will be used to choose the optimal ** ， *i*  1，， *s*



 2** 

2** 



2**  

which is equal to either

 *i* 

** 1 ， *i*  1，， *p* ，

  *E*

**** T

*E*

**** T

*E*  *i*

**** T

   

*k*

*j*

or ** 3 ， *k*  1，， *d*

. Nevertheless， in

 2**  2**  2**  

*E*

 *E* **** T *E* **** T **** T 

real application， how to simultaneously select a total of *s*

  shrinkage parameters *i* ， *i*  1，， *s*

is challenging. To

 *I*11

*I*12

*I*13 

bypass this difficulty， we follow the idea of [12，16，17]，

  and simplify the tuning parameters as

  *I*21

**0

*i*



*I*

31

*I*22

*I*32

*I*23  ，

*I*33 

1) **

 **1

， *i*  1，， *p*，

 

where

*n*

T 1

1*i*

2) **  ** 2

**0

*j*

2 *j*

， *j*  1，， *q*，

*I*11

 

*i*1

*Xi* *i* *i* *i Xi* ;

*n*

*I*22   *E*

*i*

** T

1 *i i*

*D*

;

**3

3) **3*k* 

**0

*k*

， *k*  1，， *d*

*i*1 ** **

 0 0

1 *n* T

in the numerical studies followed， where *i* ， * j* and

*I*33   *Hi Hi*

**0 are respectively the *i*th element， *j*th element and *k*th

2 *i*1

*k*

element of the unpenalized estimate

**0 ，

**0

and

**0 .

and

*I*12  *I*21  *I*13  *I*31  *I*23  *I*32  0.

Consequently， the original *s* dimensional problem about

Finally， by using the Fisher information matrix to ap-

proximate the observed information matrix， the following

*i* becomes a three dimensional problem about

**  ** ，** ，** T . ** can be selected according to the fol-

1 2 3

algorithm summarizes the computation of penalized

maximum likelihood estimators of the parameters in

lowing BIC-type criterion

JMVGLRM.

BIC 2 **ˆ，**ˆ， **ˆ   *df*

log *n*

 .

**  **

**Algorithm:**

Step 1. Take the ordinary maximum likelihood esti-

where

*n n*

0  *df*  *s* is simply the number of nonzero co-

mators (without penalty) **0 ，

** 0 ， **0

of ** ， ** ， ** as

efficients of

**ˆ .

their initial values.

Step 2. Given the current values

T

** *m*  ** *m*T ，** *m*T ， ***m*T 

update it by

From our simulation study， we found that this method works well.

## Simulation Studies

In this section we conduct simulation studies to assess

** *m*1  ** *m*

*In*

** *m* 

 *n* ** **

*m* 1

the small sample performance of the proposed proce- dures. We consider the sample size *n* = 100， 200， and 400

*U* ** *m*   *n* 

**

** *m* ** *m*.

respectively. Each subject is supposed to be measured by

*mi* times with

*mi* 1  Binomial11， 0.8 . In the simu-

Step 3. Repeat Step 2 above until certain convergence criteria are satisfied.

### 4.2. Choosing the Tuning Parameters

lation study， 1000 repetitions of random samples are

generated by using the above data generation procedure. For each simulated data set， the proposed variable selec- tion procedures for finding out penalized maximum like- lihood estimators with SCAD and adaptive lasso

The penalty function

controls

*p* *l*   involves the tuning pa-

(ALASSO) penalty functions [17] are considered. The

rameters

** *l*  *l* ha1t， 2， 3 ** t

the amount of

unknown tuning parameters

** *l*  ，

*l*  1， 2， 3

for the

penalty. Many selection criteria， such as CV， GCV， AIC penalty functions are chosen by BIC criterion in the

and BIC selection can be used to select the tuning pa- simulation. The performance of estimator **ˆ ， **ˆ and

rameters. Wang *et al.* [16] suggested using the BIC for **ˆ will be assessed by the mean square error (MSE)， ththee SCADSCAD estiestimatomatorr inin linlinearear mmooddeelsls anandd partiallypartially linlinearear defined as

32 D. K. XU *ET AL*.

MSE**ˆ   *E* **ˆ  **0  **ˆ  **0 ，

T

**1** means the average number of zero regression coeffi- cients that are correctly estimated as zero， and “Incor-

T

MSE(**ˆ)  *E* **ˆ  ** 0  **ˆ  **  ，

0

T

MSE(**ˆ)  *E* **ˆ  **0  **ˆ  **0 .

### Example 1: Linear Mean Model for JMVGLRM

In this example， we first consider the linear model for mean parameters as a special JMVGLRM. We choose the true values of the mean parameters， moving average parameters and log-innovation variances to be

**   ** ， ** ，， ** T with **  1 ， **  0.5 ， **  0.5 ，

1 2 10 1 2 4

rect” depicts the average number of non-zero regression coefficients that are erroneously set to zero.

From **Table 1**， we can make the following observa- tions. Firstly， the performances of variable selection pro- cedures with different penalty functions become better and better as *n* increases. For example， the values in the column labeled “Correct” become more and more closer to the true number of zero regression coefficients in the models. Secondly， the SCAD and ALASSO penalty methods perform similarly in the sense of correct vari- able selection rate， which significantly reduces the model uncertainty and complexity. Thirdly， for the designed

settings， the overall performance of the variable selection

**  ** ， ** ，， ** T with **  0.5 ， **  0.4 ， and

1 2 7 2 3

procedure is satisfactory.

**  ** ，** ，，** T with **  0.3 ， **  0.3 ， respec-

Next， we compare the two decomposition methods

1 2 7 1 2

tively， while the remaining coefficients， corresponding to the irrelevant variables， are given by zeros. In the models

*x*  1， *x*T T with *x* is generated from a multivariate

*ij* 1*ij* 1*ij*

under two data generating processes， autoregressive (AR)

decomposition [1] and moving average (MA) decompo- sition [6]. The main measurements for comparison are differences between the fitted mean **ˆ*i* and the true



normal distribution with mean zero， marginal variance 1

mean *i* ， and the fitted covariance matrix ˆ *i* to the

and all correlations 0.5. We take

*h*   *x* 7

and

true

 . In particular， we define two relative errors as

*ij ijt t* 1

*i*

1 *n *ˆ  ** ˆ 1

ˆ

*i i*

*i*

 

  

*n*

*zijk*  1， *tij*  *tik* ， *tij*  *tik* 

2

6 T

，， *tij*  *tik*

RERR  **ˆ    *i i* ， RERR    .

*n i*1 *i*

*n i*1

and the measurement times *tij* are generated from the uniform distribution *U* 0， 2 . Using these values， the

Here *A* denotes the largest singular value of *A*. We compute the averages of these two relative errors for

mean *i*

and covariance matrix *i*

are constructed

1000 replications with *n* = 100 and 200. **Table 2** gives

through the modified Cholesky decomposition described the averages of relative errors for the MA decomposition

in Section 2. The responses *yi* are then drawn from the

and AR decomposition， when the data are generated from

multivariate normal distribution

*N*  *i* ， *i* ， *i*  1，， *n*.

our model under different true covariance matrix. In **Ta-**

The average number of the estimated zero coefficients

for the parametric components， with 1000 simulation runs， is reported in **Table 1**. Note that “Correct” in **Table**

**ble 2**， “MA.data” (“AR.data”) means that the true co-

variance matrix follows the moving average structure (autoregressive structure). “MA.fit” (“AR.fit”) means we

**Table 1. Variable selection for JMVGLRM (linear mean model) using different penalties and sample size.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | n |  | SCAD |  |  | ALASSO |  |
|  |  | MSE | Correct | Incorrect | MSE | Correct | Incorrect |
|  | 100 | 0.0012 | 6.9340 | 0 | 0.0012 | 7.0000 | 0 |
| ** | 200 | 7.8107e−004 | 6.9870 | 0 | 8.3486e−004 | 7.0000 | 0 |
|  | 400 | 0.0005 | 6.9990 | 0 | 0.0006 | 7.0000 | 0 |
|  | 100 | 1.3369e−004 | 4.9080 | 0 | 1.2259e−004 | 4.9880 | 0 |
| ** | 200 | 6.5800e−005 | 4.9850 | 0 | 7.4626e−005 | 5.0000 | 0 |
|  | 400 | 4.6587e−005 | 4.9980 | 0 | 5.3295e−005 | 5.0000 | 0 |
|  | 100 | 0.0417 | 4.8700 | 0.0010 | 0.0356 | 4.9750 | 0.0010 |
| ** | 200 | 0.0254 | 4.9380 | 0 | 0.0246 | 4.9970 | 0 |
|  | 400 | 0.0218 | 4.9940 | 0 | 0.0190 | 5.0000 | 0 |

D. K. XU *ET AL*. 33

**Table 2. Average of relative errors using different methods and sample size.**

MA.fit AR.fit

True Method n

RERR  **ˆ 

RERR ˆ 

RERR  **ˆ 

RERR ˆ 

100 0.0159 0.1999 0.0609 0.8528

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | SCAD | 200 | 0.0127 | 0.1719 | 0.0554 | 0.8245 |
| MA.data |  | 100 | 0.0161 | 0.1548 | 0.0696 | 0.8158 |
|  | ALASSO | 200 | 0.0131 | 0.1442 | 0.0646 | 0.8047 |
|  | SCAD | 100  200 | 0.0495  0.0427 | 0.6636  0.6527 | 0.0404  0.0356 | 0.3061  0.2639 |
| AR.data |  | 100 | 0.0541 | 0.6960 | 0.0400 | 0.2370 |
|  | ALASSO | 200 | 0.0436 | 0.6473 | 0.0362 | 0.2253 |

decompose the covariance matrix by MA decomposition

** 0 is a *q*-dimensional vector of parameters with

(AR decomposition) to fit data. We see that when the

true covariance matrix follows the moving average

*q*  2*n*1 3   2 and **

0



is a *d*-dimensional vector of

structure， the errors in estimating ** and  both in-

parameters with

*d*  3*n*1 3   3 *x*  1， *x*T T with *x*

  *ij*

1*ij*

1*ij*

crease when incorrectly decomposing the covariance

matrix using the autoregressive structure， and vice versa. However， for this simulation study， model misspecifica- tion seems to affect the MA decomposition less than AR decomposition.

is generated from a multivariate normal distribution with mean zero， marginal variance 1 and all correlations 0.5.

We take

*h*   *x* *d*

### Example 2: Generalized Linear Mean Model

*ij ijt t* 1



 2

 *q*1 T

### for JMVGLRM

*zijk* 

1， *tij*  *tik* ， *tij*  *tik*

，， *tij*  *tik* ，

Consider the following logistic link function to model the

where the measurement times

*t*a*ij*re generated from the

mean component in the JMVGLRM， then we have

uniform distribution *U* 0， 2 .

#   T

The true coefficient vectors are

T

log *it*

*ij*

 *xij * .

**0 

1，-0.5， 0.5， 0 *p*3 

We use the settings in example 1 to assess the per-

T

formance of the proposed variable selection procedures， and the simulation results are reported in **Table 3**.

**0  -0.4， 0.4， 0

*d* 2 

The results in **Table 3** show that under different sam-

**  -0.6， 0.6， 0 T ，

ple size， the proposed variable selection methods have

0 *q*2

the desired performance， which is substantively similar

and， where 0*m* denotes a *m*-vector of 0’s. Using these

to the previous example.

values， the mean *i* and covariance matrix *i* are

constructed through the modified Cholesky decomposi-

### Example 3: High-Dimensional Setup for JMVGLRM

tion described in Section 2. Then， the responses *yi* are then drawn from the multivariate normal distribution

*N*  *i* ， *i* ， *i*  1，， *n*. The summary of simulation re-

In this example， we discuss how the proposed variable

selection procedures can be applied to the “large *n*， di- verging *s*” setup for JMVGLRM. We consider the fol- lowing high-dimensional logistic mean model in JMVGLRM:

sults are reported in **Table 4**.

It is easy to see from **Table 4** that， the proposed vari- able selection method is able to correctly identify the true submodel， and works remarkably well， even if it is the “large *n*， diverging *s*” setup for JMVGLRM.

  T T  2  T

log *it*

*ij*

 *xij *0 ， *lijk*  *zijk* 0 ， log * ij*

 *hij *0 ，

1. **Acknowledgements**

where **0 is a *p*-dimensional vector of parameters with

This work is supported by grants from the National Na-

*p*  4*n*1 3   4



for *n* = 100， 200 and 400， and *u*

tural Science Foundation of China (10971007， 11271039，

denotes the largest integer not greater than *u*. In addition， 11261025); Funding Project of Science and Technology

34 D. K. XU *ET AL*.

**Table 3. Variable selection for JMVGLRM (generalized linear mean model) using different penalties and sample size.**

Model n

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SCAD ALASSO | | | | |
|  | MSE Correct Incorrect MSE Correct Incorrect | | | | |
| 100 | 0.1346 | 6.8820 | 0.0300 | 0.1591 6.9580 | 0.0700 |
| ** 200 0.1028 | | 6.9920 | 0 | 0.0886 6.9980 | 0.0010 |
| 400 | 0.0838 7.0000 | | 0 | 0.0727 7.0000 | 0 |
| 100 1.2997e−004 | | 4.8480 | 0 | 1.4948e−004 4.9900 | 0 |
| ** 200 7.2503e−005 | | 4.9720 | 0 | 8.3386e−005 5.0000 | 0 |
| 400 2.5737e−005 | | 4.9820 | 0 | 5.9863e−005 5.0000 | 0 |
| 100 0.0149 | | 4.9270 | 0 | 0.0297 4.9980 | 0.0030 |
| ** 200 | 0.0086 4.9940 | | 0 | 0.0178 5.0000 | 0 |
| 400 | 0.0059 5.0000 | | 0 | 0.0135 5.0000 | 0 |

**Table 4. Variable selection for high-dimensional JMVGLRM (generalized linear mean model) using different penalties and sample size.**

Model (n， p/q/d)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | SCAD ALASSO | | | | | | |
|  | MSE Correct Incorrect MSE Correct Incorrect | | | | | | |
| (100， 14) |  | 0.0053 | 10.7840 | 0.0090 0.0063 | | 11.0000 | 0.0090 |
| ** (200， 18) |  | 0.0004 | 14.9900 0 | | 0.0011 | 15.0000 0 | |
| (400， 24) |  | 0.0002 | 20.9980 0 | | 0.0005 | 21.0000 0 | |
| (100， 7) |  | 0.0022 | 4.8690 | 0.0060 | 0.0022 | 4.9990 0.0060 | |
| ** (200， 9) |  | 6.2065e−006 | 6.8200 | 0 | 5.4613e−006 | 6.9820 | 0 |
| (400， 12) |  | 1.8671e−005 9.7170 | | 0 | 3.1547e−006 9.8960 | | 0 |
| (100， 10) |  | 0.0151 | 7.8060 | 0.0060 | 0.0276 | 7.9910 0.0060 | |
| ** (200， 14) |  | 0.0117 | 11.9260 0 | | 0.0225 | 12.0000 0 | |
| (400， 18) | 0.0071 | 15.988 0 0 | | | 0.0105 | 16.0000 0 | |

Research Plan of Beijing Education Committee (JC- 006790201001); Beijing municipal key disciplines (No. 006000541212010).

## REFERENCES

[1] M. Pourahmadi， “Joint Mean-Covariance Models with Applications to Lontidinal Data: Unconstrained Parame- terisation，” *Biometrika*， Vol. 86， No. 3， 1999， pp. 677-

690. [doi:10.1093/biomet/86.3.677](http://dx.doi.org/10.1093/biomet/86.3.677)

[[2] M. Pourahmadi， “Maximum Like](http://dx.doi.org/10.1093/biomet/86.3.677)lihood Estimation for Generalised Linear Models for Multivariate Normal Co- variance Matrix，” *Biometrika*， Vol. 87， No. 2， 2000， pp.

425-435. [doi:10.1093/biomet/87.2.425](http://dx.doi.org/10.1093/biomet/87.2.425)

[[3] P. T. Diggle and A. Verbyla， “Nonpara](http://dx.doi.org/10.1093/biomet/87.2.425)metric Estimation of Covariance Structure in Longitudinal Data，” *Biomet- rics*， Vol. 54， No. 2， 1998， pp. 401-415. [doi:10.2307/3109751](http://dx.doi.org/10.2307/3109751)

[4] J. Q. Fan， T. Huang and R. Li， “Analysis of Longitudinal [Data with Semiparam](http://dx.doi.org/10.2307/3109751)etric Estimation of Covariance Function，” *Journal of the American Statistical Associa- tion*， Vol. 102， No. 478， 2007， pp. 632-641. [doi:10.1198/016214507000000095](http://dx.doi.org/10.1198/016214507000000095)

[[5] J. Q. Fan and Y. Wu， “Semiparametric E](http://dx.doi.org/10.1198/016214507000000095)stimation of [Covariance Matrices for Longitudin](http://dx.doi.org/10.1198/016214507000000095)al Data，” *Journal of the American Statistical Association*， Vol. 103， No. 484， 2008， pp. 1520-1533. [doi:10.1198/016214508000000742](http://dx.doi.org/10.1198/016214508000000742)

[[6] A. J. Rothman， E. Levina and J. Zhu， “A New Approach](http://dx.doi.org/10.1198/016214508000000742) [to Cholesky-Based Covariance Regularization in High](http://dx.doi.org/10.1198/016214508000000742) Dimensions，” *Biometrika*， Vol. 97， No. 3， 2010， pp. 539-

550. [doi:10.1093/biomet/asq022](http://dx.doi.org/10.1093/biomet/asq022)

[[7] W. P. Zhang and C. L. Leng， “A Moving Average Cho-](http://dx.doi.org/10.1093/biomet/asq022) [lesky Factor Model in Covarianc](http://dx.doi.org/10.1093/biomet/asq022)e Modeling for Longitu- dinal Data，” *Biometrika*， Vol. 99， No. 1， 2012， pp. 141-

150. [doi:10.1093/biomet/asr068](http://dx.doi.org/10.1093/biomet/asr068)

[[8] L. Breiman， “Better Subset Selection Using Nonnegative](http://dx.doi.org/10.1093/biomet/asr068) [Garrote，”](http://dx.doi.org/10.1093/biomet/asr068) *[Techonometrics](http://dx.doi.org/10.1093/biomet/asr068)*[， Vol. 37， No. 4](http://dx.doi.org/10.1093/biomet/asr068)， 1995， pp. 373-

D. K. XU *ET AL*. 35

384. [doi:10.1080/00401706.1995.10484371](http://dx.doi.org/10.1080/00401706.1995.10484371)

[[9] R. Tibshirani， “Regression Shrinkage and Selection via](http://dx.doi.org/10.1080/00401706.1995.10484371) [the LASSO，”](http://dx.doi.org/10.1080/00401706.1995.10484371) *[Journal of Royal Statistical So](http://dx.doi.org/10.1080/00401706.1995.10484371)ciety*， *Series B*， Vol. 58， No. 1， 1996， pp. 267-288.

[10] W. J. Fu， “Penalized Regression: The Bridge versus the LASSO，” *Journal of Computational and Graphical Sta- tistics*， Vol. 7， No. 3， 1998， pp.397-416.

[11] J. Q. Fan and R. Li， “Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties，” *Journal of American Statistical Association*， Vol. 96， No. 456， 2001， pp. 1348-1360. [doi:10.1198/016214501753382273](http://dx.doi.org/10.1198/016214501753382273)

[[12] H. Zou and R. Li， “One-Step Sparse Estimates in Non-](http://dx.doi.org/10.1198/016214501753382273) [concave Penalized Likelihood Models，”](http://dx.doi.org/10.1198/016214501753382273) *[The Annals of](http://dx.doi.org/10.1198/016214501753382273) Statistics*， Vol. 36， No. 4， 2008， pp. 1509-1533. [doi:10.1214/009053607000000802](http://dx.doi.org/10.1214/009053607000000802)

[[13] Z. Z. Zhang and D. R. Wang， “Simultaneous](http://dx.doi.org/10.1214/009053607000000802) Variable [Selection for Heteroscedastic Regr](http://dx.doi.org/10.1214/009053607000000802)ession Models，” *Sci- ence China Mathematic*， Vol. 54， No. 3， 2011， pp. 515-

530. [doi:10.1007/s11425-010-4147-8](http://dx.doi.org/10.1007/s11425-010-4147-8)

[[14] P. X. Zhao and L. G. Xue， “Variable](http://dx.doi.org/10.1007/s11425-010-4147-8) Selection in Semi- parametric Regression Analysis for Longitudinal Data，” *Annals of the Institute of Statistical Mathematics*， Vol. 64， No. 1， 2012， pp. 213-231.

[doi:10.1007/s10463-010-031](http://dx.doi.org/10.1007/s10463-010-0312-7)2-7

[[15] H. J. Ye and J. X. Pan， “Modelli](http://dx.doi.org/10.1007/s10463-010-0312-7)ng of Covariance Struc- tures in Generalized Estimating Equations for Longitudi- nal Data，” *Biometrika*， Vol. 93， No. 4， 2006， pp. 927-941. [doi:10.1093/biomet/93.4.927](http://dx.doi.org/10.1093/biomet/93.4.927)

[[16] H. Wang， G. Li and C. L. T](http://dx.doi.org/10.1093/biomet/93.4.927)sai， “Tuning Parameter Se- lectors for the Smoothly Clipped Absolute Deviation Me- thod，” *Biometrika*， Vol. 94， No. 3， 2007， pp. 553-568. [doi:10.1093/biomet/asm053](http://dx.doi.org/10.1093/biomet/asm053)

[[17] H. Zou， “The Adaptive Lasso](http://dx.doi.org/10.1093/biomet/asm053) and Its Oracle Properties，” *Journal of American Statistical Association*， Vol. 101， No. 476， 2006， pp. 1418-1429. [doi:10.1198/0162145060000](http://dx.doi.org/10.1198/016214506000000735)00735