HHS Public Access

Author manuscript

Comput Stat Data Anal. Author manuscript; available in PMC 2018 November 01.

Published in final edited form as:

Comput Stat Data Anal. 2017 November ; 115: 267–280. doi:10.1016/j.csda.2017.05.001.

ARMA Cholesky Factor Models for the Covariance Matrix of Linear Models

Keunbaik Lee[1], **Changryong Baek**[1], and **Michael J. Daniels**[2] **[Professor]**

[1]Department of Statistics, Sungkyunkwan University, Seoul, 110-745, Korea

[2]Department of Statistics & Data Sciences and Department of Integrative Biology, University of Texas at Austin, Austin, TX

Summary

In longitudinal studies, serial dependence of repeated outcomes must be taken into account to make correct inferences on covariate effects. As such, care must be taken in modeling the covariance matrix. However, estimation of the covariance matrix is challenging because there are many parameters in the matrix and the estimated covariance matrix should be positive definite. To overcomes these limitations, two Cholesky decomposition approaches have been proposed: modified Cholesky decomposition for autoregressive (AR) structure and moving average Cholesky decomposition for moving average (MA) structure, respectively. However, the correlations of repeated outcomes are often not captured parsimoniously using either approach separately. In this paper, we propose a class of flexible, nonstationary, heteroscedastic models that exploits the structure allowed by combining the AR and MA modeling of the covariance matrix that we denote as ARMACD. We analyze a recent lung cancer study to illustrate the power of our proposed methods.

Keywords

Cholesky decomposition; longitudinal data; heteroscedastic

1. Introduction

In a longitudinal study, an outcome variable is observed repeatedly over a period of time. Since the repeated observations are not independent, their dependence must be taken into account to make proper inference on covariate effects in terms of correct standard errors (Diggle et al., 2002). In addition, in the presence of missing data (which is common in longitudinal studies), correctly modelling the covariance matrix is necessary to avoid bias in (mean) covariate effects (Daniels and Hogan, 2008). However, estimation of the covariance matrix is challenging because there are many parameters and the estimated covariance matrix needs to be positive definite. To remove these restrictions, simple structures for the

covariance matrix are often assumed such as exchangeable or autoregressive order one. However, such statistical approaches do not permit more general forms of the correlation structure and cannot explain heterogeneous covariances (Pourahmadi, 1999). To overcome these limitations, joint mean-variance modeling approaches have been proposed for the covariance matrix (Pourahmadi, 1999; Chiu, Leonard, and Tsui, 1996; Zhang and Leng, 2012, Hoff and Niu, 2012) and the correlation matrix (Daniels and Pourahmadi, 2002; Zhang et al., 2015). In this paper, we focus on modeling of the covariance matrix. Pourahmadi (1999) and Zhang and Leng (2012) proposed Cholesky-type decomposition approaches for modeling of covariance matrix of longitudinal data, the modified Cholesky decomposition and the moving average Cholesky decomposition, respectively.

In the modified Cholesky decomposition (MCD), the inverse of the covariance matrix is decomposed into generalized autoregressive parameters (GARPs) and innovation variances (IVs); the positive-definite restriction corresponds to the IVs being positive (Pourahmadi, 1999, 2000). Bayesian modeling using the MCD was proposed in Daniels and Pourahmadi (2002) and Daniels and Zhao (2003). Pan and MacKenzie (2003, 2006) extended the MCD to deal with unbalanced longitudinal data and proposed the BIC as the optimal approach to identify the degree of the mean, innovation variance and GARP parameters under an assumption that each follows a polynomial model (with the GARP parameters also assumed to just depend on lag). Lee et al. (2012) and Lee (2013) used the MCD to model the random effects covariance matrix in GLMMs and Lee and Sung (2014) extended these to marginalized random effects models (Heagerty, 1999).

In the moving average Cholesky decomposition (MACD), the covariance matrix is decomposed into generalized moving average parameters (GMAPs) and IVs (Zhang and Leng, 2012). This approach uses the moving average parameterization of the covariance matrix as an alternative to the autoregressive one. Similar to the MCD, the covariance matrix is positive definite when the IVs are positive. Lee and Yoo (2014) used this decomposition for modeling of the random effects covariance matrix to analyze long series of longitudinal binary data.

When a high-order AR structure of the covariance matrix is required to capture the dependence structure, practitioners often consider autoregressive moving average (ARMA) models (Judge et al., 1980). The advantage of the ARMA models is that the models provide for a wide variety of structures in the covariance matrix but can be specified using a small number of parameters (Rochen and Helms, 1989). Rochen (1992) extended Rochen and Helms's (1989) model, which had a homogeneous ARMA covariance matrix, to allow for a heterogeneous ARMA covariance matrix. In this paper, we also consider the heterogeneous ARMA covariance matrix by ‘combining’ the MCD and MACD to create a more flexible decomposition of the covariance matrix. This new decomposition will still provide simple conditions for positive definiteness (in terms of the IVs) and allow flexible nonstationary and heteroscedastic models that can be more parsimonious than models based on either the MCD or MACD alone. We emphasize the models proposed here are much more general (and flexible) than stationary ARMA models used in time series; we make connections to ‘standard’ ARMA models to emphasize that we try to exploit structure on both autoregressive *and* moving average type parameters in a non-stationary setting.

This paper is organized as follows. In Section 2, we propose covariance models using the ARMA Cholesky decomposition (ARMACD). We show that the resulting maximum likelihood estimator of the mean and the parameters of the covariance matrix are consistent and asymptotically normally distributed. In Section 3, we examine the bias of mean parameter estimates to misspecification of covariance matrix with and without missing data in simulation studies. In Section 4, we apply our proposed models to data from a recent lung cancer study. Finally, we summarize and propose future work in Section 5.

2. ARMA Cholesky Factor Models for Covariance Matrix

We propose a new class of models for the covariance matrix in longitudinal data that relies on the parameters in the new ARMACD.

2.1 Proposed model

Let $y_i = (y_{i1}, \cdots, y_{in_i})^T$ be the longitudinal response vector of the ith subject ($i = 1, \cdots, N$); y_{it} is the response at time g_{it} for subject i; we assume the observation times are equally spaced, so without loss of generality, we can set $g_{it} = t$. We assume that the responses for different subjects are independent. Let x_{it} indicate covariates corresponding to y_{it}. We assume the y_{it} follows the linear model which is given by

$$y_{it} - x_{it}^T\beta = \sum_{l=1}^{t-1}\phi_{it,t-l}(y_{it-l} - x_{it-l}^T\beta) + \sum_{j=1}^{t-1} l_{it,t-j}e_{it-j} + e_{it}, \quad (1)$$

where β is a $p \times 1$ coefficients vector of x_{it}, $e_{i1} = y_{i1} - x_{i1}^T\beta$, $e_i = (e_{i1}, \cdots, e_{in_i})^T$, $e_i \sim N(0, D_i)$, and $D_i = \mathrm{diag}(\sigma_{i1}^2, \cdots, \sigma_{in_i}^2)$. The parameters $\phi_{it,t-l}$, $l_{it,t-j}$, and σ_{ij}^2 are generalized autoregressive parameters (GARPs), generalized moving average parameters (GMAPs), and innovation variances (IVs), respectively. The GARPs and GMAPs are unconstrained. We implicitly assume that model (1) is well-defined.

We can rewrite (1) in matrix form as

$$T_i(y_i - X_i\beta) = L_i e_i \quad (2)$$

where T_i is a unit lower triangular matrix having ones on its diagonal and $-\phi_{i,tj}$ at its (t, j)th position for $j < t$, and L_i is a unique lower triangular matrix having ones on its diagonal and $l_{i,tj}$ at its (t, j)th position for $j < t$, and $X_i = (x_{i1}, \ldots, x_{in_i})^T$ is the design matrix. In this paper we assume that X_i is of full rank. From (2), we have

$$T_i \sum\nolimits_i T_i^T = L_i D_i L_i^T. \quad (3)$$

where Σ_i is the covariance matrix for y_i.

Theorem. 1. Σ_i in (3) is positive definite if the diagonal elements of D_i (i.e., the IVs) are all positive.

Proof: Let x be an arbitrary nonzero vector. Since $\sum_i = T_i^{-1} L_i D_i L_i^T T_i^{-T}$, $x^T \sum_i x = x T_i^{-1} L_i D_i L_i^T T_i^{-T} x$. For $z = L_i^T T_i^{-T} x$, $x^T \Sigma_i x = z^T D_i z$. Therefore, Σ_i is positive definite if $\sigma_{it}^2 > 0$ for $t = 1, \cdots, n_i$.

The GARPs, GMAPs, and IVs can be modeled using time and/or subject-specific covariate vectors $w_{i,tj}$ and $h_{i,t}$ by setting

$$\begin{aligned} \phi_{i,tj} &= w_{i,tj}^T \alpha, \; l_{i,tj} = z_{i,tj}^T \gamma, \\ \log(\sigma_{i,t}^2) &= h_{i,t}^T \lambda, \end{aligned} \tag{4}$$

where α, γ, and λ are $a \times 1$, $b \times 1$, and $c \times 1$ vectors of unknown parameters, respectively. Note that we jointly model the mean and covariance structures of responses in terms of the generalized linear models from (1) and (4).

The design vectors $w_{i,tj}$, $z_{i,tj}$, and $h_{i,t}$ are used to model the GARP/GMAP/IV parameters. These design vectors can include subject-specific covariates (so heteroscedastic regression) and/or to incorporate structure. For the latter, $h_{i,t} = (1, t)$ corresponds to the innovation variances change (log-)linearly with time and corresponds to a nonstationary covariance structure. Parsimonious higher lag models can be specified by including time lag, $|t - j|$ in the design vectors $w_{i,tj}$ and/or $z_{i,tj}$. Finally, structured nonstationary (in the GARP and GMAP) models can be specified; for example, a nonstationary first order moving average process could be specified using $z_{i,tj} = (I_{|t-j|=1}, I_{(t-j)=1} \times t)$. Overall, the covariance matrix with these models (4) includes heteroscedastic and nonstationary processes for Y_{it}.

Note that the parameters in the decomposition are not identifiable without any structure of the covariance matrix. When there are no subject-specific covariates in the co-variance matrix, the matrices T and L in (2) have $T(T-1)/2 + T(T-1)/2$ parameters where $T = max(n_i)$. However, with a specific structure of the covariance matrix such as those based on the ARMA Cholesky structure, both matrices can be identified and the identifiability is easily assessed by checking invertibility of the Hessian matrix discussed in Section 2.3. Also note that our ARMACD models have the parameters of the MCD and MACD as special cases; so our model extends previous models and provides a unified framework for modeling the covariance structure of linear models in longitudinal studies.

The advantage of ARMA modeling in time series analysis is well known. For example, parsimony in parameterization makes the model interpretation easy and provides stable estimation of parameters. In addition, it also provides better forecasting performance than competing higher order AR or MA models. For example, Hansen and Lunde (2005) pointed out that including both AR and MA structure in the financial GARCH model leads to no

worse performance than other sophisticated models in forecasting. Along similar lines, we expect ARMACD to also enjoy many useful advantages of ARMA modeling. In addition, in our setting of longitudinal data, we can construct more complex (but still parsimonious) 'ARMA' types models that allow nonstationary and heteroscedasticity as described above.

2.2 Maximum Likelihood Estimation

We derive the likelihood function for the model specified in Section 2.1. Let $\theta = (\beta^T, \alpha^T, \gamma^T, \lambda^T)$. The likelihood function is a product of multivariate normal probability densities,

$$\begin{aligned} L(\theta;y) &= \prod_{i=1}^{N} (2\pi)^{-\frac{n_i}{2}} |\textstyle\sum_i|^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}(y_i - X_i\beta)^T \textstyle\sum_i^{-1} (y_i - X_i\beta)\right\} \\ &= \prod_{i=1}^{N} (2\pi)^{-\frac{n_i}{2}} \left(\prod_{i=1}^{n_i} \sigma_{it}^2\right)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}(y_i - X_i\beta)^T T_i^T L_i^{-T} D_i^{-1} L_i^{-1} T_i (y_i - X_i\beta)\right\}. \end{aligned} \quad (5)$$

Recall from (3) that $\sum_i = T_i^{-1} L_i D_i L_i^T T_i^{-T}$

The log likelihood is given by

$$\log L(\theta;y) = \sum_{i=1}^{N} \left\{ -\frac{n_i}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n_i} h_{it}^T \lambda - \frac{1}{2}(r_i - W_i\alpha)^T L_i^{-T} D_i^{-1} L_i^{-1} (r_i - W_i\alpha) \right\},$$

where $r_i = y_i - X_i\beta$ and

$$W_i = \begin{pmatrix} 0^T \\ (y_{i1} - x_{i1}^T\beta) w_{i21}^T \\ \vdots \\ \sum_{j=1}^{t-1} (y_{ij} - x_{ij}^T\beta) w_{itj}^T \\ \vdots \\ \sum_{j=1}^{n_i-1} (y_{ij} - x_{ij}^T\beta) w_{in_ij}^T \end{pmatrix}$$

with 0^T being $1 \times a$ vector of zeros.

Maximizing the log-likelihood with respect to θ yields the likelihood equations

$$\sum_{i=1}^{N} \frac{\partial \log L(\theta;y_i)}{\partial \theta} = 0$$

where

$$\frac{\partial \log L(\theta;y_i)}{\partial \beta}=x_i^T\sum\nolimits_i^{-1}(y_i - X_i\beta), \quad (6)$$

$$\frac{\partial \log L(\theta;y_i)}{\partial \alpha}=W_i^T L_i^{-T} D_i^{-1} L_i^{-1}(r_i - W_i\alpha), \quad (7)$$

$$\frac{\partial \log L(\theta;y_i)}{\partial \gamma_j}=-\frac{1}{2}(y_i - X_i\beta)^T\left(T_i^T\frac{\partial L_i^{-T}}{\partial \gamma_j}D_i^{-1}L_i^{-1}T_i+T_i^T L_i^{-T}D_i^{-1}\frac{\partial L_i^{-1}}{\partial \gamma_j}T_i\right)(y_i-X_i\beta), \quad (8)$$

$$\frac{\partial \log L(\theta;y_i)}{\partial \lambda_l}=-\frac{1}{2}\left\{\sum_{t=1}^{n_i}h_{itl}+(y_i - X_i\beta)^T T_i^T L_i^{-T}\frac{\partial D_i^{-1}}{\partial \lambda_l}L_i^{-1}T_i(y_i - X_i\beta)\right\}, \quad (9)$$

with

$$\frac{\partial L_i^{-1}}{\partial \gamma_j}=-L_i^{-1}\frac{\partial L_i}{\partial \gamma_j}L_i^{-1},$$

$$\frac{\partial L_i}{\partial \gamma_j}=\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ z_{i21j} & 0 & 0 & \cdots & 0 \\ z_{i31j} & z_{i32j} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{in_i1j} & z_{in_i2j} & z_{in_i3j} & \cdots & 0 \end{pmatrix},$$

$$\frac{\partial D_i^{-1}}{\partial \lambda_l}=-\operatorname{diag}\left\{\frac{h_{i1l}}{\sigma_{i1}^2},\frac{h_{i2l}}{\sigma_{i2}^2},\cdots,\frac{h_{in_il}}{\sigma_{in_i}^2}\right\},$$

for $j = 1, \cdots, b$ and $l = 1, \cdots, c$.

From (6), we have the maximum likelihood estimator (MLE) for β which is given by

$$\hat{\beta}=\left(\sum_{i=1}^{N}X_i^T\hat{\sum}_i^{-1}X_i\right)^{-1}\sum_{i=1}^{N}X_i^T\hat{\sum}_i^{-1}y_i,$$

where $\hat{\Sigma}_i = \Sigma_i(\hat{\alpha}, \hat{\gamma}, \hat{\lambda})$ and $(\hat{\alpha}, \hat{\gamma}, \hat{\lambda})$ are the ML estimator of $(\alpha, \gamma, \lambda)$. From (7), we also have the MLE for α which is given by

$$\hat{\alpha}=\left(\sum_{i=1}^{N}\hat{W}_i^T\hat{L}_i^{-T}\hat{D}_i^{-1}\hat{L}_i^{-1}\hat{W}_i\right)^{-1}\sum_{i=1}^{N}\hat{W}_i^T\hat{L}_i^{-1}\hat{D}_i^{-1}\hat{L}_i^{-1}\hat{r}_i,$$

where $\hat{W}_i = W_i(\hat{\beta})$, $\hat{L}_i = \hat{L}_i(\hat{\gamma})$, $\hat{D}_i = D_i(\hat{\lambda})$, and $\hat{r}_i = y_i - X_i\hat{\beta}$. Because we do not have closed forms of the solutions of γ and λ from (8) and (9), we use an iterative procedure to find the MLE.

Since the analytic forms of second derivatives of the observed data log-likelihood in (8) and (9) are not available in closed form, we use quasi-Newton methods to solve the likelihood equations. Let $\omega = (\gamma^T, \lambda^T)$. Then the $(c + 1)$th iteration $\omega_1^{(c+1)}$ is updated using

$$\omega^{(c+1)}=\omega^{(c)}+[H(\omega^{(c)};y)]^{-1}\frac{\partial \log L}{\partial \omega^{(c)}},$$

where

$$\frac{\partial \log L}{\partial \omega}=\left(\sum_{i=1}^{N}\frac{\partial \log L(\theta;y_i)}{\partial \gamma^T},\sum_{i=1}^{N}\frac{\partial \log L(\theta;y_i)}{\partial \lambda^T}\right)^T,$$
$$H(\omega;y)=\mathrm{diag}\left\{E\left(-\frac{\partial^2 \log L}{\partial\gamma\partial\gamma^T}\right),E\left(-\frac{\partial^2 \log L}{\partial\lambda\partial\lambda^T}\right)\right\},$$

with

$$E\left(-\frac{\partial^2 \log L}{\partial\gamma_{j'}\partial\gamma_j}\right)=\frac{1}{2}\mathrm{tr}(B_i\Sigma_i),$$
$$E\left(-\frac{\partial^2 \log L}{\partial\lambda_{j'}\partial\lambda_j}\right)=\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{n_i}h_{itj'}h_{itj}, \quad (10)$$

having

$$B_i=T_i^T\frac{\partial^2 L_i^{-T}}{\partial\gamma_{j'}\partial\gamma_j}D_i^{-1}L_i^{-1}T_i+T_i^T\frac{\partial L_i^{-1}}{\partial\gamma_j}D_i^{-1}\frac{\partial L_i^{-1}}{\partial\gamma_{j'}}T_i+T_i^T\frac{\partial L_i^{-T}}{\partial\gamma_{j'}}D_i^{-1}\frac{\partial L_i^{-1}}{\partial\gamma_j}T_i+T_i^TL_i^{-T}D_i^{-1}\frac{\partial^2 L_i^{-1}}{\partial\gamma_{j'}\partial\gamma_j}T_i.$$

2.3 Asymptotic properties

Since we use maximum likelihood for estimation, the resulting estimators are efficient. Let $\theta = (\beta^T, \omega^T, \lambda^T)^T$, where $\omega = (\alpha, \gamma)$, and $\theta_0=(\beta_0^T,\omega_0^T,\lambda_0^T)^T$ is the true value of θ. Also let $I(\theta_0) = diag(I(\beta_0), I(\omega_0), I(\lambda_0))$ where

$$I(\beta)=\sum_{i=1}^{N}X_i^T\sum_i^{-1}X_i, I(\omega)$$
$$=\begin{pmatrix} E\left(-\frac{\partial^2\log L}{\partial\alpha\partial\alpha^T}\right) & E\left(-\frac{\partial^2\log L}{\partial\gamma\partial\alpha^T}\right)^T \\ E\left(-\frac{\partial^2\log L}{\partial\gamma\partial\alpha^T}\right) & E\left(-\frac{\partial^2\log L}{\partial\gamma\partial\gamma^T}\right)^T \end{pmatrix}, I(\lambda)$$
$$=E\left(-\frac{\partial^2\log L}{\partial\lambda\partial\lambda^T}\right).$$

Finally, $E\left(-\frac{\partial^2\log L}{\partial\alpha\partial\alpha^T}\right)=\sum_{i=1}^{N}E\{W_i^TL_i^{-T}D_i^{-1}L_i^{-1}\}$,
$E\left(-\frac{\partial^2\log L}{\partial\gamma_{j'}\partial\alpha_j}\right)=\frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}(A_i\sum_i)$, $E\left(-\frac{\partial^2\log L}{\partial\gamma_{j'}\partial\gamma_j}\right)=\frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}(D_iC_i)$ and
$E\left(-\frac{\partial^2\log L}{\partial\lambda_{j'}\partial\lambda_j}\right)$ is given in (10), where

$$A_i=\frac{\partial T_i^T}{\partial\alpha_a}\frac{\partial L_i^{-T}}{\partial\gamma_b}D_i^{-1}L_i^{-1}T_i+\frac{\partial T_i^T}{\partial\alpha_a}L_i^{-T}D_i^{-1}\frac{\partial L_i^{-1}}{\partial\gamma_b}T_i+T_i^T\frac{\partial L_i^{-T}}{\partial\gamma_b}D_i^{-1}L_i^{-1}\frac{\partial T_i}{\partial\alpha_a}+T_i^TL_i^{-T}D_i^{-1}\frac{\partial L_i^{-1}}{\partial\gamma_b}\frac{\partial T_i}{\partial\gamma_a},$$
$$C_i=L_i^T\left\{\frac{\partial}{\partial\gamma_{j'}}\left(\frac{\partial L_i^{-1}}{\partial\gamma_j}\right)D_i^{-1}L_i+\frac{\partial L_i^{-1}}{\partial\gamma_{j'}}D_i^{-1}\frac{\partial L_i^{-1}}{\partial\gamma_j}+\frac{\partial L_i^{-1}}{\partial\gamma_j}D_i^{-1}\frac{\partial L_i^{-1}}{\partial\gamma_{j'}}+L_i^{-T}D_i^{-1}\frac{\partial}{\partial\gamma_j}\left(\frac{\partial L_i^{-1}}{\partial\gamma_{j'}}\right)\right\}L_i.$$

To establish the theoretical properties, we assume the following regularity conditions hold.

C1 The dimensions p, a, b, and c of covariates x_{it}, $w_{it,j}$, $z_{i,tj}$, and h_{it}, and $\max_i n_i$ are fixed.

C2 The true value value $\theta_0=(\beta_0^T,\alpha_0^T,\gamma_0^T,\lambda_0^T)^T$ is in the interior of the parameter space Θ, which is a compact subset of $\mathbb{R}^{p+a+b+c}$.

C3 The covariates x_{it}, $w_{it,j}$, $z_{i,tj}$, and h_{it} are all bounded, meaning that all elements of the vectors are bounded.

C4 When $N\to\infty$, $I(\theta_0)/N$ converges to a positive definite matrix $\mathcal{I}(\theta_0)$.

Then we have the following results:

Theorem. 2. Under regularity conditions given in (C1)-(C4), the following three results hold as $N\to\infty$:

1. $$E\left(-\frac{\partial^2\log L}{\partial\alpha\partial\beta^T}\right)=0, E\left(-\frac{\partial^2\log L}{\partial\gamma\partial\beta^T}\right)=0, E\left(-\frac{\partial^2\log L}{\partial\lambda\partial\beta^T}\right)=0,$$
$$E\left(-\frac{\partial^2\log L}{\partial\lambda\partial\gamma^T}\right)=0, E\left(-\frac{\partial^2\log L}{\partial\lambda\partial\alpha^T}\right)=0;$$

2. 2. The maximum likelihood estimator $\hat{\theta}=(\hat{\beta}^T,\hat{\alpha}^T,\hat{\gamma}^T,\hat{\lambda}^T)^T$ is strongly consistent for $\theta_0=(\beta^T,\alpha^T,\gamma^T,\lambda^T)^T$;

3. The maximum likelihood estimator is asymptotically normal

$$\sqrt{N}\left(\hat{\theta}-\theta_0\right)\rightarrow^{d}N(0,\mathscr{I}^{-1}(\theta_0)), \quad (11)$$

where $\mathscr{I}(\theta_0) = diag(\mathscr{I}(\beta_0),\mathscr{I}(\omega_0),\mathscr{I}(\lambda_0))$ with $\mathscr{I}(\beta_0) \in \mathbb{R}^{p}$, $\mathscr{I}(\omega_0) \in \mathbb{R}^{a+b}$, $\mathscr{I}(\lambda_0) \in \mathbb{R}^{c}$

Proof: See the Appendix.

We note that (C4) in the above theorem corresponds to parameter identifiability. (C4) will hold if the following three conditions hold:

1. the design matrix for the mean parameters is full column rank
2. the design matrix for the (innovation) variance parameters is full column rank
3. $I(\omega_0)$ is invertible.

For the third condition, if there are just GMAP or just GARP parameters and the corresponding design matrix is full column rank, then $I(\omega_0)$ is invertible. When the models include both GMAP and GARP parameters, the necessary and sufficient condition is less clear.

Notice that result 1 in Theorem 2 implies that $\hat{\beta}$, $\hat{\omega}= (\hat{a}^T, \hat{\gamma}^T)^T$, and $\hat{\lambda}$ are asymptotically independent. Since $\hat{\beta}$, $\hat{a}$, $\hat{\gamma}$, and $\hat{\lambda}$ are consistent estimators, $\mathscr{I}$ in the asymptotic covariance is consistently estimated by a block diagonal matrix with block components

$$\hat{\mathscr{I}}(\beta)=N^{-1}I(\hat{\beta}),$$
$$\hat{\mathscr{I}}(\omega)=N^{-1}I(\hat{\omega}),$$
$$\hat{\mathscr{I}}(\lambda)=N^{-1}I(\hat{\lambda}),$$

However, we note that the observed information matrix is not orthogonal unless all the n_i's are equal. In Section 3, we examine the operating characteristics of our models in the common scenario of dropout for longitudinal data, for which the orthogonality in Theorem 2 between the mean and dependence parameters does *not* hold since the n_i's are not equal (Daniels and Hogan, 2008). Next, we suggest a graphical approach to find a reasonable structured model.

2.4 Model Building and Selection

In this section we explore the covariance structure using a graphical method which extends that in Pourahmadi (1999) and Pourahmadi and Daniels (2002). We first assume a particular mean structure ($x_{it}^T\beta$). Then we propose a procedure to explore the covariance structure using the following steps.

Step 1 Compute the MLEs of β and Σ (assuming no structure), $\hat{\beta}$ and $\hat{\Sigma}$ respectively. Then compute the estimated GARPs from $\hat{\Sigma}$ using the following formula (Pourahmadi, 1999):

$$\phi_t = \sum_t^{-1} \sigma_t', \quad (12)$$

where Σ_t is the $(t-1) \times (t-1)$ leading principal submatrix of Σ and σ_t' is the column vector composed of the first $t-1$ entries of the tth column of Σ. Now plot regressograms as in Pourahmadi (1999) and Pourahmadi and Daniels (2002) to find parsimonious structure. These would be plots of the GARPs vs. lag and/or plots of particular lag GARPs vs. time. See the implementation in the data example in Section 4.

Step 2 Based on the structure chosen from the exploratory analysis of the GARP in Step 1, compute the mle of β and Σ. Here we denote the mle of Σ under this structure as $\tilde{\Sigma}$. Compute $\tilde{T}_i$ from this and create transformed data, $\tilde{T}_i(y_i - X_i\beta)$. Now compute the mle of β and the new Σ using the transformed data.

Compute the estimated GMAPs from the mle of the new Σ by calculating the new T_i from the mle of the new Σ and then setting $L_i = T_i^{-1}$ to obtain the estimated GMAPs.

Plot regressograms similar to Step 1. to explore parsimonious structures for the GMAPs.

Step 3 Refit the model using the parimonious structures identified for T_i and L_i in Steps 1 and 2. Plot the diagonal elements of D vs. time to explore structure in the IVs (vs. time).

Note that we can pick a covariance structure heuristically using above procedure. Then we can more formally select the best model using likelihood ratio tests and/or penalized model selection criteria such as AIC and BIC. And also note that the exploratory analysis can 'choose' slightly different models if Steps 1 and 2 are switched.

3. Simulation Study

We conducted several simulation studies to examine the bias, efficiency, and coverage in estimating the coefficients in the mean model in the settings of both true and misspecified structure of the covariance matrix and under both ignorable (MAR) missingness and complete data.

Study 1. We simulated longitudinal data under our proposed model (1) and (4) with two covariates, group (2 levels) and $c_{it} = (t - 10/2) + \varepsilon_{it}$, $\varepsilon_{it} \sim N(0, 0.5^2)$. The model was specified as

$$y_{it} - x_{i,t}^T\beta = \phi_{it,t-1}(y_{it-l} - x_{i,t-1}^T\beta) + l_{it,t-1}e_{it-1} + e_{it}, \quad (13)$$

for $t = 1, \cdots, 10$ with $x_{i,t}^T\beta = \beta_0 + \beta_1 \text{Group}_i + \beta_2 c_{it} + \beta_3 \text{Group}_i \times c_{it}$. The true mean parameter values were $(\beta_0, \beta_1, \beta_2, \beta_3) = (0.1, -0.1, 0.1, 0.1)$; the 'group' covariate, Groupi equals 0 or

1 with probability 0.5, $\phi_{itj} = \alpha \times I_{|t-j|=1}$, $I_{itj} = \gamma \times I_{|t-j|=1}$, and $\log\sigma_{it}^2 = \lambda$ with $(\alpha, \gamma, \lambda) = (.7, .7, .3)$.

We generated 500 simulated data sets each with a sample size of 200 under complete data or under MAR missingness. We fit 3 models. Model 1 is the true model (with ARMA(1,1) structure of the covariance matrix); Model 2 has an AR(1) structure of the covariance matrix; Model 3 has an MA(1) structure of the covariance matrix. For missingness, we specified the following MAR dropout model,

$$\text{logit}P(\text{dropout}=t|\text{dropout} \geq t) = -2.3 + 0.3Y_{it-1}.$$

Based on this specification, the observed dropout rates were approximately 50 percent.

Table 2 presents the mean point estimates and the percent relative biases (RB) of the parameters. In the presence of MAR missingness, the estimates in Model 1 had smaller biases than those in Models 2 and 3. Coverage probabilities for all parameters for Model 1 were approximately 95%. In particular, the percent relative biases and coverage probabilities of β_0 and β_2 show that the true model (Model 1) fit much better than Models 2 and 3. These results demonstrate that the incorrect covariance structure can induce potentially large biases in mean parameters. To assess the accuracy of our standard errors (cf: Section 2.2), we compared the sample standard deviation (SD) of the 500 parameter estimates to the sample average (SE) of the 500 standard errors using (11); they were very close for the true model which suggests the standard error in (11) is accurate. We also calculated the estimated covariance matrix using the mean of 500 fitted values of α, γ, and λ given in Table 2. Then we calculated the sum of absolute differences (SAD) between the estimated and true covariance matrices. The SAD for Model 1 was the smallest among the three models (as expected). In addition, we calculated AICs for all three models and the proportion of times each model had the minimum AIC (among the three models). The proportion for Model 1 was 1.00, which demonstrates the potential utility of the AIC for model selection (at least in these simulation settings).

When there were no missing data, the estimates were essentially unbiased for the three models. The coverage probabilities of mean parameters in Models 1 and 2 were all close to 95% with Model 2 somewhat conservative. However, the coverage probabilities in Model 3 show significant undercoverage. As we expected, the SAD for Model 1 was the smallest and the the proportion of times the AIC chose Model 1 was 1.00.

Study 2. We also generated 500 longitudinal datasets with a sample size of 200 using model (13) with a heteroscedastic ARM A(1,1) structure of the covariance matrix by setting $\phi_{itj} = \alpha \times I_{|t-j|=1}$, $I_{itj} = \gamma \times I_{|t-j|=1}$, and $\log\sigma_{it}^2 = \lambda_0 + \lambda_1 \text{Group}_i$ with $(\alpha, \gamma, \lambda_0, \lambda_1) = (.7, .7, .2, .1)$.

We again fit three models. Model 1 is true model (a heteroscedastic ARMA(1,1) structure of the covariance matrix); Model 2 has an AR(1) structure of the covariance matrix; Model 3 has an MA(1) structure of the covariance matrix. Table 3 shows the mean point estimates of the parameters for the three models as well as standard errors, relative bias, and coverage.

Similar to Study 1, the estimates in Model 1 had smaller biases than those in Models 2 and 3 when there were MAR missingness. Coverage probabilities for all parameters for Model 1 were all approximately 95% whereas coverage for Models 2 and 3 had some substantial undercoverage. The percent relative biases of β_0 and β_2 in Models 2 and 3 were large. The SAD for Model 1 was again the smallest among the three models. In addition, proportion of times Model 1 had the minimum AIC was 1.00.

When there were no missing data, the estimates were essentially unbiased for the three models and results were similar to those in Study 1. However, there was slight overcoverage for Model 2 and significant undercoverage for Model 3 due to the misspecified covariance structure. The SAD for Model 1 was again the smallest. The proportion of times Model 1 had the minimum AIC remained at 1.00.

We performed additional simulations to check the flexibility of ARMACD model (results are in the Web Appendix). We simulated data under an AR(1) structure of covariance matrix without MAR dropout, and then fitting an ARMA(1,1) model. We saw the average point estimates of the mean parameters and the GARP, GMAP, and IV parameters were very close to true values. Type I error for $H_0 : \gamma = 0$ (AR(1) parameter is zero) was zero when AR(1) was the true model. We also conducted the same simulation under a MA(1) structure of covariance matrix and then fitting an ARMA(1,1) model. We saw similar results.

Overall, the simulation results emphasize the importance of correctly modeling the covariance structure on estimation of the mean parameters in the presence of ignorable missing data. In addition, the ARMACD models appear to have reasonable operating characteristics when simpler nested models are the true models. Finally, in the limited simulations, the LRT seems to work well (though a bit conservative) in testing nested ARMACD models.

4. Example

4.1 Lung cancer study

Lung cancer is a leading cause of cancer-related deaths world-wide, with a 5-year survival of less than 15%, because most patients are diagnosed with advanced stage disease. Recently, a longitudinal study on lung cancer was designed as a prospective open-label randomized non-comparative parallel study in a single institution (Kim et al., 2012). A total of 96 patients with lung cancer were randomly assigned to one of two treatments (GEFITINIB or ERLOTINIB). The main goal of the study was to evaluate the response rate (the percentage of patients whose cancer shrinks or disappears after treatment) for each arm. The overall response rates in the GEFITINIB and ERLOTINIB arms were 47.9% (95% CI, 33.8-62.0%) and 39.6% (95% CI, 25.7-53.4%), respectively (Kim et al., 2012); these were not significantly different. Given the non-significant treatment effect on the response rate, it was of interest to see if there was a negative impact of treatments on a patient's quality of life (QOL).

The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Lung Cancer (EORTC QLQ-LC13) is considered as a standard instrument to

assess the quality of life of lung cancer patients. The EORTC QLQ-LC13 is a 13-item lung cancer-specific questionnaire module comprising both multi-item and single-item measures of lung cancer-associated symptoms (i.e. coughing, haemoptysis, dyspnoea and pain) and side-effects from conventional chemo- and radiotherapy (i.e. hair loss, neuropathy, sore mouth and dysphagia). Patients completed the questionnaire before receiving the first dose of treatment at baseline, on day 1 of each subsequent 29-day cycle (4 weeks), and at the end of the study. Each questionnaire was originally scored on a 4-point scale ranging from normal, 'Not at All' (1) to abnormal, 'Very Much' (4). In this paper, we focus on sum of the 13-item questionnaire module (SQOL).

Covariates considered included type of treatment (ARM=1 for GEFITINIB, 0 for ERLOTINIB) and week number, TIME (0, 4, 8, ⋯, 92); TIME was re-scaled by dividing by 10. We assumed the missing responses (mostly due to dropout) were MAR in these analyses. The Akaike Information Criteria (AIC) (Akaike, 1974) and likelihood ratio tests were used as the model selection criteria.

4.2 Model Fit

We explore the covariance structure using the procedure in Section 2.4. From Step 1, we graph the GARPs versus lag (Figure 1). We also calculated means of GARPs by lag (Table 1). These suggest that the GARPs decreased approximately quadratically in lag. So we considered a quadratic structure in lag for GARPs. We also draw the plot for GMAPs from Step 2 (Figure 2) and calculated the means of GMAPs by lag (Table 1). These suggest that the GMAPs also decreased roughly quadratically in lag. Finally, we calculated IVs from Step 3 and examined the plot of IVs versus time (Figure 3). It suggested a quadratic trend in time.

Based on the exploratory analysis described in the previous paragraph, we considered various covariance structures. In particular, we fit the five models for Σ_i specified in Table 4. We use the following notation: ARMA-??? where the '?' correspond to the polynomial in the GARP, GMAP, and IV respectively or the AR and MAR order, respectively. So an ARMA-QQQ would correspond to a quadratic in lag for the GARP and GMAP and quadratic in time for the IV. And an ARMA-20Q would correspond to an AR(2) model with a quadratic in time for the IV.

We used R version 3.1.2 to fit the models and R code is available in supplementary materials. We initialize the fitting algorithm by setting $\Sigma_i = \sigma^2 I$ and using ordinary linear model estimation methods using R to obtain $\beta^{(0)}$. We choose $\alpha^{(0)} = 0$, $\gamma^{(0)} = 0$, and $\lambda^{(0)} = 0$ (i.e., $\Sigma_i = I$) as starting values for α, γ, and λ, respectively. We iterate until the sum of absolute differences in the parameters is less than 10^{-4}. Computational burden was not high. Computing time on a Intel Core i5-2450M CPU with a 2.50 GHz processor took about 20, 60, 300 seconds for the ARMA-20Q (ARMA-02Q), the ARMA-11Q, ARMA-QQQ, respectively.

4.3 Results

Maximized loglikelihoods and AICs are presented in Table 5. The model ARMA-QQQ fit better than ARMA-QCQ using LRTs (p-value=0.655). The AIC for the model ARMA-11Q was the smallest of the models considered and is the model we use for inference.

Table 6 presents maximum likelihood estimates for models ARMA-20Q, ARMA-02Q, ARMA-11Q, ARMA-QQQ, and ARMA-QCQ. In the model ARMA-11Q, the parameter of GARP was not significant ($\hat{\alpha_0}$ = −0.03, SE= 0.02, p-value= 0.134) and that of GMAP was significant ($\hat{\gamma_0}$ = 0.52, SE= 0.04, p-value< 0.001). The parameters for the linear and the quadratic of time were significant ($\hat{\lambda_1}$ = –0.43, SE= 0.08, p-value< 0.001; $\hat{\lambda_2}$ = 0.05, SE= 0.01, p-value< 0.001). It indicates that log(IV) decreased quadratically over time. The parameters of coefficients of covariates in the mean model were not significant. Thus treatment did not appear to differentially impact the mean. Note that there were fairly big differences in the estimated coefficients of covariates in the mean models in Table 6. The estimated parameters for the effect of treatment $\hat{\beta}_1$ in ARMA-02Q, ARMA-QQQ, and ARMA-QCQ were noticeably different from those in the other models. This is consistent with the simulations that show that a misspecified covariance matrix results in biased estimates of parameters in mean model when there is missing data.

Figure 4 presents the fitted expected values and their 95% confidence intervals under the model ARMA-11Q. Since two confidence intervals were overlapping, there was not much evidence of any difference between the two arms.

5. Conclusions

We have proposed linear models with a covariance matrix that is modeled using the parameters of a new ARMA Cholesky decomposition (ARMACD). These models allow for nonstationary and heteroscedasticity and are more flexible (and potentially more parsimonious) than models using only the parameters of the Modified Cholesky decomposition or Moving Average Cholesky decomposition.

Simulation studies showed the importance of correctly modeling the covariance structure on mean parameters in the presence of ignorable missing data. We also confirmed that our proposed model worked well even when the true data were generated from models with a simpler structure.

Analysis of lung cancer data shows that the covariance matrix was homoscedastic with the ARMA(1,1) structure having innovation variance parameters quadratic in time. There was no significant difference of quality of life between two treatment arms. Fairly big differences in the estimated covariates effects in the mean models indicate that a misspecified covariance matrix results in biased estimates of parameters in mean model.

Future work includes exploring robustness of inference to non-Gaussian continuous outcomes, automated model selection with this class of models, and extensions to non-continuous data by using this class of models for the random effects covariance matrix in generalized linear mixed models (GLMMs).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Drs. Myung-Ju Ahn and Insuk Sohn of Samsung Medical Center in South Korea for providing the data and for their help in data collection and clarifying some issues with data. This project was supported by Basic Science Research Program through the National Research Foundation of Korea (KRF) funded by the Ministry of Education, Science and Technology (NRF-2014R1A1A2054997, NRF-2016R1D1A1B03930343). MJD was partially supported by US NIH grants CA85295 and CA183854.

Appendix

Proof of Theorem

1. The derivatives of (6) with respect to α, γ, and λ, respectively, have the following forms.

$$\begin{aligned}\frac{\partial^2 \log L(\theta;y_i)}{\partial \alpha_j \partial \beta^T} &= X_i^T \frac{\partial}{\partial \alpha_j} \sum_i^{-1} (y_i - \mu_i(\beta)), \text{for } j=1,\cdots,a;\\ \frac{\partial^2 \log L(\theta;y_i)}{\partial \gamma_j \partial \beta^T} &= X_i^T \frac{\partial}{\partial \gamma_j} \sum_i^{-1} (y_i - \mu_i(\beta)), \text{for } j=1,\cdots,b;\\ \frac{\partial^2 \log L(\theta;y_i)}{\partial \lambda_j \partial \beta^T} &= X_i^T \frac{\partial}{\partial \lambda_j} \sum_i^{-1} (y_i - \mu_i(\beta)), \text{for } j=1,\cdots,c.\end{aligned}$$

The expectations of each of these is clearly zero since $E[Y_i - \mu_i(\beta)] = 0$.

To examine the other partial derivatives, we can rewrite the log likelihood as follows,

$$\log L(\theta;y) = \left\{ \sum_{i=1}^{N} - \frac{n_i}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{n_i} h_{it}^T \lambda - \frac{1}{2} - \frac{1}{2} e_i^T D^{-1} e_i \right\}.$$

Using this form, we take a derivative with respect to γ_j for $j = 1, \cdots, c$ to obtain

$$\frac{\partial \log L(\theta;y)}{\partial \gamma_j} = - \sum_{i=1}^{N} \frac{\partial e_i^T}{\partial \gamma_j} D_i^{-1} e_i,$$

where $\frac{\partial e_i^T}{\partial \gamma_j}$ is a vector with $\frac{\partial e_{it}}{\partial \gamma_j} = - \sum_{k=1}^{t-1} (z_{itk} e_{ik} + l_{itk} \frac{\partial e_{ik}}{\partial \gamma_j})$ for $t = 1, \cdots, n_i$. Now taking another derivative with respect to $\lambda_{j'}$, we obtain

$$\frac{\partial^2 \log L(\theta;y)}{\partial \lambda_{j'} \partial \gamma_j} = - \sum_{i=1}^{N} \frac{\partial e_i^T}{\partial \gamma_j} \frac{\partial}{\partial \lambda_{j'}} D_i^{-1} e_i = - \sum_{i=1}^{N} \sum_{t=1}^{n_i} \frac{\partial e_{it}}{\partial \gamma_j} \frac{\partial}{\partial \lambda_{j'}} \left(\frac{1}{\sigma_{it}^2} \right) e_{it}.$$

Since $\frac{\partial e_{it}}{\partial \gamma_j}$ consists of e_{ik} for $k = 1, \cdots, t-1$, e_{it} and $\frac{\partial e_{it}}{\partial \gamma_j}$ are independent. Thus, $E\left(-\frac{\partial^2 \log L(\theta;y)}{\partial \lambda_{j'} \partial \gamma_j}\right) = 0$.

Finally, we have

$$\frac{\partial^2 \log L}{\partial \lambda_j \partial \alpha_{j'}} = -\frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\left((y_i - \mu_i(\beta))(y_i - \mu_i(\beta))^T \frac{\partial^2}{\partial \lambda_j \partial \alpha_{j'}} \sum^{-1}\right).$$

Since $L_i^{-1} T_i$ is a lower triangular matrix and D_i is a diagonal matrix, by Theorem 1 in Pourahmadi (2007), $E\left(-\frac{\partial^2 \log L}{\partial \lambda_j, \partial \alpha_{j'}}\right) = 0$.

2. Let $f(y_i; \theta)$ is the normal pdf of Y_i and let $l_i = \log f(y_i; \theta)$ for $i = 1, \cdots, N$. Then,

$$l_i = -\frac{n_i}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n_i} h_{it}^T \lambda - \frac{1}{2}(y_i - x_i\beta)^T \sum_i^{-1} (y_i - x_i\beta).$$

When $\theta = \theta_0$, the mean and variance of l_i are respectively

$$\begin{aligned}
E_0(l_i) &= -\tfrac{n_i}{2}\log(2\pi) - \tfrac{1}{2}E_0\left\{\sum_{t=1}^{n_i} h_{it}^T\lambda - \tfrac{1}{2}(y_i - x_i\beta)^T\sum_i^{-1}(y_i - x_i\beta)\right\} \\
&= -\tfrac{n_i}{2}\log(2\pi) - \tfrac{1}{2}E_0\sum_{t=1}^{n_i} h_{it}^T\lambda - \tfrac{1}{2}E\{(y_i - x_i\beta_0)^T\sum_i^{-1}(y_i - x_i\beta_0)\} - \tfrac{1}{2}(x_i\beta - x_i\beta_0)^T\sum_i^{-1}(x_i\beta - x_i\beta_0) \\
&= -\tfrac{n_i}{2}\log(2\pi) - \tfrac{1}{2}E_0\sum_{t=1}^{n_i} h_{it}^T\lambda - \tfrac{1}{2}\mathrm{tr}(\textstyle\sum_i^{-1}\sum_{0i}) - \tfrac{1}{2}(x_i\beta - x_i\beta_0)^T\sum_i^{-1}(x_i\beta - x_i\beta_0), \\
\mathrm{var}_0(l_i) &= \tfrac{1}{4}\mathrm{var}_0\{(y_i - x_i\beta)^T\sum_i^{-1}(y_i - x_i\beta)\} \\
&= \tfrac{1}{2}\mathrm{tr}\left\{(\textstyle\sum_i^{-1}\sum_{0i})^2\right\} + (x_i\beta - x_i\beta_0)^T\sum_i^{-1}\sum_{0i}\sum_i^{-1}(x_i\beta - x_i\beta_0),
\end{aligned}$$

where $\sum_i = T_i^{-1} L_i D_i L_i^T L_i^{-T}$ and $\Sigma_{0i} = \Sigma_i(\theta_0)$. It follows from the compactness of the parameter space in (C2) and boundedness of the covariates in (C3) that $var_0(l_i) \le K$ for all i where K is a finite constant; here, we also need a fixed $\max n_i$ in (C1) to satisfy Lemma 2 in Chiu et al. (1996). By Kolmogorov's strong law of large numbers, we have that

$$\frac{1}{N}\sum_{i=1}^{N} l_i - \frac{1}{N}\sum_{i=1}^{N} E_0(l_i) \to 0$$

almost surely. Notice that the above constant K is independent of θ and that

$$-2\sum_{i=1}^{N}E_0(l_i)/N=\frac{1}{N}\left[\sum_{i=1}^{N}n_i\log(2\pi)+\sum_{i=1}^{N}\sum_{t=1}^{n_i}h_{it}^T\lambda+\sum_{i=1}^{N}\mathrm{tr}(\sum_i^{-1}\sum_{0i})+\sum_{i=1}^{N}(x_i\beta-x_i\beta_0)^T\sum_i^{-1}(x_i\beta-x_i\beta_0)\right].$$

From the proof of Theorem 1 in Chiu et al. (1996), it is easy to show the consistency of $\hat{\theta}$ using (C2) and (C3).

3. The proof here is essentially the same as that of Theorem 2 in Chiu et al. (1996). Note that Theorem 2 of Chiu et al. (1996) also relies on results in Roussas (1968), and requires several regularity conditions for asymptotic normality. These are all satisfied from (C2)-(C4).

References

Akaike H. A new look at the Statistical Model Identification. IEEE Transactions Automatic Control. 1974; 19:716–723.

Chiu TYM, Leonard T, Tsui KW. The matrix-logarithm covariance matrix. Journal of the American Statistical Association. 1996; 91:198–210.

Daniels MJ, Pourahmadi M. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. Biometrika. 2002; 89:553–566.

Daniels MJ, Zhao YD. Modelling the random effects covariance matrix in longitudinal data. Statistics in Medicine. 2003; 22:1631–1647. [PubMed: 12720301]

Daniels, MJ., Hogan, JW. Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis. Chapman & Hall/CRC; 2008.

Daniels MJ, Pourahmadi M. Modeling covariance matrices via partial autocorrelations. Journal of Multivariate Analysis. 2009; 100:2352–2363. [PubMed: 20161018]

Diggle, PJ., Heagerty, P., Liang, KY., Zeger, S. Analysis of Longitudinal Data. 2nd. Oxford University Press; 2002.

Hansen PR, Lunde A. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? Journal of Applied Econometrics. 2005; 20:873–889.

Heagerty PJ. Marginally specified logistic-normal models for longitudinal binary data. Biometrics. 1999; 55:688–698. [PubMed: 11314994]

Hoff PD, Niu X. A covariance regression model. Statistica Sinica. 2012; 22:729–753.

Judge, GG., Griffiths, WE., Hill, RC., Lee, TC. The Theory and Practice of Ecomometrics. New York: Wiley; 1980.

Kim ST, Uhm JE, Lee J, Sun J, Sohn I, Kim SW, Jung S, Park Y, Ahn JS, Park K, Ahn M. Randomized phase study of gefitinib versus erlotinib in patients with advanced non-small cell lung cancer who failed previous chemotherapy. Lung Cancer. 2012; 75:82–88. [PubMed: 21684626]

Lee K, Yoo JK, Lee J, Hagan J. Modeling the random effects covariance matrix for the generalized linear mixed models. Computational Statistics & Data Analysis. 2012; 56:1545–1551.

Lee K. Bayesian modeling of random effects covariance matrix for generalized linear mixed models. Communications for Statistical Applications and Methods. 2013; 20:235–240.

Lee K, Sung S. Autoregressive Cholesky factor modeling for marginalized random effects models. Communications for Statistical Applications and Methods. 2014; 21:169–181.

Lee K, Yoo J. Bayesian Cholesky factor models in random effects covariance matrix for generalized linear mixed models. Computational Statistics & Data Analysis. 2014; 80:111–116.

Pan JX, MacKenzie G. On modelling mean-covariance structures in longitudinal studies. Biometrika. 2003; 90:239–244.

Pan JX, MacKenzie G. Regression models for covariance structures in longitudinal studies. Statistical Modelling. 2006; 6:43–57.

Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. Biometrika. 1999; 86:677–690.

Pourahmadi M. Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. Biometrika. 2000; 87:425–435.

Pourahmadi M. Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance-correlation parameters. Biometrika. 2007; 94:1006–1013.

Rochen J, Helms RW. Maximum likelihood estimation for incomplete repeated-measures experiments under an ARMA covariance structure. Biometrics. 1989; 45:207–218. [PubMed: 2655730]

Rochen J. ARMA covariance structures with time heteroscedasticity for repeated measures experiments. Journal of the American Statistical Association. 1992; 87:777–784.

Roussas GG. Asymptotic normality of the maximum likelihood estimate in Markov process. Metrika. 1968; 14:62–70.

Zhang W, Leng C. A moving average Cholesky factor model in covariance modelling for longitudinal data. Biometrika. 2012; 99:141–150.

Zhang W, Leng C, Tang CY. A joint modeling approach for longitudinal studies. Journal of Royal Statistical Society Series B. 2015; 77:219–238.

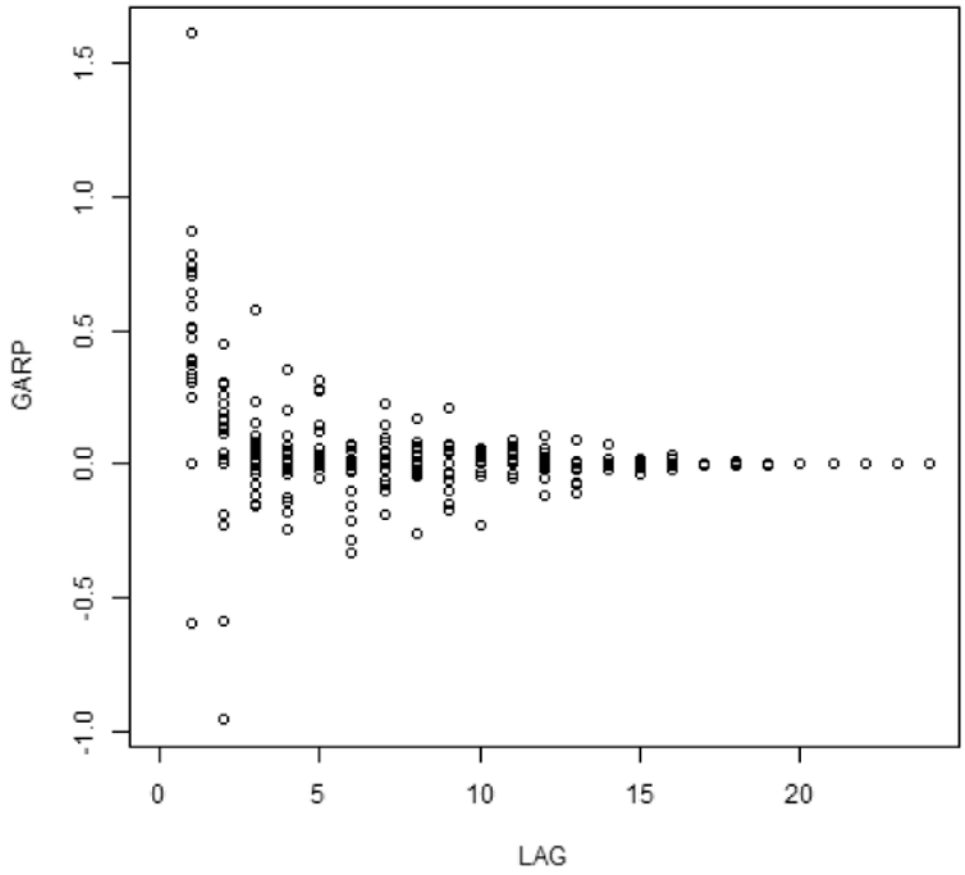

Figure 1.
GARP versus lag for the lung cancer study.

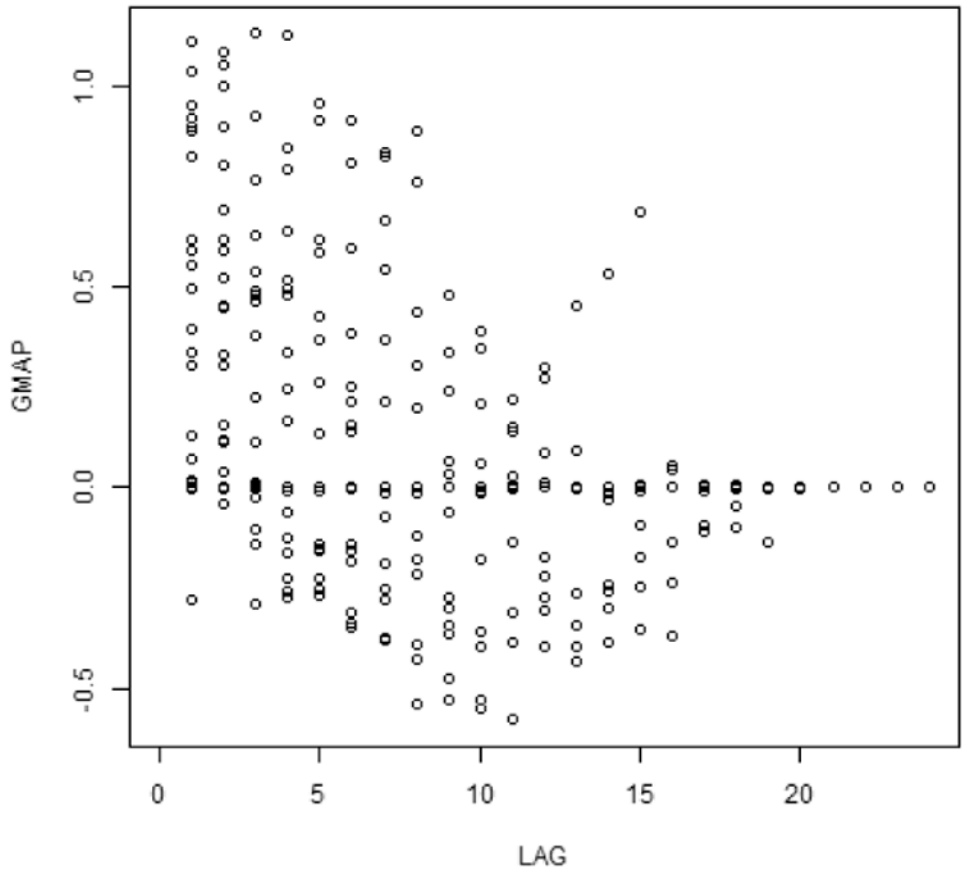

Figure 2.
GMAP versus lag for the lung cancer study.

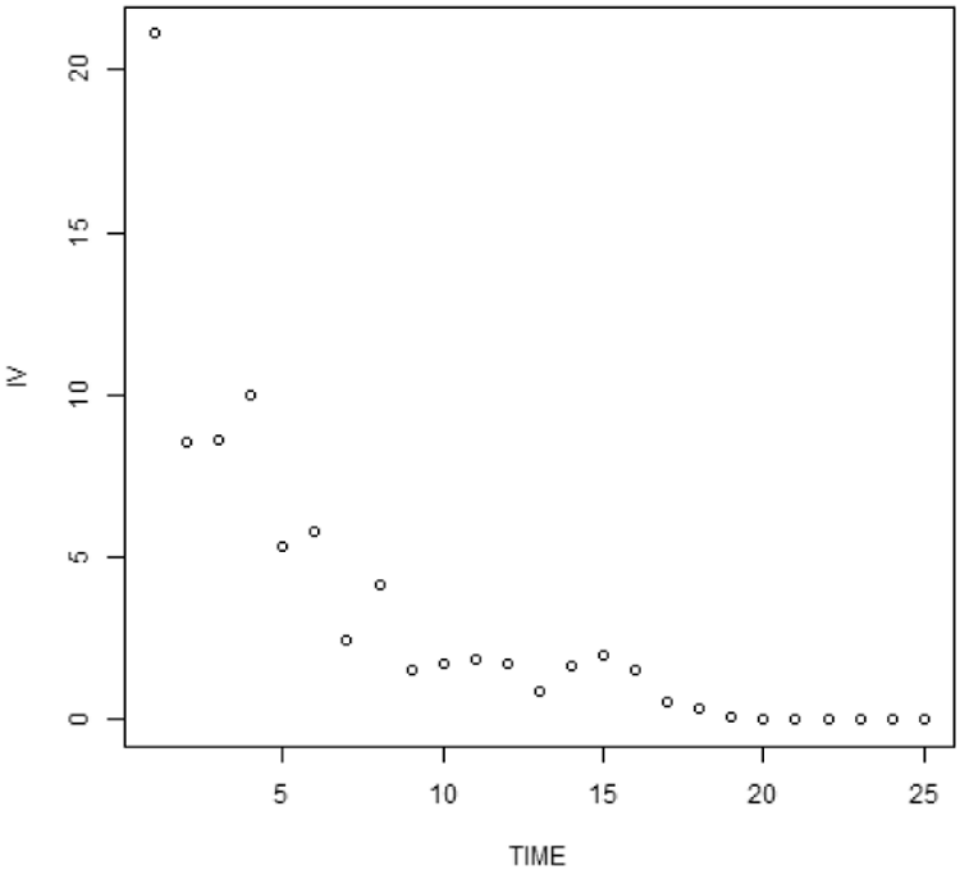

Figure 3.
IV versus Time for the lung cancer study.

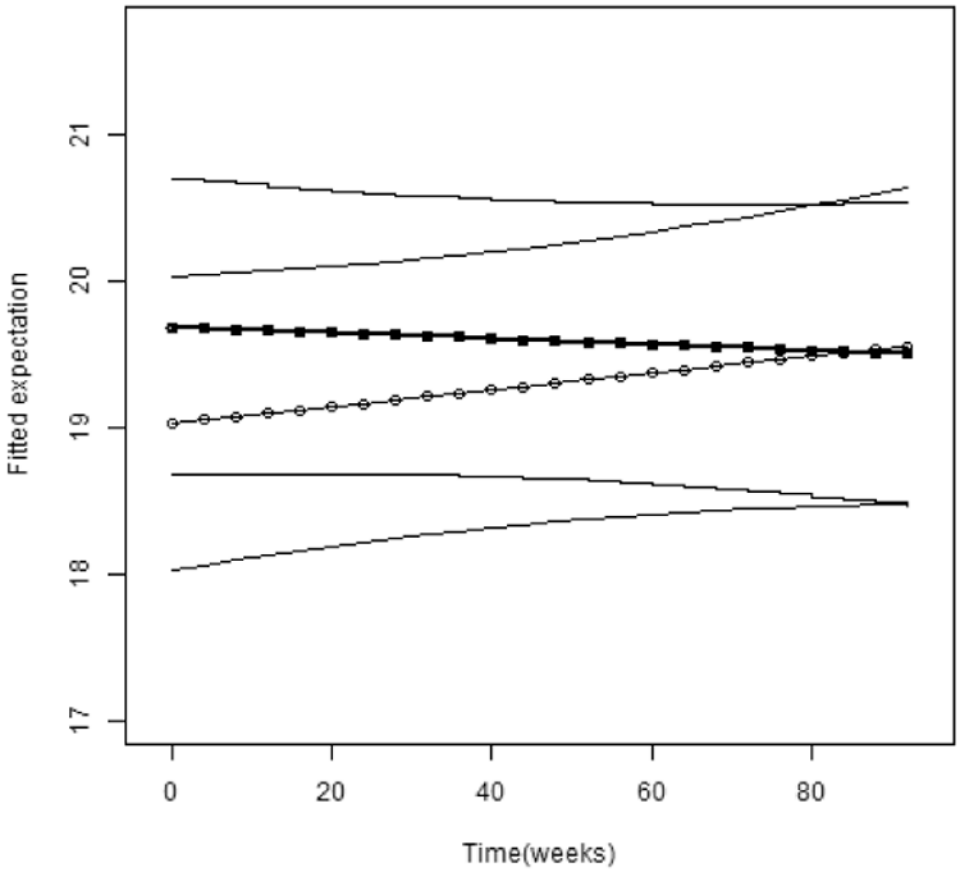

Figure 4.
Fitted expected values and 95% confidence intervals under the two treatment arms (Gefitinib (circle) and Erlotinib (dark square)).

Table 1

Means of GARPs and GMAPs by lag for the lung cancer study.

LAG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
GARP	0.463	0.036	0.040	0.009	0.062	-0.044	0.013	0.003	-0.007	-0.001	0.018	0.001	-0.018	0.004	-0.003	0.002	0.001	0.002	-0.002	0.000	0.000	0.000	0.000	0.000
GMAP	0.449	0.397	0.254	0.215	0.152	0.104	0.104	0.041	-0.075	-0.093	-0.084	-0.053	-0.103	-0.063	-0.019	-0.071	-0.025	-0.020	-0.023	-0.001	0.000	0.000	0.000	0.000

Table 2

Simulation results for parameter estimates in Study 1. Bias of Models 1-3 for sample sizes of 200 and $T = 10$ for 500 simulated data sets under mis-specification of the dependence. Displayed is the average regression coefficient estimate (Mean), the percent relative bias (RB) ($(\bar{\beta} - \beta)/\beta \times 100$), the average standard error (SE), the sample standard deviation (SD) of 500 estimates, the coverage probabilities (COVER), sum of absolute difference (SAD) between the estimated and true covariance matrices, and the proportion of times that AIC chose the specified model as best.

	MAR						Complete					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
	Mean	RB	Mean	RB	Mean	RB	Mean	RB	Mean	RB	Mean	RB
	SE_{SD}	COVER	SE_{SD}	COVER	SE_{SD}	COVER	SE_{SD}	COVER	SE_{SD}	COVER	SE_{SD}	COVER
β_0 (0.1)	0.15	50.0	0.23	130.0	0.06	−40.0	0.10	0.0	0.10	0.0	0.11	10.0
	$0.13_{0.12}$	0.95	$0.16_{0.14}$	0.90	$0.08_{0.15}$	0.71	$0.10_{0.10}$	0.95	$0.14_{0.12}$	0.97	$0.07_{0.13}$	0.76
β_1 (−0.1)	−0.10	0.0	−0.10	0.0	−0.09	−10.0	−0.10	0.0	−0.10	0.0	−0.10	10.0
	$0.18_{0.18}$	0.95	$0.23_{0.20}$	0.98	$0.12_{0.21}$	0.71	$0.14_{0.14}$	0.95	$0.19_{0.17}$	0.97	$0.10_{0.18}$	0.73
β_2 (0.1)	0.11	10.0	0.13	30.0	0.08	−20.0	0.10	0.0	0.10	0.0	0.10	0.0
	$0.03_{0.02}$	0.95	$0.03_{0.03}$	0.84	$0.02_{0.03}$	0.82	$0.01_{0.02}$	0.94	$0.02_{0.02}$	0.96	$0.01_{0.02}$	0.86
β_3 (0.1)	0.10	0.0	0.10	0.0	0.10	0.0	0.10	0.0	0.10	0.0	0.10	0.0
	$0.04_{0.03}$	0.96	$0.05_{0.04}$	0.96	$0.03_{0.04}$	0.81	$0.02_{0.02}$	0.94	$0.03_{0.03}$	0.98	$0.02_{0.03}$	0.86
α (0.7)	0.71	1.4	0.89	27.1			0.70	0.0	0.87	24.3		
	$0.02_{0.02}$	094	$0.02_{0.01}$	0.00			$0.02_{0.02}$	095	$0.01_{0.01}$	0.00		
γ (0.7)	0.70	0.0			0.96	37.1	0.70	0.0			0.95	35.7
	$0.02_{0.02}$	0.95			$0.01_{0.01}$	0.000	$0.02_{0.02}$	0.96			$0.01_{0.01}$	0.000
λ (0.3)	0.30	0.0	0.57	90.0	0.71	136.7	0.30	0.0	0.58	93.3	0.72	140.0
	$0.03_{0.04}$	0.94	$0.03_{0.04}$	0.00	$0.03_{0.05}$	0.00	$0.03_{0.03}$	0.97	$0.03_{0.04}$	0.00	$0.03_{0.05}$	0.00
SAD	12.89		52.16		104.18		10.31		35.27		103.98	
P(AIC)	1.00		0.00		0.00		1.00		0.00		0.00	

Table 3

Simulation results for parameter estimates in Study 2. Bias of Models 1-3 for sample sizes of 200 and $T = 10$ for 500 simulated data sets under mis-specification of the dependence. Displayed is the average regression coefficient estimate (Mean), the percent relative bias (RB) ($(\bar{\beta} - \beta)/\beta \times 100$), the average standard error (SE), the sample standard deviation (SD) of 500 estimates, the coverage probabilities (COVERAGE), sum of absolute difference (SAD) between the estimated and true covariance matrices, and the proportion of times that AIC chose the specified model as best.

	MAR						Complete					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
	Mean	RB	Mean	RB	Mean	RB	Mean	RB	Mean	RB	Mean	RB
	SE_{SD}	COVER	SE_{SD}	COVER	SE_{SD}	COVER	SE_{SD}	COVER	SE_{SD}	COVER	SE_{SD}	COVER
β_0 (0.1)	0.14	40.0	0.22	120.0	0.06	−40.0	0.10	0.0	0.10	0.0	0.10	0.0
	$0.11_{0.11}$	0.94	$0.15_{0.13}$	0.91	$0.08_{0.13}$	0.73	$0.09_{0.09}$	0.96	$0.13_{0.11}$	0.96	$0.07_{0.12}$	0.60
β_1 (−0.1)	−0.09	−10.0	−0.09	−10.0	−0.09	−10.0	−0.10	0.0	−0.10	0.0	−0.11	10.0
	$0.17_{0.17}$	0.96	$0.22_{0.19}$	0.98	$0.11_{0.20}$	0.75	$0.13_{0.14}$	0.96	$0.19_{0.17}$	0.98	$0.10_{0.18}$	0.72
β_2 (0.1)	0.11	10.0	0.13	30.0	0.08	−20.0	0.10	0.0	0.10	0.0	0.10	0.0
	$0.02_{0.02}$	0.93	$0.03_{0.03}$	0.85	$0.02_{0.03}$	0.80	$0.01_{0.01}$	0.97	$0.02_{0.02}$	0.97	$0.01_{0.02}$	0.84
β_3 (0.1)	0.10	0.0	0.10	0.0	0.10	0.0	0.10	0.0	0.10	0.0	0.10	0.0
	$0.04_{0.04}$	0.95	$0.05_{0.04}$	0.96	$0.03_{0.04}$	0.82	$0.02_{0.02}$	0.95	$0.03_{0.03}$	0.97	$0.02_{0.02}$	0.87
α (0.7)	0.71	1.4	0.89	27.1			0.70	0.0	0.87	24.3		
	$0.02_{0.02}$	0.95	$0.02_{0.01}$	0.00			$0.02_{0.02}$	0.95	$0.01_{0.01}$	0.00		
γ (0.7)	0.70	0.0			0.96	37.1	0.70	0.0			0.97	38.6
	$0.02_{0.02}$	0.94			$0.01_{0.01}$	0.00	$0.02_{0.02}$	0.93			$0.01_{0.01}$	0.00
λ_0 (0.2)	0.20	0.0	0.47	135.0	0.60	200.0	0.20	0.0	0.48	140.0	0.62	210.0
	$0.05_{0.05}$	0.94	$0.05_{0.06}$	0.00	$0.05_{0.07}$	0.00	$0.04_{0.04}$	0.95	$0.04_{0.05}$	0.00	$0.04_{0.06}$	0.00
λ_1 (0.1)	0.10	0.0	0.10	0.0	0.10	0.0	0.10	0.0	0.10	0.0	0.11	0.0
	$0.07_{0.07}$	0.97	$0.07_{0.08}$	0.93	$0.07_{0.10}$	0.83	$0.06_{0.06}$	0.97	$0.06_{0.07}$	0.92	$0.06_{0.09}$	0.85
SAD	12.47		46.62		99.40		9.94		39.14		99.30	
P(AIC)	1.00		0.00		0.00		1.00		0.00		0.00	

Table 4

Models for ϕ_{itj}, I_{itj}, and $\log\sigma_{it}$ in the lung cancer example.

Model	GARP	GMAP	IV
ARMA-20Q	$\phi_{itj} = \alpha_0 I_{(\lvert t-j\rvert=1)} + \alpha_1 I_{(\lvert t-j\rvert=2)}$		$\log\sigma_{it}^2=\lambda_0+\lambda_1 t+\lambda_2 t^2$
ARMA-02Q		$I_{itj} = \alpha_0 I_{(\lvert t-j\rvert=1)} + \alpha_1 I_{(\lvert t-j\rvert=2)}$	$\log\sigma_{it}^2=\lambda_0+\lambda_1 t+\lambda_2 t^2$
ARMA-11Q	$\phi_{itj} = \alpha_0 I_{(\lvert t-j\rvert=1)}$	$I_{itj} = \gamma_0 I_{(\lvert t-j\rvert=1)}$	$\log\sigma_{it}^2=\lambda_0+\lambda_1 t+\lambda_2 t^2$
ARMA-QQQ	$\phi_{itj} = \alpha_0 + \alpha_1\lvert t-j\rvert + \alpha_2\lvert t-j\rvert^2$	$I_{itj} = \gamma_0 + \gamma_1\lvert t-j\rvert + \gamma_2\lvert t-j\rvert^2$	$\log\sigma_{it}^2=\lambda_0+\lambda_1 t+\lambda_2 t^2$
ARMA-QCQ	$\phi_{itj} = \alpha_0 + \alpha_1\lvert t-j\rvert + \alpha_2\lvert t-j\rvert^2$	$I_{itj} = \gamma_0 + \gamma_1\lvert t-j\rvert + \gamma_2\lvert t-j\rvert^2 + \gamma_3\lvert t-j\rvert^3$	$\log\sigma_{it}^2=\lambda_0+\lambda_1 t+\lambda_2 t^2$

Table 5

Maximized loglikelihoods and AICs for the models considered. edit model ‘labels’ based on my suggestion in section 4

Model	ARMA-20Q	ARMA-02Q	ARMA-11Q	ARMA-QQQ	ARMA-QCQ
Max. loglik.	-2040.5	-2129.2	-2034.5	-2032.8	-2032.7
AIC	4099.0	4276.4	4087.0	4091.6	4093.4

Table 6

Maximum likelihood estimates for models ARMA-20Q, ARMA-02Q, ARMA11Q, ARMA-QQQ, and ARMA-QCQ fit to the lung cancer data. Parameter estimates with standard errors in the parentheses.

	AR-20Q	ARMA-02Q	ARMA-11Q	ARMA-QQQ	ARMA-QCQ
β_0 (Int.)	19.76* (0.54)	19.55* (0.46)	19.69* (0.52)	19.69* (0.51)	19.71* (0.51)
β_1 (Arm)	−0.70 (0.76)	−0.86 (0.66)	−0.66 (0.73)	−0.74 (0.71)	−0.75 (0.72)
β_2 (Time)	−0.14 (0.18)	−0.23 (0.15)	−0.08 (0.17)	−0.10 (0.18)	−0.10 (0.18)
β_3 (Arm × Time)	0.37 (0.27)	0.43 (0.22)	0.31 (0.26)	0.31 (0.26)	0.31 (0.26)
α_0	0.54 (0.03)		−0.03 (0.02)	−0.23 (0.17)	−0.20 (0.19)
α_1				0.17 (0.09)	0.07 (0.32)
α_2				0.00 (0.01)	0.03 (0.08)
γ_0	0.33* (0.04)	0.57* (0.03)	0.52* (0.04)	0.65* (0.11)	0.63* (0.12)
γ_1		0.45* (0.04)		0.05 (0.18)	0.10 (0.25)
γ_2				−0.09 (0.06)	−0.05 (0.11)
γ_3					−0.01 (0.03)
λ_0	2.79* (0.10)	2.79* (0.10)	2.81* (0.10)	2.80* (0.10)	2.80* (0.10)
λ_1	−0.41* (0.08)	−0.26* (0.08)	−0.43* (0.08)	−0.42* (0.08)	−0.42* (0.08)
λ_2	0.05* (0.01)	0.03* (0.01)	0.05* (0.01)	0.05* (0.01)	0.05* (0.01)

* indicates significance with 95% confidence level.