LINEAR MODELS WITH A GENERALIZED AR(1) COVARIANCE STRUCTURE FOR LONGITUDINAL AND SPATIAL DATA

Sean L. Simpson

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics, School of Public Health.

Chapel Hill
2008

Approved by:

Advisor: Lloyd J. Edwards

Reader: Keith E. Muller

Reader: Pranab K. Sen

Reader: Paul W. Stewart

Reader: Martin A. Styner

© 2008
Sean L. Simpson
ALL RIGHTS RESERVED

ABSTRACT

SEAN L. SIMPSON. Linear Models with a Generalized AR(1) Covariance Structure for Longitudinal and Spatial Data.
(Under the direction of Dr. Lloyd J. Edwards)

Cross-sectional and longitudinal imaging studies are moving increasingly to the forefront of medical research due to their ability to characterize spatial and spatiotemporal features of biological structures across the lifespan. With Gaussian data, such designs require the general linear model for repeated measures data when standard multivariate techniques do not apply. A key advantage of this model lies in the flexibility of modeling the covariance of the outcome as well as the mean. Proper specification of the covariance model can be essential for the accurate estimation of and inference about the means and covariates of interest.

Many repeated measures settings have within-subject correlation decreasing exponentially in time or space. Even though observed correlations often decay at a much slower or much faster rate than the AR(1) structure dictates, it sees the most use among the variety of correlation patterns available. A three-parameter generalization of the continuous-time AR(1) structure, termed the *generalized autoregressive* (GAR) covariance structure, accommodates much slower and much faster correlation decay patterns. Special cases of the GAR model include the AR(1) and equal correlation (as in compound symmetry) models. The flexibility achieved with three parameters makes the GAR structure especially attractive for the High Dimension, Low Sample Size case so common in medical imaging and various kinds of "-omics" data. Excellent analytic and numerical properties help make the GAR model a valuable addition to the suite of parsimonious covariance structures for repeated measures data.

The accuracy of inference about the parameters of the GAR model in a moderately large sample context is assessed. The GAR covariance model is shown to be far more robust to misspecification in controlling fixed effect test size than the AR(1) model. It is as robust to misspecification as another comparable model, the *damped exponential* (DE), while possessing better statistical and convergence properties.

The GAR model is extended to the multivariate repeated measures context via the development of a Kronecker product GAR covariance structure. This structure allows modeling data in which the correlation between measurements for a given subject is induced by two factors (e.g., spatio-temporal data). A key advantage of the model lies in the ease of interpretation in terms of the independent contribution of every repeated factor to the overall within-subject covariance matrix. The proposed model allows for an imbalance in both dimensions across subjects.

Analyses of cross-sectional and longitudinal imaging data as well as strictly longitudinal data demonstrate the benefits of the proposed models. Simulation studies further illustrate the advantages of the methods. The demonstrated appeal of the models make it important to pursue a variety of unanswered questions, especially in the areas of small sample properties and covariance model robustness.

To Lorenzo, Jr. and Adelaide, for past enablements and present encouragement; to Lorenzo, Sr. and Bessie, for modeling commitment; to Gail for exemplifying selflessness and perseverance; to Marsha and Dennis for unmitigated support; and to Eva with a future in the making.

ACKNOWLEDGMENTS

I must first thank my family whose love, support, and guidance have made my successes possible and my obstacles surmountable. I am eternally indebted to my parents for their selfless devotion to nurturing my interests and abilities.

I am extremely fortunate to have had Lloyd Edwards as my dissertation advisor. His wisdom and insight have been immeasurable in facilitating my growth as a researcher and preparing me for an academic career in biostatistics. His genuine advice and concern as a mentor throughout my graduate career is a rarity for which I am immensely grateful. I owe much gratitude to Keith Muller who also served as an advisor and mentor during my time at Chapel Hill. He has been instrumental in fostering my development as a collaborator and methodologist. I must also thank the other members of my committee: Pranab Sen, Paul Stewart, and Martin Styner. They have each been more than willing to provide guidance and insight as necessary. I thank Dale Zimmerman from the Department of Statistics and Actuarial Science at the University of Iowa for the observation that the model can be reparameterized to give further insight into its properties. I am also thankful to Chirayath Suchindran for giving me the opportunity to serve as a trainee on NICHHD Training Grant #T32-HD 007237 ("Research Training in Population Statistics").

TABLE OF CONTENTS

List Of Tables

List Of Figures

CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Repeated measures studies are commonly utilized in biomedical research. Such data allow examining how particular correlated outcomes vary over time or in space. In longitudinal studies, measurements are taken on the same subjects at various points in time. For instance, Rijcken et al. (1987) was interested in identifying risk factors for pulmonary function loss. Therefore they recorded many baseline characteristics thought to be possible risk factors, and attempted to measure FEV_1 (forced expiratory volume in 1 second) values every 3 years for 21 years to assess pulmonary function loss in terms of the rate of decline.

In spatial studies, and more particularly in medical imaging studies, measurements are taken on the same subjects at various points in space. The image of an organ in a given subject is often characterized by the correlated (repeated) values resulting from the mathematical parameterization of its shape. For example, Gerig et al. (2003) analyzed changes of the hippocampal structure in schizophrenics as compared to matched controls via MRI.

As technological advances continually reduce the financial and logistic obstacles associated with the use of imaging equipment, imaging studies are becoming increasingly important in medical research. This increasing importance was made evident by the creation of the National Institute of Biomedical Imaging and Bioengineering at the National Institutes of Health in the United States in September 2000. Medical imaging

research units and centers have also been established by several universities in recent years.

Current imaging research often utilizes a third hybrid type of repeated measures data, namely spatio-temporal, that has both spatial and longitudinal components. The goal of these studies is to analyze images taken over time. For instance, much of the current Autism research involves examining the development of children's brains (via neuroimaging) over time. Due to the frequency of the three types of repeated measures data, the development of accurate analytic techniques for these studies is of immense importance.

In repeated measures data, both the mean and the covariance are modeled. Though modeling the mean is the primary focus of many studies, modeling the covariance structure is one of the most fundamental and important concerns in the analysis of these data. As noted by Louis (1988), there is a tradeoff between including additional covariates in the mean model and increasing the complexity of the covariance structure. In other words, the covariance structure may be able to account for the effects of unmeasured covariates. Properly specifying the covariance model leads to more accurate estimation of and inference about the covariates of interest. Having an accurately specified covariance structure also allows for a better understanding of the biological process under investigation.

While much work has been done on estimation, inference, and diagnostics of the mean model (fixed effects), relatively little has been done in these areas for the covariance model. Proper specification of the covariance model can be essential for the accurate estimation of and inference about the means and covariates of interest. Muller et al. (2007) showed that there can be severe test size inflation in fixed effect inference if the covariance structure is badly misspecified. This heavy dependence of fixed-effects inference accuracy on the proper specification of the covariance model indicates that the

amount of work on covariance models has not been commensurate with their level of importance.

Many repeated measures settings have within-subject correlation decreasing exponentially in time or space. Even though observed correlations often decay at a much slower or much faster rate than the AR(1) structure dictates, it sees the most use among the variety of correlation patterns available. In this dissertation I will propose a three-parameter covariance model that is a generalization of the continuous-time AR(1) structure which accommodates much slower and much faster correlation decay patterns.

1.2 Repeated Measures Model

Consider the following general linear model for repeated measures data:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}_i \tag{1.1}$$

where $\boldsymbol{y}_i$ is a $p_i \times 1$ vector of p_i observations on the i^{th} subject $i = 1, ..., N$, $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed and unknown population parameters, $\boldsymbol{X}_i$ is a $p_i \times q$ fixed and known design matrix corresponding to the fixed effects, and $\boldsymbol{e}_i$ is a $p_i \times 1$ vector of random error terms. We assume $\boldsymbol{e}_i \sim N_{p_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e))$ and is independent of $\boldsymbol{e}_{i'}$ for $i \neq i'$. It follows that $\boldsymbol{y}_i \sim N_{p_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e))$ and is independent of $\boldsymbol{y}_{i'}$ for $i \neq i'$. We also assume that $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e)$ is a $p_i \times p_i$ positive-definite symmetric covariance matrix whose elements are twice differentiable functions of a finite number of fixed, unknown parameters $\boldsymbol{\tau}_e = \{\boldsymbol{\tau}_{e1}, ..., \boldsymbol{\tau}_{ek}\}$, $\boldsymbol{\tau}_e \in \boldsymbol{T}$, where $\boldsymbol{T}$ is the set of all parameters for which $\boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e)$ is positive-definite. Also, the parameters in $\boldsymbol{\tau}_e$ are functionally independent of those in $\boldsymbol{\beta}$. The model may be abbreviated as $\boldsymbol{y}_i \sim N_{p_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_{ei})$.

The frequently used general linear mixed model, detailed by Laird and Ware (1982), can be viewed as a special case of this model in which $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\tau}) = \boldsymbol{Z}_i\boldsymbol{\Sigma}_{bi}(\boldsymbol{\tau}_b)\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_{ei}(\boldsymbol{\tau}_e)$, which may be abbreviated as $\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i\boldsymbol{\Sigma}_{bi}\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_{ei}$. We have that $\boldsymbol{Z}_i$ is a $p_i \times m$ fixed and known design matrix corresponding to the random effects, $\boldsymbol{b}_i$, $\boldsymbol{\Sigma}_{bi}$ is the $m \times m$ positive-definite symmetric covariance matrix of the random

effects, $\boldsymbol{b}_i$, which are mutually independent, and $\boldsymbol{\Sigma}_{ei}$ is the $p_i \times p_i$ positive-definite symmetric covariance matrix of the random errors, $\boldsymbol{e}_i$, which are mutually independent. It is assumed that $\boldsymbol{b}_i$ is independent of $\boldsymbol{e}_i$.

1.3 Exponentially Decreasing Correlation

1.3.1 Introduction

The continuous-time first-order autoregressive covariance structure, often denoted as AR(1), is the classic model applied to longitudinal and spatial data when the within subject correlation is believed to decrease exponentially in time or space. For longitudinal data this means that the correlation between measurements on a given subject is assumed to decrease exponentially as the time between measurement occasions increases. Analogously, this correlation is assumed to decrease exponentially as the distance between measurement locations increases for spatial data. For $\boldsymbol{\Sigma}_{ei} = \{\sigma_{ei;jk}\}$, the AR(1) covariance structure is

$$\sigma_{ei;jk} = \mathcal{V}(y_{ij}, y_{ik}) = \sigma_e^2 \begin{cases} \rho_e^{\,d(t_{ij}, t_{ik})} & j \neq k \\ 1 & j = k\,, \end{cases} \tag{1.2}$$

where $\mathcal{V}(\,\cdot\,)$ is the covariance operator, $d(t_{ij}, t_{ik})$ is the distance between measurement times or locations, σ_e^2 is the variability of the measurements at each time or location, and ρ_e is the correlation between observations separated by one unit of time or distance. This two-parameter model was briefly discussed in Louis (1988) in an article surveying methods for analyzing repeated measures data. It is a special case of the model presented in Diggle (1988).

The AR(1) covariance model is particularly appealing when there are only a small number of subjects with many observations per subject. It is also able to accommodate *incomplete*, *unbalanced* (within subject), *inconsistently-spaced*, and/or *irregularly-spaced* data which are all common issues in longitudinal studies. These problems may arise from having missing observations due to failure to meet scheduled appointments, unequal

lengths of follow-up due to staggered entry or early withdrawal, and intentional unequal spacing of measurements over time. Irregular-spacing may also be inherent in the biological process being studied. For instance, a longitudinal study of pulmonary function in children with cystic fibrosis may necessarily have this issue. Spatial data, and more specifically medical imaging data, tend to be complete and balanced due to the tight control typical of this area of research. However, irregular-spacing is again an issue with these data that is easily handled with the AR(1) model. For this dissertation, it is assumed that missing data is missing completely at random (MCAR).

Despite the utility of the AR(1) covariance model, there are situations in which it may not be flexible enough to accurately model the correlation pattern induced by repeatedly taking measurements over time or in space. For instance, as noted in Munoz et al. (1992), longitudinal studies in epidemiologic settings tend to have within-subject correlations that decay at a slower rate than that imposed by the AR(1) structure. Conversely, there are also many situations in both longitudinal and imaging data in which these correlations decay at a faster rate. Therefore, developing a more flexible version of the AR(1) model would be extremely beneficial to the scientific community.

1.3.2 AR(1) Generalizations

LaVange and Muller (1992) described a three-parameter generalization of the AR(1) structure as a tool for power analysis in repeated measures studies, but did not discuss any properties or consider estimation. They defined the model in order to be able to generate a realistic suite of possible repeated measures covariance models in terms of only three parameters. The appeal of the model led the author of NQuery® power software to embed the model in the software as a study planning tool.

Munoz et al. (1992) presented a three-parameter generalization of the AR(1) structure, namely the *damped exponential* (DE) structure, which allows for an attenuation or acceleration of the exponential decay rate imposed by the AR(1) structure. Murray

(1990) described the same model in an unpublished dissertation. For $\mathbf{\Sigma}_{ei} = \{\sigma_{ei;jk}\}$, the *damped exponential* (DE) covariance structure is

$$\sigma_{ei;jk} = \mathcal{V}(y_{ij}, y_{ik}) = \sigma_e^2 \begin{cases} \rho_e^{[d(t_{ij}, t_{ik})]^{\nu_e}} & j \neq k \\ 1 & j = k \,, \end{cases} \tag{1.3}$$

where $\mathcal{V}(\,\cdot\,)$ is the covariance operator, $d(t_{ij}, t_{ik})$ is the distance between measurement times or locations, σ_e^2 is the variability of the measurements at each time or location, ρ_e is the correlation between observations separated by one unit of time or distance, and ν_e is the decay speed. They assume $0 \leq \rho_e < 1$ and $0 \leq \nu_e$.

Implicit in this model formulation is the presence of both a stationary variance and correlation structure. The AR(1) covariance model is a special case of this model for which $\nu_e = 1$. For values of $\nu_e > 1$, the correlation between measurements on a given subject decreases in time or space at a faster rate than for $\nu_e = 1$. As $\nu_e \to \infty$, this model approaches the moving average model of order 1, MA(1). For values of ν_e such that $0 < \nu_e < 1$, the correlation between measurements on a given subject decreases in time or space at a slower rate than for $\nu_e = 1$. When $\nu_e = 0$, this model reduces to the well known compound symmetric covariance model for which the correlation between measurements on a given subject is fixed at ρ_e no matter how far apart in time or space the measurements were taken. Though values of $\nu_e < 0$ yield valid autocorrelation functions for which the correlation between measurements on a given subject would increase with increasing time or distance between measurements, this is rare in biostatistical applications. Therefore the parameter space was restricted for reasons of practicality.

Nunez-Anton and Woodworth (1994) also proposed a three-parameter covariance model to deal with *incomplete*, *unbalanced*, and *irregularly-spaced* longitudinal data. For $\mathbf{\Sigma}_{ei} = \{\sigma_{ei;jk}\}$, their covariance structure is

$$\sigma_{ei;jk} = \mathcal{V}(y_{ij}, y_{ik}) = \sigma_e^2 \begin{cases} \rho_e^{(t_{ij}^{\lambda_e} - t_{ik}^{\lambda_e})} & j \neq k, \lambda_e \neq 0 \\ \rho_e^{log(t_{ij}/t_{ik})} & j \neq k, \lambda_e = 0 \\ 1 & j = k\,, \end{cases} \tag{1.4}$$

where $\mathcal{V}(\,\cdot\,)$ is the covariance operator, t_{ij} is the time or location of the j^{th} measurement for subject i, σ_e^2 is the variability of the measurements at each time or location, ρ_e is the correlation between observations separated by one unit of time or distance when $\lambda_e = 1$, and λ_e is the power transformation of time parameter. They assume $0 \leq \rho_e < 1$.

This model differs from DE model in that it involves power transformations of time rather than of time intervals. The AR(1) covariance model is a special case of this model for which $\lambda_e = 1$. The models appeal lies in the fact that the power transformation can produce nonstationary correlation structures for the within-subject errors, with stationary correlation structures as a special case. In other words, the model can accommodate data where the correlation between measurements for a given subject does not just depend on the differences in time or location between two measurements, but also on the absolute time or location of each individual measurement. However, the model lacks the flexibility of the DE model when a stationary structure is assumed.

1.4 Maximum Likelihood Estimation

1.4.1 Overview

The goal of estimation is to estimate the parameters that characterize a given model based on the available data. The basic approach of maximum likelihood (ML) estimation is to find the parameter values that maximize the probability, or "likelihood", that the observations would end up being equal to the observed data. In other words, find the parameter values that are best supported by the observed data. An ML estimate, $\widehat{\boldsymbol{\theta}}$, of a vector of parameters, $\boldsymbol{\theta}$, maximizes the likelihood function of the parameters given the data. That is, if the likelihood function is denoted $L(\boldsymbol{y}; \boldsymbol{\theta})$, then the maximum value of the function is attained at $L(\boldsymbol{y}; \widehat{\boldsymbol{\theta}})$ over all possible values of $\boldsymbol{\theta}$.

Under model 1.1, the set of measurements on a given individual are assumed to follow a multivariate normal distribution with $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\tau}_e\}$, thus the likelihood function of the parameters is given by

$$L(\boldsymbol{y}_i; \boldsymbol{\beta}, \boldsymbol{\tau}_e) = (2\pi)^{-n_i/2}|\boldsymbol{\Sigma}_i(\boldsymbol{\tau}_e)|^{-1/2}\exp\{-(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i(\boldsymbol{\tau}_e)^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/2\}.$$

Since the measurements from different subjects are assumed to be mutually independent, the joint likelihood function of all the measurements is the product of the N individual likelihood functions. That is,

$$\begin{aligned} L(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\tau}_e) &= \prod_{i=1}^{N} L(\boldsymbol{y}_i; \boldsymbol{\beta}, \boldsymbol{\tau}_e) \\ &= \prod_{i=1}^{N} (2\pi)^{-n_i/2}|\boldsymbol{\Sigma}_i(\boldsymbol{\tau}_e)|^{-1/2}\exp\{-(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i(\boldsymbol{\tau}_e)^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})/2\}. \end{aligned}$$

The log-likelihood function, which is commonly used to derive the estimators, is thus

$$\begin{aligned} \ln L &= l(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\tau}_e) \\ &= -\sum_{i=1}^{N} \frac{n_i}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln|\boldsymbol{\Sigma}_i(\boldsymbol{\tau}_e)| - \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i(\boldsymbol{\tau}_e)^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}). \end{aligned} \tag{1.5}$$

The MLE of $\boldsymbol{\beta}$ is then

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{y}_i,$$

where $\widehat{\boldsymbol{\Sigma}}_i$ is the covariance matrix for $\boldsymbol{y}_i$ with the MLE of $\boldsymbol{\tau}_e$, $\widehat{\boldsymbol{\tau}}_e$, inserted. Since it is generally not possible to express $\widehat{\boldsymbol{\tau}}_e$ in a closed form, the solution for $\widehat{\boldsymbol{\beta}}$ given above is not a closed form expression either, and thus their values must be found via numerical algorithms.

1.4.2 Iterative Algorithms

Since the likelihood equations cannot usually be solved analytically for the general linear model for repeated measures data, iterative numerical procedures must be

employed. Many algorithms for this purpose have been proposed; however, the three most commonly used are the Newton-Raphson, Fisher scoring, and EM algorithm. These iterative techniques will be discussed in this section.

1.4.2.1 Newton-Raphson and Fisher Scoring Algorithms

Both the Newton-Raphson and Fisher scoring (also referred to as the method of scoring) algorithms are based on the first and second-order partial derivatives of the log-likelihood function. As detailed in Jennrich and Schluchter (1986) the Newton-Raphson algorithm is an iterative procedure that computes new parameter values from current values, with the $(r + 1)^{st}$ iterate under model 1.1 with $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\tau}_e\}$ being:

$$\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} - \boldsymbol{H}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} \boldsymbol{s}_{\boldsymbol{\theta}} \tag{1.6}$$

where

$$\boldsymbol{H}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{bmatrix} \partial^2 l/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta} & \partial^2 l/\partial\boldsymbol{\beta}\partial\boldsymbol{\tau}_e \\ \partial^2 l/\partial\boldsymbol{\tau}_e\partial\boldsymbol{\beta} & \partial^2 l/\partial\boldsymbol{\tau}_e\partial\boldsymbol{\tau}_e \end{bmatrix}$$

$$\boldsymbol{s}_{\boldsymbol{\theta}} = \begin{bmatrix} \partial l/\partial\boldsymbol{\beta} \\ \partial l/\partial\boldsymbol{\tau}_e \end{bmatrix}$$

which are referred to as the Hessian matrix (or observed information matrix) and score vector respectively. The Fisher scoring algorithm is identical to that of the Newton-Raphson method except that the Hessian matrix is replaced by its expectation. This matrix of expectations is commonly referred to as the expected information matrix.

As noted in Murray (1990) the Fisher scoring method may outperform the Newton-Raphson procedure when the expected information matrix has a block diagonal structure. Jennrich and Schluchter (1986) mentioned that the Fisher scoring method is also preferred when the second derivatives of $\boldsymbol{\Sigma}_i$ are nonzero since the Newton-Raphson algorithm may require substantially more computation. However, the Newton-Raphson method is more generally applicable since calculating the expected information matrix

may not always be computationally feasible. For both methods, constrained optimization must normally be employed since the procedures may converge to maxima outside the boundaries that ensure the positive-definiteness of $\mathbf{\Sigma}_i$.

1.4.2.2 EM Algorithm

The EM algorithm, described by Dempster et al. (1977) is a two-step procedure consisting of an expectation step (the E step) and a maximization step (the M step). The conditional expectation of the log-likelihood given the observed data is taken in the E step in order to estimate sufficient statistics for the complete data. The M step then maximizes this conditional expectation in order to produce maximum likelihood estimates of the parameters based on the sufficient statistics derived in the E step. The method then iterates between the two steps until the estimates converge.

As noted by Wu (1983) the computation involved in the EM algorithm is often relatively easy due to the nice form of the complete-data likelihood function. As a result, the algorithm often uses less computation time per iteration than either the Newton-Raphson or Fisher scoring method. The EM procedure also tends to use less computer storage space than the other two methods. However, as mentioned by Murray (1990), the EM algorithm does not directly produce covariance estimates for the estimators. It may also require many more iterations to converge than the other two methods in many practical situations. Jennrich and Schluchter (1986) proposed a hybrid algorithm, termed the EM scoring algorithm, combining elements of the EM and Fisher scoring algorithms. However, this method is only advantageous when $\mathbf{\Sigma}_i$ is a function of a large number of parameters (e.g. large, unstructured $\mathbf{\Sigma}_i$). Lindstrom and Bates (1988) concluded that a well-implemented Newton-Raphson algorithm is generally preferable to the EM algorithm.

1.4.3 Finite Difference Approximations

All three commonly used iterative algorithms require derivative calculations. The Newton-Raphson and Fisher scoring methods utilize first and second-order partial derivatives of the log-likelihood function, while the EM method employs the first partial derivatives of the conditional expectation of the log-likelihood given the observed data. The finite difference approximations of these derivatives are useful when exact analytical derivatives are unwieldy or difficult to obtain. The forward and central difference derivative approximations are the two main types of these approximations.

As noted by SAS documentation (SAS Institute, 1999), the forward difference derivative approximations are often not as precise as the central difference derivative approximations, though they do use less computer time. For first-order derivatives the forward difference derivative approximations require n additional objective function calls, where n is the number of parameters. These approximations are computed as

$$g_i = \frac{\partial f}{\partial \theta_i} \approx \frac{f(\boldsymbol{\theta} + h_i \boldsymbol{e}_i) - f(\boldsymbol{\theta})}{h_i},$$

where e_i is the i^{th} unit vector and h_i is the i^{th} step size, $i = 1, ..., n$. Dennis and Schnabel (1983) detailed the forward difference approximation formulas for the Hessian matrix. An additional $n + n^2/2$ function calls are required for second-order derivative approximations based on only the objective function, which are calculated as

$$\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \approx \frac{f(\boldsymbol{\theta} + h_i \boldsymbol{e}_i + h_j \boldsymbol{e}_j) - f(\boldsymbol{\theta} + h_i \boldsymbol{e}_i) - f(\boldsymbol{\theta} + h_j \boldsymbol{e}_j) + f(\boldsymbol{\theta})}{h_i h_j}.$$

Second-order derivative approximations based on the gradient vector, which only require n additional gradient calls, are computed as

$$\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \approx \frac{g_i(\boldsymbol{\theta} + h_j \boldsymbol{e}_j) - g_i(\boldsymbol{\theta})}{2h_j} + \frac{g_j(\boldsymbol{\theta} + h_i \boldsymbol{e}_i) - g_j(\boldsymbol{\theta})}{2h_i}.$$

The more precise, but more computationally intensive central difference derivative approximation formulas are presented below. For first-order derivatives the approximations, which take an additional $2n$ objective function calls, are computed as

$$g_i = \frac{\partial f}{\partial \theta_i} \approx \frac{f(\boldsymbol{\theta} + h_i \boldsymbol{e}_i) - f(\boldsymbol{\theta} - h_i \boldsymbol{e}_i)}{2h_i}.$$

Abramowitz and Stegun (1972) detailed the central difference approximation formulas for the Hessian matrix. An additional $2n + 2n^2$ function calls are required for second-order derivative approximations based on only the objective function, which are calculated as

$$\frac{\partial^2 f}{\partial \theta_i^2} \approx \frac{-f(\boldsymbol{\theta} + 2h_i \boldsymbol{e}_i) + 16f(\boldsymbol{\theta} + h_i \boldsymbol{e}_i) - 30f(\boldsymbol{\theta}) + 16f(\boldsymbol{\theta} - h_i \boldsymbol{e}_i) - f(\boldsymbol{\theta} - 2h_i \boldsymbol{e}_i)}{12h_i^2},$$

$$\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \approx \frac{f(\boldsymbol{\theta}+h_i \boldsymbol{e}_i+h_j \boldsymbol{e}_j)-f(\boldsymbol{\theta}+h_i \boldsymbol{e}_i-h_j \boldsymbol{e}_j)-f(\boldsymbol{\theta}-h_i \boldsymbol{e}_i+h_j \boldsymbol{e}_j)+f(\boldsymbol{\theta}-h_i \boldsymbol{e}_i-h_j \boldsymbol{e}_j)}{4h_i h_j}.$$

Second-order derivative approximations based on the gradient vector, requiring $2n$ additional gradient calls, are computed as

$$\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \approx \frac{g_i(\boldsymbol{\theta} + h_j \boldsymbol{e}_j) - g_i(\boldsymbol{\theta} - h_j \boldsymbol{e}_j)}{4h_j} + \frac{g_j(\boldsymbol{\theta} + h_i \boldsymbol{e}_i) - g_j(\boldsymbol{\theta} - h_i \boldsymbol{e}_i)}{4h_i}.$$

The step sizes, h_i, for the forward difference approximations of the first-order derivatives based on objective function calls and the second-order derivatives based on gradient calls are defined as

$$h_i = \sqrt[2]{\eta}(1 + |\theta_i|).$$

For the forward difference approximations of second-order derivatives based only on objective function calls and for all central difference formulas the step sizes are

$$h_i = \sqrt[3]{\eta}(1 + |\theta_i|).$$

The value of η can either be user or computer defined.

1.5 Kronecker Product Covariance Structures

1.5.1 Overview

Multivariate repeated measures studies are characterized by data that have more than one set of correlated outcomes or repeated factors. Spatio-temporal data fall into this more general category since the outcome variables are repeated in both space and time. When analyzing multivariate repeated measures data, it is often advantageous to model the covariance separately for each repeated factor. This method of modeling the covariance utilizes the Kronecker product to combine the factor specific covariance structures into an overall covariance model.

Kronecker product covariance structures, also known as separable covariance models, were first introduced by Galecki (1994). A covariance matrix is separable if and only if it can be written as $\boldsymbol{\Sigma} = \boldsymbol{V} \otimes \boldsymbol{S}$. He noted that one of the model's main advantages is the ease of interpretation in terms of the independent contribution of every repeated factor to the overall within-subject covariance matrix. The model also allows for covariance matrices with nested parameter spaces and factor specific within-subject variance heterogeneity. Galecki (1994), along with Mitchell et al. (2006) and Naik and Rao (2001), detailed the numerous computational advantages of the Kronecker product covariance structure. Since all calculations can be performed on the smaller dimensional factor specific models, the computation of the partial derivatives, inverse, and the Cholesky decomposition of the overall covariance matrix is relatively easier.

Despite the many benefits of separable covariance models, they have limitations. As mentioned by Cressie and Huang (1999), the interaction among the various factors cannot be modeled when utilizing a Kronecker product structure. Galecki (1994), Huizenga et al. (2002), and Mitchell et al. (2006) all noted that a lack of identifiability can result with this model. This indeterminacy stems from the fact that if $\boldsymbol{\Sigma} = \boldsymbol{\Gamma} \otimes \boldsymbol{\Omega}$ is the overall within-subject covariance matrix, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$ are not unique since for $a \neq 0$,

$a\boldsymbol{\Gamma} \otimes (1/a)\boldsymbol{\Omega} = \boldsymbol{\Gamma} \otimes \boldsymbol{\Omega}$. However, this nonidentifiability can be fixed by rescaling one of the factor specific covariance matrices so that one of its diagonal nonzero elements is equal to 1. With homogeneous variances, this rescaled matrix is a correlation matrix.

Several tests of separability have been developed to determine whether a separable or nonseparable covariance model is most appropriate. Shitan and Brockwell (1995) constructed an asymptotic chi-square test for separability. A likelihood ratio test for separability was derived by Mitchell et al. (2006) and Lu and Zimmerman (2005). Fuentes (2006) developed a test for separability of a spatio-temporal process utilizing spectral methods. All of these tests were developed for complete and balanced data. Given that in many situations multivariate repeated measures data are unbalanced in at least one of the factors, more general tests for separability need to be developed.

There is also a lack of literature on the estimation of separable covariance models when there is an imbalance in at least one dimension. Only Naik (2001) examined the case in which data may be unbalanced in one of the factors, namely $\boldsymbol{\Sigma}_i = \boldsymbol{V} \otimes \boldsymbol{S}_i$. This situation arises often in spatio-temporal studies since the data tend to be balanced in space, but not in time. However, there are multivariate repeated measures studies in which the data are unbalanced in both dimensions, namely $\boldsymbol{\Sigma}_i = \boldsymbol{V}_i \otimes \boldsymbol{S}_i$. This case has yet to be examined in the literature.

With the assumptions of covariance model separability and homoscedasticity, an equal variance Kronecker product structure has great appeal. In this case the overall within-subject covariance matrix is defined as $\boldsymbol{\Sigma}_i = \sigma^2\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i$. This formulation has several advantages. The reduction in the number of parameters leads to computational benefits. The model is also identifiable since $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Omega}_i$ will necessarily be correlation matrices.

1.5.2 Multivariate Repeated Measures Model with Kronecker Product Covariance

Consider the following general linear model with structured covariance for repeated measures data:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}_i \tag{1.7}$$

where $\boldsymbol{y}_i$ is a $t_i s_i \times 1$ vector of $t_i s_i$ observations on the i^{th} subject $i = 1, ..., N$, $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed and unknown population parameters, $\boldsymbol{X}_i$ is a $t_i s_i \times q$ fixed and known design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$, and $\boldsymbol{e}_i$ is a $t_i s_i \times 1$ vector of random error terms. We assume $\boldsymbol{e}_i \sim N_{t_i s_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_{ei}(\sigma_e^2, \boldsymbol{\tau}_e) = \sigma_e^2[\boldsymbol{\Gamma}_{ei}(\boldsymbol{\tau}_{e\gamma}) \otimes \boldsymbol{\Omega}_{ei}(\boldsymbol{\tau}_{e\omega})])$ and is independent of $\boldsymbol{e}_{i'}$ for $i \neq i'$. It follows that $\boldsymbol{y}_i \sim N_{t_i s_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_{ei}(\sigma_e^2, \boldsymbol{\tau}_e) = \sigma_e^2[\boldsymbol{\Gamma}_{ei}(\boldsymbol{\tau}_{e\gamma}) \otimes \boldsymbol{\Omega}_{ei}(\boldsymbol{\tau}_{e\omega})])$ and is independent of $\boldsymbol{y}_{i'}$ for $i \neq i'$. We assume that $\boldsymbol{\Gamma}_{ei}(\boldsymbol{\tau}_{e\gamma})$ is a $t_i \times t_i$ positive-definite symmetric correlation matrix whose elements are twice differentiable functions of a finite number of fixed, unknown parameters $\boldsymbol{\tau}_{e\gamma} = \{\boldsymbol{\tau}_{e\gamma 1}, ..., \boldsymbol{\tau}_{e\gamma k}\}$, $\boldsymbol{\tau}_{e\gamma} \epsilon \boldsymbol{T}_\gamma$, where $\boldsymbol{T}_\gamma$ is the set of all parameters for which $\boldsymbol{\Gamma}_{ei}(\boldsymbol{\tau}_{e\gamma})$ is positive-definite. We also assume that $\boldsymbol{\Omega}_{ei}(\boldsymbol{\tau}_{e\omega})$ is a $s_i \times s_i$ positive-definite symmetric correlation matrix whose elements are twice differentiable functions of a finite number of fixed, unknown parameters $\boldsymbol{\tau}_{e\omega} = \{\boldsymbol{\tau}_{e\omega 1}, ..., \boldsymbol{\tau}_{e\omega k}\}$, $\boldsymbol{\tau}_{e\omega} \epsilon \boldsymbol{T}_\omega$, where $\boldsymbol{T}_\omega$ is the set of all parameters for which $\boldsymbol{\Omega}_{ei}(\boldsymbol{\tau}_{e\omega})$ is positive-definite. This implies that $\boldsymbol{\Sigma}_{ei}(\sigma_e^2, \boldsymbol{\tau}_e)$ is an $t_i s_i \times t_i s_i$ positive-definite symmetric covariance matrix whose elements are twice differentiable functions of a finite number of fixed, unknown parameters σ_e^2 and $\boldsymbol{\tau}_e = \{\boldsymbol{\tau}_{e\gamma}; \boldsymbol{\tau}_{e\omega}\}$, $\boldsymbol{\tau}_e \epsilon \boldsymbol{T}$, where $\boldsymbol{T}$ is the set of all parameters for which $\boldsymbol{\Sigma}_{ei}(\sigma_e^2, \boldsymbol{\tau}_e)$ is positive-definite. The parameters in $\{\sigma_e^2, \boldsymbol{\tau}_e\}$ are assumed to be functionally independent of those in $\boldsymbol{\beta}$. The model may be abbreviated as $\boldsymbol{y}_i \sim N_{t_i s_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_{ei} = \sigma_e^2[\boldsymbol{\Gamma}_{ei} \otimes \boldsymbol{\Omega}_{ei}])$.

1.6 Inference

Likelihood-based statistics are generally used for inference concerning the covariance parameters in the general linear model for repeated measures data. The three most common methods of this type being the Likelihood Ratio, Wald, and Score tests. Chi and Reinsel (1989) utilized the Score test for making inferences about the correlation parameter in the random errors of a model for longitudinal data with random effects and an AR(1) random error structure. However, this test was only used for reasons of simplicity in the context of testing for the presence of autocorrelation versus independence in the random errors of a mixed model. Jones and Boadi-Boateng (1991) employed both the Wald and Likelihood Ratio tests to make inferences about the nonlinear covariance parameters of *irregularly-spaced* longitudinal data with an AR(1) random error structure. They noted that the Likelihood Ratio method is the most effective way of testing the significance of the nonlinear parameters. The Wald test does not always perform as well since the covariance matrix of these parameters is estimated using a numerical approximation of the information matrix. It was also noted in Verbeke and Molenberghs (2000) that the Wald test is valid for large samples, but that it can be unreliable for small samples and for parameters known to have skewed or bounded distributions, as is the case for most covariance parameters. They too recommend use of the Likelihood Ratio Test for covariance parameter inference. Yokoyama (1997) proposed a modified Likelihood Ratio test for performing inference on the covariance parameters. However, this was only employed in lieu of the normal Likelihood Ratio Test because its computation was too complex in their context of having a multivariate random-effects covariance structure in a multivariate growth curve model with differing numbers of random effects.

Neyman and Pearson (1928) first proposed the Likelihood Ratio Test. We generally wish to test hypotheses of the form:

$$H_0 : \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t,0}$$

$$H_1 : \boldsymbol{\theta}_t \neq \boldsymbol{\theta}_{t,0} \tag{1.8}$$

where $\boldsymbol{\theta}_{t,0}$ is a vector of constants and $\boldsymbol{\theta} = (\boldsymbol{\theta}_t', \boldsymbol{\theta}_n')'$. The $\boldsymbol{\theta}_t$ vector contains the test parameters, the parameters of interest on which inference is to be performed. The $\boldsymbol{\theta}_n$ vector contains the nuisance parameters which must be estimated in order to utilize the Likelihood Ratio Method. Denoting the parameter spaces under H_0 (the null hypothesis) and H_1 (the alternative hypothesis) as γ and Γ, respectively, the test statistic for the Likelihood Ratio Method is

$$\Lambda = \frac{\max_{\boldsymbol{\theta}\epsilon\gamma} L(\boldsymbol{y};\boldsymbol{\theta})}{\max_{\boldsymbol{\theta}\epsilon\Gamma} L(\boldsymbol{y};\boldsymbol{\theta})} = \frac{L(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_\gamma)}{L(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_\Gamma)} \tag{1.9}$$

where $\widehat{\boldsymbol{\theta}}_\gamma$ and $\widehat{\boldsymbol{\theta}}_\Gamma$ are the estimates of the parameters of the model under the null and alternative hypotheses respectively. Often this statistic is written as

$$-2\ln\Lambda = 2[l(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_\Gamma) - l(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_\gamma)] \tag{1.10}$$

which, under most conditions, is asymptotically distributed as a χ^2_r random variable. The degrees of freedom, r, is equal to the number of linearly independent restrictions imposed on the parameter space by the null hypothesis. As noted by Self and Liang (1987), if the model resides on the boundary of the covariance parameter space under the null hypothesis, the asymptotic distribution of the test statistic in equation 1.10 becomes a mixture of χ^2 distributions.

1.7 Summary

Repeated measures designs with *incomplete*, *unbalanced* (within subject), *inconsistently-spaced*, and *irregularly-spaced* data are quite prevalent in biomedical research. These studies are often employed to examine longitudinal, spatial, or spatio-

temporal data, with medical imaging data being at the forefront of spatial and spatio-temporal research. With Gaussian data, such designs require the general linear model for repeated measures data when standard multivariate techniques do not apply.

Many repeated measures settings have within-subject correlation decreasing exponentially in time or space. Even though observed correlations often decay at a much slower or much faster rate than the AR(1) structure dictates, it sees the most use among the variety of correlation patterns available. Munoz et al. (1992) presented a parsimonious three-parameter generalization of the AR(1) structure, namely the DE structure, which allows for an attenuation or acceleration of the exponential decay rate imposed by the AR(1) structure. Nunez-Anton and Woodworth (1994) also proposed a three-parameter covariance model for repeated measures data that allows for nonstationary correlation structures for the within-subject errors; however, it lacks the flexibility of the DE model when the data are assumed to have a stationary structure, which is often the case.

In this dissertation I will propose a three-parameter covariance model that is a generalization of the continuous-time AR(1) structure. I will show that this new model, termed the *generalized autoregressive* (GAR) covariance structure, is more appropriate for many types of data than the AR(1) model. I also will make evident that the GAR model is more flexible than the DE covariance structure. However, merely showing that the two models are different is sufficient for the relevance of the GAR since the aim is not to replace other models, but to add to the suite of parsimonious covariance structures for repeated measures data. The GAR model is proposed in Chapter 2 along with the derivations of the estimates of its parameters and the variances of those estimates for confidence interval construction. Chapter 3 examines inference about the parameters of the model. Chapter 4 extends the GAR model to the multivariate repeated measures

setting in which a Kronecker product covariance model is employed. Chapter 5 summarizes the preceding chapters and discusses possible future research.

CHAPTER 2. ESTIMATION IN LINEAR MODELS WITH A GENERALIZED AR(1) COVARIANCE STRUCTURE

2.1 Introduction

2.1.1 Motivation

Repeated measures designs with *incomplete*, *unbalanced* (within subject), *inconsistently-spaced*, and *irregularly-spaced* data are quite prevalent in biomedical research. These studies are often employed to examine longitudinal, spatial, or spatio-temporal data, with medical imaging data being at the forefront of spatial and spatio-temporal research. Accurately modeling the covariance structure of these types of data can be of immense importance for proper analyses to be conducted. Louis (1988) noted that there is a tradeoff between including additional covariates and increasing the complexity of the covariance structure. The covariance structure may be able to account for the effects of unmeasured fixed effect covariates. Proper specification of the covariance model can be essential for the accurate estimation of and inference about the means and covariates of interest. Muller et al. (2007) showed that there can be severe test size inflation in fixed effect inference if the covariance structure is badly misspecified.

Many repeated measures settings have within-subject correlation decreasing exponentially in time or space. The continuous-time first-order autoregressive covariance structure, denoted AR(1), sees the most utilization among the variety of correlation patterns available in this context. This two-parameter model was briefly examined by Louis (1988) and is a special case of the model described by Diggle (1988). The AR(1) covariance model is very widely used; however, there are situations in which it may not be flexible enough to accurately model the correlation pattern induced by repeatedly

taking measurements over time or in space. In both longitudinal and imaging studies the within-subject correlations often decay at a slower or faster rate than that imposed by the AR(1) structure. Thus, a more flexible version of the AR(1) model is needed. Due to the desire to maintain parsimony, only three-parameter generalizations are considered.

The emphasis on covariance models with a modest number of parameters reflects the desire to accommodate a variety of kinds of real data. Missing and mis-timed data typically occur in longitudinal research, while High Dimension, Low Sample Size (HDLSS) data occur with imaging, metabolomics, genomics, etc. The inevitable and ever more common scientific use of longitudinal studies of imaging only increases the pressure to expand the suite of flexible and credible covariance models based on a small number of parameters.

A three-parameter generalization of the continuous-time AR(1) structure, termed the *generalized autoregressive* (GAR) covariance structure, which is more appropriate for many types of data than comparable models is proposed in this chapter. Special cases of the GAR model include the AR(1) and equal correlation (as in compound symmetry) models. The flexibility achieved with three parameters makes the GAR structure especially attractive for the High Dimension, Low Sample Size case so common in medical imaging and various kinds of "-omics" data. Excellent analytic and numerical properties help make the GAR model a valuable addition to the suite of parsimonious covariance structures for repeated measures data.

Section 2.2 provides a formal definition of the GAR model as well as parameter estimators and their variances. Graphical comparisons across parameter values illustrate the flexibility of the GAR model. Simulation studies in section 2.3 help compare the AR(1), DE, and GAR models. I discuss the analysis of an example dataset in section 2.4 and conclude with a summary discussion including planned future research in section 2.5.

2.1.2 Literature Review

LaVange and Muller (1992) described a three-parameter generalization of the AR(1) structure as a tool for power analysis in repeated measures studies, but did not discuss any properties or consider estimation. They defined the model in order to be able to generate a realistic suite of possible repeated measures covariance models in terms of only three parameters. The appeal of the model led the author of NQuery® power software to embed the model in the software as a study planning tool. Munoz et al. (1992) presented a three-parameter generalization of the AR(1) structure, namely the *damped exponential* (DE) structure, which allows for an attenuation or acceleration of the exponential decay rate imposed by the AR(1) structure. Murray (1990) described the same model in an unpublished dissertation. As noted by Grady and Helms (1995), the DE model has issues with convergence due to its parameterization. Nunez-Anton and Woodworth (1994) proposed a three-parameter covariance model that allows for nonstationary correlation for the within-subject errors. The model lacks the flexibility of the DE model when a stationary structure is assumed.

Proper estimation of fixed effect and covariance parameters in a repeated measures model requires iterative numerical algorithms. Jennrich and Schluchter (1986) detailed the Newton-Raphson and Fisher scoring algorithms for maximum likelihood (ML) estimation of model parameters. Dempster et al. (1977) first proposed the EM algorithm for parameter estimation. The Newton-Raphson method is more widely applicable than the Fisher scoring method due to the computational infeasibility of calculating the expected information matrix in certain contexts. As noted by Lindstrom and Bates (1988), the method is also generally preferable to the EM procedure.

The benefits of profiling out σ^2 and optimizing the resulting profile log-likelihood to derive model parameter estimates via the Newton-Raphson algorithm was discussed by Lindstrom and Bates (1988). They noted that this optimization will generally require

fewer iterations, will have simpler derivatives, and the convergence will be more consistent. There are also certain situations in which the Newton-Raphson algorithm may fail to converge when optimizing the original log-likelihood but converge with ease utilizing the profile log-likelihood.

2.2 GAR Covariance Model

2.2.1 Notation

Consider the following general linear model for repeated measures data with the GAR covariance structure:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}_i \tag{2.1}$$

where $\boldsymbol{y}_i$ is a $p_i \times 1$ vector of p_i observations on the i^{th} subject $i = 1, ..., N$, $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed and unknown population parameters, $\boldsymbol{X}_i$ is a $p_i \times q$ fixed and known design matrix corresponding to the fixed effects, and $\boldsymbol{e}_i$ is a $p_i \times 1$ vector of random error terms. We assume $\boldsymbol{e}_i \sim N_{p_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_{ei})$ and is independent of $\boldsymbol{e}_{i'}$ for $i \neq i'$. It follows that $\boldsymbol{y}_i \sim N_{p_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_{ei})$ and is independent of $\boldsymbol{y}_{i'}$ for $i \neq i'$.

For $\boldsymbol{\Sigma}_{ei} = \{\sigma_{ei;jk}\}$, the *generalized autoregressive* (GAR) covariance structure is

$$\sigma_{ei;jk} = \mathcal{V}(y_{ij}, y_{ik}) = \sigma_e^2 \begin{cases} \rho_e^{\,1 + [(d(t_{ij}, t_{ik}) - 1)\delta_e/(D-1)]} & j \neq k \\ 1 & j = k\,, \end{cases} \tag{2.2}$$

where $\mathcal{V}(\,\cdot\,)$ is the covariance operator, $d(t_{ij}, t_{ik})$ is the distance between measurement times or locations, D is a computational flexibility *constant* that can be specified and by default it is set to the maximum number of distance units, σ_e^2 is the variability of the measurements at each time or location, ρ_e is the correlation between observations separated by one unit of time or distance, and δ_e is the decay speed. We assume $0 \leq \rho_e < 1, 0 \leq \delta_e$, and $D > 1$.

Implicit in this model formulation is the presence of both a stationary variance and correlation structure. The AR(1) covariance model is a special case of this model for

which $\delta_e = D - 1$. For values of $\delta_e > D - 1$, the correlation between measurements on a given subject decreases in time or space at a faster rate than for $\delta_e = D - 1$. As $\delta_e \to \infty$, this model approaches the moving average model of order 1, MA(1). For values of δ_e such that $0 < \delta_e < D - 1$, the correlation between measurements on a given subject decreases in time or space at a slower rate than for $\delta_e = D - 1$. When $\delta_e = 0$, this model reduces to the well known compound symmetric covariance model for which the correlation between measurements on a given subject is fixed at ρ_e no matter how far apart in time or space the measurements were taken. Though values of $\delta_e < 0$ yield valid autocorrelation functions for which the correlation between measurements on a given subject would increase with increasing time or distance between measurements, this is rare in biostatistical applications. Therefore the parameter space is restricted for reasons of practicality.

2.2.2 Plots

Graphical depictions of the GAR structure help to provide insight into the types of correlation patterns that can be modeled. Figures 2.2-2.4 show a subset of the correlation patterns that can be modeled with the GAR model.

2.2.3 Maximum Likelihood Estimation

In order to estimate the parameters of the model defined in equation 2.1, σ_e^2 is first profiled out of the likelihood. The first and second partial derivatives of the profile log-likelihood are then derived to compute ML estimates of parameters, via the Newton-Raphson algorithm, and the variance-covariance matrix of those estimates. The resulting estimates are then used to compute the value and variance of $\widehat{\sigma}_e^2$. A SAS IML (SAS Institute, 2002) computer program has been written implementing this estimation procedure for the general linear model (GLM) with GAR covariance structure and is available upon request.

Setting $\boldsymbol{\Sigma}_{ei} = \sigma_e^2 \boldsymbol{\Gamma}_{ei}(\boldsymbol{\tau}_e)$, where $\boldsymbol{\tau}_e = \{\delta_e, \rho_e\}$, the log-likelihood function of the parameters given the data under the model is:

$$\begin{aligned} l(\boldsymbol{y}; \boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\tau}_e) &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln|\sigma_e^2 \boldsymbol{\Gamma}_{ei}| - \frac{1}{2\sigma_e^2}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\big(p_i \ln(\sigma_e^2) + \ln|\boldsymbol{\Gamma}_{ei}|\big) - \frac{1}{2\sigma_e^2}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}), \end{aligned} \tag{2.3}$$

where $n = \sum_{i=1}^{N} p_i$. Taking the first partial derivative with respect to σ_e^2 gives

$$\frac{\partial l}{\partial \sigma_e^2} = -\frac{1}{2}\sum_{i=1}^{N} p_i \sigma_e^{-2} + \frac{1}{2\sigma_e^4}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}).$$

Setting the derivative to zero and solving for an estimate of the variance yields

$$\widehat{\sigma}_{eML}^2(\boldsymbol{\beta}, \boldsymbol{\tau}_e) = \frac{1}{n}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}).$$

Substituting $\widehat{\sigma}_{eML}^2(\boldsymbol{\beta}, \boldsymbol{\tau}_e)$ into the log-likelihood function in equation 2.3 leads to the following profile log-likelihood:

$$l_p(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\tau}_e) = -\frac{1}{2}\sum_{i=1}^{N}\ln|\boldsymbol{\Gamma}_{ei}| - \frac{1}{2}n\ln\left[\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right] - \frac{1}{2}n\ln\left(\frac{1}{n}\right) - \frac{n}{2}\,. \tag{2.4}$$

The first partial derivative with respect to $\boldsymbol{\beta}$ in equation 2.4 is

$$\partial l_p/\partial \boldsymbol{\beta} = n\Big(\textstyle\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\Big)^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}).$$

Setting the previous equation to zero and solving implies $\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i = \boldsymbol{0}$. If $\Big(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\Big)^{-1} = 0$ the likelihood would be degenerate. Therefore

$$\widehat{\boldsymbol{\beta}}_{ML}(\widehat{\boldsymbol{\tau}}_e) = \Big(\textstyle\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}(\widehat{\boldsymbol{\tau}_e})^{-1}\boldsymbol{X}_i\Big)^{-1}\Big(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}(\widehat{\boldsymbol{\tau}_e})^{-1}\boldsymbol{y}_i\Big).$$

The remaining first partial derivatives are as follows:

$$\frac{\partial l_p}{\partial \rho_e} = -\frac{1}{2}\sum_{i=1}^{N}\text{tr}\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\right) + \frac{n}{2}\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})$$

$$\frac{\partial l_p}{\partial \delta_e} = -\frac{1}{2}\sum_{i=1}^{N}\text{tr}\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e}\right) + \frac{n}{2}\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}),$$

where $\partial\boldsymbol{\Gamma}_{ei}/\partial\rho_e$ and $\partial\boldsymbol{\Gamma}_{ei}/\partial\delta_e$ are $p_i \times p_i$ matrices with $(j,k)^{th}$ element for $j,k \in \{1,...,p_i\}$:

$$\left(\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\right)_{jk} = \begin{cases} \left(1 + \dfrac{[d(t_{ij},t_{ik})-1](\delta_e)}{(D-1)}\right)\rho_e^{[(d(t_{ij},t_{ik})-1)\delta_e/(D-1)]} & j \neq k \\ 0 & j = k \end{cases},$$

$$\left(\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e}\right)_{jk} = \begin{cases} \left(\rho_e^{1} + [(d(t_{ij},t_{ik})-1)\delta_e/(D-1)]\right)(\ln\rho_e)\dfrac{[d(t_{ij},t_{ik})-1]}{(D-1)} & j \neq k \\ 0 & j = k\,. \end{cases}$$

The ML estimates of the model parameters are found by utilizing the Newton-Raphson algorithm detailed in section 1.4.2.1 which requires the first and second partial derivatives. The second partial derivatives of the parameters are also employed to determine the asymptotic variance-covariance matrix of the estimators.

The second partial derivatives with respect to the profile log-likelihood are:

$$\frac{\partial^2 l_p}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}} = n\left[2\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-2}\left(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)\left(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)' - \left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1}\left(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{X}_i\right)\right]$$

$$\frac{\partial^2 l_p}{\partial \rho_e^2}=-\frac{1}{2}\sum_{i=1}^{N}\frac{\partial^2 \ln|\boldsymbol{\Gamma}_{ei}|}{\partial \rho_e^2}-\frac{n}{2}\left\{\frac{\partial^2 \ln\left[\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right]}{\partial\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^2}\left[\frac{\partial\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)}{\partial\rho_e}\right]^2+\right.$$
$$\left.\frac{\partial \ln\left[\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right]}{\partial\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)}\frac{\partial^2\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)}{\partial\rho_e^2}\right\}$$
$$=\frac{1}{2}\sum_{i=1}^{N}\text{tr}\left[\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\right)'\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\rho_e}-\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial^2\boldsymbol{\Gamma}_{ei}}{\partial\rho_e^2}\right]+$$
$$\frac{n}{2}\left\{\left[\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-2}\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^2\right]+\right.$$
$$\left.\left[\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\frac{\partial\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\right)}{\partial\rho_e}\boldsymbol{r}_i(\boldsymbol{\beta})\right]\right\}.$$

Similarly,

$$\frac{\partial^2 l_p}{\partial \delta_e^2}=\frac{1}{2}\sum_{i=1}^{N}\text{tr}\left[\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\right)'\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\delta_e}-\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial^2\boldsymbol{\Gamma}_{ei}}{\partial\delta_e^2}\right]+$$
$$\frac{n}{2}\left[\left\{\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-2}\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^2\right\}+\right.$$
$$\left.\left\{\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\frac{\partial\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\right)}{\partial\delta_e}\boldsymbol{r}_i(\boldsymbol{\beta})\right\}\right].$$

$$\frac{\partial^2 l_p}{\partial\boldsymbol{\beta}\partial\rho_e}=n\left[\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-2}\left(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})-\right.$$
$$\left.\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1}\left(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)\right].$$

$$\frac{\partial^2 l_p}{\partial\boldsymbol{\beta}\partial\delta_e}=n\Bigg[\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-2}\left(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})-$$
$$\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1}\left(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)\Bigg].$$

$$\begin{aligned}\frac{\partial^2 l_p}{\partial\rho_e\partial\delta_e}&=-\frac{1}{2}\sum_{i=1}^{N}\frac{\partial^2\ln|\boldsymbol{\Gamma}_{ei}|}{\partial\rho_e\partial\delta_e}-\frac{n}{2}\Bigg\{\frac{\partial^2\ln\left[\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right]}{\partial\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^2}\times\\&\frac{\partial\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)}{\partial\rho_e}\frac{\partial\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)}{\partial\delta_e}+\\&\frac{\partial\ln\left[\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right]}{\partial\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)}\frac{\partial^2\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)}{\partial\rho_e\partial\delta_e}\Bigg\}\\&=\frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}\left[\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\right)'\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\delta_e}-\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial^2\boldsymbol{\Gamma}_{ei}}{\partial\rho_e\partial\delta_e}\right]+\\&\frac{n}{2}\Bigg\{\left[\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-2}\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)\times\right.\\&\left.\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)\right]+\\&\left[\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\frac{\partial\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial\boldsymbol{\Gamma}_{ei}}{\partial\delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\right)}{\partial\rho_e}\boldsymbol{r}_i(\boldsymbol{\beta})\right]\Bigg\},\end{aligned}$$

where $\partial^2\boldsymbol{\Gamma}_{ei}/\partial\rho_e^2$, $\partial^2\boldsymbol{\Gamma}_{ei}/\partial\delta_e^2$, and $\partial^2\boldsymbol{\Gamma}_{ei}/\partial\rho_e\partial\delta_e$ are the $p_i\times p_i$ matrices whose $(j,k)^{th}$ element is:

$$\left(\frac{\partial^2 \boldsymbol{\Gamma}_{ei}}{\partial \rho_e^2}\right)_{jk} = \begin{cases} \left\{1 + \frac{[d(t_{ij},t_{ik})-1](\delta_e)}{(D-1)}\right\}\left\{\frac{[d(t_{ij},t_{ik})-1](\delta_e)}{(D-1)}\right\}\rho_e^{[(d(t_{ij},t_{ik})-1)\delta_e/(D-1)]-1} & j \neq k \\ 0 & j = k \end{cases},$$

$$\left(\frac{\partial^2 \boldsymbol{\Gamma}_{ei}}{\partial \delta_e^2}\right)_{jk} = \begin{cases} \left\{\rho_e^{1+[(d(t_{ij},t_{ik})-1)\delta_e/(D-1)]}\right\}(\ln \rho_e)^2\left\{\frac{[d(t_{ij},t_{ik})-1]}{(D-1)}\right\}^2 & j \neq k \\ 0 & j = k \end{cases},$$

$$\left(\frac{\partial^2 \boldsymbol{\Gamma}_{ei}}{\partial \rho_e \partial \delta_e}\right)_{jk} = \begin{cases} \frac{[d(t_{ij},t_{ik})-1]}{(D-1)}\rho_e^{[(d(t_{ij},t_{ik})-1)\delta_e/(D-1)]}\left(\left\{1 + \frac{[d(t_{ij},t_{ik})-1](\delta_e)}{(D-1)}\right\}(\ln \rho_e) + 1\right) & j \neq k \\ 0 & j = k \end{cases},$$

with $j, k = 1, ..., p_i$. Also, using results from Magnus and Neudecker (1999), we have

$$\begin{aligned} \frac{\partial\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\right)}{\partial \rho_e} &= \left(\frac{\partial \boldsymbol{\Gamma}_{ei}^{-1}}{\partial \rho_e}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e} + \boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial^2 \boldsymbol{\Gamma}_{ei}}{\partial \rho_e^2}\right)\boldsymbol{\Gamma}_{ei}^{-1} + \boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\frac{\partial \boldsymbol{\Gamma}_{ei}^{-1}}{\partial \rho_e} \\ &= \left(-\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e} + \boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial^2 \boldsymbol{\Gamma}_{ei}}{\partial \rho_e^2}\right)\boldsymbol{\Gamma}_{ei}^{-1} + \\ &\qquad \boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\left(-\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\right) \\ &= \boldsymbol{\Gamma}_{ei}^{-1}\left(-\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e} + \frac{\partial^2 \boldsymbol{\Gamma}_{ei}}{\partial \rho_e^2} - \frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\right)\boldsymbol{\Gamma}_{ei}^{-1}. \end{aligned}$$

Similarly,

$$\begin{aligned} \frac{\partial\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\right)}{\partial \delta_e} &= \boldsymbol{\Gamma}_{ei}^{-1}\left(-\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e} + \frac{\partial^2 \boldsymbol{\Gamma}_{ei}}{\partial \delta_e^2} - \frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e}\right)\boldsymbol{\Gamma}_{ei}^{-1}, \\ \frac{\partial\left(\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\right)}{\partial \rho_e} &= \boldsymbol{\Gamma}_{ei}^{-1}\left(-\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e} + \frac{\partial^2 \boldsymbol{\Gamma}_{ei}}{\partial \rho_e \partial \delta_e} - \frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_e}\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_e}\right)\boldsymbol{\Gamma}_{ei}^{-1}. \end{aligned}$$

After obtaining the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}_e$ utilizing the Newton-Raphson algorithm with the above derivatives, an estimate of σ_e^2 is calculated by substituting in these estimates into $\widehat{\sigma}_{eML}^2(\boldsymbol{\beta}, \boldsymbol{\tau}_e)$. The derivation of an estimate of variance for $\widehat{\sigma}_{eML}^2(\boldsymbol{\beta}, \boldsymbol{\tau}_e)$, assuming that $\boldsymbol{\beta}$ and $\boldsymbol{\tau}_e$ are known, is as follows. We have that

$$\widehat{\sigma}_e^2 = n^{-1}\sum\nolimits_{i=1}^{N} \boldsymbol{r}_i(\boldsymbol{\beta})' \boldsymbol{\Gamma}_{ei}^{-1} \boldsymbol{r}_i(\boldsymbol{\beta}),$$

where $\boldsymbol{r}_i(\boldsymbol{\beta}) = (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \sim \boldsymbol{N}_{p_i}(\boldsymbol{0}, \sigma_e^2 \boldsymbol{\Gamma}_{ei})$. Quadratic forms theory gives $\boldsymbol{r}_i(\boldsymbol{\beta})' \boldsymbol{\Gamma}_{ei}^{-1} \boldsymbol{r}_i(\boldsymbol{\beta}) / \sigma_e^2 \sim \chi_{p_i}^2$. In turn,

$$\begin{aligned}\mathcal{V}(\widehat{\sigma}_e^2) &= \mathcal{V}\Big[\textstyle\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})/n\Big] \\ &= \textstyle\sum_{i=1}^{N}\mathcal{V}[\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta})]/n^2\end{aligned}$$

since $\boldsymbol{y}_i \perp \boldsymbol{y}_{i'}$ for $i \neq i'$. If $Q_i \sim \chi^2(p_i)$, then $\mathcal{V}\big(\widehat{\sigma}_e^2\big) = \sum_{i=1}^{N}\mathcal{V}(\sigma_e^2 Q_i)/n^2 = 2\sigma_e^4/n$. Therefore $\widehat{\mathcal{V}}\big[\widehat{\sigma}_{eML}^2(\boldsymbol{\beta}, \boldsymbol{\tau}_e)\big] = 2\widehat{\sigma}_e^4/n$.

The asymptotic variance-covariance matrix of the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}_e$ is given by the following Hessian matrix (the observed information matrix):

$$\boldsymbol{H}_{\boldsymbol{\beta}\boldsymbol{\tau}_e} = \begin{bmatrix} \dfrac{\partial^2 l_p}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}} & \dfrac{\partial^2 l_p}{\partial\boldsymbol{\beta}\partial\delta_e} & \dfrac{\partial^2 l_p}{\partial\boldsymbol{\beta}\partial\rho_e} \\ \dfrac{\partial^2 l_p}{\partial\boldsymbol{\beta}\partial\delta_e} & \dfrac{\partial^2 l_p}{\partial\delta_e^2} & \dfrac{\partial^2 l_p}{\partial\rho_e\partial\delta_e} \\ \dfrac{\partial^2 l_p}{\partial\boldsymbol{\beta}\partial\rho_e} & \dfrac{\partial^2 l_p}{\partial\rho_e\partial\delta_e} & \dfrac{\partial^2 l_p}{\partial\rho_e^2} \end{bmatrix}. \tag{2.5}$$

The estimated asymptotic variance-covariance matrix of the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}_e$ is simply calculated by substituting the ML estimates of these parameters into equation 2.5 and taking its inverse.

From the previous derivations we can see that $\widehat{\sigma}_e^2$ is consistent (i.e., $\widehat{\sigma}_e^2 \rightarrow_p \sigma_e^2$) since we have that $E\big(\widehat{\sigma}_e^2\big) = \sigma_e^2$ and that $\mathcal{V}\big(\widehat{\sigma}_e^2\big) \rightarrow 0$ as $n \rightarrow \infty$. To establish the consistency of the remaining estimators, an examination of the asymptotic properties of the inverse of the observed information matrix is necessary $\big(\boldsymbol{H}_{\boldsymbol{\beta}\boldsymbol{\tau}_e}^{-1}\big)$. As noted in Vonesh and Chinchilli (1997), it is well known that under normal theory likelihood estimation we have that the asymptotic distributions of the ML estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}_e$ are

$$\sqrt{N}\Big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\Big) \rightarrow_d N_q(\mathbf{0}, \boldsymbol{\Sigma}_\beta)$$

and

$$\sqrt{N}(\widehat{\boldsymbol{\tau}}_e - \boldsymbol{\tau}_e) \rightarrow_d N_2(\mathbf{0}, \boldsymbol{\Sigma}_{\tau_e})$$

provided that the limits, $\boldsymbol{\Sigma}_\beta = \lim_{N\rightarrow\infty}\big(N\boldsymbol{H}_{\boldsymbol{\beta}}^{-1}\big) = \left[\lim_{N\rightarrow\infty}\left(N^{-1}\frac{\partial^2 l_p}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}}\right)\right]^{-1}$ and

$$\boldsymbol{\Sigma}_{\tau_e} = \lim_{N\to\infty}\left(N\boldsymbol{H}_{\tau_e}^{-1}\right) = \left[\lim_{N\to\infty}\left(N^{-1}\begin{bmatrix} \dfrac{\partial^2 l_p}{\partial\delta_e^2} & \dfrac{\partial^2 l_p}{\partial\rho_e\partial\delta_e} \\ \dfrac{\partial^2 l_p}{\partial\rho_e\partial\delta_e} & \dfrac{\partial^2 l_p}{\partial\rho_e^2}\end{bmatrix}\right)\right]^{-1}$$ exist. For "reasonable" values of p_i the limits should exist, though further investigation is needed. If the limits exist, the estimators will be consistent and fully asymptotically efficient.

2.2.4 Computational Issues and Parameter Constraints

A computational issue that may arise when implementing the GAR covariance model is that the model may give an estimated correlation greater than 1 if $\left[(d(t_{ij},t_{ik})-1)\widehat{\delta}_e/(D-1)\right] < -1$. Due to the parameter restrictions, this is only possible when $\widehat{\delta}_e/(D-1) > 1$ (faster than an AR(1) decay) and $d(t_{ij},t_{ik}) < 1$. This complication can be avoided by simply choosing a scale so that $\min_i[d(t_{ij},t_{ik})] \geq 1$.

Both the DE, which has $\mathcal{V}(y_{ij},y_{ik}) = \sigma_e^2 \cdot \rho_e^{d(t_{ij},t_{ik})^{\nu_e}}$, and GAR covariance matrices can be negative definite (or indefinite), a complication which appears not to have been addressed in any detail for the DE model in the literature. This may occur when there is a faster decay rate than that imposed by the AR(1) model coupled with a 'large' ρ. The acceptable ρ for this situation depends on the number of observations per subject, the spacing of the observations, and the decay speed. The appendix contains a proof of the positive definiteness of the GAR for decay speeds slower than or equal to that of the AR(1) model ($\delta_e \leq D-1$) based on Hadamard product theory. For $0 < \delta_e < D-1$, the GAR model has the nice statistical property that it can be reparameterized as the Hadamard product of a compound symmetric and a continuous-time AR(1) model (Appendix). The DE model does not have this feature, though enumeration studies show that it is also positive definite for decay speeds slower than or equal to that of the AR(1) model ($\nu_e \leq 1$).

To deal with the issue of negative definiteness (and indefiniteness) in both the GAR and DE for decay speeds greater than that of the AR(1) model ($\delta_e > D-1; \nu_e > 1$),

restricting $\rho_e \leq 0.5$ provides a conservative approach to ensure positive definiteness. This follows from the restriction for the MA(1) model (the limiting structure for the GAR and DE as $\delta_e, \nu_e \to \infty$) discussed in Diggle (1990). A more practical and less restrictive solution for the GAR is to rescale the distances since this complication seems to occur only when $\min_i[d(t_{ij}, t_{ik})] = 1$. In fact, enumeration studies show that with equally-spaced intervals of two units or greater the GAR covariance matrix is positive definite for every δ_e, ρ_e, and p_i (at least with $p_i \leq 1000$). This is not true for the DE model. In this case, for the GAR model, it can be empirically shown that the $\det(\mathbf{\Gamma}_{ei}) = 1 - f(\rho_e; \delta_e)$, where $f(\rho_e; \delta_e) \in (0, 1)$ and all eigenvalues of $\mathbf{\Gamma}_{ei}$ are positive. The function $f(\rho_e; \delta_e)$ decreases in δ_e and interval size, and increases in ρ_e and p_i. Thus, in extreme cases (e.g., $p_i = 500$ and $\rho_e = 0.99$), $f(\rho_e; \delta_e) \approx 1$ leading to the possibility of a computationally singular covariance matrix despite being theoretically nonsingular.

2.3 Simulations

To assess the empirical performance of the GAR covariance model, data were generated under the general linear model for repeated measures data with the GAR covariance structure and then fit with the GAR, DE, and AR(1) covariance models. Only the complete and balanced case was considered with $N = 100$ subjects and $p_i = p \in \{10, 50, 75\}$ observations each at unit distance intervals. The fixed effects included an intercept, a dummy variable indicating membership in one of two groups (50 subjects per group), and a continuous repeated covariate, with $\boldsymbol{\beta} = [1, 1, 1]'$. Preliminary simulation results showed that the GAR and DE model fits varied the most when the simulated data were highly correlated with a decay rate slower than that of the AR(1) model. Thus, the data were generated with the GAR covariance parameters $\rho_e = 0.5$ or 0.9 and $\delta_e = D/10, D/2,$ or $9D/10$ (with $D =$ maximum number of distance units $+ 1$). The scale parameter was set to $\sigma_e^2 = 10$. Each simulation for $p \in \{10, 50\}$ and the varying correlation parameters consisted of $10,000$ realizations. Due to computational

intensity the simulations for $p = 75$ and the varying correlation parameters consisted of $1,000$ realizations. All model fits employed ML estimation with the profile likelihood as discussed in 2.2.3. Both the Akaike Information Criterion (AIC, Akaike 1974) and the Bayesian Information Criterion (BIC, Schwarz 1978) were utilized to assess model fits. However, because they gave almost identical results, only the BIC results are reported for the sake of brevity. The BIC is $BIC = -2l(\boldsymbol{y}_i; \boldsymbol{\beta}, \boldsymbol{\Sigma}_i) + (q + w)\ln(n)$,where q is the number of fixed effect parameters, w is the number of unique covariance parameters, and n is the total number of observations.

Table 2.1 shows the results of the simulations for $p \in \{10, 50, 75\}$ with varying covariance parameters. The table displays the percent of realizations the GAR model better fits the data according to the BIC, and the median and maximum percent difference in BIC when this occurs. The percent differences allow gauging the relative disparity in BIC values between the GAR and comparable models when the GAR is selected as the best model.

As evidenced by the simulation results in Table 2.1, the GAR covariance model is more appropriate than either the DE or continuous-time AR(1) models when the data truly have a GAR correlation pattern. The GAR is selected as the better model more often in all conditions. According to the model selection percentages, and the median and maximum percent differences, the usefulness of the GAR tends to be most pronounced in situations in which the data exhibit a high within-subject correlation with a slow decay. The relative fit of the GAR covariance model also improves as the number of observations per subject increases. This performance has implications in many medical imaging studies in which the dimensionality of the data tends to be high. For instance, in shape analysis the image of an organ in a given subject is often characterized by the many correlated (repeated) values giving the mathematical parameterization of its shape.

As the true correlation pattern approaches that of the AR(1) model, the relative performance of the GAR covariance model tends to decline since both the GAR and DE models reduce to the AR(1) model as a special case. In Table 2.1 for $p = 10$, $\delta_e = (9D)/10$ exactly corresponds to this special case. Whereas, for $p = 50$ and $p = 75$ in Table 2.1, $\delta_e = (9D)/10$ corresponds to a decay rate close to, but slightly slower than that of the AR(1) model. The simulations make a strong case for the addition of the GAR covariance model to the suite of parsimonious covariance structures for repeated measures data.

2.4 Example

Several pediatric neurological disorders affect the myelin sheath of the nervous system. For example, infantile Krabbe disease, an inherited neurodegenerative disorder, causes demyelination of nerve cells leading to a rapid degeneration of mental and motor skills and death within the first 2-4 years of life. The understanding of myelination patterns in people with and without neurological conditions is critical in the radiologic assessment of disease progression and treatment response. Diffusion-tensor magnetic resonance imaging (DTI) allows gauging the degree of myelination via a proxy measure called fractional anisotropy (FA). McGraw et al. (2005) give a more detailed description of DTI and Krabbe disease.

Our analysis includes DTI scans of fibers of the cortico-spinal tracts (shown in Figure 1) associated with motor functioning for 46 control neonates. FA values were obtained at 20 locations, spaced 3 millimeters apart, along fiber tracts. FA values can theoretically range between 0 and 1, with higher values representing a more myelinated and mature nerve cell, but are typically between 0 and 0.6 for neonates. The gender, race, birth weight, gestational age at birth, and the gestational age at the time of the scan were also recorded. The study hypothesized that the older neonates (those with a higher

gestational age at the time of the scan) would have higher FA values. Gilmore et al. (2007) provide a more detailed description of the data.

We model the control neonate data with the general linear model for repeated measures data. The initial full model is as follows:

$$\boldsymbol{y}_i = \beta_0 + \beta_1 \boldsymbol{X}_{i,\text{loc}} + \beta_2 \boldsymbol{X}^2_{i,\text{loc}} + \beta_3 \boldsymbol{X}_{i,\text{gab}} + \beta_4 \boldsymbol{X}_{i,\text{gas}} + \beta_5 \boldsymbol{X}_{i,\text{gen}} + \beta_6 \boldsymbol{X}_{i,\text{rac}} + \beta_7 \boldsymbol{X}_{i,\text{bwt}} + \boldsymbol{e}_i. \tag{2.6}$$

The FA values for each of the 20 locations for each subject are contained in $\boldsymbol{y}_i$. The vectors $\boldsymbol{X}_{i,\text{gen}}$ and $\boldsymbol{X}_{i,\text{rac}}$ indicate the gender and race of the i^{th} neonate respectively. The gestational ages at birth and at the time of the scan are contained in $\boldsymbol{X}_{i,\text{gab}}$ and $\boldsymbol{X}_{i,\text{gas}}$, while $\boldsymbol{X}_{i,\text{bwt}}$ contains the birthweights. Preliminary analyses showed that FA values are a quadratic function of the fiber location, thus $\boldsymbol{X}_{i,\text{loc}}$ and $\boldsymbol{X}^2_{i,\text{loc}}$ were included to represent this trend. The location variable was shifted to start at 0, and all continuous covariates were centered about their respective means.

We model the covariance of the within-subject errors with the continuous-time AR(1), DE, and GAR structures in order to assess the best model via the BIC. All three models are given initial parameter values such that they correspond at the beginning of the optimization process. The GAR covariance model best fits the data with a BIC value of $-4{,}793$, while the AR(1) model yields a BIC value of $-4{,}704$ and the DE model fails to converge. The nonconvergence of the DE structure may be the result of the complexity in modeling a decay speed parameter that is nonlinear in the exponent of ρ (i.e. $\rho^{d(t_{ij},t_{ik})^{\nu}}$, where ν is the decay speed parameter). In contrast, the GAR structure, as defined in equation 2.2, has a decay speed parameter that is linear in the exponent. The computational flexibility of the GAR model due to the specified constant D is another possible reason for the convergence disparity.

We continue the analysis employing the GAR covariance model. In order to obtain a parsimonious model, the full model defined in equation 2.6 is reduced via backward

selection with $\alpha = 0.10$. The final model after reduction is

$$\boldsymbol{y}_i = \beta_0 + \beta_1 \boldsymbol{X}_{i,\text{loc}} + \beta_2 \boldsymbol{X}^2_{i,\text{loc}} + \beta_4 \boldsymbol{X}_{i,\text{gas}} + \boldsymbol{e}_i. \tag{2.7}$$

The resulting parameter estimates and p-values (based on the standard approximate F-test) associated with each of the covariates are presented in Table 2.2.

As expected, neonates who are older at the time of the scan have significantly higher FA values, and thus are likely to have more developed cortico-spinal fiber tracts. Figure 2.5 shows the predicted FA values as a function of location at the minimum and maximum gestational ages at scan (39.6 and 48.1 weeks respectively). The predicted FA values as a function of the neonates gestational age at the time of the scan for the first and middle locations are displayed in Figure 2.6.

The within-subject error variance estimate and correlation parameter estimates of the GAR (defined in equation 2.2) and AR(1) covariance models for the final data model are also given in Table 2.2. The flexibility of the GAR covariance structure allows it to model a correlation function in which the correlation is high for close measurements, but then decays at a faster rate than that imposed by the AR(1) structure as the measurements become farther apart. Figure 2.7 shows the predicted correlation as a function of the distance between measurement locations for both the best fitting GAR model and the AR(1) model.

2.5 Discussion and Conclusions

As shown by the simulations in section 2.3, the GAR covariance model performs much better than either the AR(1) or DE models for populations with a GAR structure. Its utility becomes even more pronounced in high dimensional settings which are prevalent in many areas of imaging research. As evidenced in section 2.2.4, the GAR has better statistical properties than the DE model. The appeal of the GAR covariance model is also shown in the example of section 2.4 in which the AR(1) model fits the data more poorly and the DE model fails to converge. Accurate estimation of and inference about

the fixed effect parameters of interest can be heavily dependent on the proper specification of the covariance matrix. A better fit of the covariance model gives more confidence in the results of the analysis of the neonate data.

There are many possible directions for future research with the proposed GAR covariance model. Any combination of allowing for the estimation of the D as a parameter and including higher order polynomial functions in the exponent of the model would further increase its flexibility. An assessment of the model's small sample performance and robustness to misspecification is a priority for future investigation. Also, introducing a nonstationary GAR covariance model may prove extremely useful in neuroimaging studies of the developing brain since the variability of brain characteristics tends to change over time. The model could have a nonstationary variance and/or correlation structure. For spatio-temporal data, or any data that have within-subject correlations induced by more than one factor, the development of a Kronecker product GAR covariance model would be beneficial.

Table 2.1 Summary of Model Fits:
100 subjects; 10, 50, and 75 observations per subject

	Simulated Model		GAR vs. DE			GAR vs. AR(1)		
p[a]	ρ_e	δ_e	GAR[b]	Median[c]	Max[d]	GAR[b]	Median[c]	Max[d]
10	0.5	$D/10$	72	0.09	0.30	100	3.26	5.27
		$D/2$	76	0.04	0.21	100	0.41	1.38
		$(9D)/10$	57	0.01	0.04	100	0.02	0.32
	0.9	$D/10$	81	0.29	0.83	100	12.75	17.41
		$D/2$	86	0.21	0.78	100	1.92	4.36
		$(9D)/10$	55	0.02	0.14	100	0.03	0.60
50	0.5	$D/10$	100	0.27	0.48	100	4.56	5.72
		$D/2$	94	0.05	0.19	100	0.65	1.14
		$(9D)/10$	63	0.00	0.01	100	0.01	0.08
	0.9	$D/10$	100	1.41	2.41	100	19.60	22.74
		$D/2$	100	0.47	0.77	100	3.30	4.17
		$(9D)/10$	75	0.02	0.09	100	0.08	0.27
75	0.5	$D/10$	100	0.33	0.60	100	4.81	6.04
		$D/2$	94	0.04	0.14	100	0.65	1.04
		$(9D)/10$	59	0.00	0.02	100	0.01	0.10
	0.9	$D/10$	100	1.74	2.54	100	20.10	22.57
		$D/2$	100	0.46	0.98	100	3.42	4.75
		$(9D)/10$	74	0.02	0.09	100	0.10	0.3

[a]10,000 realizations for $p = 10$ and 50 observations per subject; 1000 realizations for $p = 75$ observations per subject

[b]Percent of realizations GAR model selected by BIC as the better model fit (SE < 1.6)

[c]Median percent difference in BIC when GAR model is selected

[d]Maximum percent difference in BIC when GAR model is selected

Table 2.2 Neurological Data:
Final Mean Model Estimates, Standard Errors, P-values,
and Covariance Parameter Estimates

	GAR			AR(1)		
Parameter	Estimate	SE	P-value	Estimate	SE	P-value
β_1	0.0303	0.0023	< 0.0001	0.0299	0.0022	< 0.0001
β_2	-0.0015	0.0001	< 0.0001	-0.0015	0.0001	< 0.0001
β_4	0.0137	0.0030	< 0.0001	0.0140	0.0032	< 0.0001
σ_e^2	0.0047	0.0002	–	0.0046	0.0002	–
ρ_e	0.9875	0.0040	–	0.9109	0.0083	–
$\delta_e/(D-1)$	9.4107	3.0536	–	–	–	–

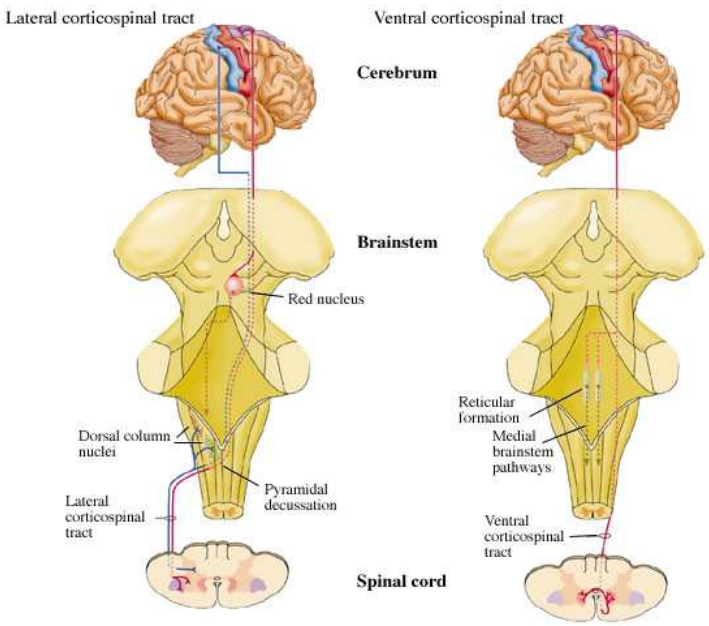

Figure 2.1. The Lateral and Ventral Cortico-spinal Tracts In the Human Nervous System.

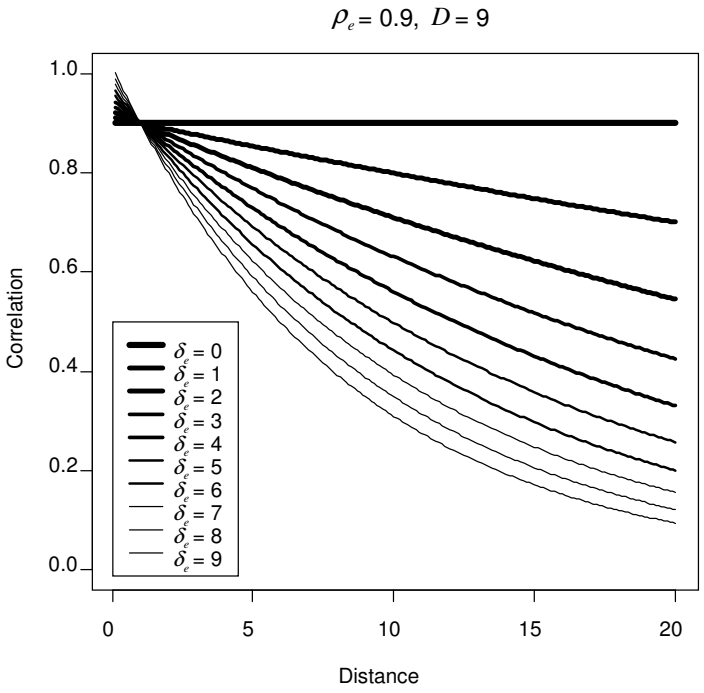

Figure 2.2. Plot of the various correlation patterns that can be obtained by varying the parameter δ_e while keeping ρ_e and D constant. $\delta_e = 8$ corresponds to an AR(1) decay rate, while $\delta_e = 0$ corresponds to compound symmetry.

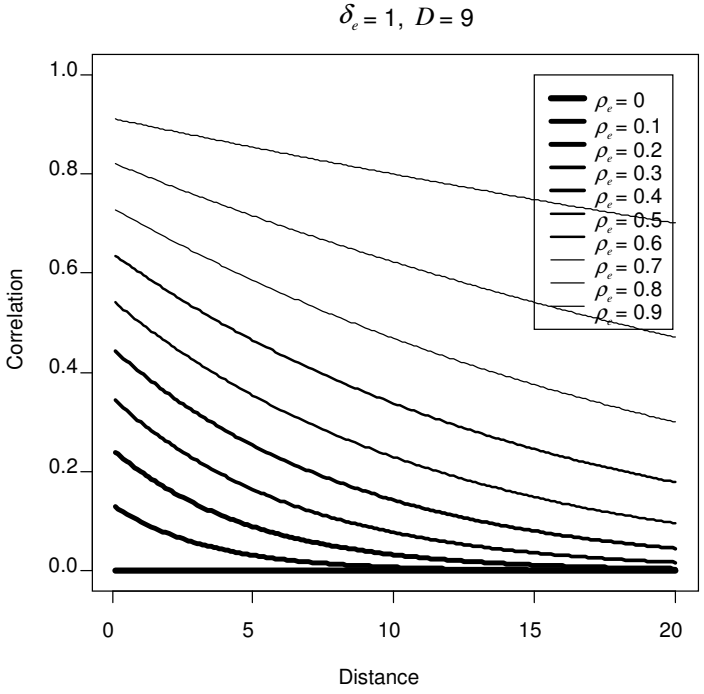

Figure 2.3. Plot of the various correlation patterns that can be obtained by varying the parameter ρ_e while keeping δ_e and D constant.

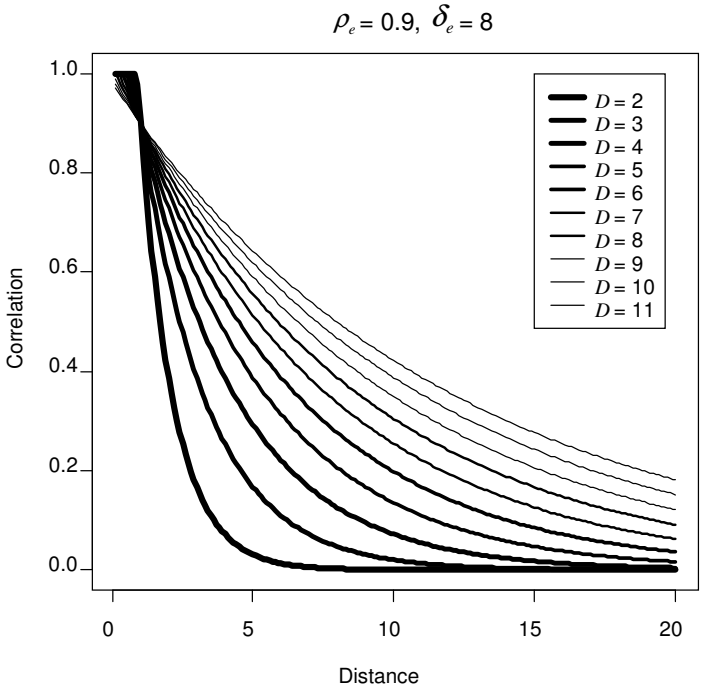

Figure 2.4. Plot of the various correlation patterns that can be obtained by varying the constant D while keeping δ_e and ρ_e constant. $D = 9$ corresponds to an AR(1) decay rate.

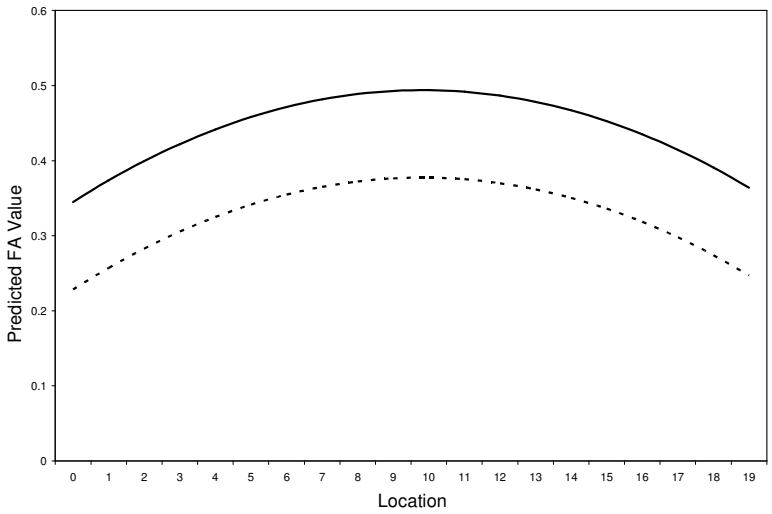

Figure 2.5. Predicted FA values for the neonates by location at the minimum (dashed line) and maximum (solid line) gestational ages at the time of the scan.

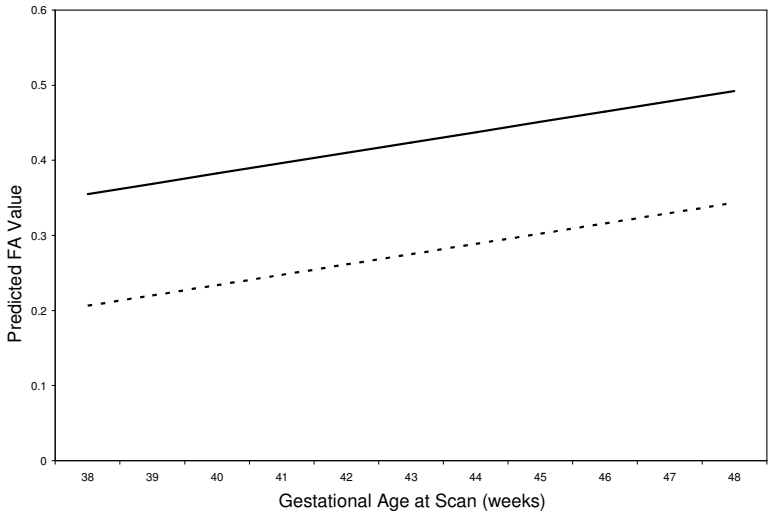

Figure 2.6. Predicted FA values for the neonates by the gestational age at the time of the scan at the first (dashed line) and middle (solid line) locations.

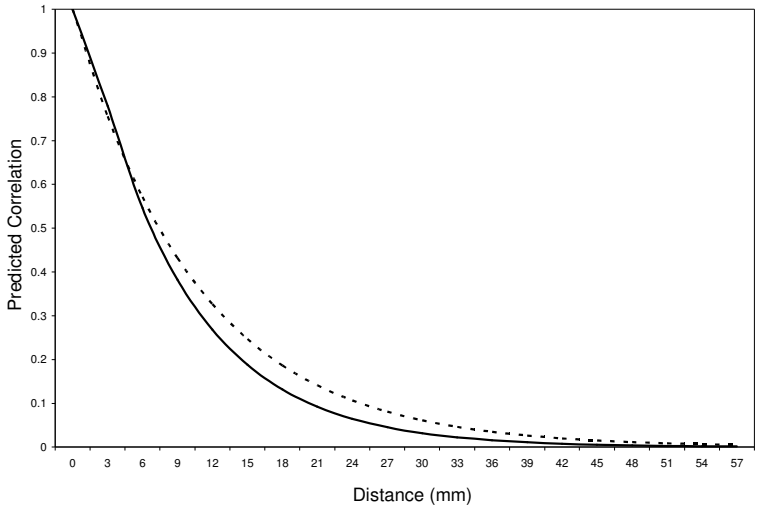

Figure 2.7. Predicted correlation curve for the best fitting GAR model (solid line) and AR(1) model (dashed line) as a function of the distance between measurements.

References

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transaction on Automatic Control*, AC-19, 716-723.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood for Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

Diggle, P. J. (1988), "An Approach to the Analysis of Repeated Measures," *Biometrics*, 44, 959-971.

– – (1990), *Time Series: A Biostatistical Introduction*, Oxford: Clarendon Press.

Gilmore, J. H., Lin, W., Corouge, I., Vetsa, Y. S. K., Smith, J. K., Kang, C., Gu, H., Hamer, R. M., Lieberman, J. A., and Gerig, G. (2007), "Early Postnatal Development of Corpus Callosum and Corticospinal White Matter Assessed With Quantitative Tractography," *American Journal of Neuroradiology*, 28, 1789-1795.

Grady, J. J., and Helms, R. W. (1995), "Model Selection Techniques for the Covariance Matrix for Incomplete Longitudinal Data," *Statistics In Medicine*, 14, 1397-1416.

Jennrich, R. I., and Schluchter, M. D. (1986), "Unbalanced Repeated-Measures Models With Structured Covariance Matrices," Biometrics, 42, 805-820.

LaVange, L. M., and Muller, K. E. (1992), "Using Power Calculations in Designing Repeated Measures Studies," *Unpublished Presentation*, ENAR Meeting, Biometric Society.

Lindstrom, M. J., and Bates, D. M. (1988), "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," *Journal of the American Statistical Association*, 83, 1014-1022.

Louis, T. A. (1988), "General Methods for Analyzing Repeated Measures," *Statistics in Medicine*, 7, 29-45.

Magnus, J. R., and Neudecker, H. (1999), *Matrix differential calculus with applications in statistics and econometrics*, New York: John Wiley & Sons.

McGraw, P., Liang, L., Escolar, M., Mukundan, S., Kurtzberg, J., and Provenzale, J. M. (2005), "Krabbe Disease Treated With Hematopoietic Stem Cell Transplantation: Serial Assessment of Anisotropy Measurements--Initial Experience," *Radiology*, 236, 221-230.

Muller, K. E., Edwards, L. J., Simpson, S. L., and Taylor, D. J. (2007), "Statistical Tests With Accurate Size and Power for Balanced Linear Mixed Models," *Statistics in Medicine*, 26, 3639-3660.

Munoz, A., Carey, V., Schouten, J. P., Segal, M., and Rosner, B. (1992), "A Parametric Family of Correlation Structures for the Analysis of Longitudinal Data," *Biometrics*, 48, 733-742.

Murray, S. C. (1990), "Linear Models With Generalized AR(1) Covariance Structure for Irregularly-Timed Data," *Unpublished Dissertation*, University of North Carolina at Chapel Hill.

Nunez-Anton, V., and Woodworth, G. (1994), "Analysis Of Longitudinal Data With Unequally Spaced Observations And Time-Dependent Correlated Errors," *Biometrics*, 50, 445-456.

SAS Institute. (2002), *SAS/IML, Version 9*, SAS Institute, Inc.: Cary, NC.

Schwarz, S. R. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461-464.

Vonesh, E. F., and Chinchilli, V. M. (1997), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, New York: Marcel Dekker, Inc.

CHAPTER 3. INFERENCE IN LINEAR MODELS WITH A GENERALIZED AR(1) COVARIANCE STRUCTURE

3.1 Introduction and Literature Review

It is widely known that the Likelihood Ratio Test (LRT) performs well in Gaussian linear models for inference about fixed effects when utilizing maximum likelihood (ML) estimation in a large sample context. Likelihood-based statistics are also generally used for inference concerning the covariance parameters in the general linear model for repeated measures data. The three most common methods of this type being the Likelihood Ratio, Wald, and Score tests (Verbeke and Molenberghs, 2000).

Chi and Reinsel (1989) utilized the Score test for making inferences about the correlation parameter in the random errors of a model for longitudinal data with random effects and an AR(1) random error structure. However, this test was only used for reasons of simplicity in the context of testing for the presence of autocorrelation versus independence in the random errors of a mixed model. Jones and Boadi-Boateng (1991) employed both the Wald and Likelihood Ratio tests to make inferences about the nonlinear covariance parameters of *irregularly-spaced* longitudinal data with an AR(1) random error structure. They noted that the Likelihood Ratio method is the most effective way of testing the significance of the nonlinear parameters. As they pointed out, the Wald test does not always perform as well since the covariance matrix of these parameters is estimated using a numerical approximation of the information matrix. It was also noted in Verbeke and Molenberghs (2000) that the Wald test is valid for large samples, but that it can be unreliable for small samples and for parameters known to have

skewed or bounded distributions, as is the case for most covariance parameters. They too recommend use of the Likelihood Ratio Test for covariance parameter inference.

Yokoyama (1997) proposed a modified Likelihood Ratio test for performing inference on the covariance parameters. However, this was only employed in lieu of the normal Likelihood Ratio Test because the latter's computation was too complex in their context of a multivariate random-effects covariance structure in a multivariate growth curve model.

In this chapter, I consider fixed effect and covariance parameter inference in linear models with a *generalized autoregressive* (GAR) covariance structure. Section 3.2 provides a formal definition of the GAR model. Section 3.3 discusses inference about the parameters of the model. Simulation studies in section 3.4 help compare inference accuracy in the AR(1), DE, and GAR models. I discuss the analyses of two example datasets in section 3.5 and conclude with a summary discussion including planned future research in section 3.6.

3.2 GAR Covariance Model

Again consider the following general linear model for repeated measures data with the GAR covariance structure:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}_i \tag{3.1}$$

where $\boldsymbol{y}_i$ is a $p_i \times 1$ vector of p_i observations on the i^{th} subject $i = 1, ..., N$, $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed and unknown population parameters, $\boldsymbol{X}_i$ is a $p_i \times q$ fixed and known design matrix corresponding to the fixed effects, and $\boldsymbol{e}_i$ is a $p_i \times 1$ vector of random error terms. We assume $\boldsymbol{e}_i \sim N_{p_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_{ei})$ and is independent of $\boldsymbol{e}_{i'}$ for $i \neq i'$. It follows that $\boldsymbol{y}_i \sim N_{p_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_{ei})$ and is independent of $\boldsymbol{y}_{i'}$ for $i \neq i'$.

For $\mathbf{\Sigma}_{ei} = \{\sigma_{ei;jk}\}$, the *generalized autoregressive* (GAR) covariance structure is

$$\sigma_{ei;jk} = \mathcal{V}(y_{ij}, y_{ik}) = \sigma_e^2 \begin{cases} \rho_e^{1 + [(d(t_{ij}, t_{ik}) - 1)\delta_e/(D-1)]} & j \neq k \\ 1 & j = k\,, \end{cases} \tag{3.2}$$

where $\mathcal{V}(\,\cdot\,)$ is the covariance operator, $d(t_{ij}, t_{ik})$ is the distance between measurement times or locations, D is a computational flexibility *constant* that can be specified and by default it is set to the maximum number of distance units, σ_e^2 is the variability of the measurements at each time or location, ρ_e is the correlation between observations separated by one unit of time or distance, and δ_e is the decay speed. We assume $0 \leq \rho_e < 1, 0 \leq \delta_e$, and $D > 1$.

Implicit in this model formulation is the presence of both a stationary variance and correlation structure. The AR(1) covariance model is a special case of this model for which $\delta_e = D - 1$. For values of $\delta_e > D - 1$, the correlation between measurements on a given subject decreases in time or space at a faster rate than for $\delta_e = D - 1$. As $\delta_e \to \infty$, this model approaches the moving average model of order 1, MA(1). For values of δ_e such that $0 < \delta_e < D - 1$, the correlation between measurements on a given subject decreases in time or space at a slower rate than for $\delta_e = D - 1$. When $\delta_e = 0$, this model reduces to the well known compound symmetric covariance model for which the correlation between measurements on a given subject is fixed at ρ_e no matter how far apart in time or space the measurements were taken. Though values of $\delta_e < 0$ yield valid autocorrelation functions for which the correlation between measurements on a given subject would increase with increasing time or distance between measurements, this is rare in biostatistical applications. Therefore the parameter space is restricted for reasons of practicality.

3.3 Inference

Neyman and Pearson (1928) first proposed the Likelihood Ratio Test. We generally wish to test hypotheses of the form:

$$H_0 : \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t,0} \tag{3.3}$$
$$H_1 : \boldsymbol{\theta}_t \neq \boldsymbol{\theta}_{t,0} \tag{3.4}$$

where $\boldsymbol{\theta}_{t,0}$ is a vector of known constants. The $\boldsymbol{\theta}_t$ vector $(a \times 1)$ contains the test parameters, the parameters of interest on which inference is to be performed. We define $\boldsymbol{\theta} = [\, \boldsymbol{\theta}_t \quad \boldsymbol{\theta}_n' \,]'$ $(c \times 1)$, where $\boldsymbol{\theta}_n$ is a $(b \times 1)$ vector that contains the nuisance parameters which must be estimated in order to utilize the Likelihood Ratio Method ($a + b = c$ is the total number of parameters). Denoting the parameter spaces under H_0 (the null hypothesis) and H_1 (the alternative hypothesis) as $\gamma \epsilon R^c$ and $\Gamma \epsilon R^c$, respectively, the test statistic for the Likelihood Ratio Method is

$$\Lambda = \frac{\underset{\boldsymbol{\theta}\epsilon\gamma}{\max} L(\boldsymbol{y};\boldsymbol{\theta})}{\underset{\boldsymbol{\theta}\epsilon\Gamma}{\max} L(\boldsymbol{y};\boldsymbol{\theta})} = \frac{L(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_\gamma)}{L(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_\Gamma)} \tag{3.5}$$

where $\widehat{\boldsymbol{\theta}}_\gamma$ $(c \times 1)$ and $\widehat{\boldsymbol{\theta}}_\Gamma$ $(c \times 1)$ are the estimates of the parameters of the model under the null and alternative hypotheses respectively. Often this statistic is written as

$$-2\ln\Lambda = 2[l(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_\Gamma) - l(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_\gamma)] \tag{3.6}$$

which, under regularity conditions, is asymptotically distributed as a χ^2_r random variable. The degrees of freedom, r, is equal to the number of linearly independent restrictions imposed on the parameter space by the null hypothesis. As noted by Self and Liang (1987), if the model resides on the boundary of the covariance parameter space under the null hypothesis, the asymptotic distribution of the test statistic in equation 3.6 becomes a mixture of χ^2 distributions.

Setting $\boldsymbol{\Sigma}_{ei} = \sigma^2_e \boldsymbol{\Gamma}_{ei}(\boldsymbol{\tau}_e)$, where $\boldsymbol{\tau}_e = \{\delta_e, \rho_e\}$, we have the following log-likelihood function for model 3.1:

$$l(\boldsymbol{y};\boldsymbol{\beta},\sigma_e^2,\boldsymbol{\tau}_e) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln|\sigma_e^2\boldsymbol{\Gamma}_{ei}| - \frac{1}{2\sigma_e^2}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}) \tag{3.7}$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\big(p_i\ln(\sigma_e^2) + \ln|\boldsymbol{\Gamma}_{ei}|\big) - \frac{1}{2\sigma_e^2}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\boldsymbol{\Gamma}_{ei}^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}),$$

where $n = \sum_{i=1}^{N} p_i$. Thus, tests of hypotheses about the parameters utilize the test statistic

$$-2\ln\Lambda = 2[l(\boldsymbol{y};\widehat{\boldsymbol{\beta}}_\Gamma,\widehat{\sigma}^2_{e\Gamma},\widehat{\boldsymbol{\tau}}_{e\Gamma}) - l(\boldsymbol{y};\widehat{\boldsymbol{\theta}}_\gamma,\widehat{\boldsymbol{\beta}}_\gamma,\widehat{\sigma}^2_{e\gamma},\widehat{\boldsymbol{\tau}}_{e\gamma})] \tag{3.8}$$

where the restricted and unrestricted estimates are computed via the ML methods discussed in Chapter 2.

3.4 Simulations

3.4.1 Fixed Effect Inference

To assess the relative robustness of the GAR covariance model to misspecification, data were generated under the general linear model for repeated measures data with various exponentially decaying covariance structures and then fitted with the GAR, DE, and AR(1) covariance models. Simulated fixed effect test size was examined for the three covariance model fits. Only the complete and balanced case was considered with $N = 100$ subjects and $p_i = p \in \{5, 20\}$ observations each at two-unit distance intervals. The fixed effects included an intercept, and three dummy variables indicating membership in one of four groups (25 subjects per group), with $\boldsymbol{\beta} = [1, 1, 1, 0]'$ in order to empirically evaluate test size for $H_0 : \beta_4 = 0$. The simulated covariance model was constructed as a weighted sum of the GAR and DE models with a scale parameter set to $\sigma_e^2 = 1$, namely $\boldsymbol{\Sigma}(\rho_e, \delta_e, \nu_e) = \tau\boldsymbol{\Sigma}_{\text{GAR}}(\rho_e, \delta_e) + (1-\tau)\boldsymbol{\Sigma}_{\text{DE}}(\rho_e, \nu_e)$ with $\tau \in \{0, 0.25, 0.5, 0.75, 1\}$. Four parameter sets were considered: 1) $(\delta_e, \nu_e) = (0, 0)$ corresponding to compound symmetry (CS), 2) $(\delta_e, \nu_e) = ((D-1)/4, 0.5)$ corresponding to a slower than AR(1) decay rate, 3) $(\delta_e, \nu_e) = (D-1, 1)$ corresponding to an AR(1) decay rate, and 4) $(\delta_e, \nu_e) = (1.5 \cdot (D-1), 1.2)$ corresponding to a faster

than AR(1) decay rate. All four cases had $\rho_e = 0.8$. For sets 1) and 3), all τ's lead to the same model. Each simulation for $p \in \{5, 20\}$ and the varying correlation parameters consisted of $5,000$ realizations. All model fits employed ML estimation with the profile likelihood as discussed in Chapter 2.

Table 3.1 shows the results of the simulations for $p \in \{5, 20\}$ respectively. The table contains simulated test size (target $\alpha = 0.05$) for the Likelihood Ratio Test (LRT) of $H_0 : \beta_4 = 0$ for GAR, DE, and AR(1) covariance model fits. The LRT was employed due to the utilization of ML estimation in a moderately large sample context.

For situations in which the true within subject correlation is constant (a misspecification for the AR(1) model, but not for the GAR or DE models) both the GAR and DE models control test size far better than the AR(1) model. This test size inflation with the AR(1) fit increases drastically as the number of observations per subject increases. When all three models are misspecifications for correlation decay rates slower than that of AR(1), both the GAR and DE models control test size better than the AR(1) model. Again, this disparity increases in p. Test size is fairly well controlled for all three models for misspecifications with a faster than AR(1) decay rate. The relatively smaller amount of overall variation in this context may mitigate the effects of misspecifying the covariance with an AR(1) model. It is important to note that for this scenario in which the true decay rate is faster than that of the AR(1), a Quasi-Newton algorithm had to be used to ensure consistent convergence with the DE model for $p = 20$. Even when the data were generated from the DE structure, the model rarely converged when implemented with the Newton-Raphson method. The GAR covariance model is as robust to misspecification in controlling fixed effect test size as the DE model, while possessing better convergence properties. The GAR covariance model is far more robust to misspecification than the AR(1) model. These results further strengthen the case for the

addition of the GAR covariance model to the suite of parsimonious covariance structures for repeated measures data.

3.4.2 Covariance Parameter Inference

Hypothesis tests concerning the decay speed parameter of the GAR covariance model were examined to assess the ability of the commonly used LRT to discern the model from its special cases. More specifically, simulated test size was examined for tests of $H_0 : \delta_e = 0$ (corresponding to Compound Symmetry) and $H_0 : \delta_e = D - 1$ (corresponding to the AR(1) model). Only the complete and balanced case was considered with $N = 100$ subjects and $p_i = p \in \{5, 20\}$ observations each at two-unit distance intervals. Three fixed effect scenarios were considered corresponding to signal strengths of none, moderate and high respectively: 1) $\beta = 0$ (one group with mean 0), 2) $\boldsymbol{\beta} = [0.3, 0.3, 0.3, 0.3]'$ (one reference group with three additional groups; 25 subjects per group), 3) $\boldsymbol{\beta} = [0.6, 0.6, 0.6, 0.6]'$ (one reference group with three additional groups; 25 subjects per group). Here I use the term 'signal strength' to denote the importance of the fixed effects in the model. All three cases had $\rho_e \in \{0.5, 0.9\}$ and $\sigma_e^2 = 1$. Power for an overall multivariate test (all four commonly used tests coincide in this context; Muller and Stewart, 2006, give details) of the mean model was used as a proxy for signal strength. It is important to note that this power is an underestimate of the true power given the structured nature of the covariance matrices. Scenarios 2 and 3 had an average power of 0.489 and 0.952 across all parameter combinations respectively. Each simulation for $p \in \{5, 20\}$ and the varying mean and correlation parameters consisted of $5,000$ realizations. All model fits employed ML estimation with the profile likelihood as discussed in Chapter 2.

Table 3.2 shows the results of the simulations for $p \in \{5, 20\}$ respectively. The table contains simulated test size (target $\alpha = 0.05$) for the Likelihood Ratio Test (LRT) of $H_0 : \delta_e = 0$ and $H_0 : \delta_e = D - 1$.

Test size is consistently controlled for $H_0 : \delta_e = 0$ across all parameter combinations. Neither the signal strength (proxied by mean model MULTIREP test power) nor the number of observations per subject has any bearing on this control. With a high initial correlation, $\rho = 0.9$, test size for $H_0 : \delta_e = D - 1$ is fairly well controlled regardless of the signal strength or number of observations per subject. However, with a relatively low initial correlation, $\rho = 0.5$, test size is reasonably well controlled for lower signal strengths and fewer observations per subject, but becomes increasingly inflated with stronger signals and more observations per subject.

It is important to note that a boundary issue can arise when conducting an LRT of $H_0 : \delta_e = 0$ employing the estimation procedure of Chapter 2. On occasion infeasible negative values of the test statistic are produced. This complication can be avoided by allowing δ_e to be slightly negative (e.g., $-1 \times 10^{-2} \leq \delta_e$). This relaxation of the constraint will be examined in future investigations.

3.5 Examples

3.5.1 Neonate Neurological Development

The analysis of the neonate data in Chapter 2 found that the GAR covariance model fit the data better than the AR(1) model according to the BIC. A more formal evaluation of whether the two model fits differ would be to conduct a Likelihood Ratio Test of $H_0 : \delta_e = D - 1$. This test results in a p-value < 0.0001, thus corroborating the disparate fits. Given the strength of this result along with the simulation results of section 3.4.2 ($\widehat{\rho}_e = 0.91$ when data fit with AR(1) model), we can be confident in concluding that the fits are truly different.

3.5.2 Diet and Hypertension

The Dietary Approaches to Stop Hypertension (DASH) trial was a multicenter, randomized, parallel arm feeding study that tested the effects of dietary patterns on blood pressure. The three diets were a control diet (low in fruits, vegetables, and dairy

products, with a fat content typical of the average diet in the United States), a diet rich in fruits and vegetables (a diet similar to the control except it provided more fruits and vegetables and fewer snacks and sweets), and a combination diet rich in fruits, vegetables, and low-fat dairy foods and reduced in saturated fat, total fat, and cholesterol (DASH diet). Participants were healthy adults 22 years of age or older who were not taking antihypertensive medication. Ambulatory blood pressure monitoring (ABPM) was used to take blood pressure measurements on the subjects over a 24 hour period. The devices were programmed to take readings automatically every 30 minutes and to repeat a reading if it fell outside the acceptable range defined in the monitor's internal algorithm. Appel et al. (1997) and Moore et al. (1999) provide more detail on the DASH study.

Our analysis includes blood pressure measurements for 194 subjects, a subset of the those on the control and DASH diets. The data model consisting of the actual times of measurement would not converge with an AR(1) covariance model fit (though it did converge with a GAR model fit), and thus only the hour-interval data is examined. For each subject 24 measurements were constructed at hourly intervals. The main objective of the study was to determine if there is an effect due to diet. The race and age of participants are also included as covariates. The analysis of these data exemplifies the difference in fixed effect inference that can occur when modeling the covariance with the GAR instead of the AR(1) model.

We model the DASH data with the general linear model for repeated measures data. The model of interest is as follows:

$$\boldsymbol{y}_i = \beta_0 + \beta_1 \boldsymbol{X}_{i,\text{diet}} + \beta_2 \boldsymbol{X}_{i,\text{race}} + \beta_3 \boldsymbol{X}_{i,\text{age}} + \beta_4 \boldsymbol{X}_{i,\text{hour}} + \beta_5 \boldsymbol{X}^2_{i,\text{hour}} + \beta_6 \boldsymbol{X}^3_{i,\text{hour}} + \boldsymbol{e}_i. \quad (3.9)$$

The 24 blood pressure measurements for each subject are contained in $\boldsymbol{y}_i$. The vectors $\boldsymbol{X}_{i,\text{diet}}$ and $\boldsymbol{X}_{i,\text{race}}$ indicate the diet and race of the i^{th} participant respectively. Their age is contained in $\boldsymbol{X}_{i,\text{age}}$. The vectors $\boldsymbol{X}_{i,\text{hour}}$, $\boldsymbol{X}^2_{i,\text{hour}}$, and $\boldsymbol{X}^3_{i,\text{hour}}$ are included to represent the cubic trend in time present in the data.

We model the covariance of the within-subject errors with the continuous-time AR(1) and GAR structures to illustrate the difference in fixed effect inference that can occur with the disparate fits. The GAR covariance model best fits the data with a BIC value of 24,323, while the AR(1) model yields a BIC value of 24,484. A Likelihood Ratio Test of $H_0 : \delta_e = D - 1$ corroborates this difference in fit between the GAR and AR(1) models with a p-value < 0.0001. The resulting parameter estimates and p-values (based on the standard approximate F-test) associated with each of the covariates are presented in Table 3.3 for both the GAR and AR(1) covariance model fits.

For both model fits, subjects who are on the DASH diet, white, and younger have significantly lower blood pressure than others. The difference in fixed effect inference occurs for the quadratic trend in time where the AR(1) fit leads to significance at the $\alpha = 0.05$ level, while the GAR fit does not. This disparity is corroborated by a Likelihood Ratio Test of the parameter. Given the better fit of the GAR model and the simulation results of section 3.4.1, the AR(1) fit most likely leads to a type I error in this context.

The within-subject error variance estimate and correlation parameter estimates of the GAR (defined in equation 3.2) and AR(1) covariance models for the data model are also given in Table 3.3. The flexibility of the GAR covariance structure allows it to model a correlation function in which the correlation is high for close measurements, but then decays at a slower rate than that imposed by the AR(1) structure as the measurements become farther apart. Figure 3.1 shows the predicted correlation as a function of the distance between measurement locations for both the better fitting GAR model and the AR(1) model.

3.6 Discussion and Conclusions

As shown by the simulations in section 3.4, the GAR covariance model is far more robust to misspecification than the AR(1) model in terms of accurate fixed effect

inference. Relative to the DE, the GAR model is as robust to misspecification in controlling fixed effect test size, while having much better convergence properties. The utility of the GAR model becomes even more pronounced in high dimensional settings which are prevalent in many areas of imaging research. The appeal of the GAR covariance model is also shown in the examples of section 3.5. For the neonate neurological data, the AR(1) model fits the data significantly worse. An AR(1) covariance fit to the DASH data likely leads to a type I error, while the GAR fit does not. A better fit of the covariance model gives more confidence in the results of the analyses.

There are many avenues for future inference research in the GAR covariance model. Assessment of both fixed effect and covariance parameter inference in the model with small samples would have implications for both imaging and genetics research. Examining the robustness of the model to violations of the Gaussian assumption may prove useful in many contexts. The simulation results show the need for the development of tests that are unbiased even with a misspecified covariance.

Table 3.1 Simulated Fixed Effect Test Size for Target $\alpha = 0.05$
5,000 realizations, 100 subjects, 5 and 20 observations per subject

	Simulated Model $[\boldsymbol{\Sigma} = \tau\boldsymbol{\Sigma}_{\text{GAR}} + (\mathbf{1}-\tau)\boldsymbol{\Sigma}_{\text{DE}}]$		Fitted Model		
p[a]	($\boldsymbol{\Sigma}_{\text{GAR}}$, $\boldsymbol{\Sigma}_{\text{DE}}$)	τ	GAR	DE	AR(1)
5	(CS, CS)	-	0.058	0.058	0.097
	(Slow Decay, Slow Decay)	0	0.051	0.051	0.068
		0.25	0.062	0.061	0.081
		0.50	0.052	0.052	0.068
		0.75	0.052	0.051	0.070
		1	0.060	0.060	0.076
	(AR(1) Decay, AR(1) Decay)	-	0.062	0.061	0.062
	(Fast Decay, Fast Decay)	0	0.058	0.059	0.052
		0.25	0.057	0.057	0.051
		0.50	0.057	0.056	0.049
		0.75	0.056	0.056	0.050
		1	0.054	0.054	0.047
20	(CS, CS)	-	0.060	0.060	0.235
	(Slow Decay, Slow Decay)	0	0.074	0.058	0.163
		0.25	0.066	0.051	0.159
		0.50	0.063	0.054	0.148
		0.75	0.060	0.051	0.144
		1	0.057	0.045	0.127
	(AR(1) Decay, AR(1) Decay)	-	0.054	0.053	0.054
	(Fast Decay, Fast Decay)	0	0.048	0.053[b]	0.038
		0.25	0.049	0.053[b]	0.041
		0.50	0.048	0.049[b]	0.040
		0.75	0.048	0.051[b]	0.041
		1	0.049	0.051[b]	0.040

[a]For $p = 5$, standard error < 0.004; for $p = 20$, standard error < 0.006

[b]Fitted with Quasi-Newton algorithm due to rare convergence with Newton-Raphson

Table 3.2 Simulated Covariance Parameter Test Size for Target $\alpha = 0.05$
5,000 realizations, 100 subjects, 5 and 20 observations per subject

	Simulated Model		Null Hypothesis	
p[a]	Signal Strength	ρ_e	$\delta_e = 0$	$\delta_e = (D-1)$
5	None	0.5	0.023	0.034
		0.9	0.023	0.055
	Moderate	0.5	0.022	0.042
		0.9	0.021	0.052
	High	0.5	0.022	0.109
		0.9	0.026	0.052
20	None	0.5	0.028	0.051
		0.9	0.022	0.049
	Moderate	0.5	0.022	0.083
		0.9	0.022	0.050
	High	0.5	0.022	0.416
		0.9	0.025	0.061

[a]For $p = 5$, standard error < 0.005; for $p = 20$, standard error < 0.007

Table 3.3 DASH Data:
Final Mean Model Estimates, Standard Errors, P-values,
and Covariance Parameter Estimates

	GAR			AR(1)		
Parameter	Estimate	SE	P-value	Estimate	SE	P-value
β_1	3.5053	0.8588	< 0.0001	3.4971	0.7060	< 0.0001
β_2	2.4448	0.7368	0.0009	2.5004	0.6062	< 0.0001
β_3	0.1221	0.0412	0.0031	0.1107	0.0338	0.0011
β_4	$-$ 1.8010	0.0947	< 0.0001	$-$ 1.7992	0.0991	< 0.0001
β_5	$-$ 0.0096	0.0058	0.0976	$-$ 0.0144	0.0062	0.0198
β_6	0.0129	0.0008	< 0.0001	0.0130	0.0009	< 0.0001
σ_e^2	128.15	2.6559	$-$	127.75	2.6477	$-$
ρ_e	0.6853	0.0131	$-$	0.6853	0.0106	$-$
$\delta_e/(D-1)$	0.4786	0.0322	$-$	$-$	$-$	$-$

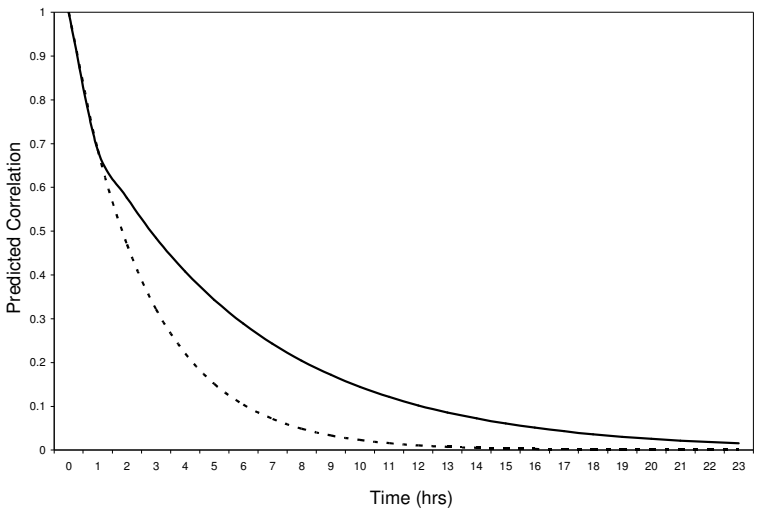

Figure 3.1 Predicted correlation curve for the better fitting GAR model (solid line) and AR(1) model (dashed line) as a function of the time between measurements.

References

Appel, L. J., Moore, T. J., Obarzanek, E., Vollmer, W. M., Svetkey, L. P., Sacks, F. M., Bray, G. A., Vogt, T. M., Cutler, J. A., Windhauser, M. M., Lin, P. H., and Karanja, N. (1997), "A Clinical Trial of the Effects of Dietary Patterns on Blood Pressure. DASH Collaborative Research Group," *New England Journal of Medicine*, 336, 1117-1124.

Chi, E. M., and Reinsel, G. C. (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association* **84**, 452-459.

Jones, R. H., and Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics* **47**, 161-175.

Moore, T. J., Vollmer, W. M., Appel, L. J., Sacks, F. M., Svetkey, L. P., Vogt, T. M., Conlin, P. R., Simons-Morton, D. G., Carter-Edwards, L., and Harsha, D. W. (1999), "Effect of Dietary Patterns on Ambulatory Blood Pressure: Results From the Dietary Approaches to Stop Hypertension (DASH) Trial," *Hypertension*, 34, 472-477.

Muller, K. E., and Stewart, P. W. (2006), *Linear Model Theory*, New Jersey: John Wiley & Sons.

Neyman, J., and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part II. *Biometrika* **20A**, 263-294.

Self, S. G., and Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605-610.

Verbeke, G., and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, New York: Springer.

Yokoyama, T. (1997). Tests for a family of random-effects covariance structures in a multivariate growth curve model. *Journal of Statistical Planning and Inference* **65**, 281-292.

CHAPTER 4. LINEAR MODELS WITH A KRONECKER PRODUCT GENERALIZED AR(1) COVARIANCE STRUCTURE

4.1 Introduction

4.1.1 Motivation

Longitudinal imaging studies are moving increasingly to the forefront of medical research due to their ability to characterize spatio-temporal features of biological structures across the lifespan. For instance, much of current Autism research involves examining the development of children's brains (via neuroimaging) over time (Cody et al., 2002). With Gaussian data, such designs require the general linear model for multivariate repeated measures data when standard multivariate techniques do not apply. A key advantage of this model lies in the flexibility of modeling the covariance of the outcome as well as the mean. While much work has been done on estimation, inference, and diagnostics of the mean model (fixed effects), relatively little has been done in these areas for the covariance model. The heavy dependence of fixed-effects inference accuracy on the proper specification of the covariance model indicates that the amount of work on covariance models has not been commensurate with their level of importance.

The *generalized autoregressive* (GAR) covariance model, introduced in chapter 1, is a flexible three-parameter covariance model that can be applied in situations in which the within subject correlation is believed to decrease exponentially in time or space. It allows for an attenuation or acceleration of the exponential decay rate imposed by the commonly used continuous-time AR(1) structure. In this chapter I propose the Kronecker product GAR covariance structure for multivariate repeated measures data in which the correlation between measurements for a given subject is induced by two factors. The

model allows for an imbalance in both dimensions across subjects. However, it is important to note that within a given subject both factors must have *consistently-spaced* measurements. In the context of spatio-temporal data this means that at each time point a subject must have the same number of measurements taken at the same spatial locations. This five-parameter structure is especially attractive for the High Dimension, Low Sample Size cases so common in medical imaging and various kinds of "-omics" data. Excellent analytic and numerical properties make the Kronecker product GAR model a valuable addition to the suite of parsimonious covariance structures for multivariate repeated measures data.

Section 4.2 provides a formal definition of the Kronecker Product GAR model as well as parameter estimators and their variances. Simulation studies in section 4.3 assess the performance of the estimation procedure (detailed in 4.2.3) for the Kronecker product GAR covariance model in a moderately large-sample context. I discuss the analysis of an example set of data in section 4.4 and conclude with a summary discussion including planned future research in section 4.5.

4.1.2 Literature Review

Kronecker product covariance structures, also known as separable covariance models, were first introduced by Galecki (1994). A covariance matrix is separable if and only if it can be written as $\mathbf{\Sigma} = \mathbf{\Gamma} \otimes \mathbf{\Omega}$, where $\mathbf{\Gamma}$ and $\mathbf{\Omega}$ are factor specific covariance matrices (e.g. the covariance matrices for the temporal and spatial dimensions of spatio-temporal data respectively). A key advantage of this model lies in the ease of interpretation in terms of the independent contribution of every repeated factor to the overall within-subject error covariance matrix. The model also accommodates covariance matrices with nested parameter spaces and factor specific within-subject variance heterogeneity. Galecki (1994), along with Mitchell et al. (2006) and Naik and Rao (2001), detailed the numerous computational advantages of the Kronecker product

covariance structure. Much easier computations of the partial derivatives, inverse, and Cholesky decomposition of the overall covariance matrix can be performed on the smaller dimensional factor specific models.

Despite the many benefits of separable covariance models, they have limitations. As mentioned by Cressie and Huang (1999), patterns of interaction among the various factors cannot be modeled when utilizing a Kronecker product structure. Galecki (1994), Huizenga et al. (2002), and Mitchell et al. (2006) all noted that a lack of identifiability can result with such a model. The indeterminacy stems from the fact that if $\boldsymbol{\Sigma} = \boldsymbol{\Gamma} \otimes \boldsymbol{\Omega}$ is the overall within-subject error covariance matrix, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Omega}$ are not unique since for $a \neq 0$, $a\boldsymbol{\Gamma} \otimes (1/a)\boldsymbol{\Omega} = \boldsymbol{\Gamma} \otimes \boldsymbol{\Omega}$. However, this nonidentifiability can be fixed by rescaling one of the factor specific covariance matrices so that one of its diagonal nonzero elements is equal to 1. With homogeneous variances, this rescaled matrix is a correlation matrix.

Several tests have been developed to determine the appropriateness of a separable covariance model. Shitan and Brockwell (1995) constructed an asymptotic chi-square test for separability. A likelihood ratio test for separability was derived by Lu and Zimmerman (2005) and Mitchell et al. (2006). Fuentes (2006) developed a test for separability of a spatio-temporal process utilizing spectral methods. All of these tests were developed for complete and balanced data. Given that in many situations multivariate repeated measures data are unbalanced in at least one of the factors, more general tests for separability need to be developed.

There is also a lack of literature on the estimation of separable covariance models when there is an imbalance in at least one dimension. Only Naik and Rao (2001) examined the case in which data may be unbalanced in one of the factors, namely $\boldsymbol{\Sigma}_i = \boldsymbol{\Gamma} \otimes \boldsymbol{\Omega}_i$. This situation arises often in spatio-temporal studies since the data tend to be balanced in space, but not in time. Even though multivariate repeated measures

studies often have data with an imbalance in both dimensions, namely $\boldsymbol{\Sigma}_i = \boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i$, this case has yet to be examined in the literature.

With the assumptions of covariance model separability and homoscedasticity, an equal variance Kronecker product structure has great appeal. In this case the overall within-subject error covariance matrix is defined as $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i$. This formulation has several advantages. The reduction in the number of parameters leads to computational benefits. The model is also identifiable since $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Omega}_i$ will necessarily be correlation matrices.

4.2 Kronecker Product GAR Covariance Model

4.2.1 Notation

Consider the following general linear model for multivariate repeated measures data with the Kronecker product GAR covariance structure:

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{e}_i \tag{4.1}$$

where $\boldsymbol{y}_i$ is an $t_i s_i \times 1$ vector of $t_i s_i$ observations (e.g., t_i temporal measurements and s_i spatial measurements) on the i^{th} subject $i = 1, ..., N$, $\boldsymbol{\beta}$ is a $q \times 1$ vector of fixed and unknown population parameters, $\boldsymbol{X}_i$ is a $t_i s_i \times q$ fixed and known design matrix corresponding to the fixed effects, and $\boldsymbol{e}_i$ is a $t_i s_i \times 1$ vector of random error terms. We assume $\boldsymbol{e}_i \sim N_{t_i s_i}(\mathbf{0}, \boldsymbol{\Sigma}_{ei}(\sigma_e^2, \boldsymbol{\tau}_e) = \sigma_e^2[\boldsymbol{\Gamma}_{ei}(\boldsymbol{\tau}_{e_\gamma}) \otimes \boldsymbol{\Omega}_{ei}(\boldsymbol{\tau}_{e_\omega})])$ and is independent of $\boldsymbol{e}_{i'}$ for $i \neq i'$. It follows that $\boldsymbol{y}_i \sim N_{t_i s_i}(\boldsymbol{X}_i \boldsymbol{\beta}, \sigma_e^2[\boldsymbol{\Gamma}_{ei}(\boldsymbol{\tau}_{e_\gamma}) \otimes \boldsymbol{\Omega}_{ei}(\boldsymbol{\tau}_{e_\omega})])$ and is independent of $\boldsymbol{y}_{i'}$ for $i \neq i'$.

For $\boldsymbol{\Gamma}_{ei} = \{\rho_{ei\gamma;jk}\}$ and $\boldsymbol{\Omega}_{ei} = \{\rho_{ei\omega;lm}\}$, the factor specific *generalized autoregressive* (GAR) correlation structures are

$$\rho_{ei\gamma;jk} = \mathcal{C}(y_{ijl}, y_{ikl}) = \begin{cases} 1 + \left[(d(t_{ijl}, t_{ikl}) - 1)\delta_{e_\gamma}/(D_\gamma - 1)\right] & j \neq k, \\ \rho_{e_\gamma} & \\ 1 & j = k \end{cases} \tag{4.2}$$

$$\rho_{ei\omega;lm} = \mathcal{C}(y_{ijl}, y_{ijm}) = \begin{cases} 1 + [(d(s_{ijl}, s_{ijm}) - 1)\delta_{e_\omega}/(D_\omega - 1)] & l \neq m, \\ \rho_{e_\omega} & \\ 1 & l = m \end{cases} \quad (4.3)$$

where $\mathcal{C}(\,\cdot\,)$ is the correlation operator, $d(t_{ijl}, t_{ikl})$ and $d(s_{ijl}, s_{ikl})$ are the distances between measurement times and locations respectively, D_γ and D_ω are *constants* that can be specified and by default they are set to the maximum number of time and spatial distance units respectively, σ_e^2 is the variability of the measurements at each time-location pair, ρ_{e_γ} and ρ_{e_ω} are the correlations between observations separated by one unit of time and distance respectively, and δ_{e_γ} and δ_{e_ω} are the decay speeds. We assume $0 \leq \rho_{e_\gamma}, \rho_{e_\omega} < 1;\ 0 \leq \delta_{e_\gamma}, \delta_{e_\omega};$ and $D_\gamma, D_\omega > 1$. Ensuring that the factor specific matrices $\mathbf{\Gamma}_{ei}$ and $\mathbf{\Omega}_{ei}$ are positive definite (as discussed in Chapter 2) is sufficient for ensuring the positive definiteness of $\mathbf{\Sigma}_{ei}$. This follows from Corollary 4.1.1 in the Appendix.

Implicit in this model formulation is the presence of both a stationary variance and correlation structure. The AR(1) correlation model is a special case of the factor specific GAR models for which $\delta_e = D - 1$. For values of $\delta_e > D - 1$, the correlation between measurements on a given subject decreases in time or space at a faster rate than for $\delta_e = D - 1$. As $\delta_e \to \infty$, these factor specific models approach the moving average model of order 1, MA(1). For values of δ_e such that $0 < \delta_e < D - 1$, the correlation between measurements on a given subject decreases in time or space at a slower rate than for $\delta_e = D - 1$. When $\delta_e = 0$, these models reduce to the well known compound symmetric correlation model for which the correlation between measurements on a given subject is fixed at ρ_e no matter how far apart in time or space the measurements are taken. Though values of $\delta_e < 0$ yield valid autocorrelation functions for which the correlation between measurements on a given subject would increase with increasing time or distance between measurements, this is rare in biostatistical applications. Therefore the parameter space is restricted for reasons of practicality.

4.2.2 Plots

Graphical depictions of the Kronecker product GAR structure help to provide insight into the types of correlation patterns that can be modeled. A correlation pattern in which both of the factor specific matrices (e.g. spatial and temporal matrices) have a decay rate that is slower than that of the AR(1) model is illustrated in Figure 4.1. Figures 4.2 and 4.3 exhibit patterns with dual AR(1) and faster than AR(1) decay rates respectively.

4.2.3 Maximum Likelihood Estimation

Proper estimation of fixed effect and covariance parameters in a multivariate repeated measures model requires iterative numerical algorithms. Jennrich and Schluchter (1986) detailed the Newton-Raphson and Fisher scoring algorithms for maximum likelihood (ML) estimation of model parameters. Dempster et al. (1977) first proposed the EM algorithm for parameter estimation. The Newton-Raphson method is more widely applicable than the Fisher scoring method due to the computational infeasibility of calculating the expected information matrix in certain contexts. As noted by Lindstrom and Bates (1988), the method is also generally preferable to the EM procedure.

The benefits of profiling out σ^2 and optimizing the resulting profile log-likelihood to derive model parameter estimates via the Newton-Raphson algorithm was discussed by Lindstrom and Bates (1988). They noted that this optimization will generally require fewer iterations, will have simpler derivatives, and the convergence will be more consistent. There are also certain situations in which the Newton-Raphson algorithm may fail to converge when optimizing the original log-likelihood but converge with ease when utilizing the profile log-likelihood.

In order to estimate the parameters of the model defined in equation 4.1, σ_e^2 is first profiled out of the likelihood. The first and second partial derivatives of the profile log-likelihood are then employed to compute ML estimates of parameters, via the Newton-

Raphson algorithm, and the variance-covariance matrix of those estimates. The resulting estimates are then used to compute the value and variance of $\widehat{\sigma}_e^2$. A SAS IML (SAS Institute, 2002) computer program has been written implementing this estimation procedure for the general linear model with a Kronecker product GAR covariance structure and is available upon request.

Setting $\boldsymbol{\Sigma}_{ei} = \sigma_e^2[\boldsymbol{\Gamma}_{ei}(\boldsymbol{\tau}_{e_\gamma}) \otimes \boldsymbol{\Omega}_{ei}(\boldsymbol{\tau}_{e_\omega})]$, where $\boldsymbol{\tau}_e = \{\boldsymbol{\tau}_{e_\gamma}; \boldsymbol{\tau}_{e_\omega}\} = \{\delta_{e_\gamma}, \rho_{e_\gamma}; \delta_{e_\omega}, \rho_{e_\omega}\}$, the log-likelihood function of the parameters given the data under the model is:

$$\begin{aligned} l(\boldsymbol{y}; \boldsymbol{\beta}, \sigma_e^2, \boldsymbol{\tau}_e) &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln|\sigma_e^2\boldsymbol{\Gamma}_{ei} \otimes \boldsymbol{\Omega}_{ei}| - \frac{1}{2\sigma_e^2}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei} \otimes \boldsymbol{\Omega}_{ei})^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\left(t_i s_i \ln(\sigma_e^2) + \ln|\boldsymbol{\Gamma}_{ei} \otimes \boldsymbol{\Omega}_{ei}|\right) - \\ &\quad \frac{1}{2\sigma_e^2}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei} \otimes \boldsymbol{\Omega}_{ei})^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}), \end{aligned} \tag{4.4}$$

where $n = \sum_{i=1}^{N} t_i s_i$. Taking the first partial derivative with respect to σ_e^2 gives

$$\frac{\partial l}{\partial \sigma_e^2} = -\frac{1}{2}\sum_{i=1}^{N} t_i s_i \sigma_e^{-2} + \frac{1}{2\sigma_e^4}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta}).$$

Setting the derivative to zero and solving for an estimate of the variance yields

$$\widehat{\sigma}_{eML}^2(\boldsymbol{\beta}, \boldsymbol{\tau}_e) = \frac{1}{n}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta}).$$

Substituting $\widehat{\sigma}_{eML}^2(\boldsymbol{\beta}, \boldsymbol{\tau}_e)$ into the log-likelihood function in equation 4.4 leads to the following profile log-likelihood:

$$\begin{aligned} l_p(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\tau}_e) = &-\frac{1}{2}\sum_{i=1}^{N}\ln|\boldsymbol{\Gamma}_{ei} \otimes \boldsymbol{\Omega}_{ei}| - \\ &\frac{1}{2}n\ln\left[\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\right] - \frac{1}{2}n\ln\left(\frac{1}{n}\right) - \frac{n}{2}. \end{aligned} \tag{4.5}$$

To avoid computational issues it is best to use the equality

$$\ln|\boldsymbol{\Gamma}_{ei} \otimes \boldsymbol{\Omega}_{ei}| = s_i \ln|\boldsymbol{\Gamma}_{ei}| + t_i \ln|\boldsymbol{\Omega}_{ei}|$$

in case $|\boldsymbol{\Gamma}_{ei} \otimes \boldsymbol{\Omega}_{ei}|$ is close to zero.

The first partial derivative with respect to $\boldsymbol{\beta}$ in equation 4.5 is

$$\partial l_p / \partial \boldsymbol{\beta} = n\Big(\sum\nolimits_{i=1}^{N} \boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\Big)^{-1} \sum\nolimits_{i=1}^{N} \boldsymbol{X}_i'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta}).$$

Setting the previous equation to zero and solving implies $\sum_{i=1}^{N} \boldsymbol{X}_i'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i = \mathbf{0}$. If $\Big(\sum_{i=1}^{N} \boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\Big)^{-1} = 0$ the likelihood would be degenerate.

Therefore

$$\widehat{\boldsymbol{\beta}}_{ML}(\boldsymbol{\tau}_e) = \Big(\sum\nolimits_{i=1}^{N} \boldsymbol{X}_i'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{X}_i\Big)^{-1} \Big(\sum\nolimits_{i=1}^{N} \boldsymbol{X}_i'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{y}_i\Big).$$

The remaining first partial derivatives are as follows:

$$\begin{aligned}\frac{\partial l_p}{\partial \rho_{e_\gamma}} = &-\frac{1}{2}\sum_{i=1}^{N} s_i \mathrm{tr}\left[\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_{e_\gamma}}\right] + \frac{n}{2}\Bigg[\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1} \times \\ &\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\left(\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_{e_\gamma}} \otimes \boldsymbol{\Omega}_{ei}\right)(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\Bigg]\end{aligned}$$

$$\begin{aligned}\frac{\partial l_p}{\partial \rho_{e_\omega}} = &-\frac{1}{2}\sum_{i=1}^{N} t_i \mathrm{tr}\left[\boldsymbol{\Omega}_{ei}^{-1}\frac{\partial \boldsymbol{\Omega}_{ei}}{\partial \rho_{e_\omega}}\right] + \frac{n}{2}\Bigg[\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1} \\ &\times \sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\left(\boldsymbol{\Gamma}_{ei} \otimes \frac{\partial \boldsymbol{\Omega}_{ei}}{\partial \rho_{e_\omega}}\right)(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\Bigg]\end{aligned}$$

$$\frac{\partial l_p}{\partial \delta_{e_\gamma}} = -\frac{1}{2}\sum_{i=1}^{N} s_i \text{tr}\left[\boldsymbol{\Gamma}_{ei}^{-1}\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_{e_\gamma}}\right] + \frac{n}{2}\left[\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1}\otimes\boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1}\times \right.$$
$$\left.\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1}\otimes\boldsymbol{\Omega}_{ei}^{-1})\left(\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_{e_\gamma}}\otimes\boldsymbol{\Omega}_{ei}\right)(\boldsymbol{\Gamma}_{ei}^{-1}\otimes\boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\right]$$

$$\frac{\partial l_p}{\partial \delta_{e_\omega}} = -\frac{1}{2}\sum_{i=1}^{N} t_i \text{tr}\left[\boldsymbol{\Omega}_{ei}^{-1}\frac{\partial \boldsymbol{\Omega}_{ei}}{\partial \delta_{e_\omega}}\right] + \frac{n}{2}\left[\left(\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1}\otimes\boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\right)^{-1}\times \right.$$
$$\left.\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1}\otimes\boldsymbol{\Omega}_{ei}^{-1})\left(\boldsymbol{\Gamma}_{ei}\otimes\frac{\partial \boldsymbol{\Omega}_{ei}}{\partial \delta_{e_\omega}}\right)(\boldsymbol{\Gamma}_{ei}^{-1}\otimes\boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})\right],$$

where $\partial\boldsymbol{\Gamma}_{ei}/\partial\rho_{e_\gamma}$ and $\partial\boldsymbol{\Gamma}_{ei}/\partial\delta_{e_\gamma}$ are $t_i \times t_i$ matrices with $(j,k)^{th}$ element for $j,k \in \{1,...,t_i\}$:

$$\left(\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \rho_{e_\gamma}}\right)_{jk} = \begin{cases} \left(1+\frac{[d(t_{ijl},t_{ikl})-1]\delta_{e_\gamma}}{(D_\gamma-1)}\right)\rho_{e_\gamma}^{\left[(d(t_{ijl},t_{ikl})-1)\delta_{e_\gamma}/(D_\gamma-1)\right]} & j\neq k \\ 0 & j=k \end{cases},$$
$$\left(\frac{\partial \boldsymbol{\Gamma}_{ei}}{\partial \delta_{e_\gamma}}\right)_{jk} = \begin{cases} \left(\rho_{e_\gamma}^{1+\left[(d(t_{ijl},t_{ikl})-1)\delta_{e_\gamma}/(D_\gamma-1)\right]}\right)\left(\ln\rho_{e_\gamma}\right)\frac{[d(t_{ijl},t_{ikl})-1]}{(D_\gamma-1)} & j\neq k \\ 0 & j=k \end{cases},$$

$\partial\boldsymbol{\Omega}_{ei}/\partial\rho_{e_\omega}$ and $\partial\boldsymbol{\Omega}_{ei}/\partial\delta_{e_\omega}$ are $s_i \times s_i$ matrices with $(l,m)^{th}$ element for $l,m \in \{1,...,s_i\}$ same as above with ρ_{e_ω}, δ_{e_ω}, and D_ω replacing ρ_{e_γ}, δ_{e_γ}, and D_γ.

The ML estimates of the model parameters are computed by utilizing the Newton-Raphson algorithm with the above first partial derivatives and the second partial derivatives of the profile log-likelihood. The second partial derivatives of the parameters, which are also employed to determine the asymptotic variance-covariance matrix of the estimators, are approximated by finite difference formulas. These derivative approximations are detailed in Abramowitz and Stegun (1972) and Dennis and Schnabel (1983). The $\binom{5}{2}+5=15$ analytic second derivatives can be derived explicitly as in Chapter 1. However, the approximations have proven very accurate.

After obtaining the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}_e$ utilizing the Newton-Raphson algorithm, an estimate of σ_e^2 is calculated by substituting the estimates into $\widehat{\sigma}_{e\,ML}^2(\boldsymbol{\beta},\boldsymbol{\tau}_e)$. An

estimator of the variance for $\widehat{\sigma}^2_{eML}(\boldsymbol{\beta}, \boldsymbol{\tau}_e)$, assuming that $\boldsymbol{\beta}$ and $\boldsymbol{\tau}_e$ are known, is as follows. We have that

$$\widehat{\sigma}^2_{eML}(\boldsymbol{\beta}, \boldsymbol{\tau}_e) = \frac{1}{n}\sum\nolimits_{i=1}^{N} \boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta}),$$

where $\boldsymbol{r}_i(\boldsymbol{\beta}) = (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \sim \boldsymbol{N}_{s_i t_i}(\boldsymbol{0}, \sigma_e^2[\boldsymbol{\Gamma}_{ei} \otimes \boldsymbol{\Omega}_{ei}])$. Quadratic forms theory gives $\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})/\sigma_e^2 \sim \chi^2_{s_i t_i}$. In turn,

$$\begin{aligned} \mathcal{V}\big(\widehat{\sigma}_e^2\big) &= \mathcal{V}\Big[\textstyle\sum_{i=1}^{N} \boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})/n\Big] \\ &= \textstyle\sum_{i=1}^{N} \mathcal{V}[\boldsymbol{r}_i(\boldsymbol{\beta})'(\boldsymbol{\Gamma}_{ei}^{-1} \otimes \boldsymbol{\Omega}_{ei}^{-1})\boldsymbol{r}_i(\boldsymbol{\beta})]/n^2 \end{aligned}$$

since $\boldsymbol{y}_i \perp \boldsymbol{y}_{i'}$ for $i \neq i'$. If $Q_i \sim \chi^2(s_i t_i)$, then $\mathcal{V}\big(\widehat{\sigma}_e^2\big) = \sum_{i=1}^{N} \mathcal{V}(\sigma_e^2 Q_i)/n^2 = 2\sigma_e^4/n$. Therefore $\widehat{\mathcal{V}}\big[\widehat{\sigma}^2_{eML}(\boldsymbol{\beta}, \boldsymbol{\tau}_e)\big] = 2\widehat{\sigma}_e^4/n$.

The asymptotic variance-covariance matrix of the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}_e$ is given by the Hessian matrix (the observed information matrix) below:

$$\boldsymbol{H}_{\boldsymbol{\beta}\boldsymbol{\tau}_e} = \begin{bmatrix} \dfrac{\partial^2 l_p}{\partial \boldsymbol{\beta}' \partial \boldsymbol{\beta}} & \dfrac{\partial^2 l_p}{\partial \boldsymbol{\beta} \partial \delta_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \boldsymbol{\beta} \partial \rho_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \boldsymbol{\beta} \partial \delta_{e\omega}} & \dfrac{\partial^2 l_p}{\partial \boldsymbol{\beta} \partial \rho_{e\omega}} \\ \dfrac{\partial^2 l_p}{\partial \boldsymbol{\beta} \partial \delta_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \delta_{e\gamma}^2} & \dfrac{\partial^2 l_p}{\partial \rho_{e\gamma} \partial \delta_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \delta_{e\omega} \partial \delta_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \rho_{e\omega} \partial \delta_{e\gamma}} \\ \dfrac{\partial^2 l_p}{\partial \boldsymbol{\beta} \partial \rho_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \rho_{e\gamma} \partial \delta_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \rho_{e\gamma}^2} & \dfrac{\partial^2 l_p}{\partial \delta_{e\omega} \partial \rho_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \rho_{e\omega} \partial \rho_{e\gamma}} \\ \dfrac{\partial^2 l_p}{\partial \boldsymbol{\beta} \partial \delta_{e\omega}} & \dfrac{\partial^2 l_p}{\partial \delta_{e\omega} \partial \delta_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \delta_{e\omega} \partial \rho_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \delta_{e\omega}^2} & \dfrac{\partial^2 l_p}{\partial \rho_{e\omega} \partial \delta_{e\omega}} \\ \dfrac{\partial^2 l_p}{\partial \boldsymbol{\beta} \partial \rho_{e\omega}} & \dfrac{\partial^2 l_p}{\partial \rho_{e\omega} \partial \delta_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \rho_{e\omega} \partial \rho_{e\gamma}} & \dfrac{\partial^2 l_p}{\partial \rho_{e\omega} \partial \delta_{e\omega}} & \dfrac{\partial^2 l_p}{\partial \rho_{e\omega}^2} \end{bmatrix}. \quad (4.6)$$

The estimated asymptotic variance-covariance matrix of the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}_e$ is simply calculated by substituting the ML estimates of these parameters into equation 4.6 and taking its inverse.

From the previous derivations we can see that $\widehat{\sigma}_e^2$ is consistent (i.e., $\widehat{\sigma}_e^2 \rightarrow_p \sigma_e^2$) since we have that $E\big(\widehat{\sigma}_e^2\big) = \sigma_e^2$ and that $\mathcal{V}\big(\widehat{\sigma}_e^2\big) \rightarrow 0$ as $n \rightarrow \infty$. To establish the consistency of the remaining estimators, an examination of the asymptotic properties of the inverse of

the observed information matrix is necessary $\left(\boldsymbol{H}_{\boldsymbol{\beta}\boldsymbol{\tau}_e}^{-1}\right)$. As noted in Vonesh and Chinchilli (1997), it is well known that under normal theory likelihood estimation we have that the asymptotic distributions of the ML estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}_e$ are

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \rightarrow_d N_q(\mathbf{0}, \boldsymbol{\Sigma}_\beta)$$

and

$$\sqrt{N}(\widehat{\boldsymbol{\tau}}_e - \boldsymbol{\tau}_e) \rightarrow_d N_2(\mathbf{0}, \boldsymbol{\Sigma}_{\tau_e})$$

provided that the limits, $\boldsymbol{\Sigma}_\beta = \lim_{N\to\infty}\left(N\boldsymbol{H}_{\boldsymbol{\beta}}^{-1}\right) = \left[\lim_{N\to\infty}\left(N^{-1}\frac{\partial^2 l_p}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}}\right)\right]^{-1}$ and

$\boldsymbol{\Sigma}_{\tau_e} = \lim_{N\to\infty}\left(N\boldsymbol{H}_{\tau_e}^{-1}\right) = \left[\lim_{N\to\infty}\left(N^{-1}\begin{bmatrix} \frac{\partial^2 l_p}{\partial\boldsymbol{\delta}_e^2} & \frac{\partial^2 l_p}{\partial\boldsymbol{\rho}_e\partial\boldsymbol{\delta}_e} \\ \frac{\partial^2 l_p}{\partial\boldsymbol{\rho}_e\partial\boldsymbol{\delta}_e} & \frac{\partial^2 l_p}{\partial\boldsymbol{\rho}_e^2}\end{bmatrix}\right)\right]^{-1}$ exist. For "reasonable" values of t_i and s_i the limits should exist, though further investigation is needed. If the limits exist, the estimators will be consistent and fully asymptotically efficient.

4.2.4 Computational Issues

A complication that may arise when implementing the Kronecker product GAR covariance model is that the proposed estimation method can produce negative variance estimates for the correlation parameters. This may occur for the parameters of either one or both of the factor specific matrices when there is a faster decay rate than that imposed by the AR(1) model coupled with a 'small' ρ_e. The instability of the second order derivatives of the objective function resulting from the small, quickly decaying correlation leads to this problem.

One approach would be to implement an estimation method which utilizes only first order derivatives such as a quasi-Newton procedure. An often used algorithm for this first order iterative procedure is an efficient modification of Powell's (1978, 1982) Variable Metric Constrained WatchDog (VMCWD) algorithm that uses a quadratic

programming subroutine which updates and downdates the Cholesky factor as detailed by Gill et al. (1984). However, quasi-Newton approaches generally have worse stability and convergence properties than the Newton-Raphson method. The estimates produced also tend not to be as accurate as those from the Newton-Raphson method. Another approach is to recognize this complication as a diagnostic tool. Since a covariance matrix of this nature is approximately equal to the identity matrix, an independence model should be fit for the factor specific structure in this situation.

4.3 Simulations

To assess the empirical performance of the estimation procedure (detailed in 4.2.3) for the Kronecker product GAR covariance model, data were generated under the general linear model for multivariate repeated measures data with the Kronecker product GAR covariance structure with four different parameter sets. Only the complete and balanced case was considered with $N = 100$ subjects, and $t = t_i = 10 \times s = s_i = 10$ observations each at two-unit distance intervals (thus $D_\gamma = D_\omega = 18$). The fixed effects included an intercept, a dummy variable indicating membership in one of two groups (50 subjects per group), and a continuous repeated covariate, with $\beta = [1, 1, 1]'$. Simulation results showed that the estimation approach performs best when both correlation matrices exhibit a pattern of high initial correlation with a slow decay rate (the 'optimal' case). Conversely, the procedure performs worst when both correlation matrices exhibit a pattern of low initial correlation with a fast decay rate (the 'suboptimal' case). I focus on these extreme cases, but include two intermediate parameter sets for completeness.

For parameter set 1, we let $\boldsymbol{\rho}_e = [\,\rho_{e_{\gamma 1}} \quad \rho_{e_{\omega 1}}\,]' = [\,0.9 \quad 0.9\,]'$ and $\boldsymbol{\delta}_e = [\,\delta_{e_{\gamma 1}} \quad \delta_{e_{\omega 1}}\,]' = [\,(D-1)/4 \quad (D-1)/4\,]' = [\,4.25 \quad 4.25\,]'$ corresponding to the optimal case. The two intermediate cases, parameter sets 2 and 3, have $\boldsymbol{\rho}_e = [\,\rho_{e_{\gamma 2}} \quad \rho_{e_{\omega 2}}\,]' = [\,0.9 \quad 0.9\,]'$ with $\boldsymbol{\delta}_e = [\,\delta_{e_{\gamma 2}} \quad \delta_{e_{\omega 2}}\,]' = [\,4(D-1) \quad 4(D-1)\,]' = [\,68 \quad 68\,]'$, and

$\boldsymbol{\rho}_e = [\,\rho_{e_{\gamma 3}} \quad \rho_{e_{\omega 3}}\,]' = [\,0.5 \quad 0.5\,]'$ with $\boldsymbol{\delta}_e = [\,\delta_{e_{\gamma 3}} \quad \delta_{e_{\omega 3}}\,]' = [\,(D-1)/4 \quad (D-1)/4\,]' = [\,4.25 \quad 4.25\,]'$. For parameter set 4, corresponding to the worst case, we let $\boldsymbol{\rho}_e = [\,\rho_{e_{\gamma 4}} \quad \rho_{e_{\omega 4}}\,]' = [\,0.5 \quad 0.5\,]'$ and $\boldsymbol{\delta}_e = [\,\delta_{e_{\gamma 4}} \quad \delta_{e_{\omega 4}}\,]' = [\,D \quad D\,]' = [\,18 \quad 18\,]'$. The decay speed parameter values in set 4 are not set larger due to the complications discussed in 4.2.4. It is important to note that the combination of an 'optimal' and 'suboptimal' correlation matrix will also lead to intermediate results as in sets 2 and 3. Each simulation for the four parameter sets consisted of $10,000$ realizations with the scale parameter set to $\sigma_e^2 = 10$. All model fits employed ML estimation with the profile likelihood as discussed in 4.2.3.

The estimation procedure was assessed by evaluating the estimators, their standard deviations, and the average standard error of the estimators for the four parameter sets. Table 4.1 reports these simulation results. It is evident that the estimation approach performs well because the bias is relatively small, and the estimated standard error is close to the sample standard deviation for the estimators. Confirming preliminary results, its performance declines as the decay rate increases and the initial correlation decreases. This is evidenced by the increase in bias, standard deviation, and difference between the standard deviation and average standard error for the estimators for sets 1 and 2, 3 and 4, and 1 and 3. As a side note, it should be pointed out that there is virtually no bias in fixed effect estimation for any of the four scenarios (Demidenko, 2004, reviews proofs of fixed effect unbiasedness).

4.4 Example

Schizophrenia is a mental illness characterized by disabling manifestations of impairments in the perception or expression of reality. Longitudinal studies have shown that the pathomorphologic brain changes occurring in schizophrenics may be progressive and associated with clinical outcome. Thus, much recent work has focused on the effect of antipsychotic drugs on brain morphology. The caudate (shown in Figure 4.4), an

important part of the brain's learning and memory system, has been one target of these drugs.

Our analysis includes longitudinal MRI scans of the left caudate for 240 schizophrenia patients and 56 controls. The surface of each object was parameterized via the m-rep method as described in Styner and Gerig (2001). The caudate shape was determined as a 3 x 7 grid of mesh points (see Figure 4.5). Data were reduced to one outcome measure: *Radius* in cm as a measure of local object width (21 locations per caudate). The distance between two radii for a given subject was calculated as the mean Euclidian distance over all images. Scans were taken up to 47 months post-baseline with the median and maximum number of scans per subject being 3 and 7 respectively. The schizophrenia patients were randomized to either haloperidol (a conventional antipsychotic) or olanzapine (an atypical antipsychotic). These two groups were combined into one treatment group for the purposes of our analysis. The other covariates of interest were age, gender, and race. The shape of the caudate, and thus the radii, have been shown to be significantly different at baseline between schizophrenics and controls. The study hypothesized that the neuroprotective effect of the drugs would lead to no overall differences in shape between the patients and controls.

We model the data with the general linear model for multivariate repeated measures data. The initial full model is as follows:

$$\boldsymbol{y}_i = \beta_0 + \beta_1 \boldsymbol{X}_{i,\text{trt}} + \beta_2 \boldsymbol{X}_{i,\text{age}} + \beta_3 \boldsymbol{X}_{i,\text{gen}} + \beta_4 \boldsymbol{X}_{i,\text{race}} + \boldsymbol{e}_i. \tag{4.7}$$

The $\log_2$(radius) values for each of the $s_i = s = 21$ locations (spatial factor) and t_i images (temporal factor) for each subject are contained in $\boldsymbol{y}_i$ $(t_i \cdot 21 \times 1)$. The vectors $\boldsymbol{X}_{i,\text{trt}}$, $\boldsymbol{X}_{i,\text{gen}}$, and $\boldsymbol{X}_{i,\text{rac}}$ indicate the treatment group (patients and controls), gender, and race of the i^{th} subject respectively. The ages at baseline are contained in $\boldsymbol{X}_{i,\text{age}}$.

We model the temporal and spatial factor specific covariances of the within-subject errors with the continuous-time AR(1), DE, and GAR structures (the three main models

for exponentially decaying correlation patterns discussed in Chapters 1-3) in order to assess the best model via the BIC. Table 4.2 contains the BIC values for all nine possible covariance model fits. Modeling both the temporal and spatial covariances with the GAR provides the best model fit.

We continue the analysis employing the Kronecker product GAR covariance model. In order to obtain a parsimonious model, the full model defined in equation 4.7 is reduced via backward selection with $\alpha = 0.20$. The final model after reduction is

$$\boldsymbol{y}_i = \beta_0 + \boldsymbol{e}_i. \tag{4.8}$$

Thus, as expected, there is no difference in caudate shape between the treated schizophrenics and the controls when taking into account all images taken over time.

The residual variance estimate and correlation parameter estimates of the Kronecker product GAR structure (defined in equation 4.1) for the final data model are given in Table 4.3. Graphical depictions of these estimates are exhibited in Figures 4.6 and 4.7, which show the observed vs. predicted correlation patterns as a function of the months between images and millimeters between radii respectively, starting with the minimum temporal and spatial distances for the data. As evidenced by Figure 4.6, the temporal factor specific GAR covariance structure is able to model a correlation function in which the correlation remains high regardless of how far apart in time the images are taken. The spatial correlations, shown in Figure 4.7, are modest for radii that are close, and then decay slowly toward zero as they become farther apart. The predicted correlation as a function of both factors is exhibited in Figure 4.8. This illustration of the predicted overall within-subject correlation function again displays the slow spatial decay pattern and the near constant temporal pattern. The utility of the Kronecker product covariance model lies in both the flexibility of the factor specific models as well as the interpretability stemming from the Kronecker product structure.

4.5 Discussion and Conclusions

As shown by the simulations in section 4.3, we can be confident in the estimation of the Kronecker product GAR covariance model parameters. The method allows for the modeling and understanding of two factor specific correlation patterns even when there is an imbalance in both dimensions across subjects. This five-parameter structure has excellent analytic and numerical properties that make it especially attractive for the High Dimension, Low Sample Size case so common in longitudinal medical imaging and various kinds of longitudinal "-omics" data. Analysis of the caudate data nicely illustrates the interpretability of the model in a complex context.

There are many possible directions for future research with the proposed Kronecker product GAR covariance model. Any combination of allowing for the estimation of D_γ and D_ω as parameters and including higher order polynomial functions in the exponent of the factor specific models would further increase its flexibility. An assessment of the model's small sample performance and robustness to misspecification is a priority for future investigation. Also, introducing a nonstationary Kronecker product GAR covariance model may prove extremely useful in neuroimaging studies of the developing brain since the variability of brain characteristics tends to change across time and space. The model could have a nonstationary variance and/or correlation structure. For data that have within-subject correlations induced by three or more factors, as in longitudinal imaging data represented via the m-rep method (Pizer et al., 2002 has details), the generalization of the Kronecker product GAR covariance model to n repeated factors would be beneficial.

Table 4.1 Summary of Simulation Results

	Simulated Model							
Set	$\boldsymbol{\rho}_e$(value)	$\boldsymbol{\delta}_e$(value)	Mean$(\widehat{\boldsymbol{\rho}}_e)$	Mean$\left(\widehat{\boldsymbol{\delta}}_e\right)$	SD$(\widehat{\boldsymbol{\rho}}_e)$	Mean$\left(\widehat{\text{SE}}\right)$	SD$\left(\widehat{\boldsymbol{\delta}}_e\right)$	Mean$\left(\widehat{\text{SE}}\right)$
1	$\rho_{e_{\gamma 1}}$(.9)	$\delta_{e_{\gamma 1}}$(4.25)	0.900	4.256	0.004	0.004	0.227	0.226
	$\rho_{e_{\omega 1}}$(.9)	$\delta_{e_{\omega 1}}$(4.25)	0.900	4.255	0.004	0.004	0.227	0.226
2	$\rho_{e_{\gamma 2}}$(.9)	$\delta_{e_{\gamma 2}}$(68)	0.901	70.027	0.011	0.011	11.650	11.180
	$\rho_{e_{\omega 2}}$(.9)	$\delta_{e_{\omega 2}}$(68)	0.900	69.602	0.011	0.011	11.143	11.045
3	$\rho_{e_{\gamma 3}}$(.5)	$\delta_{e_{\gamma 3}}$(4.25)	0.500	4.282	0.013	0.013	0.384	0.382
	$\rho_{e_{\omega 3}}$(.5)	$\delta_{e_{\omega 3}}$(4.25)	0.500	4.274	0.013	0.013	0.376	0.382
4	$\rho_{e_{\gamma 4}}$(.5)	$\delta_{e_{\gamma 4}}$(18)	0.509	19.744	0.054	0.053	9.490	6.578
	$\rho_{e_{\omega 4}}$(.5)	$\delta_{e_{\omega 4}}$(18)	0.509	19.698	0.054	0.053	7.364	6.414

Table 4.2 BIC Values for all Combinations of Factor Specific Covariance Model Fits for the Initial Caudate Data Model

	Spatial Model		
Temporal Model	GAR	DE	AR(1)
GAR	− 14,213.3	− 13,818.5	− 12,639.6
DE	− 14,210.2	− 13,815.5	− 12,636.8
AR(1)	− 10,300.2	− 9,906.1	− 8,698.7

Table 4.3 Final Kronecker Product GAR Structure Covariance Model Estimates for Caudate Data

Factor	Parameter	Estimate	SE
–	σ_e^2	0.4047	0.0045
Time	ρ_{e_γ}	0.9915	0.0002
	$\delta_{e_\gamma}/(D_\gamma - 1)$	0.0026	0.0012
Space	ρ_{e_ω}	0.3806	0.0108
	$\delta_{e_\omega}/(D_\omega - 1)$	0.0402	0.0039

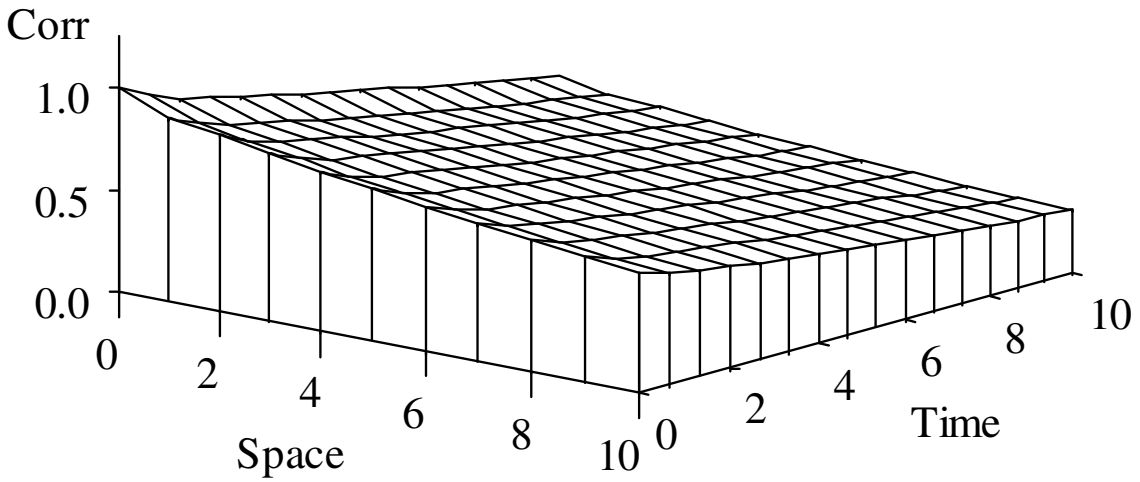

Figure 4.1. Plot of correlation as a function of spatial and temporal distance when both factor specific matrices have a decay rate that is slower than that of the AR(1) model. The correlation parameters are $\{\rho_{e_{\gamma 1}} = 0.9, \rho_{e_{\omega 1}} = 0.9\}$ and $\{\delta_{e_\gamma}/(D_\gamma - 1) = 0.5, \delta_{e_\omega}/(D_\omega - 1) = 0.5\}$.

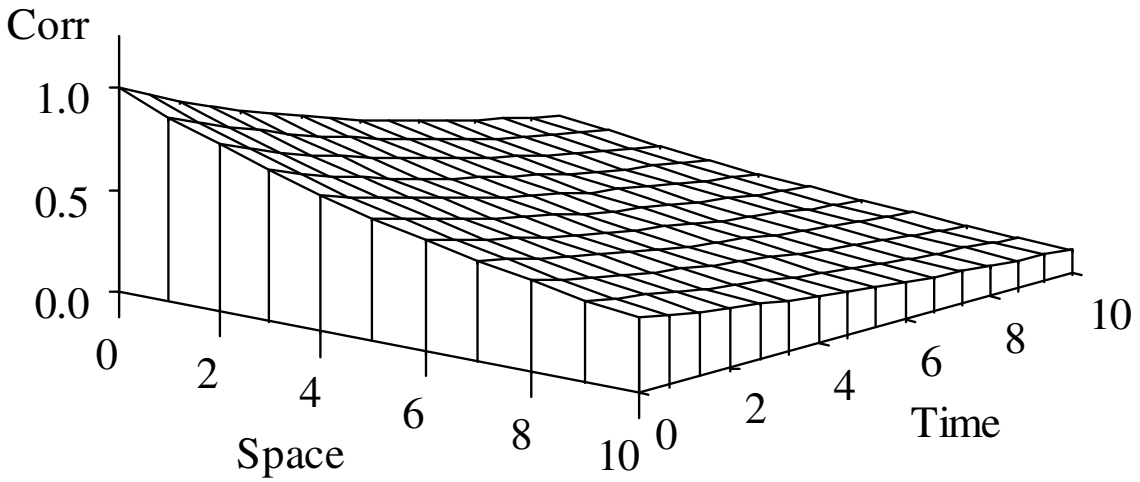

Figure 4.2. Plot of correlation as a function of spatial and temporal distance when both factor specific matrices have an AR(1) decay rate. The correlation parameters are $\{\rho_{e_{\gamma 1}} = 0.9, \rho_{e_{\omega 1}} = 0.9\}$ and $\{\delta_{e_\gamma}/(D_\gamma - 1) = 1, \delta_{e_\omega}/(D_\omega - 1) = 1\}$.

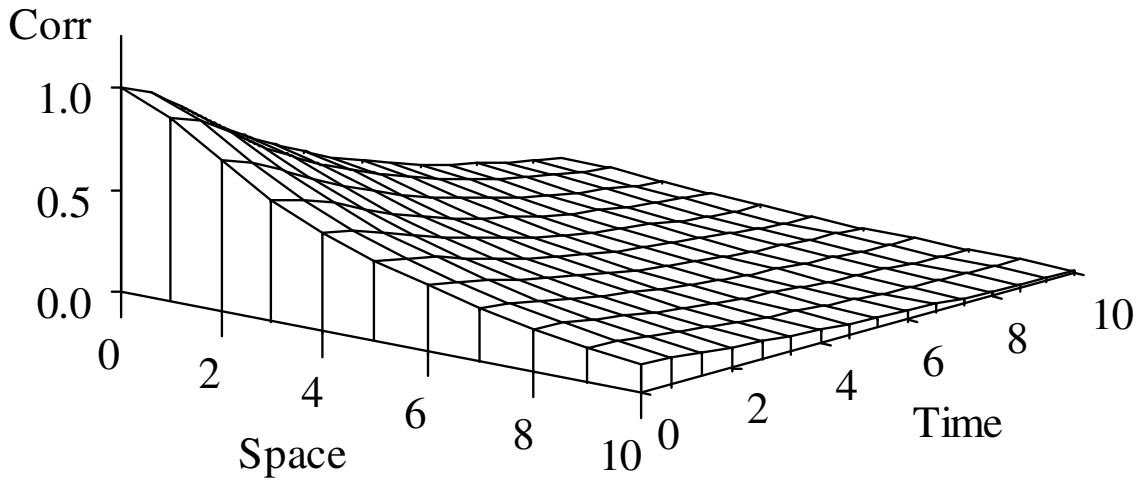

Figure 4.3. Plot of correlation as a function of spatial and temporal distance when both factor specific matrices have a decay rate that is faster than that of the AR(1) model. The correlation parameters are $\{\rho_{e_{\gamma 1}} = 0.9, \rho_{e_{\omega 1}} = 0.9\}$ and $\{\delta_{e_\gamma}/(D_\gamma - 1) = 2, \delta_{e_\omega}/(D_\omega - 1) = 2\}$.

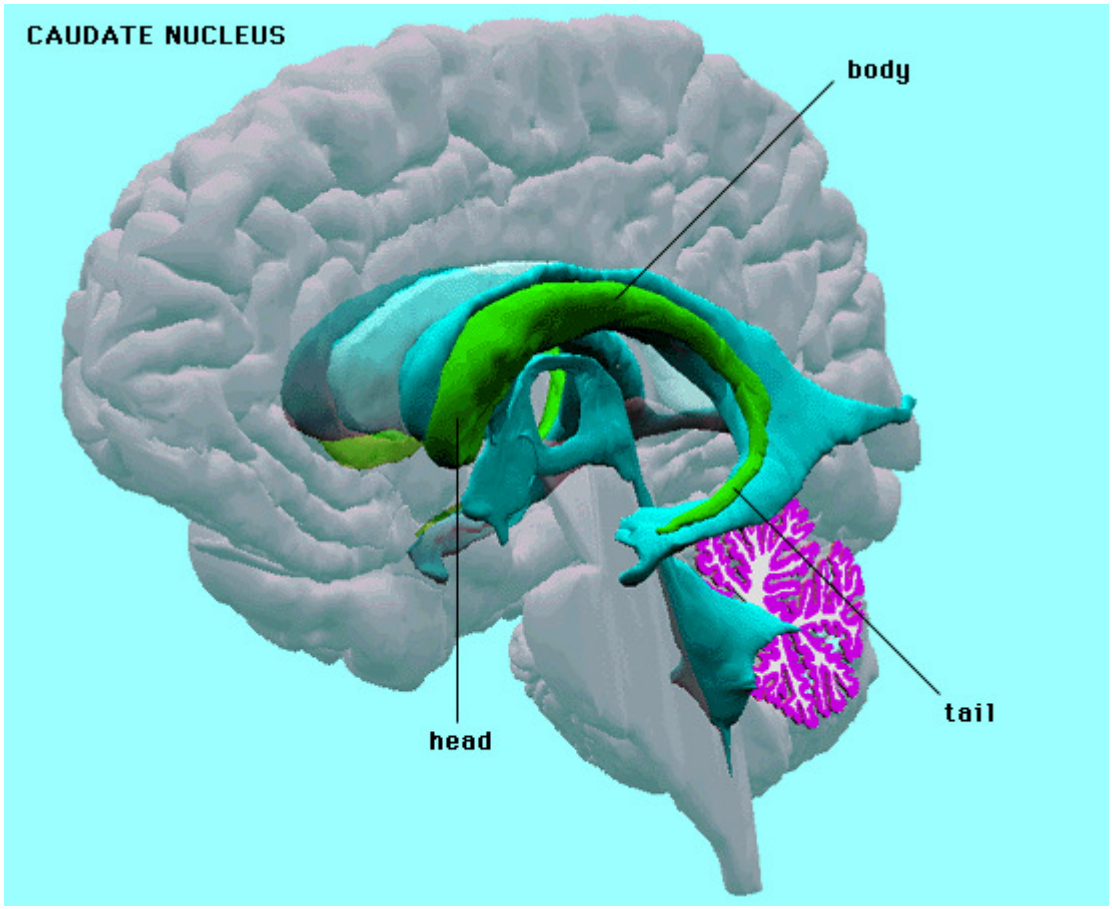

Figure 4.4 The Caudate Nuclei in the Human Brain.

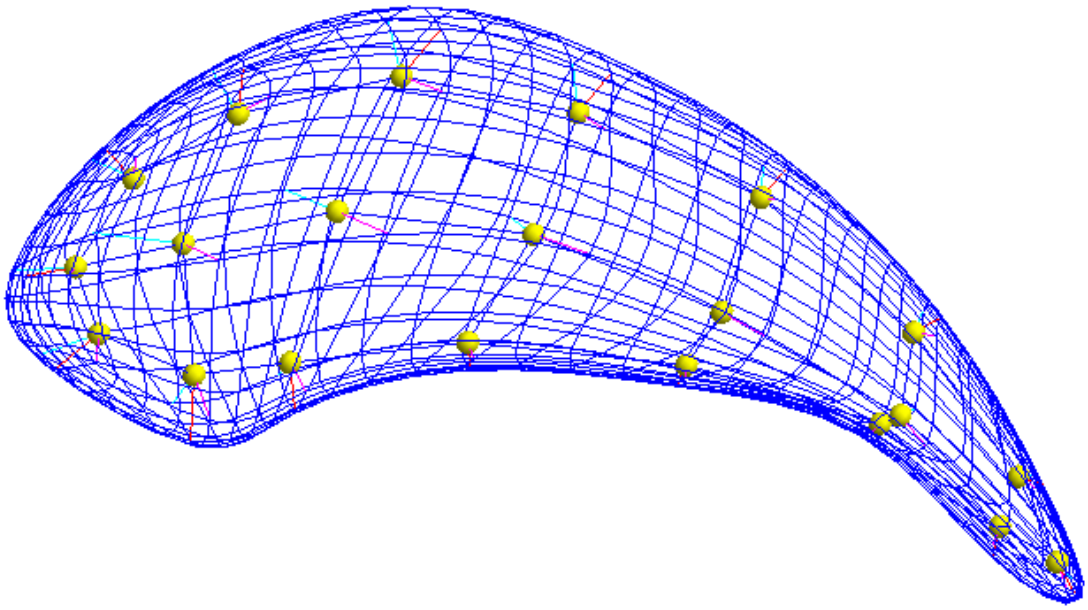

Figure 4.5 M-rep Shape Representation Model of the Caudate.

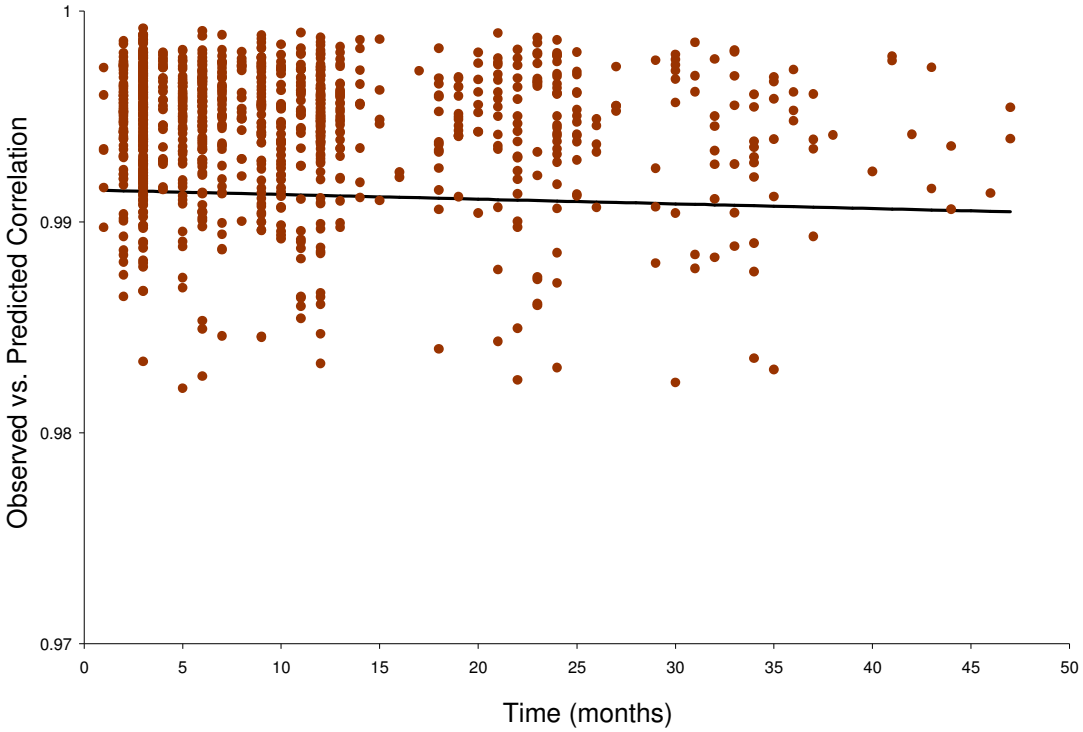

Figure 4.6 Observed vs. Predicted correlation as a function of the time between images.

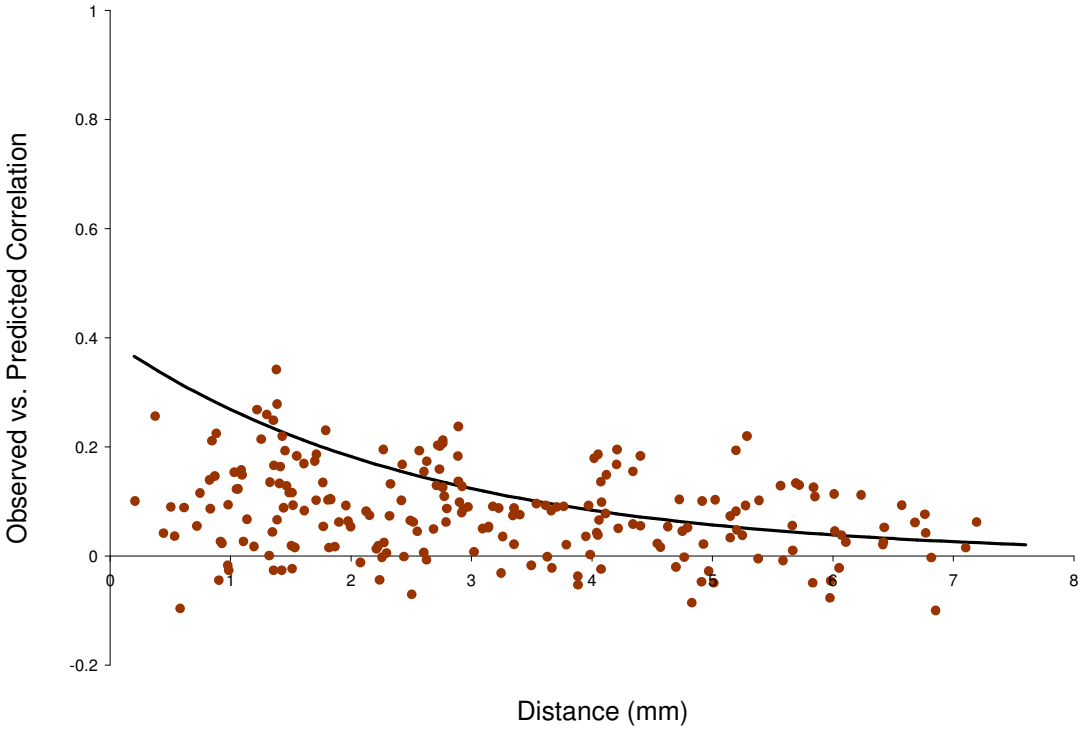

Figure 4.7 Observed vs. Predicted correlation as a function of the distance between radius locations.

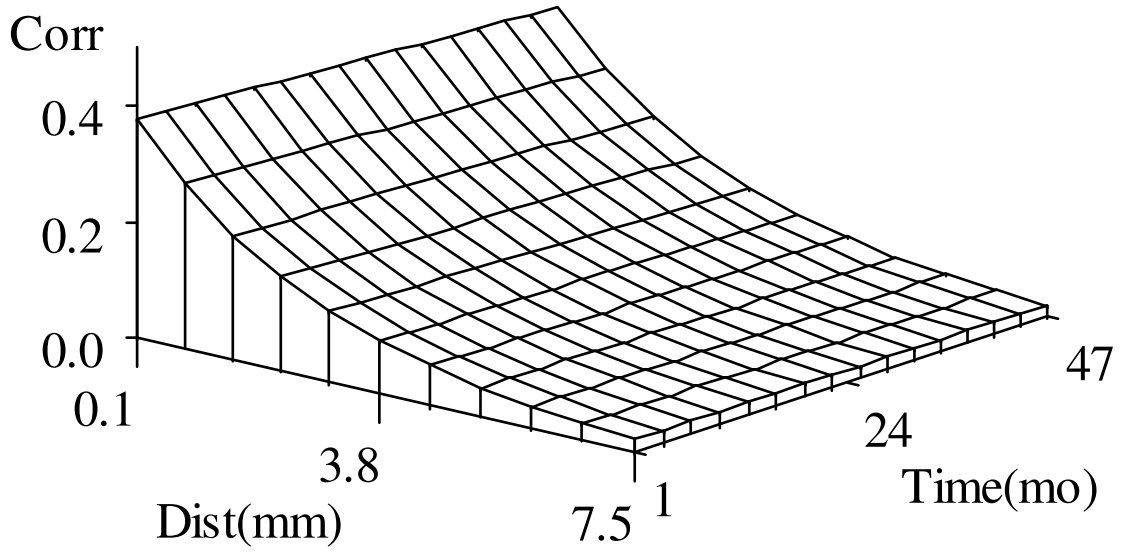

Figure 4.8 Predicted correlation as a function of the distance between radius locations and time between images.

References

Abramowitz, M., and Stegun, I. A. (1972), *Handbook of Mathematical Functions*, New York: Dover Publications, Inc.

Cody, H., Pelphrey, K., and Piven, J. (2002), "Structural and Functional Magnetic Resonance Imaging of Autism," *International Journal of Developmental Neuroscience*, 20, 421-438.

Cressie, N., and Huang, H. (1999), "Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions," *Journal of the American Statistical Association*, 94, 1330-1340.

Demidenko, E. (2004), *Mixed Models Theory and Applications*, New York: Wiley.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood for Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society, Series B, 39, 1-38.

Dennis, J. E., and Schnabel, R. B. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, New Jersey: Prentice-Hall.

Diggle, P. J. (1988), "An Approach to the Analysis of Repeated Measures," Biometrics, 44, 959-971.

Fuentes, M. (2006), "Testing for separability of spatial-temporal covariance functions," Journal of Statistical Planning and Inference, 136, 447-466.

Galecki, A. T. (1994), "General Class of Covariance Structures for Two or More Repeated Factors in Longitudinal Data Analysis," *Communications In Statistics-Theory and Method*s, 23, 3105-3119.

Gill, E. P., Murray, W., Saunders, M. A., and Wright, M. H. (1984), "Procedures for Optimization Problems With a Mixture of Bounds and General Linear Constraints," *ACM Transactions on Mathematical Software*, 10, 282-298.

Gilmore, J. H., Lin, W., Corouge, I., Vetsa, Y. S. K., Smith, J. K., Kang, C., Gu, H., Hamer, R. M., Lieberman, J. A., and Gerig, G. (2007), "Early Postnatal Development of Corpus Callosum and Corticospinal White Matter Assessed With Quantitative Tractography," American Journal of Neuroradiology, 28, 1789-1795.

Huizenga, H. M., de Munck, J. C., Waldorp, L. J., and Grasman, R. P. (2002), "Spatiotemporal EEG/MEG Source Analysis Based on a Parametric Noise Covariance Model," *IEEE Transactions on Biomedical Engineering*, 49, 533-539.

Jennrich, R. I., and Schluchter, M. D. (1986), "Unbalanced Repeated-Measures Models With Structured Covariance Matrices," Biometrics, 42, 805-820.

Lindstrom, M. J., and Bates, D. M. (1988), "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," Journal of the American Statistical Association, 83, 1014-1022.

Louis, T. A. (1988), "General Methods for Analyzing Repeated Measures," Statistics in Medicine, 7, 29-45.

Lu, N., and Zimmerman, D. L. (2005), "The Likelihood Ratio Test for a Separable Covariance Matrix," *Statistics and Probability Letters*, 73, 449-457.

McGraw, P., Liang, L., Escolar, M., Mukundan, S., Kurtzberg, J., and Provenzale, J. M. (2005), "Krabbe Disease Treated With Hematopoietic Stem Cell Transplantation: Serial Assessment of Anisotropy Measurements--Initial Experience," Radiology, 236, 221-230.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2006), "A Likelihood Ratio Test for Separability of Covariances," *Journal of Multivariate Analysis*, 97, 1025-1043.

Muller, K. E., Edwards, L. J., Simpson, S. L., and Taylor, D. J. (2007), "Statistical Tests With Accurate Size and Power for Balanced Linear Mixed Models," *Statistics in Medicine*, 26, 3639-3660.

Naik, D. N., and Rao, S. S. (2001), "Analysis of Multivariate Repeated Measures Data With a Kronecker Product Structured Covariance Matrix," *Journal of Applied Statistics*, 28, 91-105.

Pizer, S. M., Fletcher, T., Thall, A., Styner, M., Gerig, G., and Joshi, S. (2002), "Object Models in Multiscale Intrinsic Coordinates via M-reps," *Proceedings of Generative Model Based Vision GMBV*.

Powell, J. M. D. (1978a), "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," *Numerical Analysis, Dundee 1977, Lecture Notes in Mathematics 630*, G. A. Watson (ed.), Berlin: Springer-Verlag, 144-175.

– – (1978b), "Algorithms for Nonlinear Constraints That Use Lagrangian Functions," *Mathematical Programming*, 14, 224-248.

– – (1982a), "Extensions to Subroutine VF02AD," *Systems Modeling and Optimization, Lecture Notes In Control and Information Sciences 38*, R. F. Drenick and F. Kozin (eds.), Berlin: Springer-Verlag, 529-538.

– – (1982b), "VMCWD: A Fortran Subroutine for Constrained Optimization," *DAMTP 1982/NA4*, Cambridge, England.

SAS Institute. (2002), SAS/IML, Version 9, SAS Institute, Inc.: Cary, NC.

Shitan, M., and Brockwell, P. J. (1995), "An Asymptotic Test for Separability of a Spatial Autoregressive Model," *Communications In Statistics-Theory and Methods*, 24, 2027-2040.

Styner, M. A., and Gerig, G. (2001), "Three-Dimensional Medial Shape Representation Incorporating Object Variability," *Computer Vision and Pattern Recognition CVPR*, 651-656.

CHAPTER 5. CONCLUSIONS AND FUTURE RESEARCH

5.1 Summary

This dissertation has focused on developing a generalization of the continuous-time AR(1) covariance model for correlated data. Even though observed correlations often decay at a much slower or much faster rate than the AR(1) structure dictates, it sees the most use among the variety of correlation patterns available. The new model, termed the *generalized autoregressive* (GAR) covariance structure, was shown to accommodate much slower and much faster correlation decay patterns. It was also shown to have excellent analytic and numerical properties making it a valuable addition to the suite of parsimonious covariance structures for repeated measures data. Given the heavy dependence of fixed-effects inference accuracy on the proper specification of the covariance model, the amount of work on covariance models has not been commensurate with their level of importance.

The first objective of this dissertation research was to develop the GAR model. Estimators were derived for the model parameters utilizing the Newton-Raphson method with the profile log-likelihood. An examination of the correlation patterns which can be modeled revealed that special cases of the GAR structure include the AR(1), equal correlation (as in compound symmetry), and MA(1) (moving average model of order 1) models. Careful study of model properties illuminated both potential complications and solutions when implementing the GAR covariance model.

An assessment of the performance of the GAR model relative to comparable models exhibited its utility. As evidenced by the simulation results in Chapter 2, the GAR covariance model is more appropriate than either the DE or continuous-time AR(1)

models when the data truly have a GAR correlation pattern. Its utility becomes even more pronounced in high dimensional settings which are prevalent in many areas of imaging research. The better statistical and convergence properties of the GAR relative to the DE model increase its appeal. Analysis of the neonate neurological DTI data illustrated these advantages.

The second focus of this research involved examining inference accuracy for both fixed effect and covariance parameters in the general linear model with a GAR covariance structure. Simulation results in Chapter 3 showed that the GAR covariance model is as robust to misspecification in controlling fixed effect test size as the DE model, while possessing better statistical and convergence properties. The GAR covariance model is far more robust to misspecification than the AR(1) model. Analysis of the DASH data exemplified the disparate, more defensible fixed effect inference results that can occur when fitting the GAR as opposed to the AR(1) covariance model. These results served to further strengthen the case for the inclusion of the GAR covariance model in the suite of parsimonious covariance structures for repeated measures data.

Hypothesis tests concerning the decay speed parameter of the GAR covariance model were examined to assess the ability of the commonly used Likelihood Ratio Test to discern the model from its special cases. More specifically, simulated test size was examined for tests of $H_0 : \delta_e = 0$ (corresponding to Compound Symmetry) and $H_0 : \delta_e = D - 1$ (corresponding to the AR(1) model). Inference about the covariance parameters provides a more formal assessment of model fit for nested models than information criteria. As evidenced by these simulation results in Chapter 3, the LRT gives an unbiased assessment of model fit for moderately large samples under most conditions. Application of the LRT to the both the DTI and DASH data corroborated the better fit of the GAR model.

Finally, the GAR model was extended to the multivariate repeated measures context by developing the Kronecker product GAR covariance structure which allows modeling correlations patterns induced by two factors (as is the case with spatio-temporal data). Simulation results in Chapter 4 made it evident that the proposed estimation approach performs well in this more complex context. Analysis of the caudate data nicely illustrated the benefits of the model in terms of its flexibility and interpretability.

5.2 Future Research

Providing a thorough investigation of the proposed GAR covariance model was the goal of this dissertation. Though not everything can be covered, this work provides a solid foundation on which to base future research. There are several possible model extensions that would lead to an even more flexible GAR structure. A better understanding of unexamined GAR model properties may have implications in numerous biomedical contexts. This research has also pointed to the need for better hypothesis tests in repeated measures settings.

Any combination of allowing for the estimation of the D as a parameter and including higher order polynomial functions in the exponent of the model would further increase the flexibility of the GAR covariance model. These modifications may lead to more computational flexibility, thus facilitating convergence and more accurate estimation. They would also allow for the modeling of an even wider range of exponentially decaying correlation patterns.

The development of a nonstationary GAR covariance model would prove extremely useful in neuroimaging studies of the developing brain since the variability of brain characteristics tends to change over time. The model could have a nonstationary variance and/or correlation structure. Though, the relative loss of parsimony may prove limiting in the High Dimension, Low Sample Size (HDLSS) context.

There are many possible avenues for future inference investigation with the proposed GAR covariance model. Assessment of both fixed effect and covariance parameter inference in the model with small samples would have implications for both imaging and genetics research. Evaluating the performance of the model under violations of the Gaussian assumption may prove useful in many contexts. Also, as evidenced by the simulation results in Chapter 3, there is a strong need for the development of tests for fixed effects that are unbiased even with a misspecified covariance.

For data that have within-subject correlations induced by three or more factors, as in longitudinal imaging data represented via the mrep method (noted in Chapter 4), the generalization of the Kronecker product GAR covariance model to n repeated factors would be beneficial. In order to determine the appropriateness of these models, more flexible tests of separability are needed. With the ultimate goal being the development of a separability test for n covariance structures of any type, the extension of current tests to allow for two unbalanced covariance models would be a natural first step.

Though often seen as a nuissance structure, properly modeling the covariance is of extreme importance for accurate fixed effect inference. Accurate covariance structure specification also aids in giving insight into the biological process under investigation. The relative dearth of current literature on covariance modeling further necessitates continued research in this area.

APPENDIX: THEOREMS AND PROOFS

Chapter 2

The following theorem is presented in Schott (1997):

Theorem 2.1. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be each be an $m \times m$ symmetric matrix. If $\boldsymbol{B}$ is positive definite and $\boldsymbol{A}$ is nonnegative definite with positive diagonal elements, then $\boldsymbol{A} \circ \boldsymbol{B}$ (Hadamard Product) is positive definite

I present the following theorem:

Theorem 2.2. $\delta_e \leq D - 1$ is a sufficient condition for the positive definiteness of the GAR model

Proof. The *generalized autoregressive* (GAR) covariance structure can be reparameterized as follows: Let

$$\left\langle \mathbf{\Gamma} \right\rangle_{jk} = \begin{cases} \rho_e^{1 - [\delta_e/(D-1)]} & j \neq k \\ 1 & j = k \,, \end{cases}$$

$$\left\langle \mathbf{\Omega} \right\rangle_{jk} = \begin{cases} \left\{ \rho_e^{[\delta_e/(D-1)]} \right\}^{d(t_{ij}, t_{ik})} & j \neq k \\ 1 & j = k \,. \end{cases}$$

Then,

$$\left(1/\sigma_e^2\right) \mathbf{\Sigma}_{ei} = \mathbf{\Gamma} \circ \mathbf{\Omega} \text{ (Hadamard Product),}$$

where

$$\mathbf{\Gamma} = \begin{cases} \text{Compound Symmetry with } \left(\rho = \rho_e^{1 - [\delta_e/(D-1)]} \right) & 0 \leq \delta_e < D - 1 \\ \mathbf{1}\mathbf{1}' & \delta_e = D - 1 \end{cases}$$

$$\mathbf{\Omega} = \begin{cases} \text{Continuous-time AR(1) with } \left(\rho = \rho_e^{[\delta_e/(D-1)]} \right) & \delta_e > 0 \\ \mathbf{1}\mathbf{1}' & \delta_e = 0 \end{cases}.$$

It is well known that the CS and AR(1) models are both positive definite for $0 \leq \rho_e < 1$.

Thus, by Theorem 2.2, $\delta_e \leq D - 1$ is a sufficient (but not necessary) condition for the positive definiteness of the GAR model.

Chapter 4

The following theorem is presented in Schott (1997):

Theorem 4.1. Let $\lambda_1, ..., \lambda_m$ be the eigenvalues of the $m \times m$ matrix $\boldsymbol{A}$, and let $\theta_1, ..., \theta_m$ be the eigenvalues of the $p \times p$ matrix $\boldsymbol{B}$. Then the set of mp eigenvalues of $\boldsymbol{A} \otimes \boldsymbol{B}$ is given by $\{\lambda_i \theta_j : i = 1, ..., m; j = 1, ..., p\}$.

I present the following corollary:

Corollary 4.1.1. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $m \times m$ and $p \times p$ matrices respectively. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are positive definite then $\boldsymbol{A} \otimes \boldsymbol{B}$ is positive definite.

Proof. This follows directly from the theorem.

Thus with $\boldsymbol{\Sigma}_{ei}(\sigma_e^2, \boldsymbol{\tau}_e) = \sigma_e^2[\boldsymbol{\Gamma}_{ei}(\boldsymbol{\tau}_{e_\gamma}) \otimes \boldsymbol{\Omega}_{ei}(\boldsymbol{\tau}_{e_\omega})]$, the positive definiteness of $\boldsymbol{\Gamma}_{ei}$ and $\boldsymbol{\Omega}_{ei}$ is sufficient for the positive definiteness of $\boldsymbol{\Sigma}_{ei}$.

REFERENCES

Abramowitz, M., and Stegun, I. A. (1972), *Handbook of Mathematical Functions*, New York: Dover Publications, Inc.

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transaction on Automatic Control*, AC-19, 716-723.

Appel, L. J., Moore, T. J., Obarzanek, E., Vollmer, W. M., Svetkey, L. P., Sacks, F. M., Bray, G. A., Vogt, T. M., Cutler, J. A., Windhauser, M. M., Lin, P. H., and Karanja, N. (1997), "A Clinical Trial of the Effects of Dietary Patterns on Blood Pressure. DASH Collaborative Research Group," *New England Journal of Medicine*, 336, 1117-1124.

Chi, E. M., and Reinsel, G. C. (1989). "Models for Longitudinal Data With Random Effects and AR(1) Errors," *Journal of the American Statistical Association* 84, 452-459.

Cody, H., Pelphrey, K., and Piven, J. (2002), "Structural and Functional Magnetic Resonance Imaging of Autism," *International Journal of Developmental Neuroscience*, 20, 421-438.

Cressie, N., and Huang, H. (1999), "Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions," *Journal of the American Statistical Association*, 94, 1330-1340.

Demidenko, E. (2004), *Mixed Models Theory and Applications*, New York: Wiley.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood for Incomplete Data via the Em Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

Dennis, J. E., and Schnabel, R. B. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, New Jersey: Prentice-Hall.

Diggle, P. J. (1988), "An Approach to the Analysis of Repeated Measures," *Biometrics*, 44, 959-971.

— — (1990), *Time Series: A Biostatistical Introduction*, Oxford: Clarendon Press.

Fuentes, M. (2006), "Testing for separability of spatial-temporal covariance functions," Journal of Statistical Planning and Inference, 136, 447-466.

Galecki, A. T. (1994), "General Class of Covariance Structures for Two or More Repeated Factors in Longitudinal Data Analysis," *Communications In Statistics-Theory and Method*s, 23, 3105-3119.

Gerig, G., Muller, K. E., Kistner, E. O., Chi, Y., Chakos, M., Styner, M., and Lieberman, J. A. (2003), "Age and Treatment Related Local Hippocampal Changes in Schizophrenia Explained by a Novel Shape Analysis Method," *MICCAI 2003 proceedings*, LNCS 2879: 653-660.

Gill, E. P., Murray, W., Saunders, M. A., and Wright, M. H. (1984), "Procedures for Optimization Problems With a Mixture of Bounds and General Linear Constraints," *ACM Transactions on Mathematical Software*, 10, 282-298.

Gilmore, J. H., Lin, W., Corouge, I., Vetsa, Y. S. K., Smith, J. K., Kang, C., Gu, H., Hamer, R. M., Lieberman, J. A., and Gerig, G. (2007), "Early Postnatal Development of Corpus Callosum and Corticospinal White Matter Assessed With Quantitative Tractography," *American Journal of Neuroradiology*, 28, 1789-1795.

Grady, J. J., and Helms, R. W. (1995), "Model Selection Techniques for the Covariance Matrix for Incomplete Longitudinal Data," *Statistics In Medicine*, 14, 1397-1416.

Huizenga, H. M., de Munck, J. C., Waldorp, L. J., and Grasman, R. P. (2002), "Spatiotemporal EEG/MEG Source Analysis Based on a Parametric Noise Covariance Model," *IEEE Transactions on Biomedical Engineering*, 49, 533-539.

Jennrich, R. I., and Schluchter, M. D. (1986), "Unbalanced Repeated-Measures Models With Structured Covariance Matrices," *Biometrics*, 42, 805-820.

Jones, R. H., and Boadi-Boateng, F. (1991), "Unequally Spaced Longitudinal Data With AR(1) Serial Correlation," *Biometrics* 47, 161-175.

Laird, N. M., and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics* 38, 963-974.

LaVange, L. M., and Muller, K. E. (1992), "Using Power Calculations in Designing Repeated Measures Studies," *Unpublished Presentation*, ENAR Meeting, Biometric Society.

Lindstrom, M. J., and Bates, D. M. (1988), "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," *Journal of the American Statistical Association*, 83, 1014-1022.

Louis, T. A. (1988), "General Methods for Analyzing Repeated Measures," *Statistics in Medicine*, 7, 29-45.

Lu, N., and Zimmerman, D. L. (2005), "The Likelihood Ratio Test for a Separable Covariance Matrix," *Statistics and Probability Letters*, 73, 449-457.

Magnus, J. R., and Neudecker, H. (1999), *Matrix differential calculus with applications in statistics and econometrics*, New York: John Wiley & Sons.

McGraw, P., Liang, L., Escolar, M., Mukundan, S., Kurtzberg, J., and Provenzale, J. M. (2005), "Krabbe Disease Treated With Hematopoietic Stem Cell Transplantation: Serial Assessment of Anisotropy Measurements--Initial Experience," *Radiology*, 236, 221-230.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2006), "A Likelihood Ratio Test for Separability of Covariances," *Journal of Multivariate Analysis*, 97, 1025-1043.

Moore, T. J., Vollmer, W. M., Appel, L. J., Sacks, F. M., Svetkey, L. P., Vogt, T. M., Conlin, P. R., Simons-Morton, D. G., Carter-Edwards, L., Harsha, D. W. (1999), "Effect of Dietary Patterns on Ambulatory Blood Pressure: Results From the Dietary Approaches to Stop Hypertension (DASH) Trial," *Hypertension*, 34, 472-477.

Muller, K. E., Edwards, L. J., Simpson, S. L., and Taylor, D. J. (2007), "Statistical Tests With Accurate Size and Power for Balanced Linear Mixed Models," *Statistics in Medicine*, 26, 3639-3660.

Muller, K. E., and Stewart, P. W. (2006), *Linear Model Theory*, New Jersey: John Wiley & Sons.

Munoz, A., Carey, V., Schouten, J. P., Segal, M., and Rosner, B. (1992), "A Parametric Family of Correlation Structures for the Analysis of Longitudinal Data," *Biometrics*, 48, 733-742.

Murray, S. C. (1990), "Linear Models With Generalized AR(1) Covariance Structure for Irregularly-Timed Data," *Unpublished Dissertation*, University of North Carolina at Chapel Hill.

Naik, D. N., and Rao, S. S. (2001), "Analysis of Multivariate Repeated Measures Data With a Kronecker Product Structured Covariance Matrix," *Journal of Applied Statistics*, 28, 91-105.

Neyman, J., and Pearson, E. S. (1928), "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part II. *Biometrika* 20A, 263-294.

Nunez-Anton, V., and Woodworth, G. (1994), "Analysis Of Longitudinal Data With Unequally Spaced Observations And Time-Dependent Correlated Errors," *Biometrics*, 50, 445-456.

Pizer, S. M., Fletcher, T., Thall, A., Styner, M., Gerig, G., and Joshi, S. (2002), "Object Models in Multiscale Intrinsic Coordinates via M-reps," *Proceedings of Generative Model Based Vision GMBV*.

Powell, J. M. D. (1978a), "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," *Numerical Analysis, Dundee 1977, Lecture Notes in Mathematics 630*, G. A. Watson (ed.), Berlin: Springer-Verlag, 144-175.

– – (1978b), "Algorithms for Nonlinear Constraints That Use Lagrangian Functions," *Mathematical Programming*, 14, 224-248.

– – (1982a), "Extensions to Subroutine VF02AD," *Systems Modeling and Optimization, Lecture Notes In Control and Information Sciences 38*, R. F. Drenick and F. Kozin (eds.), Berlin: Springer-Verlag, 529-538.

– – (1982b), "VMCWD: A Fortran Subroutine for Constrained Optimization," *DAMTP 1982/NA4*, Cambridge, England.

Rijcken, B., Schouten, J. P., Weiss, S. T., Speizer, F. E., and van der Lende, R. (1987), "The Relationship of Nonspecific Bronchial Responsiveness to Respiratory Symptoms in a Random Population Sample," *American Review of Respiratory Disease* 136, 62-68.

SAS Institute. (1999). *SAS/STAT User's Guide, Version 8*. SAS Institute, Inc.: Cary, NC.

SAS Institute. (2002), *SAS/IML, Version 9*, SAS Institute, Inc.: Cary, NC.

Schwarz, S. R. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461-464.

Self, S. G., and Liang, K. (1987), "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions," *Journal of the American Statistical Association* 82, 605-610.

Shitan, M., and Brockwell, P. J. (1995), "An Asymptotic Test for Separability of a Spatial Autoregressive Model," *Communications In Statistics-Theory and Methods*, 24, 2027-2040.

Styner, M. A., and Gerig, G. (2001), "Three-Dimensional Medial Shape Representation Incorporating Object Variability," *Computer Vision and Pattern Recognition CVPR*, 651-656.

Verbeke, G., and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, New York: Springer.

Vonesh, E. F., and Chinchilli, V. M. (1997), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, New York: Marcel Dekker, Inc.

Wu, C. F. J. (1983), "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics* 11, 95-103.

Yokoyama, T. (1997), "Tests for a Family of Random-Effects Covariance Structures in a Multivariate Growth Curve Model," *Journal of Statistical Planning and Inference* 65, 281-292.