Empirical Likelihood Based Longitudinal Studies

by

© **Tharshanna Nadarajah**

A thesis submitted to the School of Gradate Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Department of Mathematics and Statistics

Memorial University

April 2016

St. John's, Newfoundland and Labrador, Canada

Abstract

In longitudinal data analysis, our primary interest is in the regression parameters for the marginal expectations of the longitudinal responses; the longitudinal correlation parameters are of secondary interest. The joint likelihood function for longitudinal data is challenging, particularly for correlated discrete outcome data. Marginal modeling approaches such as generalized estimating equations (GEEs) have received much attention in the context of longitudinal regression. These methods are based on the estimates of the first two moments of the data and the working correlation structure. The confidence regions and hypothesis tests are based on the asymptotic normality. The methods are sensitive to misspecification of the variance function and the working correlation structure. Because of such misspecifications, the estimates can be inefficient and inconsistent, and inference may give incorrect results. To overcome this problem, we propose an empirical likelihood (EL) procedure based on a set of estimating equations for the parameter of interest and discuss its characteristics and asymptotic properties. We also provide an algorithm based on EL principles for the estimation of the regression parameters and the construction of a confidence region for the parameter of interest. We extend our approach to variable selection for high-dimensional longitudinal data with many covariates. In this situation it is necessary to identify a submodel that adequately represents the data. Including redundant variables may impact the model's accuracy and efficiency for inference. We propose a penalized empirical likelihood (PEL) variable selection based on GEEs; the variable selection and the estimation of the coefficients are carried out simultaneously. We discuss its characteristics and asymptotic properties, and present an algorithm for optimizing PEL. Simulation studies show that when the model assumptions are correct, our method performs as well as existing methods, and when the model is misspecified, it has clear advantages. We have applied the method to two case examples.

I dedicate this thesis to my wife and best friend Premni and my beautiful son Kavevarman, in gratitude for their endless love, support, and encouragement.

Acknowledgements

First, I would like to express my sincere gratitude to my advisor Dr. Asokan Mulayath Variyath, for his patience, enthusiasm, motivation, and immense knowledge. He has provided constant support during my Ph.D. research and has guided me throughout the preparation of this thesis. I could not have had a better advisor and mentor for my Ph.D.

My co-advisor, Dr. J. Concepción Loredo-Osti, has always been there to listen and offer advice. I am also thankful to him for encouraging the use of correct grammar and consistent notation and for carefully reading countless versions of this manuscript.

I gratefully acknowledge the financial support in the form of graduate fellowships and teaching assistantships provided by Memorial University of Newfoundland's School of Graduate Studies, the Department of Mathematics & Statistics, and my supervisor. I wish to thank all the faculty in our department for their support. It would not have been possible for me to complete the requirements of the graduate program without their guidance.

I would also like to thank my family, and in particular my wife and son, for their unwavering support in every way possible.

Finally, to my friends, thank you for your understanding and assistance during many, many moments of crisis. Your friendship enriches my life. I cannot list all your names here, but you are always on my mind.

Table of contents

List of tables

List of figures

Chapter 1

Introduction

1.1 Longitudinal Data

Longitudinal studies are common in areas such as epidemiology, clinical trials, economics, agriculture, and survey sampling. These studies investigate inference for data that involve repeated observations of the same subject over periods of time. The main feature of longitudinal data is that the repeated responses for each subject will likely be correlated since they relate to the same individual and consequently share the same covariates at any given point in time. In longitudinal studies, we are interested in the changes in the responses over time as a function of the covariates, generally under the assumption that observations from different individuals are independent. For example, longitudinal studies are used to characterize growth and aging, to assess the effect of risk factors on human health, and to evaluate the effectiveness of treatments. To obtain an unbiased, efficient, and reliable estimate, we must properly model the correlation between the repeated responses for each individual. However, the modelling of correlation, especially when the responses are discrete, is a challenging task even if the responses are collected over equi-spaced time points. The major methods used for the analysis of longitudinal data dealing with mixed effects, transitional, and marginal regression models and the generalized estimating equation (GEE) approach.

1.2 Analysis of Longitudinal Data

Mixed effects regression is probably the most widely used methodology for the analysis of longitudinal data. The most common models are linear mixed effects models (LMMs), nonlinear mixed effects models (NLMMs), and generalized linear mixed effects models (GLMMs). Mixed effects models incorporate the correlation within the individual responses by introducing random effects. LMMs and NLMMs are appropriate only for continuous responses. However, in practice, many types of responses follow non-Gaussian distributions, and in these cases GLMMs are appropriate. A potential disadvantage of mixed effects models is that they rely on parametric assumptions, which may lead to biased parameter estimates when a model is misspecified. Moreover, the estimation of the parameters is challenging when the random effects have a high dimension; it typically involves integrals that do not have an explicit form. In the absence of an analytical solution, Breslow and Clayton [1993] proposed the penalized quasi-likelihood (PQL) for the GLMM; it uses a Laplace approximation to find the marginal likelihood. However, the PQL often yields biased estimates of the regression parameters since the estimators of the variance components are biased, especially for discrete longitudinal data.

Generalized linear models (GLMs) often handle longitudinal data by assuming a Markov structure that incorporates the correlation-within-individual measurements in the transitional models. In these Markov structure based GLM models, the conditional distribution of each response is expressed as a function of the past responses and the covariates. These models are more difficult to apply when there are missing data and the repeated measurements are not equally spaced in time. In addition, the interpretation of the regression parameters varies with the order of the serial correlation, and the regression parameter estimates are sensitive to the assumption of time dependence. Because of the aforementioned difficulties in modelling and performing inference, we focus on marginal models in this thesis.

1.3 Marginal Models

The key component of the marginal model is that the mean response at each time point depends on the covariates through a known link function. The longitudinal observations consist of an outcome random variable y_{it} and a p-dimensional vector of covariates $\boldsymbol{x}_{it}$, observed for subjects $i = 1, \ldots, k$ at a time point t, $t = 1, \ldots, m_i$. For the ith subject, let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im_i})^T$ be the response vector, and let $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{it}, ..., \boldsymbol{x}_{im_i})^T$ be the $m_i \times p$ matrix of covariates. Marginal models assume that the conditional mean of the tth response depends only on $\boldsymbol{x}_{it}$: $\mathrm{E}(y_{it}|\boldsymbol{X}_i) = \mathrm{E}(y_{it}|\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{im_i}) = \mathrm{E}(y_{it}|\boldsymbol{x}_{it})$. However, this assumption does not hold when the covariate effect is time-dependent. As a result, special care is required when fitting marginal models with time-varying covariates. Marginal models describe only the (marginal) means of the outcome variables, ignoring the correlation or covariance structure of longitudinal observations.

Marginal models for longitudinal data can be extended to the GLM framework. The marginal density of y_{it} is assumed to follow an exponential family (McCullagh and Nelder [1989]) of the form

$$f(y_{it}) = \exp\left[(y_{it}\theta_{it} - a(\theta_{it}))\phi + b(y_{it}, \phi)\right], \tag{1.1}$$

where $\theta_{it} = h(\eta_{it})$, h is a known injective function with $\eta_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta}$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression effects of $\boldsymbol{x}_{it}$ on y_{it}, and $a(*)$ and $b(*)$ are functions that are assumed to be known. The mean and variance of y_{it} can be written

$$\mathrm{E}(y_{it}|\boldsymbol{x}_{it}) = a'(\theta_{it}) = \mu_{it} \quad \text{and} \quad \mathrm{Var}(y_{it}) = a''(\theta_{it}) = v(\mu_{it})\phi, \tag{1.2}$$

where ϕ is the unknown over-dispersion parameter and $v(*)$ is a known variance function. For simplicity, we set the nuisance scale parameter ϕ to 1 in Equation (1.1) for the rest of this thesis. Let Θ be the natural parameter space of the exponential family distributions presented in (1.1) and Θ° the interior of Θ. Let $\{a'(\theta)\}$ be a three times continuously differentiable function with $\{a''(\theta)\} > 0$ in Θ°. Also, let $h(\eta)$ be a three times continuously differentiable function with $h'(\eta) > 0$ in $g(\mathcal{M})^\circ$, where $\mathcal{M}$ is the image of $\{a'(\Theta^\circ)\}$.

When the responses are continuous, the correlation can be represented by the linear dependence among the repeated responses. However, in the absence of a convenient likelihood function for discrete data, there is no unified likelihood-based approach for marginal models. Since our main interest is in modelling the relationship between the covariates and the response, we will not precisely model within-subject correlation (McCullagh and Nelder [1989]). Assuming the existence of the first two moments, Wedderburn [1974] proposed a quasi-likelihood (QL) approach for independent data. This approach is widely used to estimate regression coefficients without fully specifying the distribution of the observed data.

1.3.1 Quasi-likelihood

When there is insufficient information about the data for us to specify a parametric model, QL is often used. We can then develop the statistical analysis without fully specifying the distribution of the observed data; we first concentrate on cases where the observations are independent. We assume that the mean μ_{it} is a function of the covariates with the regression parameters $\boldsymbol{\beta}$ and covariance diagonal matrix $\sigma^2\mathrm{V}(\mu_{it})$. To construct the QL, we start by looking at a single component y_{it} of $\boldsymbol{y}$. The QL for complete data is

$$\mathbf{Q}(\boldsymbol{\mu};\boldsymbol{y}) = \sum_{i=1}^{k}\sum_{t=1}^{m_i}\mathbf{Q}(\mu_{it};y_{it}),$$

where $\mathbf{Q}(\mu_{it};y_{it}) = \displaystyle\int_{y_{it}}^{\mu_{it}} \frac{y_{it}-t}{\sigma^2\mathrm{V}(t)}dt$. The QL estimating equations for the regression parameters $\boldsymbol{\beta}$ are obtained by differentiating $\mathbf{Q}(\boldsymbol{\mu};\boldsymbol{y})$:

$$\sum_{i=1}^{k}\sum_{t=1}^{m_i}\left[\frac{\partial a'(\theta_{it})}{\partial\boldsymbol{\beta}}\frac{(y_{it}-a'(\theta_{it}))}{\mathrm{Var}(y_{it})}\right] = \mathbf{0}.$$

For instance, in the Poisson case $\mathrm{Var}(y_{it}) = a''(\theta_{it}) = a'(\theta_{it}) = \mu_{it} = \exp(\boldsymbol{x}_{it}\boldsymbol{\beta})$.

In the longitudinal setup, the components of the response vector $\boldsymbol{y}_i$ correspond to repeated observations of the same covariates for the same subject, and they are likely correlated. Let $\boldsymbol{C}_i(\rho)$ be the $m_i \times m_i$ true correlation matrix of $\boldsymbol{y}_i$, $i = 1, \ldots, k$, which is unknown in practice. Our primary goal is to estimate $\boldsymbol{\beta}$ after taking the longitudinal correlation $\boldsymbol{C}_i(\rho)$ into account. For a known $\boldsymbol{C}_i(\rho)$, the QL estimator of $\boldsymbol{\beta}$ under (1.1) is the solution of the score equation

$$g(\boldsymbol{y};\boldsymbol{\beta}) = \sum_{i=1}^{k}\boldsymbol{X}_i^T\boldsymbol{A}_i\boldsymbol{\Sigma}_i^{-1}(\rho)(\boldsymbol{y}_i-\boldsymbol{\mu}_i) = \mathbf{0}, \tag{1.3}$$

where $\boldsymbol{A}_i = \mathrm{diag}\left[a''(\theta_{i1}), ..., a''(\theta_{it}), ..., a''(\theta_{im_i})\right]$ and $\boldsymbol{\Sigma}_i(\rho) = \boldsymbol{A}_i^{1/2}\boldsymbol{C}_i(\rho)\boldsymbol{A}_i^{1/2}$ is the true covariance of $\boldsymbol{y}_i$.

In real applications the true correlation structure is often unknown. Ignoring the correlation of the measurements for the same individual could lead to an inefficient estimate of the regression coefficients and an underestimate of the standard errors.

If the probability distribution of the response $\boldsymbol{y}_i$ is poorly characterized, then it is obvious that we cannot use the likelihood approach. Even if it is not of primary interest, the correlation among a subject's repeated measurements must be taken into account for proper inference. The joint distribution of the correlated discrete responses may not have a closed form when the correlation is taken into account. To avoid specifying the joint distribution of correlated discrete responses, Liang and Zeger [1986] proposed the GEE approach, an extension of GLMs to longitudinally correlated data analysis using QL.

1.3.2 Generalized Estimating Equation Approach

The GEE approach is a semiparametric method where the estimating equations are derived without a full specification of the joint distribution of the observed data. This approach to estimating the regression parameters allows the user to specify any structure for the correlation matrix of the outcomes $\boldsymbol{y}_i$.

Liang and Zeger [1986] introduced a "working" correlation structure based on the GEE approach to obtain consistent and efficient estimators for the regression parameter $\boldsymbol{\beta}$. They solved

$$g(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\beta)) = \sum_{i=1}^{k} \boldsymbol{X}_i^T \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i^{-1}(\hat{\alpha}) \boldsymbol{A}_i^{-1/2} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{1.4}$$

where $\boldsymbol{A}_i$ is an $m_i \times m_i$ diagonal matrix with $\mathrm{Var}(\mu_{it})$ as the tth diagonal element and $\boldsymbol{R}_i(\hat{\boldsymbol{\alpha}})$ is the $m_i \times m_i$ working correlation matrix of the m_i repeated measurements used for $\boldsymbol{C}_i(\rho)$ in Equation (1.3). For $j = 1, \ldots, m_i$ and $j^{'} = 1, \ldots, m_i$, the $(j, j^{'})^{th}$ element of $\boldsymbol{R}_i$ is the known, hypothesized, or estimated correlation. The working correlation may depend on an unknown $s \times 1$ correlation parameter vector $\boldsymbol{\alpha}$. The observation times and correlation matrix may differ from subject to subject, but the correlation matrix $\boldsymbol{R}_i(\alpha)$ for the ith subject is fully specified by $\boldsymbol{\alpha}$. The working variance-covariance matrix for $\boldsymbol{y}_i$ is $\mathrm{Var}(\boldsymbol{\alpha}) = \boldsymbol{A}_i^{1/2}\boldsymbol{R}_i(\alpha)\boldsymbol{A}_i^{-1/2}$. Some common working correlation structures are independence, autoregressive of order one (AR(1)), equally correlated (EQC), moving average of order one (MA(1)), or unstructured. When $\boldsymbol{R}_i(\alpha) = I$ in (1.4), the score equations are from a likelihood analysis, which assumes that the repeated observations from a subject are independent of one another.

Liang and Zeger [1986] established the following properties of the estimator $\boldsymbol{\beta}$ that satisfies $g(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}(\beta)) = \mathbf{0}$ under the assumption that the estimating equation is asymptotically unbiased in the sense that $\lim_{k\to\infty} E[g(\boldsymbol{\beta}_0, \hat{\boldsymbol{\alpha}}(\beta_0))] = \mathbf{0}$, $\hat{\boldsymbol{\beta}}$ is consistent, and $\mathrm{Cov}(\hat{\boldsymbol{\beta}})$ can be consistently estimated. For a given working correlation structure, $\boldsymbol{\alpha}$ can be estimated using a residual-based method of moments.

To improve the efficiency of the regression parameter estimates, Prentice and Zhao [1991] extended the GEE approach to allow for joint estimating equations for both the regression parameters $\boldsymbol{\beta}$ and the nuisance correlation parameters $\boldsymbol{\alpha}$. This approach needs the existence of the third and fourth moments of $\boldsymbol{y}_i$, $i = 1, \ldots, k$.

1.3.3 Limitations of GEE Approach

The GEE-based estimate of $\boldsymbol{\beta}$ is not necessarily consistent, as discussed by Crowder [1995] and Sutradhar and Das [1999]. Crowder [1995] demonstrated that in some situations the use of an arbitrary working correlation structure may lead to no solution for $\hat{\boldsymbol{\alpha}}$, which may break down the entire GEE methodology. Sutradhar and Das [1999] showed that the GEE approach may yield an estimator of $\boldsymbol{\beta}$, that, although consistent, is less efficient than that of the independence estimating equation approach under an arbitrary working correlation structure. To overcome this difficulty, Sutradhar [2003] proposed using a stationary lag correlation structure instead of the working correlation matrix.

The estimate for $\boldsymbol{\beta}$ is obtained by solving the following estimating equations:

$$g(\boldsymbol{\beta}, \hat{\rho}(\beta)) = \sum_{i=1}^{k} \boldsymbol{X}_i^T \boldsymbol{A}_i \boldsymbol{\Sigma}_i^{-1}(\widehat{\rho})(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{1.5}$$

where $\boldsymbol{\Sigma}_i(\hat{\rho}) = \boldsymbol{A}_i^{1/2} \boldsymbol{C}_i^*(\boldsymbol{\rho}) \boldsymbol{A}_i^{1/2}$, with $\boldsymbol{C}_i^*(\boldsymbol{\rho})$ the stationary lag correlation structure for the AR(1), MA(1), or EQC models, and

$$\boldsymbol{C}_i^*(\boldsymbol{\rho}) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdot & \cdot & \cdot & \rho_{m-1} \\ \rho_1 & 1 & \rho_1 & \cdot & \cdot & \cdot & \rho_{m-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{m-1} & \rho_{m-2} & \rho_{m-3} & \cdot & \cdot & \cdot & 1 \end{bmatrix}. \tag{1.6}$$

The stationary lag correlations can be estimated via the method of moments introduced by Sutradhar and Kovacevic [2000]:

$$\hat{\rho}_l = \frac{\sum_{i=1}^{k} \sum_{t=1}^{m-l} \tilde{y}_{it} \tilde{y}_{i,t+l} / k(m-l)}{\sum_{i=1}^{k} \sum_{t=1}^{m} \tilde{y}_{it}^2 / km}, \tag{1.7}$$

where $l = |t - t'|,\ t \neq t',\ t, t' = 1, \ldots, m$ and $\tilde{y}_{it}$ is the standardized residual, defined as $\tilde{y}_{it} = \{y_{it} - \mu_{it}\}/\{a''(\theta_{it})\}^{1/2}$. For an unequal number of time points, the correlation matrix given in (1.6) is estimated using the estimate of the lag correlation ρ_l:

$$\hat{\rho}_l = \frac{\sum_{i=1}^{k} \sum_{t=1}^{m-l} \delta_{it} \delta_{i,t+l} \tilde{y}_{it} \tilde{y}_{i,t+l} / \sum_{i=1}^{k} \sum_{t=1}^{m-l} \delta_{it} \delta_{i,t+l}}{\sum_{i=1}^{k} \sum_{t=1}^{m_i} \delta_{it} \tilde{y}_{it}^2 / \sum_{i=1}^{k} \sum_{t=1}^{m_i} \delta_{it}}, \tag{1.8}$$

where $m = \max_{1 \le i \le k} m_i$, $l = 1, \ldots, m-1$, and

$$\delta_{iu} = \begin{cases} 1, & \text{if } u \le m_i \\ 0, & \text{if } m_i < u \le m. \end{cases}$$

Sutradhar and Das [1999] showed that the stationary lag correlation approach produces regression estimates that are consistent and more efficient than those obtained from the independence-assumption-based estimating equation approach. This approach assumes a known longitudinal correlation structure even though the correlation parameters are unknown.

We conducted a small simulation study to compare GEE with a stationary lag correlation approach when the correlation structure is misspecified. We consider a stationary correlation AR(1) model for longitudinal count data discussed by McKenzie [1988] and Sutradhar [2011]; see Table 1.1. We consider the stationary covariates $\tilde{\boldsymbol{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})$, where $(\tilde{x}_{i1}, \tilde{x}_{i2})$ is generated from the normal distribution with mean 0 and variance 1, and $\boldsymbol{\beta} = (0.3, 0.2)^T$. For a given $y_{i,t-1}$, $\rho * y_{i,t-1}$ is the binomial thinning operation discussed by McKenzie [1988]. That is, $\rho * y_{i,t-1} = \sum_{j=1}^{y_{i,t-1}} b_j(\rho)$ with $\Pr[b_j(\rho) = 1] = \rho$, $\Pr[b_j(\rho) = 0] = 1 - \rho$. In our simulation we use $m = 5$ time points and $k = 100$ subjects. We simulated 1000 data sets with $\rho = 0.49$ and 0.70.

Model	Dynamic Relationship	Mean, Variance, & Correlations
AR(1)	$y_{it} = \rho * y_{i,t-1} + d_{it}, \quad t = 2, \ldots, m$	$\mathrm{E}[y_{it}] = \tilde{\boldsymbol{\mu}}_i$
	$y_{i1} \sim \mathrm{Poi}(\tilde{\boldsymbol{\mu}}_i = \exp[\tilde{\boldsymbol{x}}_i \boldsymbol{\beta}])$	$\mathrm{Var}[y_{it}] = \tilde{\boldsymbol{\mu}}_i$
	$d_{it} \sim \mathrm{Poi}[\tilde{\boldsymbol{\mu}}_i(1-\rho)], \quad t = 2, \ldots, m$	$\mathrm{corr}[y_{it}, y_{i,t+l}] = \rho_l = \rho^l$

Table 1.1: A class of stationary AR(1) correlation model for longitudinal count data.

Table 1.2 gives the average estimated values of the regression coefficients and, in parentheses, the corresponding simulated standard errors. The table also gives the coverage probabilities and the width of the confidence interval (CI) for β_1 and β_2 for the 0.95 and 0.99 confidence levels. We generated the data using an AR(1) correlation structure. We used AR(1), EQC, and MA(1) for the parameter estimation under GEEs, and we compared the results with those for GEEs with lag correlation. Table

				Coverage Probability	
True Model	Method	Parameter	Estimate	95% level	99% level
	GEE (AR(1))	β_1	0.3000	0.952	0.987
			(0.070)	(0.279)	(0.367)
		β_2	0.2009	0.950	0.988
			(0.073)	(0.286)	(0.375)
AR(1)	GEE (EQC)	β_1	0.2997	0.911	0.973
$\rho = 0.70$			(0.073)	(0.247)	(0.325)
		β_2	0.1956	0.902	0.963
			(0.076)	(0.252)	(0.332)
	GEE (lag)	β_1	0.3003	0.952	0.986
			(0.070)	(0.278)	(0.366)
		β_2	0.2007	0.950	0.988
			(0.073)	(0.284)	(0.374)
	GEE (AR(1))	β_1	0.2989	0.938	0.988
			(0.062)	(0.237)	(0.319)
		β_2	0.1956	0.940	0.981
			(0.062)	(0.243)	(0.319)
	GEE (EQC)	β_1	0.2992	0.899	0.968
			(0.061)	(0.206)	(0.272)
		β_2	0.1986	0.908	0.980
			(0.062)	(0.211)	(0.278)
AR(1)	GEE (MA(1))	β_1	0.2991	0.897	0.968
$\rho = 0.49$			(0.062)	(0.205)	(0.270)
		β_2	0.1985	0.905	0.981
			(0.062)	(0.210)	(0.276)
	GEE (lag)	β_1	0.2989	0.931	0.990
			(0.061)	(0.235)	(0.309)
		β_2	0.1955	0.936	0.992
			(0.061)	(0.241)	(0.317)

Table 1.2: Coverage probabilities of regression estimates for data from an AR(1) correlation model under different working correlation models (m=5).

1.2 shows that when we use the true working correlation structure, the coverage probabilities based on GEEs and GEEs with lag correlation are almost the same. However, under an arbitrary working correlation structure, the GEEs with lag correlation have better performance. This indicates the loss of efficiency of the GEE estimators when the correlation structures are misspecified. We therefore recommend defining a lag correlation structure for the longitudinal responses.

The correlation structure (1.6) is quite robust, and it accommodates the AR(1), EQC, and MA(1) structures. Note, however, that the structure is unknown in practice, and it is better to use a stationary lag-correlation structure to represent all three correlation structures. We did not consider all possible cases since some working correlation structure may lead to no solution for $\hat{\boldsymbol{\alpha}}$. For instance, under true exchangeable correlation with the MA(1) working correlation structure, the correlation parameter $\hat{\boldsymbol{\alpha}}$ does not exist.

The parameter $\boldsymbol{\beta}$ is defined by the estimating equations $E[g(\boldsymbol{y};\boldsymbol{\beta})] = \mathbf{0}$, where $g(\boldsymbol{y};\boldsymbol{\beta}) \in \mathcal{R}^r$ is an estimating function for $\boldsymbol{\beta} \in \mathcal{R}^p$. When $r = p$ the estimating equations $k^{-1}\sum_{i=1}^{k} g(\boldsymbol{y}_i;\boldsymbol{\beta}) = \mathbf{0}$ have a unique solution for $\boldsymbol{\beta}$. When $r > p$ we have extra information about the parameter for improved efficiency, but it may not be possible to directly solve the estimating equations. To overcome this problem, Qu, Lindsay and Li [2000] proposed an adaptive quadratic inference function of the form $Q(\boldsymbol{\beta}) = g'C^{-1}g$, where g is a set of estimating functions based on moment assumptions and C is the estimated variance of g; this does not involve direct estimation of the correlation parameter. The above approaches are robust to the working correlation assumption. However, they are not robust to model misspecification.

1.4 Variable Selection for Longitudinal Data

Variable selection is an important issue in statistical modelling. It is especially important for longitudinal data because of the high dimension of the explanatory variables or predictors that arise in large-scale studies. A large number of predictors, $(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p)$, are hypothesized to have an influence on the response variable $\boldsymbol{y}$ of interest. However, some predictors may have no influence or a weak influence, and may add noise to the estimation. Excluding these variables results in simpler model that may provide a better understanding of the underlying process.

Variable selection is the problem of identifying an optimal subset of predictor variables that adequately models the relationship between the response variable of interest and the predictors. The advantages of selecting a subset of the predictors are:

- ◇ Simpler models are easier to interpret.
- ◇ The predictive ability may be improved by eliminating irrelevant variables.
- ◇ Removing redundant predictors reduces noise.
- ◇ It is cheaper to measure fewer variables.

The main objective of variable selection is to identify the smallest adequate model. In GLMs, the submodel for a random variable $\boldsymbol{y}$ with mean $\boldsymbol{\mu}$ is a subset of components of $\boldsymbol{X}$ for which

$$g(\boldsymbol{x}; \boldsymbol{\mu}) \simeq \boldsymbol{X}(s)\boldsymbol{\beta}(s)$$

where $g(*)$ is the link function, $\boldsymbol{X}(s)$ is a subset of the components of $\boldsymbol{X}$, $\boldsymbol{\beta}(s)$ is a vector of the corresponding regression parameters, and $s \subseteq (1, 2, \ldots, p)$. The variable selection problem is to find the best subset s such that the submodel is optimal according to some criteria that gives an adequate description of the data-generating mechanism.

Several variable selection methods for GLMs have been developed. Sequential approaches such as forward selection, backward elimination, and the stepwise procedure are commonly used. These approaches are less computationally intensive than other methods, but the final model may not be optimal. The most widely used method for prediction models is the cross-validation approach (Stone [1974]). The resulting model may have a lower prediction error. However, in GLMs the concept of prediction error is not well defined (Fielding and Bell [2002]).

Two popular methods based on an information theoretic approach are Akaike's information criterion (AIC), proposed by Akaike [1973, 1974], and the Bayesian information criterion (BIC), introduced by Schwarz [1978]. In these approaches, we need to evaluate all possible submodels and identify the best. A well-defined parametric model is necessary; if a parametric likelihood is not available, the empirical likelihood (EL) versions of AIC and BIC (Variyath, Chen and Abraham [2010]) can be used.

With high-dimensional data we cannot directly apply AIC or BIC because of the computational burden. Regularization methods have been developed to overcome the computational difficulties and to achieve selection stability. There is a large literature on the penalized likelihood approach, and two important approaches are the least absolute shrinkage and selection operator (LASSO; Tibshirani [1996]) and the smoothly clipped absolute deviation (SCAD; Fan and Li [2001]). Both approaches have many desirable properties. Related methods include penalized EL-based variable selection (Variyath [2006]; Nadarajah [2011]), adaptive LASSO (Zou [2006]; Zhang and Lu [2007]), least-square approximation (Wang and Leng [2007]), and the folded concave penalty method (Lv and Fan [2009]). Tang and Leng [2010] used a penalized EL framework, which is limited to mean vector estimation and linear regression models. The above methods are applicable only to GLMs.

Variable selection for longitudinal data is challenging because of the high dimensionality of the covariates and the need to construct a convenient joint likelihood function for correlated discrete outcome data. Pan [2001] developed the QL information criterion (QIC) under the working independence model and naive and robust covariance estimates of the estimated regression coefficients. Cantoni, Flemming and Ronchetti [2005] proposed a generalized version of Mallows's C_p, suitable for use with both parametric and nonparametric models. This approach avoids a stepwise procedure, and is based on a measure of the predictive error rather than on significance testing. Wang and Qu [2009] introduced a BIC-type procedure based on the quadratic inference function; it does not require the full likelihood or a QL. The implementation of best-subset procedures requires the evaluation of all possible submodels, which becomes computationally intensive when the number of covariates is large.

The idea of penalization is useful in longitudinal modelling and particularly in high-dimensional variable selection. Fan and Li [2004] proposed an innovative class of variable selection procedures for semiparametric models with continuous responses. Wang, Li and Huang [2008] studied regularized estimation procedures for nonparametric varying-coefficient models with continuous responses; their procedures can simultaneously perform variable selection and the estimation of smooth coefficient functions. Xiao, Zhang and Zhang [2009] investigated a double-penalized likelihood approach for selecting fixed effects in semiparametric mixed models with continuous responses. Dziak, Li and Qu [2009] discussed the application of a SCAD-penalized quadratic inference function. Xu, Wang and Zhu [2010] investigated a GEE-based shrinkage estimator with an artificial objective function. Xue, Qu and Zhou [2010] considered procedures for a generalized additive model where responses from the same cluster are correlated.

The above methods assume that the dimension of the predictors is small, and some of them are applicable only to continuous responses. To avoid specifying the full joint likelihood for correlated discrete data, Wang, Zhou and Qu [2012] proposed an approach based on penalized generalized estimating equations (PGEEs) with a nonconvex penalty function. This approach requires only the specification of the first two marginal moments and a working correlation matrix. In the next section, we will discuss PGEEs for variable selection in the context of longitudinal data analysis.

1.4.1 Penalized Generalized Estimating Equations

The approach of Wang et al. [2012] requires only the specification of the first two marginal moments and the correlation structure. This method provides computational efficiency and stability. It simultaneously performs the variable selection and the estimation of the regression parameters. That is, insignificant variables are removed by setting their regression parameters to zero. The method works reasonably well for high-dimensional problems.

In this method a penalized generalized estimating function is defined to be

$$\mathcal{U}(\boldsymbol{\beta}) = g(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\beta)) - k * p'_\delta(|\boldsymbol{\beta}|)\mathrm{sign}(\boldsymbol{\beta}), \tag{1.9}$$

where $g(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\beta)) = \sum_{i=1}^{k} \boldsymbol{X}_i^T \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i^{-1}(\hat{\alpha}) \boldsymbol{A}_i^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)$ are the GEEs given in (1.4), $p'_\delta(*)$ is the first derivative of the penalty function, $\mathrm{sign}(\boldsymbol{\beta}) = (\mathrm{sign}(\beta_1), \ldots, \mathrm{sign}(\beta_p))^T$ with $\mathrm{sign}(t) = \mathrm{I}(t > 0) - \mathrm{I}(t < 0)$, and δ is the tuning parameter. Different penalty functions can be used. According to Fan and Li [2001], a good penalty function results in an estimator with the following three oracle properties:

1. Unbiasedness: The estimator is nearly unbiased when the true unknown parameter is large.

2. Sparsity: This is a thresholding rule that automatically sets small estimated

coefficients to zero to reduce the model complexity.

3. Continuity: This property eliminates unnecessary variation in the model prediction.

A suitable penalty function is the SCAD penalty (Fan [1997]). Its first derivative is

$$p'_\delta(\theta) = \delta \left\{ \mathrm{I}(\theta \leq \delta) + \frac{(a\delta - \theta)_+}{(a-1)\delta} \mathrm{I}(\theta > \delta) \right\} \quad \text{for some } a > 2 \text{ and } \theta > 0. \tag{1.10}$$

Necessary conditions for the unbiasedness, sparsity, and continuity of the SCAD penalty have been proved by Antoniadis and Fan [2001]. This penalty function involves two unknown parameters, a and δ. Under some regularity conditions, Wang et al. [2012] show that the estimator based on the SCAD penalty satisfies the oracle properties for a certain choice of a and δ.

PGEEs work reasonably well in high-dimensional problems, but the GEE approach gives inconsistent estimators of $\boldsymbol{\beta}$ under an arbitrary working correlation structure, and model misspecification limits the application of this method (see Section 1.3.3). Nadarajah, Variyath and Loredo-Osti [2016] developed penalized generalized QL (PGQL) variable selection based on the stationary lag correlation structure given in (1.6). We perform a simulation study to compare PGEEs and PGQL when the correlation structure is misspecified.

We consider the stationary AR(1) correlation model of Table 1.1. We use five covariates $\tilde{\boldsymbol{X}}_i = (\tilde{x}_{i_1}, \ldots, \tilde{x}_{i_5})$, where $\tilde{x}_{i1} \sim$ Bernoulli(0.5) and x_{i2} to x_{i5} are generated from a multivariate normal distribution with mean zero, the correlation between x_{il} and x_{jl} is $0.5^{|i-j|}$, $l = 2, \ldots, 5$. We set $\boldsymbol{\beta} = (0.5, 0.5, 0.6, 0, 0)^T$. We report (i) the median of the relative model error (MRME) and (ii) the average number of correct zero and nonzero coefficients. We also give the average estimated values of the nonzero coefficients and the corresponding simulated standard errors. The model error (ME) is defined to be $\mathrm{ME}(\hat{\boldsymbol{\beta}}) = E_x\left\{\mu(\boldsymbol{X}\boldsymbol{\beta}) - \mu(\boldsymbol{X}\hat{\boldsymbol{\beta}})\right\}^2$, where $\mu(\boldsymbol{X}\boldsymbol{\beta}) = E(\boldsymbol{y}|\boldsymbol{X})$, and the relative model error is $\mathrm{RME} = \mathrm{ME}/\mathrm{ME}_{\mathrm{full}}$, where $\mathrm{ME}_{\mathrm{full}}$ is the model error when fitting the data with the full model and ME is the model error of the selected model. We generated a sample with $k = 100$ individuals and $m = 5$ time points with three different correlation structures.

Table 1.3 shows that when we use the true working correlation structure, the MRME of the PGEEs is very close to that of PGQL. The average number of zero coefficients is close to the target of two, and the nonzero regression parameter estimates are close to the true values. However, under an arbitrary working correlation structure the PGEEs have a larger MRME, and the average number of zero coefficients is not close to the target of two. When the working correlation is misspecified, PGQL performs better than the PGEEs. We repeated the simulation with different scenarios, and the conclusions were similar, so these results are omitted. However, the PGQL is not robust to misspecification. To handle possible misspecification of the mean, variance function, and correlation structure, a nonparametric method should be used.

1.5 Motivation and Proposed Approach for Longitudinal Data Analysis

Marginal models or GEE approaches require only the specification of the first two marginal moments and a correlation structure. GEE estimators are consistent and asymptotically normal as long as the mean, variance, and correlation structure are correctly specified. Marginal models have satisfactory performance when the assumptions are satisfied. Misspecification is a concern. Moreover, if the covariates are time-dependent the assumption $\lim_{k\to\infty} E[g(\boldsymbol{\beta}_0, \hat{\boldsymbol{\alpha}}(\beta_0))] = \mathbf{0}$ might not hold for an

True Model	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
AR(1) $\rho = 0.70$	PGEE (IND)	86.86	1.25	0.0	0.5002	0.5023	0.5909
					(0.068)	(0.075)	(0.076)
	PGEE (AR(1))	65.60	1.85	0.0	0.5029	0.5058	0.5930
					(0.066)	(0.073)	(0.071)
	PGEE (EQC)	69.76	1.84	0.0	0.5030	0.5047	0.5935
					(0.066)	(0.073)	(0.072)
	PGQL	66.90	1.85	0.0	0.5034	0.5052	0.5930
					(0.066)	(0.073)	(0.071)
AR(1) $\rho = 0.49$	PGEE (AR(1))	63.60	1.80	0.0	0.5003	0.5025	0.5968
					(0.056)	(0.056)	(0.057)
	PGEE (MA(1))	70.57	1.65	0.0	0.5011	0.5021	0.5986
					(0.060)	(0.061)	(0.062)
	PGQL	67.64	1.79	0.0	0.5014	0.5029	0.5973
					(0.059)	(0.060)	(0.061)
EQC $\rho = 0.70$	PGEE (IND)	77.95	1.23	0.0	0.5059	0.5053	0.5913
					(0.074)	(0.076)	(0.080)
	PGEEs (EQC)	61.43	1.87	0.0	0.5022	0.5066	0.5921
					(0.073)	(0.076)	(0.074)
	PGEE (AR(1))	63.37	1.70	0.0	0.5025	0.5066	0.5914
					(0.074)	(0.076)	(0.076)
	PGQL	62.61	1.87	0.0	0.5026	0.5069	0.5916
					(0.073)	(0.076)	(0.073)
EQC $\rho = 0.49$	PGEE (EQC)	65.39	1.82	0.0	0.5023	0.5057	0.5922
					(0.062)	(0.065)	(0.064)
	PGEE (MA(1))	75.50	1.59	0.0	0.4996	0.5022	0.5980
					(0.065)	(0.068)	(0.073)
	PGQL	66.40	1.82	0.0	0.5017	0.5046	0.5938
					(0.064)	(0.068)	(0.069)
MA(1) $\rho = 0.67$	PGEE (IND)	70.29	1.54	0.0	0.5002	0.5006	0.5994
					(0.052)	(0.054)	(0.054)
	PGEEs (MA(1))	63.56	1.72	0.0	0.5004	0.4993	0.5981
					(0.052)	(0.053)	(0.052)
	PGEE (AR(1))	69.39	1.78	0.0	0.5018	0.5013	0.5963
					(0.052)	(0.056)	(0.054)
	PGEE (EQC)	71.37	1.71	0.0	0.5006	0.5008	0.5974
					(0.052)	(0.056)	(0.058)
	PGQL	65.20	1.75	0.0	0.5005	0.5004	0.5970
					(0.051)	(0.053)	(0.053)

Table 1.3: Performance measures for count data with stationary covariates (m=5)

arbitrary working correlation structure, and so the GEE estimate of $\boldsymbol{\beta}$ is not necessarily consistent; see Hu [1993], Pepe and Anderson [1994], Emond, Ritz and Oakes

[1997], Pan, Louis and Connett [2000], and Diggle, Heagerty, Liang and Zeger [2002]. The GEE estimator of $\boldsymbol{\beta}$ with the independent working correlation is always consistent so Pepe and Anderson [1994] recommended using this correlation as a safe choice. The independent working correlation is often efficient for the estimation of coefficients associated with time-independent covariates (Fitzmaurice [1995]). However, it may be much less efficient when the covariates are time-dependent (Fitzmaurice [1995]).

In practice, the true correlation structure is unknown, and using an arbitrary working correlation structure limits the application of marginal models. Misspecification can cause estimates based on marginal models to be inefficient and inconsistent, and inference in this situation can be completely inappropriate. Confidence regions and hypothesis tests are based on asymptotic normality, which may not hold since the finite-sample distribution may not be symmetric. These problems motivate us to investigate a nonparametric likelihood method. Instead of using marginal models, we define a subject-wise EL, based on a set of GEEs for the parameter of interest.

1.5.1 Proposed Approach to Longitudinal Data Analysis

Owen [1988] introduced the EL. The EL is a nonparametric method for statistical inference; that is, we need not assume that the data come from a particular distribution. The EL combines the reliability of nonparametric methods with the flexibility and effectiveness of the likelihood approach. The EL has many nice properties parallel to those of parametric likelihood, including the ability to carry out hypothesis tests and construct confidence intervals without estimating the variance. The shape of EL confidence regions automatically reflects the emphasis of the observed data set. These regions are invariant under transformations and often behave better than confidence regions based on asymptotic normality when the sample size is small. The EL method also offers advantages in parameter estimation and the formulation of goodness-of-fit tests. Moreover, it is possible to have more estimating equations than the number of parameters, i.e., $r > p$, where $g(\boldsymbol{y};\boldsymbol{\beta}) \in \mathcal{R}^r$ is an estimating function for the parameter $\boldsymbol{\beta} \in \mathcal{R}^p$. For instance, let $y_1, \ldots, y_n$ be independent and identically distributed univariate observations from a member of a semiparametric family F for which the first two moments are equal. If our aim is to estimate θ, the information about F is expressed in the form of the estimating equations $g_1(\boldsymbol{y}, \theta) = \boldsymbol{y} - \theta$; $g_2(\boldsymbol{y}, \theta) = \boldsymbol{y}^2 - \theta - \theta^2$. In this example, $r = 2 > p = 1$. In this situation, we can estimate θ by maximizing the EL subject to the constraint $E[g(\boldsymbol{y}, \theta)] = \mathbf{0}$. The EL has been successfully applied in areas such as linear models, GLMs, survey sampling, variable selection, survival analysis, and time series.

We investigate the use of a nonparametric EL in longitudinal data analysis. We explore the asymptotic properties of the method and assess the performance of the method based on a large number of simulations. Our procedure provides consistent estimators, and has comparable performance to marginal models when the model assumptions are correct. It is superior to marginal models when the variance function and correlation structure are misspecified.

This result motivates us to extend the EL to the penalized EL-based variable selection with carrying out the estimation of the coefficients simultaneously. Simulation studies show that when the model assumptions are correct, its performance is comparable to that of existing methods, and when the model is misspecified, our method has clear advantages.

1.6 Outline of Thesis

The main goal of this thesis is to explore longitudinal data analysis based on a nonparametric approach. We focus on the EL via a set of GEEs. In Chapter 2, we develop the subject-wise EL via a set of GEEs of the parameter of interest, and discuss its characteristics and asymptotic properties. We also provide an algorithm based on EL principles for the estimation of the regression parameters and the construction of a confidence region for the parameter of interest. In Chapter 3 we present a performance analysis of our method in the context of count and continuous longitudinal data. In Chapter 4, we extend this EL to penalized EL variable selection for high-dimensional longitudinal data. We discuss its characteristics and asymptotic properties, and provide an algorithm. We also present a performance analysis of the penalized EL variable selection in the context of count and continuous longitudinal data. In Chapter 5 we apply our method to two case examples. In Chapter 6 we provide concluding remarks, and discuss future research.

Chapter 2

Empirical Likelihood

The EL method is a powerful inference tool with promising applications in many areas of statistics. It is a nonparametric-likelihood-based approach, introduced by Owen [1988] that is an alternative to parametric likelihood and bootstrap methods. This method enables us to fully employ the information available from the data for making asymptotically efficient inference about the population parameters. In this chapter, we introduce the basic concept of EL for a mean vector and discuss EL-based longitudinal modelling.

2.1 Empirical Likelihood for Mean

For a given random sample $y_1, y_2, \ldots, y_n$ from a known density $f(\boldsymbol{y}, \boldsymbol{\mu})$, let $L(\boldsymbol{\mu}) = \prod_{i=1}^{n} f(y_i, \boldsymbol{\mu})$ be the likelihood function for the parameter $\boldsymbol{\mu}$, and let $\hat{\boldsymbol{\mu}} = \arg\max_{\mu} L(\boldsymbol{\mu})$ be the maximum likelihood estimator. Suppose we wish to test the hypothesis $H_0: \ \boldsymbol{\mu} = \boldsymbol{\mu}_0$. Let

$$R(\boldsymbol{\mu}_0) = \frac{L(\boldsymbol{\mu}_0)}{L(\hat{\boldsymbol{\mu}})}$$

be the likelihood ratio statistic. Wilks' theorem (Wilks [1938]) states that under some regularity conditions $-2\log R(\boldsymbol{\mu}_0)$ converges in distribution to a chi-square with degrees of freedom equal to the dimension of $\boldsymbol{\mu}$.

Consider a random sample $y_1, y_2, \ldots, y_n$ from an unknown distribution function $F(y)$ with $p_i = Pr(Y_i = y_i)$, $i = 1, \ldots, n$, where $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$. Since $Pr(Y_1 = y_1, \ldots, Y_n = y_n) = p_1 \ldots p_n$, the likelihood function of F can be written

$$L_n(F) = \prod_{i=1}^{n} p_i,$$

which is called an EL. The maximum EL estimator (MELE) for F gives an equal mass probability $1/n$ for the n observed values. The corresponding cumulative empirical distribution function of $y_1, y_2, \ldots, y_n$ is

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n} I(y_i \leq y),$$

where $I(*)$ is the indicator function and the inequality is expressed componentwise. The log EL is of the form

$$l_n(F) = \sum_{i=1}^{n} \log(p_i), \tag{2.1}$$

subject to the constraints $\sum\limits_{i=1}^{n} p_i = 1$ and $p_i \geq 0, i = 1, 2, \ldots, n$.

Suppose that we are interested in the parameter $\mu = T(F)$ under the assumption that F is a member of a unknown distribution family $\mathcal{F}$, for some functional T of the distribution. The purpose of the profile likelihood is to find the F at which the EL attains its maximum value over the set $\{F : T(F) = \boldsymbol{\mu}\}$. Define the profile EL function to be

$$L_n(\boldsymbol{\mu}) = \sup\{L_n(F) \mid T(F) = \boldsymbol{\mu},\ F \in \mathcal{F}\}.$$

We can make likelihood inference on $\boldsymbol{\mu}$ based on $L_n(\boldsymbol{\mu})$. This likelihood has similar properties to its parametric counterpart. Since $L_n(\boldsymbol{\mu}) \leq n^{-n}$, it is convenient to standardize $L_n(\boldsymbol{\mu})$ by defining the likelihood ratio function to be

$$R(F) = n^n L_n(\boldsymbol{\mu}),$$

and it is easily shown that this can be written

$$R(F) = \prod_{i=1}^{n} np_i.$$

To obtain confidence regions for $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_d)$, we define the profile empirical log-likelihood to be

$$\ell(\boldsymbol{\mu}) = \sup\left\{ l_n(F):\ p_i \geq 0, i = 1, 2, \ldots, n;\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i(y_i - \boldsymbol{\mu}) = \mathbf{0} \right\}. \quad (2.2)$$

We can compute $\ell(\boldsymbol{\mu})$ by maximizing $l_n(F)$ given in (2.1) as a constrained optimization problem using the Lagrange multiplier method. Making use of the first-order conditions of the Lagrangian function with respect to the p_i and the constraint on p_i, we see that the likelihood is maximized when

$$\hat{p}_i = \frac{1}{n\left\{1 + \hat{\boldsymbol{\lambda}}^T (y_i - \boldsymbol{\mu})\right\}},\ i = 1, 2, \ldots, n,$$

and the Lagrange multiplier $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}(\mu)$ is the solution of

$$\sum_{i=1}^{n} \frac{(y_i - \boldsymbol{\mu})}{1 + \boldsymbol{\lambda}^T (y_i - \boldsymbol{\mu})} = \mathbf{0}.$$

Therefore, we can write the profile EL function as

$$\ell(\boldsymbol{\mu}) = -n \log(n) - \sum_{i=1}^{n} \log(1 + \hat{\boldsymbol{\lambda}}^T(\mu)(y_i - \boldsymbol{\mu})).$$

Consequently, the profile empirical log-likelihood ratio function becomes

$$W(\boldsymbol{\mu}) = \sum_{i=1}^{n} \log(n\hat{p}_i) = \sum_{i=1}^{n} \log\left[1 + \hat{\boldsymbol{\lambda}}^T(\mu)(y_i - \boldsymbol{\mu})\right].$$

Owen [1990] showed that when $\boldsymbol{\mu}_0$ is the true population mean, $2W(\boldsymbol{\mu}_0) \xrightarrow{D} \chi^2_d$ as $n \longrightarrow \infty$; similar to the parametric likelihood ratio function of Wilks [1938]. This result is useful for testing the hypothesis $H_0 : \mu_0 = T(F_0)$ and for the construction of $100(1-\alpha)\%$ confidence regions, defined by

$$\left\{\boldsymbol{\mu} : 2W(\mu) \leq \chi^2_d(1-\alpha)\right\},$$

where $\chi^2_d(1-\alpha)$ is the $(1-\alpha)$th quantile of the chi-square distribution with d degrees of freedom. These are different from the CIs based on a normal approximation. Note that there is no need to estimate the scale parameters in the construction of the CI, and the confidence regions are not necessarily symmetric because of the data-driven approach. Because of these properties, the EL method has become popular in the statistical literature.

2.2 Empirical-Likelihood-Based Longitudinal Modelling

Owen [1991] first considered the EL for linear models. EL confidence regions for regression coefficients in linear models were studied by Chen [1994]. The EL method can also be used to estimate the parameters defined by a set of estimating equations (Qin and Lawless [1994]). Owen [2001] provides a comprehensive overview of the EL and its properties. EL methods have attracted increasing attention over the last two decades, and the literature is extensive.

You, Chen and Zhou [2006] were the first to apply the EL to longitudinal data, using a subject-wise working independence model. This method ignores the within-subject correlation structure. Xue and Zhu [2007] proposed a subject-wise EL by centering the longitudinal data and obtained asymptotic normality of the MELE of the regression coefficients. They did not consider the within-subject correlation structure. It is well known that the working-independence assumption may lead to a loss of efficiency in estimation when within-subject correlation is present. Wang, Qian and Carroll [2010] showed how to incorporate the within-subject correlation structure of continuous repeated measurements into the EL. To estimate the within-subject covariance matrices, they used the nonparametric sample covariance matrix obtained from the residuals of a GEE using the working-independence assumption. In this thesis, we show how to incorporate the within-subject correlation structure of the repeated measurements into the EL.

We propose a subject-wise EL that assigns a probability p_i to subject i. For the ith subject, let $\boldsymbol{y}_i = (y_{i1}, \dots, y_{it} \dots, y_{im_i})^T$ be the response vector, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \dots, \boldsymbol{x}_{im_i})^T$ the $m_i \times p$ matrix of covariates, and $\boldsymbol{\beta} \in \mathcal{R}^p$ the vector of the regression effects of $\boldsymbol{x}$ on $\boldsymbol{y}$. We assume that all the subjects are independent and the repeated measurements

y_{it} taken on each subject are correlated.

Assuming the existence of the first two moments of $\boldsymbol{y}$, we can estimate the regression parameters using the unbiased estimating equation

$$g(\boldsymbol{Y}_i; \boldsymbol{\beta}, \hat{\rho}(\beta)) = \sum_{i=1}^{k} \boldsymbol{X}_i^T \boldsymbol{A}_i \boldsymbol{\Sigma}_i^{-1}(\widehat{\rho})(\boldsymbol{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

as given in (1.5). Following Owen [1991] and Qin and Lawless [1994], we can extend the EL inference to longitudinal data based on a set of estimating functions $g(\boldsymbol{Y}; \boldsymbol{\beta}, \hat{\rho}(\beta))$. We incorporate the within-subject correlation structure of the repeated measurements into the EL using the well-known method of moments estimators given in (1.7) and (1.8) for a given value of $\boldsymbol{\beta}$. The profile empirical log-likelihood function of $\boldsymbol{\beta}$ is defined by

$$\ell(\boldsymbol{\beta}) = \sup \left[\sum_{i=1}^{k} \log(p_i) : \; p_i \geq 0, i = 1, 2, \ldots, k; \; \sum_{i=1}^{k} p_i = 1, \; \sum_{i=1}^{k} p_i g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) = \mathbf{0} \right].$$

The EL is maximized when

$$\hat{p}_i = \frac{1}{k \left\{ 1 + \hat{\boldsymbol{\lambda}}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) \right\}}, \; i = 1, 2, \ldots, k, \tag{2.3}$$

where the Lagrange multiplier $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}(\beta)$ is the solution of

$$\sum_{i=1}^{k} \frac{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))} = \mathbf{0}. \tag{2.4}$$

This result leads to the profile empirical log-likelihood function

$$\ell(\boldsymbol{\beta}) = -k \log(k) - \sum_{i=1}^{k} \log(1 + \hat{\boldsymbol{\lambda}}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)))$$

and the profile empirical log-likelihood ratio function

$$W_l(\boldsymbol{\beta}) = -\sum_{i=1}^{k} \log(k\hat{p}_i) = \sum_{i=1}^{k} \log[1 + \hat{\boldsymbol{\lambda}}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))]. \tag{2.5}$$

Under some regularity conditions, we have $2W_l(\boldsymbol{\beta}_0) \xrightarrow{D} \chi^2_p$ as $k \longrightarrow \infty$ if

$$E\left[g(\boldsymbol{\beta}_0, \hat{\rho}(\beta_0)) g^T(\boldsymbol{\beta}_0, \hat{\rho}(\beta_0))\right]$$

is full rank where $\boldsymbol{\beta}_0$ is the true parameter value. This conclusion is similar to that for the parametric likelihood ratio function. The vector $\boldsymbol{\beta}$ can be estimated by minimizing

$$W_l(\boldsymbol{\beta}) = \sum_{i=1}^{k} \log(1 + \hat{\boldsymbol{\lambda}}^T(\beta) g(\boldsymbol{\beta}, \hat{\rho}(\beta))) \tag{2.6}$$

with respect to $\boldsymbol{\beta}$. Note that the profile log-likelihood ratio function can be minimized with respect to $\boldsymbol{\beta}$ when ρ is known. In practice, ρ is unknown, but can be consistently estimated using the method of moments; see Section 1.3.3.

The computation of the profile EL function is a key step in EL applications, and it involves constrained maximization. In some situations, the algorithm may fail because of poor initial values of the parameters. Moreover, the poor accuracy of EL confidence regions has been reported by several authors, including Qin and Lawless [1994], Hall and La Scala [1990], Corcoran, Davison and Spady [1995], Owen [2001], Tsao [2004], and Chen, Variyath and Abraham [2008]. In the next subsection we will discuss how to address these problems in the context of longitudinal data.

2.2.1 Adjusted Empirical Likelihood

The computation of the profile EL ratio function $W_l(\boldsymbol{\beta})$ given in (2.6) is a key step in EL applications. The solution for $\boldsymbol{\lambda}$ must satisfy $\{1 + \hat{\boldsymbol{\lambda}}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\} > 0$ for all $i = 1, \ldots, k$. A necessary and sufficient condition for its existence is that the vector $\mathbf{0}$ is an interior point of the convex hull of $\{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)), i = 1, \ldots, k\}$. Under some moment conditions on $g(\boldsymbol{\beta}, \hat{\rho}(\beta))$ (Owen [2001]), the convex hull contains $\mathbf{0}$ as an interior point with probability 1 as $k \to \infty$. However, when $\boldsymbol{\beta}$ is not close to the true parameter value $\boldsymbol{\beta}_0$ or when k is small, it is possible that the solution of (2.4) does not exist. To avoid this problem, Chen et al. [2008] introduced the adjusted EL (AEL). The AEL is obtained by adding a pseudo-observation to the data set. It overcomes the difficulties arising when the estimating equations for $\boldsymbol{\lambda}$ have no solution.

Let $g_i(\boldsymbol{\beta}) = g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))$ and $\overline{g}_k(\boldsymbol{\beta}) = \dfrac{1}{k} \sum_{i=1}^{k} g_i(\boldsymbol{\beta})$ for any given $\boldsymbol{\beta}$. For some positive constant b_k, by the addition of an artificial observation

$$g_{k+1}(\boldsymbol{\beta}) = -\frac{b_k}{k} \sum_{i=1}^{k} g_i(\boldsymbol{\beta}) = -b_k \overline{g}_k(\beta)$$

with $b_k = \log(k)/2$. The adjusted profile empirical log-likelihood ratio function is

$$W_l^*(\boldsymbol{\beta}) = \inf \left[-\sum_{i=1}^{k+1} \log\left[(k+1)p_i\right] : p_i \geq 0, i = 1, 2, \ldots, k+1; \sum_{i=1}^{k+1} p_i = 1, \sum_{i=1}^{k+1} p_i g_i(\boldsymbol{\beta}) = \mathbf{0} \right]$$

$$= \sum_{i=1}^{k+1} \log\left[1 + \hat{\boldsymbol{\lambda}}^T(\beta) g_i(\boldsymbol{\beta})\right]$$

with $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}(\beta)$ being the solution of $\displaystyle\sum_{i=1}^{k+1} \frac{g_i(\boldsymbol{\beta})}{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta})} = \mathbf{0}$. Note that $\mathbf{0}$ always lies inside the convex hull of $\{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)), i = 1, \ldots, k+1\}$. The adjusted profile empirical log-likelihood ratio function is well defined after adding a pseudo value $g_{k+1}(\boldsymbol{\beta})$. For a wide range of b_k, following Chen et al. [2008], we can show that the adjusted profile EL ratio function $W_l^*(\boldsymbol{\beta})$ has the same asymptotic properties as the unadjusted profile EL ratio function $W_l(\boldsymbol{\beta})$. We define the adjusted profile EL ratio estimator of $\boldsymbol{\beta}$ to be the minimizer of

$$W_l^*(\boldsymbol{\beta}) = \sum_{i=1}^{k+1} \left[\log(1 + \hat{\boldsymbol{\lambda}}^T(\boldsymbol{\beta}) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\right] \tag{2.7}$$

with respect to $\boldsymbol{\beta}$.

The adjustment is particularly useful because, even for some undesirable values of $\boldsymbol{\beta}$, the algorithm guarantees a solution. The confidence regions constructed via the AEL are found to have better coverage probabilities than those for the regular EL, and the algorithm provides a promising solution for $\boldsymbol{\lambda}$ particularly when the sample size is small. The improved coverage probability is achieved without resorting to more complex procedures such as Bartlett correction or bootstrap calibration.

2.2.2 Extended Empirical Likelihood

One of the advantages of the EL is that we can use more information about the parameters. In other words, we can use more estimating equations than the number of parameters. The extra information should improve the accuracy of the estimates. In such cases, EL-based confidence regions can have undercoverage (Qin and Lawless [1994]). This is mainly because the parameter space is $\mathcal{R}^p$, and the domain is a bounded subset of $\mathcal{R}^p$ (Tsao [2013]; Tsao and Wu [2013]). This mismatch is a result of the convex hull constraint set for the formulation of the EL. As a result, the values of $\boldsymbol{\beta} \in \mathcal{R}^p$ that violate this constraint are excluded from the domain. To overcome this problem, Tsao [2013] and Tsao and Wu [2013] expand the EL domain geometrically. This extended EL (EEL) for parameters based on estimating equations is a natural generalization of the regular EL to the full parameter space. Similar to AEL, EEL have the same asymptotic properties as the EL.

For longitudinal data the EL domain is

$$\Theta_k = \{\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathcal{R}^p \text{ such that } \sum_{i=1}^{k} p_i g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) = \mathbf{0}\},$$

where $g(\boldsymbol{\beta}, \hat{\rho}(\beta))$ is given in (1.5). Tsao and Wu [2013] expand Θ_k to $\mathcal{R}^p$ through a composite similarity mapping $h_k^C : \Theta_k \to \mathcal{R}^p$. They define $h_k^C(\beta)$ to be

$$h_k^C(\boldsymbol{\beta}) = \tilde{\boldsymbol{\beta}} + \gamma(k, W_l(\boldsymbol{\beta}))(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}),\ \boldsymbol{\beta} \in \Theta_k, \tag{2.8}$$

where $\tilde{\boldsymbol{\beta}}$ is the MELE for $\boldsymbol{\beta}$, and the function $\gamma(k, W_l(\boldsymbol{\beta}))$ is the expansion factor, given by

$$\gamma(k, W_l(\boldsymbol{\beta})) = 1 + \frac{W_l(\boldsymbol{\beta})}{k}.$$

Under the regularity conditions discussed by Tsao and Wu [2013], $h_k^C : \Theta_k \to \mathcal{R}^p$ is surjective. Thus, $s(\boldsymbol{\beta}) = \{\boldsymbol{\beta}' : h_k^C(\boldsymbol{\beta}') = \boldsymbol{\beta}\},\ \forall \boldsymbol{\beta} \in \mathcal{R}^p$ is nonempty. If $s(\boldsymbol{\beta})$ contains more than one point and h_k^C does not have an inverse, then h_k^C is surjective. Hence, a generalized inverse $h_k^{-C} : \mathcal{R}^p \to \Theta_k$ is

$$h_k^{-C}(\boldsymbol{\beta}) = \arg \min_{\boldsymbol{\beta}' \in s(\boldsymbol{\beta})} \{\|\boldsymbol{\beta}' - \boldsymbol{\beta}\|\}.$$

The extended profile empirical log-likelihood ratio function $W^{**}(\boldsymbol{\beta})$ under h_k^{-C} is defined to be

$$W_l^{**}(\boldsymbol{\beta}) = W(h_k^{-C}(\boldsymbol{\beta})) \text{ for } \boldsymbol{\beta} \in \mathcal{R}^p,$$

which is well-defined throughout $\mathcal{R}^p$. Tsao and Wu [2013] highlight the first-order version of this EEL, which is easy to use and substantially more accurate than the regular EL. It is also more accurate than available second-order EL methods. In our performance analysis in Chapter 3, we will explore the coverage probabilities based on the EEL, EL, and AEL.

In the next section, following Qin and Lawless [1994], we state and prove the results on the distributional properties of the adjusted profile EL estimates of $\hat{\boldsymbol{\beta}}$. We construct these theorems based on the GEE with lag correlation given in (1.5), since the GEE estimate of $\boldsymbol{\beta}$ under an arbitrary working-correlation structure is not necessarily consistent; see Sections 1.3.3 and 1.4.1.

2.3 Distributional Properties

In this section, we present the first-order asymptotic properties of $\hat{\boldsymbol{\beta}}$ and the adjusted profile empirical log-likelihood ratio statistics. We first introduce some notation and regularity conditions that are used in the theorems and lemma.

Regularity Conditions:

A1: $E\left\{g(\boldsymbol{\beta}_0, \hat{\rho}(\beta_0))\right\} = 0$, where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$, $g(\boldsymbol{\beta}, \hat{\rho}(\beta)) = \sum_{i=1}^{k} \boldsymbol{D}_i^T \boldsymbol{\Sigma}_i^{-1}(\widehat{\rho})(\boldsymbol{y}_i - \boldsymbol{\mu}_i)$ be the estimating function for $\boldsymbol{\beta} \in \mathcal{R}^p$ (defined in (1.5)), $\boldsymbol{D}_i = \partial\{a_i'(\theta)\}/\partial\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_i(\hat{\rho}) = \boldsymbol{A}_i^{1/2}\boldsymbol{C}_i^*(\hat{\rho})\boldsymbol{A}_i^{1/2}$, and $\boldsymbol{A}_i = \mathrm{diag}\{a_i''(\theta)\}$ for $i = 1, 2, \ldots, k$. Let $\overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta)) = \dfrac{1}{k}\sum_{i=1}^{k} g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))$ and $g_{k+1}(\boldsymbol{\beta}, \hat{\rho}(\beta)) = -b_k\overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))$, where b_k is a positive constant.

A2: $\{a'(\theta)\}$ is three times continuously differentiable and $\{a''(\theta)\} > 0$ in Θ°, where Θ be the natural parameter space of the exponential family distributions presented in (1.1) and Θ° the interior of Θ. Also, $h(\eta)$ is three times continuously differentiable and $h'(\eta) > 0$.

A3: $E_{\boldsymbol{\beta}_0}\left\{\frac{\partial g_k(\boldsymbol{\beta},\rho)}{\partial\boldsymbol{\beta}}\right\}$ and $\boldsymbol{V}_k(\boldsymbol{\beta}_0,\hat{\rho}(\beta_0)) = E_{\boldsymbol{\beta}_0}\left\{g_k(\boldsymbol{\beta},\hat{\rho}(\beta))g_k^T(\boldsymbol{\beta},\hat{\rho}(\beta))\right\}$ are positive definite.

A4: The rank of $E\left\{\frac{\partial g_k(\boldsymbol{\beta},\rho)}{\partial\boldsymbol{\beta}}\right\}$ is p in a neighbourhood of $\boldsymbol{\beta}_0$.

A5: There exist functions $G(\boldsymbol{y},\boldsymbol{X})$ such that in a neighbourhood of $\boldsymbol{\beta}_0$.

$$\left|\frac{\partial g_k(\boldsymbol{\beta},\rho)}{\partial\boldsymbol{\beta}}\right| < G(\boldsymbol{y},\boldsymbol{X}), \|g_k(\boldsymbol{y},\boldsymbol{X},\boldsymbol{\beta},\hat{\rho}(\beta))\|^3 < G(\boldsymbol{y},\boldsymbol{X})$$

with $E\left[G(\boldsymbol{y},\boldsymbol{X})\right] < \infty$.

Lemma 2.3.1 *Under regularity conditions A1-A5, suppose* $(\boldsymbol{y}_i,\boldsymbol{X}_i), i = 1,2,\ldots,k$ *is a set of independent and identically distributed random vectors. Let*

$$2W_l^*(\boldsymbol{\beta}) = 2\sum_{i=1}^{k+1}\log\left[1+\hat{\boldsymbol{\lambda}}^T(\beta)g_i(\boldsymbol{\beta},\hat{\rho}(\beta))\right] \tag{2.9}$$

be the adjusted profile empirical log-likelihood ratio function. Then, as $k\to\infty$, $\hat{\rho}(\beta)$ *is a consistent estimator in the neighbourhood of* $\boldsymbol{\beta}$*; the correlation matrix of* $\boldsymbol{y}_i$ *is* $\boldsymbol{C}_i^*(\boldsymbol{\rho})$*, defined in (1.5) and* $W_l^*(\boldsymbol{\beta})$ *attains its minimum value at some point* $\hat{\boldsymbol{\beta}}$ *in the interior of* $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\| < k^{-1/3}$ *in probability. In addition,* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\boldsymbol{\lambda}}(\beta)$ *satisfy the equations*

$$Q_{1,k+1}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\lambda}},\hat{\rho}(\beta)) = \boldsymbol{0} \quad \textit{and} \quad Q_{2,k+1}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\lambda}},\hat{\rho}(\beta)) = \boldsymbol{0}$$

where

$$Q_{1,k+1}(\boldsymbol{\beta},\boldsymbol{\lambda},\rho(\beta)) = \frac{1}{k}\sum_{i=1}^{k+1}\frac{g_i(\boldsymbol{\beta},\rho(\beta))}{1+\boldsymbol{\lambda}^T(\beta)g_i(\boldsymbol{\beta},\rho(\beta))},$$

$$Q_{2,k+1}(\boldsymbol{\beta},\boldsymbol{\lambda},\rho(\beta)) = \frac{1}{k}\sum_{i=1}^{k+1}\frac{1}{1+\boldsymbol{\lambda}^T(\beta)g_i(\boldsymbol{\beta},\rho(\beta))}\left(\frac{\partial g_i(\boldsymbol{\beta},\rho)}{\partial\boldsymbol{\beta}}\right)^T\boldsymbol{\lambda}.$$

Proof of Lemma 2.3.1:

First, we will show that the $\hat{\rho}(\beta)$ is consistent estimator in a neighborhood of $\boldsymbol{\beta}$ for the repeated count responses which can be generalized for any repeated responses. Second, we are going to show that the Lagrange multiplier $\boldsymbol{\lambda}(\beta) = O_p(k^{-1/3})$ for β such that $\|\beta - \beta_0\| \leq k^{-1/3}$ and then going to show the consistency of the minimum adjusted profile empirical likelihood ratio estimators $\hat{\boldsymbol{\beta}}$.

Let $y_{i1}, \ldots, y_{it}, \ldots, y_{im}$ be the repeated count responses with the variance and the lag 1 covariance are given by

$$\mathrm{E}(y_{it} - \mu_{it})^2 = \sigma_{itt}$$

and

$$\mathrm{E}[(y_{it} - \mu_{it})(y_{i,t+1} - \mu_{i,t+1})] = \sigma_{i,t,t+1} = \rho\mu_{it} + \mu_{it}\mu_{i,t+1}$$

respectively. Let $\tilde{y}_{it}$ be the standardized residual, defined as $\tilde{y}_{it} = \{y_{it} - \mu_{it}\}/\sqrt{\sigma_{itt}}$. Then it follows that

$$\mathrm{E}\left[\sum_{i=1}^{k}\sum_{t=1}^{m}\tilde{y}_{it}^2/km\right] = 1 \tag{2.10}$$

and

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^{k}\sum_{t=1}^{m-1}\tilde{y}_{it}\tilde{y}_{i,t+1}/k(m-1)\right] &= \rho\sum_{i=1}^{k}\sum_{t=1}^{m-1}\mu_{it}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2}/k(m-1)\\ &+\sum_{i=1}^{k}\sum_{t=1}^{m-1}\mu_{it}\mu_{i,t+1}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2}/k(m-1).\end{aligned} \tag{2.11}$$

By using (2.10) and (2.11), we can write the first order approximate expectation

$$\mathrm{E}(w_1) = \rho h_1 + f_1, \tag{2.12}$$

where

$$w_1 = \frac{\displaystyle\sum_{i=1}^{k}\sum_{t=1}^{m-1} \tilde{y}_{it}\tilde{y}_{i,t+1}/k(m-1)}{\displaystyle\sum_{i=1}^{k}\sum_{t=1}^{m} \tilde{y}_{it}^2/km}$$

is the lag 1 correlation as discussed in Section 1.3.3, and

$$h_1 = \sum_{i=1}^{k}\sum_{t=1}^{m-1} \mu_{it}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2}/k(m-1),$$

and

$$f_1 = \sum_{i=1}^{k}\sum_{t=1}^{m-1} \mu_{it}\mu_{i,t+1}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2}/k(m-1).$$

We then obtain an approximate unbiased moment estimator of ρ

$$\hat{\rho} = \frac{w_1 - f_1}{h_1}.$$

Now consider,

$$\mathrm{E}\left[\left(\frac{y_{it} - \mu_{it}}{\sqrt{\sigma_{itt}}}\right)\left(\frac{y_{i,t+1} - \mu_{i,t+1}}{\sqrt{\sigma_{i,t+1,t+1}}}\right)\right] = \sigma_{i,t,t+1}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2} \text{ for all } i \text{ and } j.$$

That is,

$$\mathrm{E}\left[\sum_{t=1}^{m-1} \tilde{y}_{it}\tilde{y}_{i,t+1} - \sigma_{i,t,t+1}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2}\right] = 0 \text{ for all } i = 1, \ldots, k,$$

where $\tilde{y}_{it}$ is the standardized residual. If μ_{ij}'s and m are bounded, it is easy to see that

$$\mathrm{E}\left[\left(\sum_{t=1}^{m-1} \left[\tilde{y}_{it}\tilde{y}_{i,t+1} - \sigma_{i,t,t+1}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2}\right]\right)^2\right] < M$$

for some $0 < M < \infty$ for all $i = 1, \ldots, k$ and $\boldsymbol{y}_i$'s are independent. By the law of large numbers for independent random variable we can conclude that

$$\sum_{i=1}^{k}\sum_{t=1}^{m-1}\left[\tilde{y}_{it}\tilde{y}_{i,t+l}-\sigma_{i,t,t+1}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2}\right]/k(m-1)\xrightarrow{P}0 \quad as \quad k\to\infty.$$

Now, (2.11) can be written

$$\begin{aligned}\sum_{i=1}^{k}\sum_{t=1}^{m-1}\tilde{y}_{it}\tilde{y}_{i,t+1}/k(m-1) &= \rho\sum_{i=1}^{k}\sum_{t=1}^{m-1}\mu_{it}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2}/k(m-1)\\ &+\sum_{i=1}^{k}\sum_{t=1}^{m-1}\mu_{it}\mu_{i,t+1}(\sigma_{itt}\sigma_{i,t+1,t+1})^{-1/2}/k(m-1)+O_p(1). \qquad (2.13)\end{aligned}$$

Similarly, consider,

$$\mathrm{E}\left[\tilde{y}_{it}^2\right]=1 \text{ for all } i \text{ and } j.$$

That is,

$$\mathrm{E}\left[\sum_{t=1}^{m}\left(\tilde{y}_{it}^2-1\right)\right]=0 \text{ for all } i=1,\ldots,k.$$

Then if μ_{ij}'s and m are bounded,

$$\mathrm{E}\left[\left(\sum_{t=1}^{m}\left[\tilde{y}_{it}^2-1\right]\right)^2\right]<M$$

for some M for all $i=1,\ldots,k$ and $\boldsymbol{y}_i$'s are independent. By the law of large numbers for independent random variable we can again conclude that

$$\sum_{i=1}^{k}\sum_{t=1}^{m-1}\left[\tilde{y}_{it}^2-1\right]/km\xrightarrow{P}0.$$

From this we can write

$$\sum_{i=1}^{k}\sum_{t=1}^{m-1}\tilde{y}_{it}^2/km=1+O_p(1). \qquad (2.14)$$

By using (2.13) and (2.14), we can obtain $w_1(1+O_p(1)) = \rho h_1 + f_1 + O_p(1) \Rightarrow \hat{\rho} = \dfrac{w_1 - f_1}{h_1} = \rho + O_p(1)$ as $k \to \infty$, where w_1, f_1, and h_1 are defined in (2.12). So

$$\hat{\rho} = \frac{w_1 - f_1}{h_1} \xrightarrow{P} \rho \; as \; k \to \infty. \tag{2.15}$$

Let $V_k(\boldsymbol{\beta}, \hat{\rho}(\beta)) = \frac{1}{k}\sum_{i=1}^{k} g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) g_i^T(\boldsymbol{\beta}, \hat{\rho}(\beta))$ and $\sigma_{1k} \geq \sigma_{2k}, \ldots, \sigma_{pk} > 0$ be eigenvalues of $V_k(\boldsymbol{\beta}_0, \hat{\rho}(\beta)))$. Without loss of generality, we assume $\sigma_{1k} \to 1$. We will claim that $\hat{\boldsymbol{\lambda}} = O_p(1)$. Let $g^*(\boldsymbol{\beta}, \hat{\rho}(\beta)) = \max_{1 \leq i \leq k} \|g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\|$. By condition A5, $g^*(\boldsymbol{\beta}, \hat{\rho}(\beta)) \leq \max\left\{G^{1/3}(\boldsymbol{y}_i, \boldsymbol{X}_i)\right\}, \; i = 1, 2, \ldots, k+1$ in the neighborhood of $\boldsymbol{\beta}_0$.
Following by Owen [2001] Lemma 11.2 (page 218), we obtain, $g^*(\boldsymbol{\beta}, \hat{\rho}(\beta)) = O_p(k^{1/3})$.
We now consider the order of $\overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))$.

$$\overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta)) = \frac{1}{k}\sum_{i=1}^{k} g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))$$

$$= \frac{1}{k}\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) + \frac{1}{k}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\sum_{i=1}^{k} \frac{\partial g_i(\boldsymbol{\beta}^*, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}},$$

where $\boldsymbol{\beta}^*$ is in the neighborhood of $\boldsymbol{\beta}_0$.
By condition A1 $E[g(\boldsymbol{\beta}_0, \hat{\rho}(\beta))] = \mathbf{0}$ and $V[g(\boldsymbol{\beta}_0, \hat{\rho}(\beta))] < \infty$, we have

$$\overline{g}_k(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) = \frac{1}{k}\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) = O_p(k^{-1/2}). \tag{2.16}$$

Also by condition A5, $\left|\dfrac{\partial g_i(\boldsymbol{\beta}, \rho)}{\partial \boldsymbol{\beta}}\right| < G(\boldsymbol{y}, \boldsymbol{X})$ such that $E\left[G(\boldsymbol{y}, \boldsymbol{X})\right] < \infty$, we have

$$\frac{1}{k}\sum_{i=1}^{k} \frac{\partial g_i(\boldsymbol{\beta}^*, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}} = O_p(1).$$

Therefore, uniformly in the region of $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| < k^{-1/3}$, we have

$$\overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta)) = O_p(k^{-1/3}).$$

Note that $\hat{\boldsymbol{\lambda}}(\boldsymbol{\beta})$ is the solution of the equation

$$\sum_{i=1}^{k+1} \frac{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}{1+\boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))} = \mathbf{0}. \tag{2.17}$$

Let $\xi = \|\hat{\boldsymbol{\lambda}}\|$ and $\boldsymbol{\vartheta} = \dfrac{\boldsymbol{\lambda}}{\xi}$. Multiplying (2.17) by $\dfrac{\boldsymbol{\vartheta}^T}{k}$, we get

$$\begin{aligned}
\mathbf{0} &= \frac{\boldsymbol{\vartheta}^T}{k} \sum_{i=1}^{k+1} \frac{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}{1+\boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))} \\
&= \frac{\boldsymbol{\vartheta}^T}{k} \sum_{i=1}^{k+1} g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) - \frac{\xi}{k} \sum_{i=1}^{k+1} \frac{\{\boldsymbol{\vartheta}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}^2}{1+\xi \boldsymbol{\vartheta}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))} && . \\
&\leq \boldsymbol{\vartheta}^T \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))(1-b_k/k) - \frac{\xi}{k(1+\xi g^*(\boldsymbol{\beta}, \hat{\rho}(\beta)))} \sum_{i=1}^{k} \{\boldsymbol{\vartheta}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}^2 && . \\
&= \boldsymbol{\vartheta}^T \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta)) - \frac{\xi}{k(1+\xi g^*(\boldsymbol{\beta}, \hat{\rho}(\beta)))} \sum_{i=1}^{k} \{\boldsymbol{\vartheta}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}^2 + O_p(k^{-4/3} b_k). && (2.18)
\end{aligned}$$

Since $1+\xi g^*(\boldsymbol{\beta}, \hat{\rho}(\beta)) \geq 0$, we have

$$\boldsymbol{\vartheta}^T \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta)) \geq \frac{\xi}{k[1+\xi \boldsymbol{\vartheta}^T g^*(\boldsymbol{\beta}, \hat{\rho}(\beta))]} \sum_{i=1}^{k} \{\boldsymbol{\vartheta}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}^2 + O_p(k^{-4/3} b_k).$$

For some $0 < \epsilon < 1$, the variance assumption on $g(\boldsymbol{\beta}, \hat{\rho}(\beta))$ in condition A3 gives

$$\frac{1}{k} \sum_{i=1}^{k} \{\boldsymbol{\vartheta}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}^2 \geq (1-\epsilon)\sigma_{1k}^2 = (1-\epsilon)$$

in probability. Therefore, as long as $b_k = o_p(k)$, (2.18) implies that

$$\frac{\xi}{[1+\xi g^*(\boldsymbol{\beta}, \hat{\rho}(\beta))]} \leq \boldsymbol{\vartheta}^T \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta)) \times \frac{1}{(1-\epsilon)} = O_p(k^{-1/3}).$$

From this, we can arrive $\xi = O_p(k^{-1/3})$ and hence $\boldsymbol{\lambda}(\beta) = O_p(k^{-1/3})$.

Next,

$$\begin{aligned}
\mathbf{0} &= \frac{1}{k}\sum_{i=1}^{k+1} \frac{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}{1+\boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))} \\
&= \frac{1}{k}\sum_{i=1}^{k+1} g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) - \frac{1}{k}\sum_{i=1}^{k+1} \boldsymbol{\lambda}^T(\beta)\left[g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) g_i^T(\boldsymbol{\beta}, \hat{\rho}(\beta))\right] + o_p(k^{-1/3}) \\
&= \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))(1-b_k/k) - \boldsymbol{\lambda}^T(\beta)\boldsymbol{V}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))(1+b_k^2/k) + o_p(k^{-1/3}),
\end{aligned} \tag{2.19}$$

where $g_{k+1}(\boldsymbol{\beta}, \hat{\rho}(\beta)) = -b_k \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))$ and b_k is a positive constant. Hence, when $k \to \infty$,

$$\boldsymbol{\lambda}(\boldsymbol{\beta}) = \boldsymbol{V}_k^{-1}(\boldsymbol{\beta}, \hat{\rho}(\beta)) \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta)) + o_p(k^{-1/3}). \tag{2.20}$$

This result corresponds to Lemma 1 in Qin and Lawless [1994] which is about the consistency of maximum empirical likelihood estimates for independent and identically distributed data. By following Qin and Lawless [1994], under the regularity conditions A2-A5 and using (2.20), we can obtain, as $k \to \infty$, with probability tending to 1 the equations $Q_{1,k+1}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho(\beta))$ and $Q_{2,k+1}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \rho(\beta))$ has a solution within the open ball $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| < k^{-1/3}$. It is noted that rest of the proof is similar to the proof of Lemma 1 in Qin and Lawless [1994] and the details are omitted here. This completes the proof.

Theorem 2.3.2 *In addition to the regularity conditions A1-A5, suppose that* $\dfrac{\partial^2 g(\boldsymbol{\beta}, \rho)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$

is bounded by some integrable function $G(\boldsymbol{y}, \boldsymbol{X})$ in the neighbourhood. Then, there exists a sequence of adjusted profile EL estimates $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ such that

$$\sqrt{k}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Delta}),$$

$$\sqrt{k}\left(\hat{\boldsymbol{\lambda}} - \mathbf{0}\right) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{U})$$

where

$$\boldsymbol{\Delta} = \left[E_{\boldsymbol{\beta}_0}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T \left\{E_{\boldsymbol{\beta}_0}\left\{g(\boldsymbol{\beta}, \hat{\rho}(\beta)) g^T(\boldsymbol{\beta}, \hat{\rho}(\beta))\right\}^{-1}\right\} E_{\boldsymbol{\beta}_0}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}\right]^{-1}$$

and

$$\boldsymbol{U} = \left\{E_{\boldsymbol{\beta}_0}\left\{g(\boldsymbol{\beta}, \hat{\rho}(\beta)) g^T(\boldsymbol{\beta}, \hat{\rho}(\beta))\right\}^{-1}\right\}$$

$$\left[I - E_{\boldsymbol{\beta}_0}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\} \boldsymbol{\Delta} E_{\boldsymbol{\beta}_0}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T \left\{E_{\boldsymbol{\beta}_0}\left\{g(\boldsymbol{\beta}, \hat{\rho}(\beta)) g^T(\boldsymbol{\beta}, \hat{\rho}(\beta))\right\}^{-1}\right\}\right].$$

Proof of Theorem 2.3.1:

Theorem 2.3.1 shows that the adjusted profile empirical likelihood ratio estimators $\boldsymbol{\beta}$ is consistent and asymptotically normally distributed. Since $g(\boldsymbol{\beta}, \hat{\rho}(\beta))$ is a smooth function this implies local minimum $\hat{\boldsymbol{\beta}}$ satisfied the following equations

$$Q_{1,k+1}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \hat{\rho}(\beta)) = \frac{1}{k} \sum_{i=1}^{k+1} \frac{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}{1 + \boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))} = \mathbf{0}.$$

$$Q_{2,k+1}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \hat{\rho}(\beta)) = \frac{1}{k} \sum_{i=1}^{k+1} \frac{1}{1 + \boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))} \left(\frac{\partial g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right)^T \boldsymbol{\lambda} = \mathbf{0}.$$

Now by using (2.19), $Q_{1,k+1}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \hat{\rho}(\beta))$ can be written

$$Q_{1,k+1}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \hat{\rho}(\beta)) = \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))(1 - b_k/k) - \boldsymbol{\lambda}^T(\beta) \boldsymbol{V}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))(1 + b_k^2/k) + o_p(k^{-1/3}).$$

The partial derivatives of the above equations are

$$\frac{\partial Q_{1,k+1}(\boldsymbol{\beta}_0, \mathbf{0}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}} = \frac{1}{k}\sum_{i=1}^{k}\frac{\partial g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}},$$

$$\frac{\partial Q_{1,k+1}(\boldsymbol{\beta}_0, \mathbf{0}, \hat{\rho}(\beta))}{\partial \boldsymbol{\lambda}} = -\frac{1}{k}\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) g_i^T(\boldsymbol{\beta}_0, \hat{\rho}(\beta)),$$

$$\frac{\partial Q_{2,k+1}(\boldsymbol{\beta}_0, \mathbf{0}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}} = \mathbf{0},$$

$$\frac{\partial Q_{2,k+1}(\boldsymbol{\beta}_0, \mathbf{0}, \hat{\rho}(\beta))}{\partial \boldsymbol{\lambda}} = \frac{1}{k}\sum_{i=1}^{k+1}\left\{\frac{\partial g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T = \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\partial g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T \left[1 - \frac{b_k}{k}\right].$$

Now by following Qin and Lawless [1994] in Theorem 1, expanding $Q_{1,k+1}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}, \hat{\rho}(\beta))$ and $Q_{2,k+1}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}, \hat{\rho}(\beta))$ at $(\boldsymbol{\beta}_0, \mathbf{0})$ under the conditions A2-A5, which leads to

$$\begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{bmatrix} = S_k^{-1} \begin{bmatrix} -Q_{1,k+1}(\boldsymbol{\beta}_0, \mathbf{0}, \hat{\rho}(\beta)) + o_p(\delta_k) \\ o_p(\delta_k) \end{bmatrix}$$

where

$$S_k = \begin{bmatrix} \frac{\partial Q_{1,k+1}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \hat{\rho}(\beta))}{\partial \boldsymbol{\lambda}^T} & \frac{\partial Q_{1,k+1}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}} \\ \frac{\partial Q_{2,k+1}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \hat{\rho}(\beta))}{\partial \boldsymbol{\lambda}^T} & \mathbf{0} \end{bmatrix}_{(\boldsymbol{\beta}_0, \mathbf{0})}$$

$$\to \begin{bmatrix} -E\left\{g(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) g^T(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right\} & E\left\{\frac{\partial g(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\} \\ E\left\{\frac{\partial g(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T & \mathbf{0} \end{bmatrix}$$

and $\delta_k = \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \|\hat{\boldsymbol{\lambda}}\|$.

Since by (2.16), $Q_{1,k+1}(\boldsymbol{\beta}_0, \mathbf{0}, \hat{\rho}(\beta)) = \overline{g}_k(\boldsymbol{\beta}_0, \hat{\rho}(\beta))(1 - b_k/k) = O_p(k^{-1/2})$, we can easily show that $\delta_k = O_p(k^{-1/2})$. Then by central limit theorem, we have

$$\sqrt{k}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Delta}),$$

$$\sqrt{k}\left(\hat{\boldsymbol{\lambda}} - \mathbf{0}\right) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{U})$$

where

$$\boldsymbol{\Delta} = \left[E_{\boldsymbol{\beta}_0}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T \left\{E_{\boldsymbol{\beta}_0}\left\{g(\boldsymbol{\beta}, \hat{\rho}(\beta))g^T(\boldsymbol{\beta}, \hat{\rho}(\beta))\right\}^{-1}\right\} E_{\boldsymbol{\beta}_0}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}\right]^{-1},$$

and

$$\boldsymbol{U} = \left\{E_{\boldsymbol{\beta}_0}\left\{g(\boldsymbol{\beta}, \hat{\rho}(\beta))g^T(\boldsymbol{\beta}, \hat{\rho}(\beta))\right\}^{-1}\right\}$$

$$\left[I - E_{\boldsymbol{\beta}_0}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\} \boldsymbol{\Delta} E_{\boldsymbol{\beta}_0}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T \left\{E_{\boldsymbol{\beta}_0}\left\{g(\boldsymbol{\beta}, \hat{\rho}(\beta))g^T(\boldsymbol{\beta}, \hat{\rho}(\beta))\right\}^{-1}\right\}\right].$$

This completes the proof.

Theorem 2.3.3 *Under regularity conditions A1-A5, the adjusted profile empirical log-likelihood ratio function $2W_l^*(\boldsymbol{\beta}_0)$, where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$, is asymptotically chi-squared distributed with degrees of freedom p.*

Proof of Theorem 2.3.3:

Now consider,

$$2W_l^*(\boldsymbol{\beta}_0) = 2\sum_{i=1}^{k+1} \log\{1 + \boldsymbol{\lambda}^T(\beta_0)g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\}$$

$$= 2\sum_{i=1}^{k+1} \boldsymbol{\lambda}^T(\beta_0)g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) - \sum_{i=1}^{k+1}\left[\boldsymbol{\lambda}^T(\beta_0)g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]^2 + o_p(1)$$

By (2.17), we have

$$\mathbf{0} = \sum_{i=1}^{k+1} \frac{\boldsymbol{\lambda}^T(\beta_0)g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{1 + \boldsymbol{\lambda}^T(\beta_0)g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}$$

$$= \sum_{i=1}^{k+1} \boldsymbol{\lambda}^T(\beta_0)g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) - \sum_{i=1}^{k+1}\left[\boldsymbol{\lambda}^T(\beta_0)g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]^2 + o_p(1)$$

Now we have

$$\sum_{i=1}^{k+1} \boldsymbol{\lambda}^T(\beta_0) g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) = \sum_{i=1}^{k+1} \left[\boldsymbol{\lambda}^T(\beta_0) g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]^2 + o_p(1)$$

From this we can write

$$2W_l^*(\boldsymbol{\beta}_0) = \sum_{i=1}^{k+1} \left[\boldsymbol{\lambda}^T(\beta_0) g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]^2 + o_p(1)$$

$$= \sum_{i=1}^{k+1} \boldsymbol{\lambda}^T(\beta_0) g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))^T \boldsymbol{\lambda}(\boldsymbol{\beta}_0) + o_p(1)$$

Substituting the expansion of $\boldsymbol{\lambda}(\beta)$ in (2.20), we get that

$$\begin{aligned}
2W_l^*(\boldsymbol{\beta}_0) &= \left[\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]^T \left[\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \rho) g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))^T\right]^{-1} \left[\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right] + o_p(1) \\
&= \left\{\left[\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))^T\right]^{-1/2} \left[\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]\right\}^T \\
&\quad \left\{\left[\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))^T\right]^{-1/2} \left[\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]\right\} + o_p(1) \\
&= k\left\{\left[\boldsymbol{V}_k(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]^{-1/2} \left[\overline{g}_k(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]\right\}^T \left\{\left[\boldsymbol{V}_k(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]^{-1/2} \left[\overline{g}_k(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right]\right\} + o_p(1) \qquad (2.21)
\end{aligned}$$

Note that, by condition A3 as $k \to \infty$, $k^{-1}\left[\sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right] \to \mathbf{N}(0, \boldsymbol{V})$, where $\boldsymbol{V} = \lim_{k\to\infty} k^{-1} \sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))^T$. Hence adjusted profile empirical log-likelihood ratio function $2W_l^*(\boldsymbol{\beta}_0) \xrightarrow{D} \chi_p^2$. This completes the proof.

We develop an efficient algorithm for estimating the regression parameters. We also construct confidence region for the parameter of interest using EL principles. These topics will be discussed in the next section.

2.4 Algorithm

To implement our method, we need an efficient algorithm. We minimize the profile EL ratio function $W_l(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ using a Newton–Raphson algorithm. At each Newton–Raphson iteration, we compute the Lagrange multiplier for updated values of $\boldsymbol{\beta}$ and $\hat{\rho}(\beta)$. Chen, Sitter and Wu [2002] proposed a modified Newton–Raphson algorithm for computing the Lagrange multiplier for a given value of the parameter. We implemented this method, which is numerically stable. The algorithm given in Sections 2.4.1, 2.4.2, and 2.4.3 can easily be extended to the AEL by the addition of a pseudo-value $g_{k+1}(\boldsymbol{\beta}) = -b_k\overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))$, where b_k is a positive constant.

2.4.1 Computation of Lagrange Multiplier

The Lagrange multiplier $\boldsymbol{\lambda}$ is estimated by solving the equation

$$\sum_{i=1}^{k} \frac{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))} = \mathbf{0}$$

for a given set of vectors $g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))$, $i = 1, 2, \dots k$. Note that the above equation is the derivative of R with respect to $\boldsymbol{\lambda}$ for a given $\boldsymbol{\beta}$, where

$$R = \sum_{i=1}^{k} \log\left\{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\right\}. \tag{2.22}$$

In the EL problem, the solution must satisfy

$$1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) > 0, \quad i = 1, 2, \dots k.$$

The modified Newton–Raphson algorithm for estimating $\boldsymbol{\lambda}$ for a given value of $\boldsymbol{\beta}$ and $\hat{\rho}(\beta)$ is as follows:

1. Set $\boldsymbol{\lambda}^c = \mathbf{0}$, $c = 0$, $\gamma^c = 1$, $\epsilon = 1e^{-08}$, $\rho = \rho^0$, and $\boldsymbol{\beta} = \boldsymbol{\beta}^0$.

2. Let $R^{\boldsymbol{\lambda}}$ and $R^{\boldsymbol{\lambda}\boldsymbol{\lambda}}$ be the first and second partial derivatives of R (given in (2.22)) with respect to $\boldsymbol{\lambda}$:

$$R^{\boldsymbol{\lambda}} = \sum_{i=1}^{k} \left[\frac{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}} \right],$$

$$R^{\boldsymbol{\lambda}\boldsymbol{\lambda}} = -\sum_{i=1}^{k} \left[\frac{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) g_i^T(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}^2} \right].$$

   Compute $R^{\boldsymbol{\lambda}}$ and $R^{\boldsymbol{\lambda}\boldsymbol{\lambda}}$ for $\boldsymbol{\lambda} = \boldsymbol{\lambda}^c$ and let $\Delta(\boldsymbol{\lambda}^c) = -\left[R^{\boldsymbol{\lambda}\boldsymbol{\lambda}}\right]^{-1} R^{\boldsymbol{\lambda}}$.

   If $\|\Delta(\boldsymbol{\lambda}^c)\| < \epsilon$ stop the algorithm and report $\boldsymbol{\lambda}^c$; otherwise continue.

3. Calculate $\boldsymbol{\delta}^c = \gamma^c \Delta(\boldsymbol{\lambda}^c)$. If $1 + (\lambda^c - \delta^c) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) \leq 0$ for some i, set $\gamma^c = \dfrac{\gamma^c}{2}$ and go to Step 2.

4. Set $\boldsymbol{\lambda}^{c+1} = \boldsymbol{\lambda}^c - \boldsymbol{\delta}^c$, $c = c+1$, and $\gamma^{c+1} = (c+1)^{-\frac{1}{2}}$ and go to Step 2. Step 2 will guarantee that $p_i > 0$ and the optimization is carried out in the right direction.

2.4.2 Algorithm for Optimizing Profile Empirical Likelihood Ratio Function

Let $\hat{\boldsymbol{\lambda}}(\beta)$ be the estimated value of $\boldsymbol{\lambda}$ for a given $\boldsymbol{\beta}$. We minimize the profile EL ratio function defined in (2.6) over $\boldsymbol{\beta}$. The Newton–Raphson algorithm is as follows:

1. Set $\boldsymbol{\beta} = \boldsymbol{\beta}^0$, $h = 0$, and $\epsilon = 1e^{-08}$.

2. Let $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}(\beta)$ and $\hat{\rho}(\beta)$ be the estimated values of $\boldsymbol{\lambda}$ and ρ.

3. Compute the new estimate of $\boldsymbol{\beta}$ via

$$\boldsymbol{\beta}^{(h+1)} = \boldsymbol{\beta}^{(h)} - \left\{ W_l^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}^h) \right\}^{-1} \left\{ W_l^{\boldsymbol{\beta}}(\boldsymbol{\beta}^h) \right\} \tag{2.23}$$

where $W_l(\boldsymbol{\beta})$ is the profile empirical log-likelihood ratio function defined in (2.6), with

$$W_l^{\boldsymbol{\beta}} = \frac{\partial W_l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \; W_l^{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{\partial^2 W_l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}.$$

Note that to compute $W_l^{\boldsymbol{\beta}}$ and $W_l^{\boldsymbol{\beta}\boldsymbol{\beta}}$, we need to estimate the Lagrange multiplier $\hat{\boldsymbol{\lambda}}(\beta)$ as in Section 2.4.1. In practice, ρ is unknown, and the correlations can be consistently estimated using the method of moments.

4. If $\min \left|\boldsymbol{\beta}^{(h+1)} - \boldsymbol{\beta}^{(h)}\right| < \epsilon$ stop the algorithm and report $\boldsymbol{\beta}^{(h+1)}$; otherwise set $h = h + 1$ and go to Step 3.

The simplified expressions for $W_l^{\boldsymbol{\beta}}$ and $W_l^{\boldsymbol{\beta}\boldsymbol{\beta}}$ are as follows. Let $R^{\boldsymbol{\beta}}$, $R^{\boldsymbol{\beta}\boldsymbol{\beta}}$, and $R^{\boldsymbol{\beta}\boldsymbol{\lambda}}$ be the first and second partial derivatives of (2.22) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$

$$R^{\boldsymbol{\beta}} = \sum_{i=1}^{k} \left[\frac{g_i'(\boldsymbol{\beta}, \hat{\rho}(\beta))\boldsymbol{\lambda}}{\{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}} \right],$$

$$R^{\boldsymbol{\beta}\boldsymbol{\beta}} = \sum_{i=1}^{k} \left\{ \left[\frac{g_i''(\boldsymbol{\beta}, \hat{\rho}(\beta))\boldsymbol{\lambda}^T}{\{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta})\}} \right] - \left[\frac{g_i'(\boldsymbol{\beta}, \hat{\rho}(\beta))\boldsymbol{\lambda}\boldsymbol{\lambda}^T [g_i'(\boldsymbol{\beta}, \hat{\rho}(\beta))]^T}{\{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}^2} \right] \right\},$$

and

$$R^{\beta\lambda} = \sum_{i=1}^{k} \left[\frac{\{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta})\} g_i'(\boldsymbol{\beta}, \hat{\rho}(\beta)) - g_i'(\boldsymbol{\beta}, \hat{\rho}(\beta))\boldsymbol{\lambda}[g_i(\boldsymbol{\beta})]^T}{\{1 + \boldsymbol{\lambda}^T g_i(\boldsymbol{\beta})\}^2} \right].$$

The first derivative of $W_l(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is

$$\begin{aligned} W_l^{\boldsymbol{\beta}} &= \sum_{i=1}^{k} \left[\frac{\left[\frac{\partial \boldsymbol{\lambda}(\beta)}{\partial \boldsymbol{\beta}}\right]^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) + g_i'(\boldsymbol{\beta}, \hat{\rho}(\beta))\boldsymbol{\lambda}(\beta)}{\{1 + \boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}} \right] \\ &= \left[\frac{\partial \boldsymbol{\lambda}(\beta)}{\partial \boldsymbol{\beta}}\right]^T R^{\boldsymbol{\lambda}} + R^{\boldsymbol{\beta}}. \end{aligned}$$

Note that for $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}(\beta)$, $R^{\boldsymbol{\lambda}} = \mathbf{0}$. Therefore,

$$W_l^{\boldsymbol{\beta}} = R^{\boldsymbol{\beta}}. \tag{2.24}$$

Similarly, the second derivative of $W_l(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is

$$W_l^{\boldsymbol{\beta\beta}} = \sum_{i=1}^{k} \left[\frac{\{1+\boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\} \left\{ \left[\frac{\partial^2 \boldsymbol{\lambda}(\beta)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right] g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) + 2 g_i'(\boldsymbol{\beta}) \left[\frac{\partial \boldsymbol{\lambda}(\beta)}{\partial \boldsymbol{\beta}}\right]^T + g_i''(\boldsymbol{\beta}, \hat{\rho}(\beta)) \boldsymbol{\lambda}(\beta) \right\}}{\{1+\boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}^2} \right]$$

$$- \sum_{i=1}^{k} \left[\frac{\left\{ \left[\frac{\partial \boldsymbol{\lambda}(\beta)}{\partial \boldsymbol{\beta}}\right]^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) + g_i'(\boldsymbol{\beta}, \hat{\rho}(\beta)) \boldsymbol{\lambda}(\beta) \right\} \left\{ \left[\frac{\partial \boldsymbol{\lambda}(\beta)}{\partial \boldsymbol{\beta}}\right]^T g_i(\boldsymbol{\beta}, \hat{\rho}(\beta)) + g_i'(\boldsymbol{\beta}, \hat{\rho}(\beta)) \boldsymbol{\lambda}(\beta) \right\}^T}{\{1+\boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))\}^2} \right]$$

$$= \left[\frac{\partial \boldsymbol{\lambda}(\beta)}{\partial \boldsymbol{\beta}}\right]^T R^{\boldsymbol{\lambda\lambda}} \left[\frac{\partial \boldsymbol{\lambda}(\beta)}{\partial \boldsymbol{\beta}}\right] + 2 \left[\frac{\partial \boldsymbol{\lambda}(\beta)}{\partial \boldsymbol{\beta}}\right]^T R^{\boldsymbol{\beta}\lambda} + R^{\boldsymbol{\beta\beta}}.$$

Following Owen [2001], a local quadratic approximation to R leads to

$$\left[\frac{\partial \lambda(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right] = \left(R^{\boldsymbol{\lambda\lambda}}\right)^{-1} R^{\boldsymbol{\beta\lambda}},$$

so

$$W_l^{\boldsymbol{\beta\beta}} = R^{\boldsymbol{\beta\beta}} - R^{\boldsymbol{\beta\lambda}} \left(R^{\boldsymbol{\lambda\lambda}}\right)^{-1} R^{\boldsymbol{\lambda\beta}}. \tag{2.25}$$

2.4.3 Construction of Confidence Interval

We use the bisection method to construct the lower and upper confidence limits based on the profile EL ratio for $\boldsymbol{\beta}$. Let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)^T$ be the estimated value of $\boldsymbol{\beta}$ from Section 2.4.2.

1. Compute a reasonable lower bound L_2 for the lower confidence limit. Set $L_1 = \hat{\beta}_1$, $L_2 = \hat{\beta}_1 - a \times \mathrm{SE}(\hat{\beta}_1)$, and $\epsilon = 1e^{-05}$, where $\mathrm{SE}(\hat{\beta}_1)$ is the standard error of $\hat{\beta}_1$ using any existing method. We can choose a such that $W_l(L_2, \hat{\beta}_2) < [\chi^2_{1,1-\alpha}]/2 < W_l(L_1, \hat{\beta}_2)$, where $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$th quantile from a χ^2 distribution with one degree of freedom.

2. Compute the profile empirical log-likelihood ratio values $W_1 = 2W_l(L_1, \hat{\beta}_2)$ and $W_2 = 2W_l(L_2, \hat{\beta}_2)$, where $W_2 < \chi^2_{1,1-\alpha} < W_1$.

3. Minimize the profile EL ratio function defined in (2.6) over β_2 for a given $L_{new} = (L_1 + L_2)/2$. Let $\hat{\beta}_{2\ new}$ be the new estimate of β_2 and $W_{new} = 2W_l(L_{new}, \hat{\beta}_{2\ new})$.

4. If $W_{new} < \chi^2_{1,1-\alpha}$ set $L_1 = L_{new}$; else set $L_2 = L_{new}$ and go to Step 3.

5. If $|W_1 - W_2| < \epsilon$ stop the algorithm and report $\beta_{1,L}$; otherwise go to Step 2.

We can use this approach to construct the upper confidence limit by setting $U_1 = \hat{\beta}_1$ and $U_2 = \hat{\beta}_1 + a \times \mathrm{SE}(\hat{\beta}_1)$. Figure 2.1 illustrates the bisection search for the lower bound of the CI.

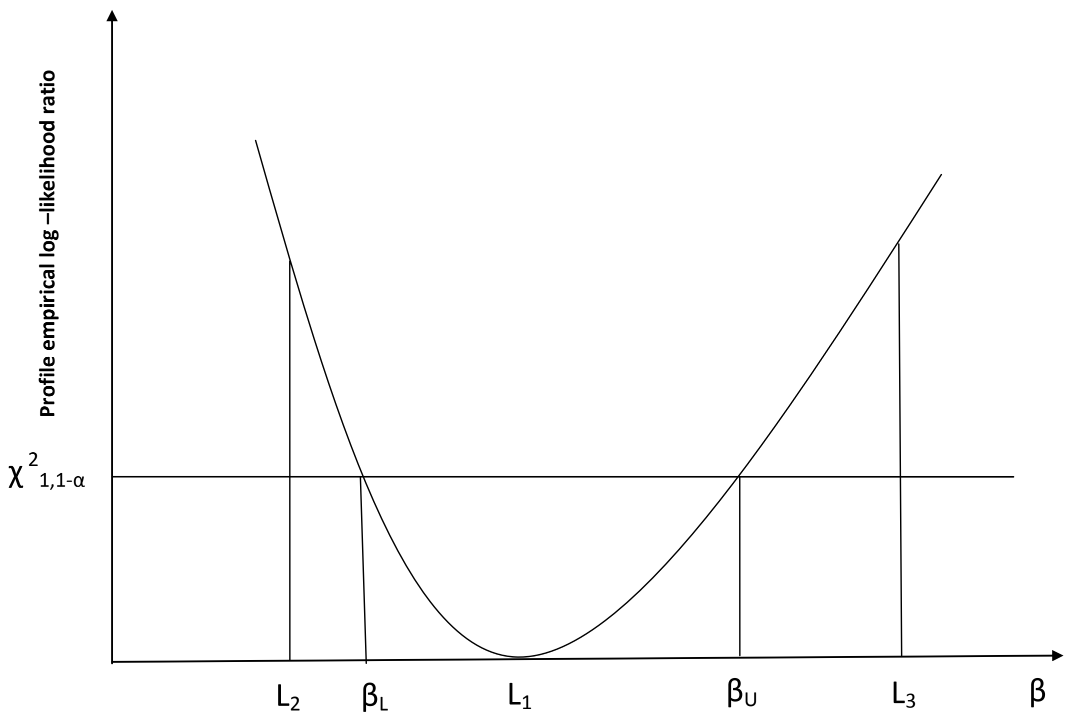

Figure 2.1: Construction of confidence interval using bisection.

Chapter 3

Performance Analysis

In this chapter, we conduct simulation studies to investigate the performance of our EL-based approach. We compute the coverage probabilities based on the ordinary EL and compare them with those of the GEE approach, which is based on a normal approximation. We use different working correlations for the comparison. We also compute the coverage probabilities based on the AEL and EEL since both approaches improve the coverage probabilities. We generate count and continuous responses with different correlation structures and compare the methods under different working correlation structures.

3.1 Correlation Models for Stationary Count Data

We consider the stationary correlation models for count data discussed by McKenzie [1988] and Sutradhar [2011]. The three models used to generate the data are

(i) Poisson Autoregressive Order 1 (AR(1)) Model

Let $y_{i1} \sim \text{Poi}(\tilde{\boldsymbol{\mu}}_i)$, where $\tilde{\boldsymbol{\mu}}_i = \exp(\tilde{\boldsymbol{x}}_i \boldsymbol{\beta})$. The repeated responses follow the AR lag 1 dynamic model given by

$$y_{it} = \rho * y_{i,t-1} + d_{it},\ t = 2, \ldots, m_i. \tag{3.1}$$

Given $y_{i,t-1}$, $\rho * y_{i,t-1}$ is the binomial thinning operation. That is,

$$\rho * y_{i,t-1} = \sum_{j=1}^{y_{i,t-1}} b_j(\rho) = z_{i,t-1},$$

where the $b_j(\rho)$ are independent and identically distributed Bernoulli(ρ) random variables. We assume that $d_{it} \sim \text{Poi}(\tilde{\boldsymbol{\mu}}_i(1-\rho))$ and it is independent of $z_{i,t-1}$. Let $\tilde{\boldsymbol{x}}_i = (\tilde{x}_{i1}, \ldots, \tilde{x}_{ip})$ be the time-independent covariate for the ith individual. It is given in Sutradhar [2011] that each response satisfying model (3.1) has a marginal Poisson distribution with

$$\mathrm{E}(y_{it}) = \mathrm{E}_{y_{i,t-1}}\mathrm{E}[y_{it}|y_{i,t-1}] = \tilde{\boldsymbol{\mu}}_i,$$

$$\mathrm{Var}(y_{it}) = \mathrm{E}_{y_{i,t-1}}\mathrm{Var}[y_{it}|y_{i,t-1}] + \mathrm{Var}_{y_{i,t-1}}\mathrm{E}[y_{it}|y_{i,t-1}] = \tilde{\boldsymbol{\mu}}_i.$$

That is,

$$\mathrm{E}(y_{it}) = \mathrm{Var}(y_{it}) = \tilde{\boldsymbol{\mu}}_i = \exp(\tilde{\boldsymbol{x}}_i\boldsymbol{\beta}).$$

Similarly, we can show that

$$\mathrm{E}(y_{it}y_{i,t-l}) = \tilde{\boldsymbol{\mu}}_i^2 + \tilde{\boldsymbol{\mu}}_i\rho^l,$$

yielding lag l correlation between y_{it} and $y_{i,t-l}$:

$$\mathrm{corr}(y_{it}, y_{i,t-l}) = \rho^l,\ l = 1, \ldots, m_i - 1.$$

This is similar to the lag l Gaussian AR(1) autocorrelation structure. However, under model (3.1) ρ must satisfy $0 \leq \rho \leq 1$, whereas for a Gaussian AR(1) structure ρ satisfies $-1 < \rho < 1$ (Sutradhar [2011]).

(ii) Poisson Moving Average Order 1 (MA(1)) Model

The repeated responses follow the MA lag 1 dynamic model given by

$$y_{it} = \rho * d_{i,t-1} + d_{it}, \; t = 2, \ldots, m_i, \tag{3.2}$$

where $\rho * d_{i,t-1} = \sum_{j=1}^{d_{i,t-1}} b_j(\rho)$ is a binomial thinning operation and $d_{it} \sim \text{Poi}\left[\dfrac{\tilde{\boldsymbol{\mu}}_i}{1+\rho}\right]$, $t = 0, \ldots, m_i$, with $\tilde{\boldsymbol{\mu}}_i = \exp(\tilde{\boldsymbol{x}}_i \boldsymbol{\beta})$. Here $t = 0$ is the initial time. Following a similar approach to that for the AR(1) process, we get

$$\text{E}(y_{it}) = \text{Var}(y_{it}) = \tilde{\boldsymbol{\mu}}_i$$

$$\text{corr}(y_{it}, y_{i,t-l}) = \begin{cases} \frac{\rho}{(1+\rho)}, & l = 1 \\ 0, & \text{otherwise.} \end{cases}$$

Again, the lag correlations are like those for the Gaussian MA(1) correlation structure, except that $0 \leq \rho \leq 1$ whereas in the Gaussian structure $-1 < \rho < 1$.

(iii) Poisson Equally Correlated Model

Let $y_{i0} \sim \text{Poi}(\tilde{\boldsymbol{\mu}}_i)$ and $d_{it} \sim \text{Poi}\left[\tilde{\boldsymbol{\mu}}_i(1-\rho)\right]$ for all $t = 1, \ldots, m_i$. The repeated responses follow the dynamic equicorrelation model given by

$$y_{it} = \rho * y_{i0} + d_{it}, \; t = 1, \ldots, m_i \tag{3.3}$$

yielding marginal properties similar to those for the AR(1) and MA(1) processes:

$$\mathrm{E}(y_{it}) = \mathrm{Var}(y_{it}) = \tilde{\boldsymbol{\mu}}_i.$$

The correlation between y_{it} and $y_{i,t-l}$ is

$$\mathrm{corr}(y_{it}, y_{i,t-l}) = \rho,$$

for all $l = 1, \ldots, m_i - 1$ with $0 \leq \rho \leq 1$. Under the Gaussian equally correlated model we have $-(1/m_i - 1) \leq \rho \leq 1$.

We simulated 1000 data sets from each of these three models. In each simulation we use the parameters $\boldsymbol{\beta} = (\beta_1, \beta_2)^T = (0.3, 0.2)^T$ and $\rho = 0.5$. We consider $k = 100$ subjects and $m = 4$ time points. For the ith subject, we generate the covariates $\tilde{\boldsymbol{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})$ from a normal distribution with mean 0 and standard deviation 1. For the analysis, we consider the working correlation to be either true correlation or a lag correlation, as discussed in Section 1.3.3.

Table 3.1 gives the average estimated values of the regression coefficients with the corresponding simulated standard errors in parentheses for the independent, AR(1), EQC, and MA(1) models. We also give the coverage probabilities for β_1 and β_2 for the 0.95 and 0.99 confidence levels with the average width of the CI in parentheses.

The results in Table 3.1 shows that the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ are close to the true values, width, and the coverage probabilities of the intervals based on the EL, EEL, and AEL are similar to those of the GEE. For instance, in the AR(1)/AR(1) case (true model/working correlation structure) the coverage probabilities of $\hat{\beta}_1$ based on the GEE, EL, EEL, and AEL are 0.947, 0.928, 0.937, and 0.937 respectively for the nominal level of 0.95. For $\hat{\beta}_2$, these probabilities are 0.954, 0.934, 0.940, and 0.942 for the same nominal level. Note that the intervals based on the EL have slight undercoverage compared with those based on GEE. The EEL and AEL give substantially better coverage probabilities. Moreover, the EEL and AEL are consistently more accurate than the EL. The results for lag correlations have similar patterns.

3.2 Misspecified Working Correlation Structure

In the simulation studies discussed in Section 3.1 we considered the correlation structure used to generate the data as the working correlation in the GEE-based modelling. However, in practice, we do not know the correlation structure of the data. As discussed in Section 1.3.3, if the working correlation is misspecified, we may lose the efficiency of the parameter estimates (Crowder [1995]; Sutradhar and Das [1999]).

We conducted a simulation study to assess the loss of efficiency. We generated repeated counts with the AR(1) correlation structure given in Section 3.1(i) with $\rho =$ 0.49 and 0.70 and $m = 5$ time points. We used three working correlation structures: EQC, MA(1), and lag correlation. Table 3.2 gives the results for the GEE, EL, EEL, and AEL. The table shows that the EL, EEL, and AEL are superior to the GEE when the correlation structure is misspecified. Note that, in this EL-based approach, we could construct CIs without estimating the variance of the parameter of interest. For example, in the AR(1)/EQC case the coverage probabilities of $\hat{\beta}_1$ based on the GEE, EL, EEL, and AEL are 0.917, 0.928, 0.934, and 0.935 respectively for the nominal 0.95 level. For $\hat{\beta}_2$, these probabilities are 0.916, 0.929, 0.937, and 0.937 for the same nominal level. In this situation, the GEE with stationary lag correlation performs better than the GEE with a misspecified working correlation, supporting the findings of Section 1.3.3. However, the EL, EEL, and AEL perform as well as the former method, despite being nonparametric methods based on a data-driven likelihood ratio function. We did not consider all possible cases, for instance a true EQC or MA(1) correlation model, since under different working correlation structures the correlation parameter $\hat{\boldsymbol{\alpha}}$ does not exist.

3.3 Over-dispersed Stationary Count Data

In this section, we consider the performance of our approach when the variance function is misspecified, in the context of stationary count data. We generate over-dispersed stationary count data y_{it} using $\tilde{\boldsymbol{\mu}}_i = u_i \exp(\tilde{\boldsymbol{x}}_i \boldsymbol{\beta})$ for the three models discussed in Section 3.1, where u_i is a random sample such that $\mathrm{E}(u_i) = 1$ and $\mathrm{Var}(u_i) = \omega$. Marginally, we have $\mathrm{E}(y_{it}) = \tilde{\boldsymbol{\mu}}_i$ and $\mathrm{Var}(y_{it}) = \tilde{\boldsymbol{\mu}}_i(1 + \tilde{\boldsymbol{\mu}}_i \omega)$. The distribution of u is chosen to be gamma with shape parameter ω and scale parameter $1/\omega$, where ω is the over-dispersion parameter. We choose over-dispersion parameter $\omega = 1/4$. However, the GEE, EL, EEL, and AEL CIs are constructed under the assumption that there is no over-dispersion. Table 3.3 gives the average estimated values of the regression coefficients, the corresponding simulated standard errors in parentheses, the coverage probabilities for β_1 and β_2 for the 0.95 and 0.99 confidence levels, and the average width of the CI in parentheses for the independent, AR(1), EQC, and MA(1) models. Table 3.3 shows that when there is over-dispersion, the EL, EEL, and AEL outperform the GEE. In the AR(1)/AR(1) case the coverage probabilities of $\hat{\beta}_1$ based on the GEE, EL, EEL, and AEL are 0.876, 0.916, 0.926, and 0.931 respectively for the nominal 0.95 level. For $\hat{\beta}_2$, these probabilities are 0.891, 0.920, 0.929, and 0.931 for the same nominal level. This indicates that the EL, EEL, and AEL are fairly robust to model misspecification. Note that the construction of the CI based on the EL, EEL, and AEL does not require the estimation of the scale parameter.

True model/ Working correlation	Parameter	Estimate	Coverage Probabilities							
			95% level				99% level			
			GEE	EL	EEL	AEL	GEE	EL	EEL	AEL
	β_1	0.2975	0.953	0.929	0.936	0.937	0.986	0.985	0.987	0.987
IND/IND		(0.046)	(0.184)	(0.175)	(0.178)	(0.179)	(0.242)	(0.232)	(0.240)	(0.238)
$\rho = 0.50$	β_2	0.1978	0.944	0.927	0.934	0.936	0.990	0.980	0.984	0.984
		(0.048)	(0.188)	(0.179)	(0.183)	(0.184)	(0.248)	(0.238)	(0.246)	(0.244)
	β_1	0.3006	0.947	0.928	0.937	0.937	0.985	0.980	0.981	0.981
AR(1)/AR(1)		(0.065)	(0.258)	(0.246)	(0.251)	(0.252)	(0.339)	(0.326)	(0.338)	(0.335)
$\rho = 0.50$	β_2	0.1978	0.954	0.934	0.940	0.942	0.990	0.980	0.984	0.983
		(0.067)	(0.264)	(0.251)	(0.256)	(0.257)	(0.347)	(0.332)	(0.343)	(0.340)
	β_1	0.3006	0.946	0.930	0.936	0.939	0.985	0.978	0.981	0.981
AR(1)/Lag		(0.066)	(0.256)	(0.246)	(0.250)	(0.252)	(0.337)	(0.325)	(0.337)	(0.334)
$\rho = 0.50$	β_2	0.1978	0.952	0.931	0.938	0.940	0.990	0.980	0.985	0.985
		(0.067)	(0.263)	(0.250)	(0.255)	(0.256)	(0.345)	(0.331)	(0.342)	(0.339)
	β_1	0.2984	0.955	0.942	0.946	0.948	0.988	0.986	0.988	0.987
EQC/EQC		(0.074)	(0.288)	(0.274)	(0.280)	(0.281)	(0.379)	(0.364)	(0.376)	(0.373)
$\rho = 0.50$	β_2	0.1936	0.950	0.939	0.941	0.941	0.990	0.986	0.986	0.986
		(0.074)	(0.295)	(0.282)	(0.288)	(0.289)	(0.387)	(0.372)	(0.386)	(0.383)
	β_1	0.2985	0.954	0.940	0.948	0.948	0.987	0.984	0.987	0.986
EQC/Lag		(0.074)	(0.288)	(0.274)	(0.280)	(0.281)	(0.379)	(0.363)	(0.376)	(0.373)
$\rho = 0.50$	β_2	0.1933	0.952	0.937	0.940	0.943	0.990	0.986	0.986	0.986
		(0.075)	(0.294)	(0.281)	(0.287)	(0.288)	(0.387)	(0.372)	(0.385)	(0.382)
	β_1	0.2989	0.943	0.926	0.929	0.931	0.989	0.982	0.985	0.983
MA(1)/MA(1)		(0.058)	(0.222)	(0.211)	(0.215)	(0.216)	(0.291)	(0.280)	(0.289)	(0.287)
$\rho = 0.50$	β_2	0.2022	0.952	0.932	0.935	0.936	0.994	0.984	0.991	0.989
		(0.056)	(0.227)	(0.216)	(0.220)	(0.221)	(0.298)	(0.285)	(0.296)	(0.293)
	β_1	0.2990	0.944	0.928	0.929	0.929	0.986	0.980	0.985	0.985
MA(1)/Lag		(0.058)	(0.220)	(0.210)	(0.215)	(0.216)	(0.290)	(0.279)	(0.289)	(0.287)
$\rho = 0.50$	β_2	0.2024	0.946	0.931	0.934	0.937	0.992	0.983	0.989	0.988
		(0.056)	(0.225)	(0.215)	(0.220)	(0.221)	(0.296)	(0.285)	(0.295)	(0.293)

Table 3.1: Coverage probabilities of regression estimates for count data with stationary covariates for the independent, AR(1), EQC, and MA(1) models.

True model/ Working correlation	Parameter	Estimate	Coverage Probabilities 95% level GEE	EL	EEL	AEL	99% level GEE	EL	EEL	AEL
	β_1	0.3001	0.952	0.938	0.944	0.946	0.987	0.983	0.984	0.984
AR(1)/AR(1)		(0.070)	(0.279)	(0.265)	(0.270)	(0.272)	(0.367)	(0.351)	(0.364)	(0.361)
$\rho = 0.70$	β_2	0.2009	0.950	0.928	0.938	0.941	0.988	0.986	0.989	0.988
		(0.073)	(0.286)	(0.0.270)	(0.275)	(0.276)	(0.375)	(0.356)	(0.369)	(0.366)
	β_1	0.2997	0.911	0.932	0.935	0.936	0.973	0.975	0.978	0.978
AR(1)/EQC		(0.073)	(0.247)	(0.273)	(0.278)	(0.280)	(0.325)	(0.361)	(0.375)	(0.371)
$\rho = 0.70$	β_2	0.1956	0.902	0.934	0.936	0.936	0.963	0.977	0.982	0.980
		(0.076)	(0.252)	(0.278)	(0.284)	(0.286)	(0.331)	(0.368)	(0.381)	(0.378)
	β_1	0.3002	0.952	0.932	0.940	0.940	0.986	0.982	0.982	0.982
AR(1)/Lag		(0.070)	(0.278)	(0.264)	(0.268)	(0.270)	(0.366)	(0.349)	(0.361)	(0.359)
$\rho = 0.70$	β_2	0.2007	0.950	0.926	0.938	0.940	0.988	0.985	0.988	0.987
		(0.073)	(0.284)	(0.268)	(0.273)	(0.275)	(0.373)	(0.354)	(0.367)	(0.364)
	β_1	0.2989	0.938	0.918	0.922	0.923	0.988	0.977	0.982	0.978
AR(1)/AR(1)		(0.062)	(0.237)	(0.225)	(0.229)	(0.231)	(0.311)	(0.298)	(0.309)	(0.306)
$\rho = 0.49$	β_2	0.1956	0.940	0.920	0.928	0.928	0.993	0.985	0.988	0.987
		(0.062)	(0.243)	(0.231)	(0.235)	(0.236)	(0.0.319)	(0.305)	(0.316)	(0.313)
	β_1	0.2992	0.899	0.931	0.936	0.936	0.968	0.978	0.985	0.984
AR(1)/EQC		(0.061)	(0.207)	(0.231)	(0.236)	(0.237)	(0.272)	(0.306)	(0.317)	(0.314)
$\rho = 0.49$	β_2	0.1987	0.908	0.945	0.948	0.950	0.980	0.987	0.992	0.990
		(0.062)	(0.212)	(0.236)	(0.241)	(0.242)	(0.278)	(0.313)	(0.324)	(0.321)
	β_1	0.2991	0.897	0.931	0.934	0.936	0.968	0.979	0.985	0.984
AR(1)/MA(1)		(0.061)	(0.205)	(0.228)	(0.232)	(0.233)	(0.270)	(0.302)	(0.313)	(0.310)
$\rho = 0.49$	β_2	0.1985	0.905	0.936	0.944	0.944	0.981	0.990	0.993	0.991
		(0.061)	(0.210)	(0.233)	(0.238)	(0.239)	(0.276)	(0.309)	(0.319)	(0.317)
	β_1	0.3006	0.946	0.930	0.936	0.939	0.985	0.978	0.981	0.981
AR(1)/Lag		(0.066)	(0.256)	(0.246)	(0.250)	(0.252)	(0.337)	(0.325)	(0.337)	(0.334)
$\rho = 0.49$	β_2	0.1978	0.952	0.931	0.938	0.940	0.990	0.980	0.985	0.985
		(0.067)	(0.263)	(0.250)	(0.255)	(0.256)	(0.345)	(0.331)	(0.342)	(0.339)

Table 3.2: Coverage probabilities of regression estimates for count data with stationary covariates when the working correlation is misspecified for an AR(1) model.

True model/ Working correlation	Parameter	Estimate	Coverage Probabilities							
			95% level				99% level			
			GEE	EL	EEL	AEL	GEE	EL	EEL	AEL
	β_1	0.2978	0.808	0.937	0.941	0.941	0.902	0.985	0.989	0.988
IND/IND		(0.074)	(0.191)	(0.273)	(0.278)	(0.280)	(0.252)	(0.361)	(0.374)	(0.370)
$\rho = 0.50$	β_2	0.1980	0.813	0.934	0.937	0.937	0.919	0.979	0.985	0.983
		(0.071)	(0.188)	(0.265)	(0.270)	(0.271)	(0.247)	(0.349)	(0.361)	(0.358)
	β_1	0.2974	0.876	0.916	0.926	0.931	0.971	0.978	0.982	0.982
AR(1)/AR(1)		(0.080)	(0.276)	(0.314)	(0.317)	(0.319)	(0.363)	(0.410)	(0.425)	(0.421)
$\rho = 0.50$	β_2	0.2016	0.898	0.924	0.929	0.932	0.973	0.983	0.987	0.986
		(0.086)	(0.282)	(0.316)	(0.323)	(0.324)	(0.370)	(0.417)	(0.432)	(0.428)
	β_1	0.2916	0.899	0.928	0.930	0.931	0.973	0.978	0.983	0.983
AR(1)/Lag		(0.085)	(0.282)	(0.310)	(0.316)	(0.318)	(0.371)	(0.408)	(0.423)	(0.419)
$\rho = 0.50$	β_2	0.1978	0.952	0.931	0.938	0.940	0.990	0.980	0.985	0.985
		(0.088)	(0.288)	(0.315)	(0.321)	(0.323)	(0.378)	(0.415)	(0.430)	(0.426)
	β_1	0.2960	0.903	0.917	0.922	0.926	0.967	0.975	0.980	0.977
EQC/EQC		(0.091)	(0.305)	(0.336)	(0.342)	(0.344)	(0.401)	(0.443)	(0.459)	(0.455)
$\rho = 0.50$	β_2	0.2000	0.892	0.905	0.912	0.913	0.963	0.970	0.979	0.977
		(0.094)	(0.311)	(0.339)	(0.346)	(0.347)	(0.409)	(0.447)	(0.463)	(0.458)
	β_1	0.2957	0.901	0.917	0.923	0.924	0.967	0.974	0.978	0.976
EQC/Lag		(0.091)	(0.305)	(0.335)	(0.341)	(0.343)	(0.400)	(0.441)	(0.457)	(0.453)
$\rho = 0.50$	β_2	0.2001	0.894	0.905	0.912	0.912	0.961	0.974	0.978	0.977
		(0.095)	(0.311)	(0.338)	(0.344)	(0.346)	(0.409)	(0.445)	(0.461)	(0.457)
	β_1	0.2980	0.859	0.928	0.933	0.935	0.949	0.980	0.983	0.983
MA(1)/MA(1)		(0.080)	(0.234)	(0.289)	(0.295)	(0.297)	(0.307)	(0.382)	(0.395)	(0.392)
$\rho = 0.50$	β_2	0.1987	0.861	0.915	0.917	0.917	0.938	0.975	0.982	0.981
		(0.080)	(0.239)	(0.289)	(0.295)	(0.296)	(0.314)	(0.381)	(0.395)	(0.391)
	β_1	0.2975	0.897	0.922	0.935	0.936	0.967	0.979	0.982	0.981
MA(1)/Lag		(0.080)	(0.257)	(0.288)	(0.294)	(0.296)	(0.337)	(0.380)	(0.394)	(0.391)
$\rho = 0.50$	β_2	0.1986	0.891	0.915	0.920	0.920	0.962	0.973	0.984	0.981
		(0.080)	(0.262)	(0.288)	(0.294)	(0.295)	(0.344)	(0.380)	(0.393)	(0.390)

Table 3.3: Coverage probabilities of regression estimates for over-dispersion of count data with stationary covariates for the independent, AR(1), EQC, and MA(1) models.

3.4 Correlation Models for Nonstationary Count Data

In this section, we assess the performance of our approach under AR(1) and EQC nonstationary correlation models for count data. When the covariates are time-dependent, it may not be reasonable to use a stationary lag correlation structure (Sutradhar [2011]).

(i) Nonstationary AR(1) Model

Let $y_{i1} \sim \text{Poi}(\mu_{it})$, where $\mu_{it} = \exp(\boldsymbol{x}_{it}\boldsymbol{\beta})$. The repeated responses follow the AR lag 1 dynamic model given by

$$y_{it} = \rho * y_{i,t-1} + d_{it},\ t = 2, \ldots, m, \tag{3.4}$$

and studied by Sutradhar [2011]. In the model (3.4), for a given $y_{i,t-1}$, $\rho * y_{i,t-1}$ is the binomial thinning operation discussed in Section 3.1(i). Furthermore, we assume that $d_{it} \sim \text{Poi}(\mu_{it} - \rho\mu_{i,t-1})$. Let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})$ be the time-dependent covariate for the ith individual.

Each response satisfying model (3.4) has a marginal Poisson distribution with

$$\text{E}(y_{it}) = \text{Var}(y_{it}) = \mu_{it} = \exp(\boldsymbol{x}_{it}\boldsymbol{\beta}).$$

For $u < t$, where $t = 2, \ldots, m$, we can derive the nonstationary correlations between y_{it} and y_{iu} as

$$\text{corr}(y_{iu}, y_{it}) = \rho^{t-u}\sqrt{\frac{\mu_{iu}}{\mu_{it}}},$$

with ρ satisfying the range restriction

$$0 < \rho < \min\left[1, \frac{\mu_{it}}{\mu_{i,t-1}}\right], \; t = 2, \ldots, m.$$

Under the nonstationary AR(1) model, the moment estimate of the lag correlation ρ is given by

$$\hat{\rho} = \frac{\sum_{i=1}^{k}\sum_{t=2}^{m}\tilde{y}_{it}\tilde{y}_{i,t-1}}{\sum_{i=1}^{K}\sum_{t=1}^{m}\tilde{y}_{it}^{2}} \frac{km}{\sum_{i=1}^{K}\sum_{t=2}^{m}[\mu_{i,t-1}/\mu_{it}]^{1/2}}$$

where the $\tilde{y}_{it}$ are the standardized residuals. Note that the formula for ρ given above was obtained by equating the lag 1 sample autocorrelation to its population counterpart (see Sutradhar [2011]).

(ii) Nonstationary Equally Correlated Model

Let $y_{i1} \sim \text{Poi}(\mu_{i1})$ and $d_{it} \sim \text{Poi}\left[\mu_{it} - \rho\mu_{i1}\right]$ for all $t = 1, \ldots, m$. The repeated responses follow the dynamic model given by

$$y_{it} = \rho * y_{i1} + d_{it}, \; t = 2, \ldots, m$$

yielding marginal properties similar to those for the AR(1) process:

$$\text{E}(y_{it}) = \text{Var}(y_{it}) = \mu_{it}.$$

Similarly, we can obtain the nonstationary correlation between y_{iu} and y_{it} as

$$\text{corr}(y_{iu}, y_{it}) = \frac{\rho\mu_{i1}}{\sqrt{\mu_{iu}\mu_{it}}},$$

with ρ satisfying the range restriction

$$0 < \rho < \min\left[1, \frac{\mu_{it}}{\mu_{i1}}\right], t = 2, \ldots, m.$$

The moment estimating equation for the lag correlation ρ parameter for the exchangeable model is similar to that of the AR(1) model. The moment formula for the lag correlation ρ under the exchangeable model is given by

$$\hat{\rho} = \frac{\sum_{i=1}^{k}\sum_{l=1}^{m-1}\sum_{t=1}^{m-l}\tilde{y}_{it}\tilde{y}_{i,t+l}}{\sum_{i=1}^{k}\sum_{l=1}^{m-1}\sum_{t=1}^{m-l}\tilde{y}_{it}^2}\frac{km}{\sum_{i=1}^{k}\sum_{l=1}^{m-1}\sum_{t=1}^{m-l}\mu_{i1}/[\mu_{it}\mu_{i,t+l}]^{1/2}},$$

where the $\tilde{y}_{it}$ are the standardized residuals (see Sutradhar, 2011).

We simulated 1000 data sets from the above two models with the same parameter set as before given in Section 3.1. Table 3.4 gives the mean estimated values of the regression coefficients together with the simulated standard errors in parentheses. We also report the simulated coverage probabilities for β_1 and β_2 for the 0.95 and 0.99 confidence levels and the average width of the CI in parentheses for the independent, AR(1), and EQC models.

Table 3.4 shows that the mean estimated values $\hat{\beta}_1$ and $\hat{\beta}_2$ are close to the true values, and the coverage probabilities of the intervals based on the EL, EEL, and AEL are similar to those of the GEE. For instance, in the EQC/EQC case the coverage probabilities of $\hat{\beta}_1$ based on the GEE, EL, EEL, and AEL are 0.949, 0.937, 0.941, and 0.942 respectively for the nominal 0.95 level. For $\hat{\beta}_2$, these probabilities are 0.952, 0.928, 0.932, and 0.933 for the same nominal level. However, the EEL and AEL give substantially better coverage probabilities than the EL.

True model/ Working correlation	Parameter	Estimate	Coverage Probabilities							
			95% level				99% level			
			GEE	EL	EEL	AEL	GEE	EL	EEL	AEL
	β_1	0.2977	0.958	0.938	0.944	0.944	0.992	0.987	0.991	0.991
IND/IND		(0.046)	(0.184)	(0.175)	(0.178)	(0.179)	(0.242)	(0.232)	(0.240)	(0.238)
$\rho = 0.50$	β_2	0.1983	0.944	0.926	0.933	0.936	0.990	0.981	0.983	0.983
		(0.049)	(0.188)	(0.179)	(0.183)	(0.184)	(0.247)	(0.238)	(0.246)	(0.244)
	β_1	0.2958	0.959	0.941	0.943	0.946	0.993	0.987	0.988	0.988
AR(1)/AR(1)		(0.063)	(0.253)	(0.240)	(0.245)	(0.246)	(0.332)	(0.318)	(0.329)	(0.327)
$\rho = 0.50$	β_2	0.1991	0.944	0.925	0.929	0.930	0.989	0.978	0.982	0.980
		(0.068)	(0.259)	(0.247)	(0.252)	(0.253)	(0.340)	(0.327)	(0.338)	(0.335)
	β_1	0.2957	0.957	0.941	0.944	0.945	0.993	0.987	0.988	0.988
AR(1)/Lag		(0.063)	(0.252)	(0.240)	(0.245)	(0.246)	(0.331)	(0.318)	(0.329)	(0.327)
$\rho = 0.50$	β_2	0.1989	0.943	0.925	0.928	0.930	0.989	0.978	0.983	0.982
		(0.068)	(0.258)	(0.247)	(0.252)	(0.253)	(0.339)	(0.327)	(0.338)	(0.335)
	β_1	0.3003	0.949	0.937	0.941	0.942	0.991	0.981	0.986	0.985
EQC/EQC		(0.073)	(0.288)	(0.272)	(0.277)	(0.279)	(0.378)	(0.360)	(0.373)	(0.369)
$\rho = 0.50$	β_2	0.1931	0.952	0.928	0.932	0.933	0.985	0.979	0.981	0.981
		(0.076)	(0.294)	(0.279)	(0.285)	(0.286)	(0.386)	(0.369)	(0.382)	(0.378)
	β_1	0.2958	0.912	0.923	0.927	0.928	0.978	0.977	0.983	0.982
EQC/Lag		(0.075)	(0.256)	(0.274)	(0.280)	(0.281)	(0.337)	(0.363)	(0.376)	(0.373)
$\rho = 0.50$	β_2	0.1952	0.919	0.935	0.940	0.940	0.977	0.982	0.985	0.985
		(0.075)	(0.262)	(0.281)	(0.287)	(0.288)	(0.344)	(0.371)	(0.384)	(0.381)

Table 3.4: Coverage probabilities of regression estimates for count data with nonstationary covariates for the independent, AR(1), and EQC models.

3.5 Over-dispersed Nonstationary Count Data

We now consider the performance of our method when the parametric model is misspecified, in the context of nonstationary count data. We generate over-dispersed nonstationary count data y_{it} using $\mu_{it} = u_{it}\exp(\boldsymbol{x}_{it}\boldsymbol{\beta})$ for the parameter set in Section 3.4 with u_{it} a random variable generated from the gamma distribution, as in Section 3.3. Table 3.5 gives the results for the independent, AR(1), and EQC models. It shows that in the presence of over-dispersion the EL, EEL, and AEL outperform the GEE. For instance, in the AR(1)/AR(1) case the coverage probabilities of $\hat{\beta}_1$ based on the GEE, EL, EEL, and AEL are 0.889, 0.921, 0.926, and 0.928 respectively for the nominal 0.95 level. For $\hat{\beta}_2$, these probabilities are 0.890, 0.923, 0.927, and 0.927 for the same nominal level. This again shows that the EL, EEL, and AEL are fairly robust to misspecification.

3.6 Correlation Models for Continuous Data

In this section, we investigate the performance of our EL approach on a class of stationary and nonstationary correlation models for longitudinal continuous data. The random errors $(\epsilon_1, ..., \epsilon_4)^T$ are generated from the multivariate normal distribution with marginal mean 0, marginal variance 1, and an auto-correlation coefficient $\rho = 0.5$. In this performance analysis, we consider three correlation models: exchangeable, AR(1), and MA(1).

(i) AR(1) Structure

For $t = 1, \ldots, m_i$, for the ith individual

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \epsilon_{it}, \tag{3.5}$$

and we assume that

$$\epsilon_{it} = \rho\epsilon_{it} + a_{it},$$

with $|\rho| < 1$ and $a_{it} \sim N(0, 1)$. The mean and variance of ϵ_{it} are given by

$$\mathrm{E}[\epsilon_{it}] = 0, \text{ and } \mathrm{Var}[\epsilon_{it}] = \frac{1}{(1-\rho^2)}, \; t = 1, \ldots, m_i,$$

respectively. Similarly, we can obtain the stationary covariance between ϵ_{it} and ϵ_{iu} as

$$\mathrm{cov}(\epsilon_{it}, \epsilon_{iu}) = \frac{\rho^{t-u}}{1-\rho^2}, \; t, u = 1, \ldots, m_i.$$

The repeated responses $y_{i1}, \ldots, y_{im_i}$ follow the AR(1) model with mean and variance, respectively,

$$\mathrm{E}[y_{it}] = \boldsymbol{x}_{it}\boldsymbol{\beta} \text{ and } \mathrm{Var}[y_{it}] = \frac{1}{(1-\rho^2)}.$$

The lag $|t-u|$ correlations $\rho_{|t-u|}$ are

$$\rho_{|t-u|} = \mathrm{corr}[y_{it}, y_{iu}] = \rho^{t-u}, \; u \neq t, \; u, t = 1, \ldots, m_i,$$

where ρ is referred to as the correlation parameter.

(ii) MA(1) Structure

The ϵ_{it} in (3.5) follow the model

$$\epsilon_{it} = \rho a_{i,t-1} + a_{it}$$

where ρ is a suitable scale parameter that does not necessarily satisfy $|\rho| < 1$, and $a_{it} \sim N(0,1)$. The mean and variance of y_{it} are

$$\mathrm{E}[y_{it}] = \boldsymbol{x}_{it}\boldsymbol{\beta}, \quad \mathrm{Var}[y_{it}] = (1+\rho^2),$$

respectively, and the lag $|t-u|$ correlations of the repeated responses are

$$\rho_{|t-u|} = \mathrm{corr}[y_{it}, y_{iu}] = \begin{cases} \frac{\rho}{(1+\rho)} & \text{if } |t-u| = 1 \\ 0, & \text{otherwise.} \end{cases}$$

(iii) Equicorrelation (EQC) Structure

The ϵ_{it} in (3.5) follow the model

$$\epsilon_{it} = \rho a_{i0} + a_{it},$$

where a_{i0} is an error value at the initial time, and ρ is a suitable correlation parameter. We assume that

$$a_{it} \sim N(0,1) \text{ and } a_{i0} \sim N(0,1),$$

and a_{it} and a_{i0} are independent for all t. The mean and variance of y_{it} are

$$\mathrm{E}[y_{it}] = \boldsymbol{x}_{it}\boldsymbol{\beta}, \quad \mathrm{Var}[y_{it}] = (1+\rho^2),$$

and the lag $|t-u|$ correlations are

$$\rho_{|t-u|} = \mathrm{corr}[y_{it}, y_{iu}] = \frac{\rho^2}{(1+\rho^2)}, \; u \neq t, \; u,t = 1,\ldots,m_i.$$

We simulated 1000 data sets from the above models under stationary and nonstationary covariates, using the parameters $\boldsymbol{\beta} = (\beta_1, \beta_2)^T = (0.4, 0.5)^T$, $\rho = 0.5$, and $m = 4$. For the ith subject, we generate the covariates $\tilde{\boldsymbol{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})$ from a normal distribution with mean 0 and standard deviation 1. Table 3.6 gives the mean estimated values of the regression coefficients, the corresponding simulated standard errors in parentheses, the simulated coverage probabilities for β_1 and β_2 for the 0.95 and 0.99 confidence levels, and the average width of the CI in parentheses for the independent, AR(1), EQC, and MA(1) models with stationary covariates. Table 3.7 gives the results for nonstationary covariates.

The coverage probabilities of the intervals based on the EL, EEL, and AEL are similar to those of the GEE. For instance, in the MA(1)/MA(1) case in Table 3.6 the coverage probabilities of $\hat{\beta}_1$ based on the GEE, EL, EEL, and AEL are 0.955, 0.945, 0.955, and 0.954 respectively for the nominal 0.95 level. For $\hat{\beta}_2$, these probabilities are 0.958, 0.944, 0.948, and 0.951 for the same nominal level. Note that the intervals based on the EL have slight undercoverage compared with those for the GEE. Also, the EEL and AEL are consistently more accurate than the EL. The lag-correlation-based coverage probabilities have similar patterns.

True model/ Working correlation	Parameter	Estimate	Coverage Probabilities							
			95% level				99% level			
			GEE	EL	EEL	AEL	GEE	EL	EEL	AEL
	β_1	0.3000	0.793	0.927	0.930	0.931	0.904	0.973	0.977	0.976
IND/IND		(0.072)	(0.184)	(0.264)	(0.269)	(0.270)	(0.242)	(0.348)	(0.360)	(0.357)
$\rho = 0.50$	β_2	0.2021	0.803	0.926	0.929	0.933	0.919	0.983	0.987	0.986
		(0.072)	(0.187)	(0.262)	(0.267)	(0.268)	(0.246)	(0.345)	(0.357)	(0.354)
	β_1	0.2931	0.889	0.921	0.926	0.928	0.963	0.979	0.982	0.982
AR(1)/AR(1)		(0.084)	(0.268)	(0.301)	(0.307)	(0.308)	(0.351)	(0.397)	(0.411)	(0.408)
$\rho = 0.50$	β_2	0.1972	0.890	0.923	0.927	0.927	0.959	0.978	0.980	0.980
		(0.085)	(0.273)	(0.305)	(0.311)	(0.312)	(0.359)	(0.403)	(0.417)	(0.413)
	β_1	0.2914	0.890	0.917	0.927	0.930	0.963	0.979	0.982	0.982
AR(1)/Lag		(0.082)	(0.267)	(0.303)	(0.306)	(0.308)	(0.351)	(0.397)	(0.411)	(0.408)
$\rho = 0.50$	β_2	0.2015	0.882	0.927	0.929	0.930	0.959	0.977	0.980	0.980
		(0.085)	(0.272)	(0.307)	(0.311)	(0.313)	(0.358)	(0.402)	(0.417)	(0.413)
	β_1	0.2978	0.905	0.925	0.937	0.937	0.968	0.977	0.982	0.981
EQC/EQC		(0.090)	(0.301)	(0.328)	(0.335)	(0.337)	(0.395)	(0.433)	(0.448)	(0.444)
$\rho = 0.50$	β_2	0.1985	0.900	0.926	0.932	0.932	0.977	0.985	0.986	0.986
		(0.091)	(0.307)	(0.332)	(0.338)	(0.340)	(0.404)	(0.437)	(0.453)	(0.448)
	β_1	0.2986	0.856	0.918	0.923	0.923	0.941	0.973	0.975	0.976
EQC/Lag		(0.092)	(0.266)	(0.333)	(0.339)	(0.341)	(0.350)	(0.439)	(0.455)	(0.451)
$\rho = 0.50$	β_2	0.1961	0.848	0.923	0.928	0.928	0.941	0.973	0.975	0.976
		(0.095)	(0.272)	(0.338)	(0.0.345)	(0.347)	(0.358)	(0.446)	(0.461)	(0.457)

Table 3.5: Coverage probabilities of regression estimates for over-dispersion count data with nonstationary covariates for the independent, AR(1), and EQC models.

True model/ Working correlation	Parameter	Estimate	Coverage Probabilities							
			95% level				99% level			
			GEE	EL	EEL	AEL	GEE	EL	EEL	AEL
	β_1	0.4010	0.955	0.947	0.950	0.951	0.989	0.988	0.988	0.988
IND/IND		(0.050)	(0.197)	(0.191)	(0.195)	(0.196)	(0.259)	(0.254)	(0.262)	(0.260)
$\rho = 0.50$	β_2	0.5015	0.953	0.936	0.942	0.942	0.993	0.985	0.990	0.989
		(0.051)	(0.197)	(0.191)	(0.195)	(0.196)	(0.259)	(0.253)	(0.262)	(0.260)
	β_1	0.4019	0.949	0.944	0.950	0.952	0.990	0.989	0.990	0.990
AR(1)/AR(1)		(0.071)	(0.278)	(0.270)	(0.276)	(0.277)	(0.365)	(0.358)	(0.371)	(0.368)
$\rho = 0.50$	β_2	0.5027	0.959	0.948	0.951	0.952	0.994	0.993	0.993	0.993
		(0.068)	(0.278)	(0.270)	(0.275)	(0.277)	(0.365)	(0.358)	(0.371)	(0.367)
	β_1	0.4018	0.946	0.944	0.949	0.950	0.990	0.988	0.989	0.989
AR(1)/Lag		(0.071)	(0.277)	(0.270)	(0.275)	(0.277)	(0.364)	(0.357)	(0.370)	(0.367)
$\rho = 0.50$	β_2	0.5026	0.959	0.948	0.950	0.953	0.994	0.992	0.994	0.994
		(0.077)	(0.310)	(0.302)	(0.308)	(0.310)	(0.408)	(0.401)	(0.415)	(0.411)
	β_1	0.4016	0.949	0.942	0.949	0.951	0.993	0.986	0.988	0.987
EQC/EQC		(0.079)	(0.310)	(0.302)	(0.308)	(0.310)	(0.408)	(0.400)	(0.415)	(0.411)
$\rho = 0.50$	β_2	0.5023	0.963	0.947	0.953	0.954	0.995	0.989	0.994	0.992
		(0.091)	(0.307)	(0.332)	(0.338)	(0.340)	(0.404)	(0.437)	(0.453)	(0.448)
	β_1	0.4015	0.948	0.945	0.948	0.949	0.992	0.986	0.988	0.987
EQC/Lag		(0.079)	(0.310)	(0.301)	(0.307)	(0.309)	(0.407)	(0.399)	(0.414)	(0.410)
$\rho = 0.50$	β_2	0.5024	0.961	0.947	0.953	0.954	0.996	0.989	0.994	0.994
		(0.077)	(0.310)	(0.301)	(0.308)	(0.309)	(0.407)	(0.400)	(0.414)	(0.410)
	β_1	0.3981	0.955	0.945	0.955	0.954	0.992	0.986	0.991	0.991
MA(1)/MA(1)		(0.065)	(0.254)	(0.247)	(0.252)	(0.253)	(0.334)	(0.327)	(0.338)	(0.335)
$\rho = 0.50$	β_2	0.4974	0.958	0.944	0.948	0.951	0.995	0.989	0.992	0.990
		(0.062)	(0.254)	(0.246)	(0.251)	(0.252)	(0.334)	(0.327)	(0.338)	(0.335)
	β_1	0.3981	0.956	0.942	0.951	0.953	0.994	0.986	0.989	0.987
MA(1)/Lag		(0.065)	(0.252)	(0.246)	(0.251)	(0.253)	(0.332)	(0.326)	(0.338)	(0.335)
$\rho = 0.50$	β_2	0.4974	0.957	0.943	0.948	0.950	0.994	0.989	0.991	0.991
		(0.062)	(0.252)	(0.246)	(0.251)	(0.252)	(0.332)	(0.326)	(0.338)	(0.335)

Table 3.6: Coverage probabilities of regression estimates for continuous data with stationary covariates for the independent, AR(1), EQC, and MA(1) models.

True model/ Working correlation	Parameter	Estimate	Coverage Probabilities							
			95% level				99% level			
			GEE	EL	EEL	AEL	GEE	EL	EEL	AEL
	β_1	0.4003	0.947	0.940	0.945	0.945	0.986	0.986	0.988	0.988
IND/IND		(0.051)	(0.196)	(0.194)	(0.198)	(0.199)	(0.258)	(0.257)	(0.266)	(0.263)
$\rho = 0.50$	β_2	0.4996	0.948	0.946	0.949	0.949	0.985	0.986	0.988	0.988
		(0.051)	(0.196)	(0.195)	(0.199)	(0.199)	(0.258)	(0.258)	(0.267)	(0.264)
	β_1	0.3998	0.939	0.926	0.936	0.937	0.989	0.987	0.989	0.989
AR(1)/AR(1)		(0.043)	(0.160)	(0.158)	(0.161)	(0.162)	(0.211)	(0.209)	(0.216)	(0.214)
$\rho = 0.50$	β_2	0.4980	0.959	0.949	0.954	0.955	0.992	0.992	0.993	0.993
		(0.040)	(0.160)	(0.159)	(0.162)	(0.163)	(0.211)	(0.210)	(0.218)	(0.215)
	β_1	0.3998	0.935	0.923	0.932	0.934	0.988	0.986	0.989	0.988
AR(1)/Lag		(0.043)	(0.159)	(0.157)	(0.160)	(0.161)	(0.209)	(0.208)	(0.216)	(0.214)
$\rho = 0.50$	β_2	0.4979	0.958	0.946	0.955	0.955	0.991	0.993	0.994	0.993
		(0.040)	(0.159)	(0.158)	(0.161)	(0.162)	(0.209)	(0.209)	(0.217)	(0.215)
	β_1	0.4018	0.954	0.947	0.955	0.956	0.992	0.988	0.989	0.989
EQC/EQC		(0.040)	(0.155)	(0.153)	(0.156)	(0.157)	(0.204)	(0.203)	(0.210)	(0.208)
$\rho = 0.50$	β_2	0.5001	0.953	0.947	0.952	0.954	0.989	0.989	0.991	0.990
		(0.040)	(0.155)	(0.153)	(0.156)	(0.157)	(0.204)	(0.203)	(0.210)	(0.208)
	β_1	0.4019	0.947	0.945	0.950	0.951	0.988	0.985	0.990	0.990
EQC/Lag		(0.040)	(0.154)	(0.153)	(0.156)	(0.156)	(0.202)	(0.202)	(0.209)	(0.207)
$\rho = 0.50$	β_2	0.5011	0.946	0.938	0.944	0.945	0.988	0.987	0.989	0.989
		(0.040)	(0.154)	(0.153)	(0.156)	(0.157)	(0.202)	(0.202)	(0.209)	(0.207)
	β_1	0.4000	0.942	0.936	0.946	0.947	0.992	0.993	0.995	0.993
MA(1)/MA(1)		(0.043)	(0.165)	(0.166)	(0.169)	(0.170)	(0.217)	(0.220)	(0.227)	(0.225)
$\rho = 0.50$	β_2	0.5027	0.938	0.931	0.939	0.939	0.987	0.979	0.986	0.983
		(0.044)	(0.165)	(0.165)	(0.168)	(0.169)	(0.217)	(0.218)	(0.226)	(0.223)
	β_1	0.4004	0.926	0.926	0.931	0.932	0.982	0.976	0.983	0.981
MA(1)/Lag		(0.038)	(0.138)	(0.136)	(0.139)	(0.140)	(0.181)	(0.180)	(0.187)	(0.185)
$\rho = 0.50$	β_2	0.5001	0.943	0.950	0.957	0.957	0.992	0.988	0.990	0.990
		(0.035)	(0.138)	(0.137)	(0.140)	(0.141)	(0.181)	(0.182)	(0.188)	(0.187)

Table 3.7: Coverage probabilities of regression estimates for continuous data with nonstationary covariates for the independent, AR(1), EQC, and MA(1) models.

3.7 Correlation Models for Misspecified Continuous Data

In this section, we compare the performances of the methods of Chapter 2 when the correlation model for continuous data is misspecified. The stationary and nonstationary correlation models for longitudinal continuous data are generated from (3.5) for the parameter set in Section 3.6, and the correlated random errors $(\epsilon_1, ..., \epsilon_4)^T$ are generated from the $\chi^2(1) - 1$ distribution instead of the normal distribution for the three correlation models:

- ◇ AR(1): $\epsilon_{it} = \rho\epsilon_{i,t-1} + a_{it},\ t = 1, \ldots, 4$
- ◇ EQC: $\epsilon_{it} = \rho a_{i,0} + a_{it},\ t = 1, \ldots, 4$
- ◇ MA(1): $\epsilon_{it} = \rho a_{i,t-1} + a_{it},\ t = 1, \ldots, 4$

However, the confidence regions for the GEE are constructed under the normality assumption.

Table 3.8 gives the mean estimated values of the coefficients and the corresponding simulated standard errors in parentheses. It also includes the coverage probability for β_1 and β_2 for the 0.95 and 0.99 confidence levels and the average width of the CI in parentheses for samples of sizes $k = 50$ and $k = 100$ for the independent, AR(1), EQC, and MA(1) models with stationary covariates. Table 3.9 gives the results for nonstationary covariates.

When the model is misspecified the EL, EEL, and AEL outperform the GEE. For example, in the AR(1)/Lag case in Table 3.8 the coverage probabilities of $\hat{\beta}_1$ based on the GEE, EL, EEL, and AEL are 0.790, 0.918, 0.931, and 0.932 respectively for the nominal 0.95 level. For $\hat{\beta}_2$, these probabilities are 0.801, 0.924, 0.937, and 0.937 for the same nominal level. Note that we do not need to estimate a scale parameter in the construction of the CI in the EL setup, and also in the EL we did not model the over-dispersion. Table 3.9 shows that when the covariates are time-dependent the GEE has substantial undercoverage compared with the results for time-independent covariates, as discussed in Section 1.5.

3.8 Summary

Our performance analysis shows that our EL, EEL, and AEL methods have consistent performance when the model assumptions are correct for longitudinal count and continuous responses. However, when the model is misspecified our EL, EEL, and AEL methods outperform the GEE. This shows that the EL, EEL, and AEL are robust to model misspecification since the CIs are constructed without estimation of the scale parameter.

True model/ Working correlation	Parameter	Estimate	Coverage Probabilities							
			95% level				99% level			
			GEE	EL	EEL	AEL	GEE	EL	EEL	AEL
	β_1	0.3971	0.836	0.906	0.909	0.913	0.930	0.963	0.975	0.971
IND/IND		(0.103)	(0.281)	(0.357)	(0.367)	(0.369)	(0.369)	(0.479)	(0.512)	(0.501)
$\rho = 0.50$	β_2	0.4988	0.838	0.904	0.926	0.926	0.928	0.966	0.978	0.976
		(0.101)	(0.281)	(0.350)	(0.370)	(0.372)	(0.369)	(0.473)	(0.513)	(0.502)
	β_1	0.3954	0.805	0.923	0.934	0.934	0.910	0.971	0.973	0.977
AR(1)/AR(1)		(0.157)	(0.397)	(0.587)	(0.610)	(0.613)	(0.522)	(0.783)	(0.827)	(0.814)
$\rho = 0.50$	β_2	0.4986	0.807	0.935	0.942	0.944	0.913	0.975	0.983	0.980
		(0.154)	(0.397)	(0.593)	(0.610)	(0.613)	(0.522)	(0.784)	(0.827)	(0.815)
	β_1	0.3949	0.790	0.918	0.931	0.932	0.902	0.971	0.980	0.977
AR(1)/Lag		(0.156)	(0.391)	(0.581)	(0.601)	(0.604)	(0.513)	(0.772)	(0.816)	(0.803)
$\rho = 0.50$	β_2	0.4983	0.801	0.924	0.937	0.937	0.911	0.977	0.986	0.983
		(0.152)	(0.391)	(0.581)	(0.602)	(0.604)	(0.513)	(0.714)	(0.818)	(0.805)
	β_1	0.3970	0.885	0.927	0.940	0.943	0.953	0.981	0.989	0.987
EQC/EQC		(0.143)	(0.444)	(0.570)	(0.590)	(0.593)	(0.584)	(0.759)	(0.803)	(0.790)
$\rho = 0.50$	β_2	0.4935	0.893	0.943	0.955	0.955	0.960	0.984	0.991	0.990
		(0.140)	(0.444)	(0.565)	(0.585)	(0.588)	(0.584)	(0.753)	(0.798)	(0.785)
	β_1	0.3973	0.811	0.928	0.935	0.941	0.902	0.981	0.989	0.987
EQC/Lag		(0.143)	(0.371)	(0.568)	(0.588)	(0.591)	(0.487)	(0.756)	(0.800)	(0.787)
$\rho = 0.50$	β_2	0.4940	0.812	0.941	0.953	0.956	0.928	0.986	0.990	0.989
		(0.140)	(0.371)	(0.562)	(0.583)	(0.586)	(0.487)	(0.750)	(0.795)	(0.782)
	β_1	0.3996	0.802	0.935	0.941	0.942	0.897	0.976	0.984	0.981
MA(1)/MA(1)		(0.144)	(0.363)	(0.563)	(0.584)	(0.587)	(0.477)	(0.750)	(0.796)	(0.782)
$\rho = 0.50$	β_2	0.4974	0.808	0.949	0.955	0.955	0.917	0.985	0.991	0.989
		(0.139)	(0.363)	(0.560)	(0.580)	(0.583)	(0.477)	(0.747)	(0.792)	(0.779)
	β_1	0.3997	0.802	0.932	0.943	0.943	0.906	0.977	0.983	0.982
MA(1)/Lag		(0.144)	(0.366)	(0.561)	(0.582)	(0.585)	(0.481)	(0.748)	(0.794)	(0.780)
$\rho = 0.50$	β_2	0.4980	0.808	0.948	0.960	0.960	0.921	0.985	0.991	0.989
		(0.138)	(0.366)	(0.558)	(0.578)	(0.581)	(0.481)	(0.745)	(0.791)	(0.777)

Table 3.8: Coverage probabilities of regression estimates for misspecified data with stationary covariates for the independent, AR(1), EQC, and MA(1) models (k=50).

True model/ Working correlation	Parameter	Estimate	Coverage Probabilities							
			95% level				99% level			
			GEE	EL	EEL	AEL	GEE	EL	EEL	AEL
	β_1	0.3992	0.843	0.930	0.943	0.943	0.945	0.983	0.990	0.989
IND/IND		(0.098)	(0.279)	(0.377)	(0.392)	(0.393)	(0.366)	(0.491)	(0.544)	(0.530)
$\rho = 0.50$	β_2	0.4936	0.846	0.926	0.934	0.935	0.936	0.979	0.987	0.985
		(0.098)	(0.279)	(0.377)	(0.392)	(0.394)	(0.366)	(0.489)	(0.544)	(0.531)
	β_1	0.3949	0.776	0.925	0.930	0.932	0.890	0.972	0.983	0.979
AR(1)/AR(1)		(0.094)	(0.221)	(0.354)	(0.368)	(0.369)	(0.291)	(0.477)	(0.512)	(0.500)
$\rho = 0.50$	β_2	0.4982	0.777	0.933	0.942	0.945	0.887	0.984	0.988	0.986
		(0.092)	(0.221)	(0.353)	(0.368)	(0.369)	(0.291)	(0.476)	(0.511)	(0.499)
	β_1	0.3990	0.750	0.919	0.931	0.933	0.870	0.981	0.988	0.985
AR(1)/Lag		(0.094)	(0.214)	(0.355)	(0.369)	(0.371)	(0.281)	(0.477)	(0.513)	(0.500)
$\rho = 0.50$	β_2	0.4952	0.762	0.917	0.927	0.928	0.864	0.972	0.982	0.979
		(0.096)	(0.214)	(0.355)	(0.369)	(0.371)	(0.281)	(0.478)	(0.514)	(0.501)
	β_1	0.3980	0.772	0.928	0.938	0.938	0.896	0.979	0.986	0.985
EQC/EQC		(0.107)	(0.256)	(0.408)	(0.424)	(0.426)	(0.336)	(0.548)	(0.588)	(0.574)
$\rho = 0.50$	β_2	0.5018	0.778	0.923	0.934	0.936	0.890	0.980	0.988	0.984
		(0.106)	(0.256)	(0.404)	(0.421)	(0.422)	(0.336)	(0.543)	(0.583)	(0.569)
	β_1	0.3981	0.751	0.921	0.932	0.934	0.888	0.979	0.983	0.983
EQC/Lag		(0.109)	(0.253)	(0.405)	(0.421)	(0.423)	(0.332)	(0.544)	(0.585)	(0.570)
$\rho = 0.50$	β_2	0.5026	0.761	0.918	0.937	0.938	0.880	0.978	0.986	0.983
		(0.107)	(0.253)	(0.402)	(0.419)	(0.420)	(0.332)	(0.541)	(0.581)	(0.566)
	β_1	0.3961	0.709	0.920	0.930	0.931	0.827	0.980	0.989	0.987
MA(1)/MA(1)		(0.100)	(0.213)	(0.369)	(0.385)	(0.386)	(0.274)	(0.497)	(0.534)	(0.521)
$\rho = 0.50$	β_2	0.5009	0.728	0.926	0.941	0.941	0.834	0.983	0.991	0.989
		(0.100)	(0.213)	(0.369)	(0.385)	(0.386)	(0.274)	(0.497)	(0.534)	(0.522)
	β_1	0.3972	0.749	0.913	0.929	0.930	0.866	0.981	0.989	0.987
MA(1)/Lag		(0.096)	(0.222)	(0.356)	(0.370)	(0.372)	(0.292)	(0.478)	(0.514)	(0.502)
$\rho = 0.50$	β_2	0.5002	0.762	0.929	0.938	0.939	0.872	0.981	0.991	0.989
		(0.096)	(0.222)	(0.356)	(0.370)	(0.372)	(0.292)	(0.478)	(0.515)	(0.502)

Table 3.9: Coverage probabilities of regression estimates for misspecified data with nonstationary covariates for the independent, AR(1), EQC, and MA(1) models (k=100).

Chapter 4

Penalized Empirical Likelihood

The penalized EL (PEL) method is a powerful inference tool with promising applications in many areas of statistics. It is a useful method for simultaneous estimation and variable selection. In this chapter, we briefly discuss PEL-based variable selection and then discuss variable selection for longitudinal data based on the penalized adjusted EL and its asymptotic properties.

4.1 Penalized Empirical Likelihood for Linear Models

Owen [1991] first considered the EL for linear models. EL confidence regions for regression coefficients in linear models were studied by Chen [1994]. Most of the existing works on the EL focus on fixed-dimensional regression parameters; see Owen [2001] and Chen and Van Keilegom [2009] for more details. Variyath [2006] and Variyath et al. [2010] introduced EL-based variable selection as an alternative to AIC and BIC. They showed that these methods perform better in situations where the parametric distributional assumptions are misspecified. In this case, a complete enumeration of all the submodels is necessary to identify the best submodel, which is difficult when there are many covariates.

The importance of high-dimensional statistical inference using the EL has recently been recognized by Variyath [2006], Hjort, McKeague and Van Keilegom [2009] and Chen, Peng and Qin [2009]. Variyath [2006] introduced the PEL for linear models, and reported some computational issues with over-penalizations. To overcome this problem, Nadarajah [2011] proposed the penalized adjusted EL (PAEL) variable selection for GLMs, a modification of the method proposed by Fan and Li [2001] to avoid the technical problem of the nonexistence of the Lagrange multiplier when the sample size is small. Tang and Leng [2010] studied variable selection using a penalty in the EL framework; it is limited to mean vector estimation and linear regression models.

Consider a linear model of the form

$$y_i = \boldsymbol{X}_i\boldsymbol{\beta} + \epsilon_i, \;\; i = 1, 2, \ldots, n,$$

where $\boldsymbol{X}_i \in \mathcal{R}^p$ is a vector of covariates and $\boldsymbol{\beta} \in \mathcal{R}^p$ a vector of parameters. Following Owen [1991] and Qin and Lawless [1994], we extend EL inference for linear models based on a set of estimating functions $g(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta})$ with dimension p. This leads to the profile empirical log-likelihood ratio function

$$W(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log(1 + \hat{\boldsymbol{\lambda}}^T(\beta) g(y_i, \boldsymbol{X}_i, \boldsymbol{\beta})),$$

and the penalized profile empirical log-likelihood ratio estimator of $\boldsymbol{\beta}$ is the minimizer of

$$\begin{aligned}\mathbf{L}(\boldsymbol{\beta}) &= W(\boldsymbol{\beta}) + n\sum_{j=1}^{p} p_\delta(|\beta_j|)\\ &= \sum_{i=1}^{n}\left[\log(1+\hat{\boldsymbol{\lambda}}^T(\beta)g(y_i,\boldsymbol{X}_i,\boldsymbol{\beta}))\right] + n\sum_{j=1}^{p} p_\delta(|\beta_j|) \qquad (4.1)\end{aligned}$$

with respect to $\boldsymbol{\beta}$, where $p_\delta(*)$ is the penalty function.

When the dimension p grows, variable selection using GEEs becomes more interesting. In this situation, the EL is challenging, both theoretically and computationally. Tang and Leng [2012] introduced the penalized EL for high-dimensional GEEs using quadratic inference functions, which is applied to incorporate correlation into the model. In this approach, it is assumed that the inverse of the working correlation can be approximated by a linear combination of several basis matrices, which are not directly involved in the estimation of the correlation parameter. In high-dimensional longitudinal data analysis, it is reasonable to expect that only a subset of the covariates are relevant. To identify the subset of influential covariates, we propose using the PEL.

4.2 Penalized Empirical Likelihood for Longitudinal Data

In this section, we introduce penalized subject-wise EL variable selection, using the method of moments to exploit the within-subject correlation. Following Owen [1991] and Qin and Lawless [1994], we extend EL inference to longitudinal data based on the set of estimating functions discussed in Section 2.2.

We now consider the subject-wise penalized profile empirical log-likelihood ratio estimator of $\boldsymbol{\beta}$ as the minimizer of

$$\begin{aligned}\mathbf{L}_l(\boldsymbol{\beta}) &= W_l(\boldsymbol{\beta}) + k\sum_{j=1}^{p} p_\delta(|\beta_j|) \\ &= \sum_{i=1}^{k}\left[\log(1+\hat{\boldsymbol{\lambda}}^T(\beta)g(\boldsymbol{y}_i;\boldsymbol{\beta},\hat{\rho}(\beta)))\right] + k\sum_{j=1}^{p} p_\delta(|\beta_j|) \qquad (4.2)\end{aligned}$$

with respect to $\boldsymbol{\beta}$, where $g(\boldsymbol{y}_i;\boldsymbol{\beta},\hat{\rho}(\beta))$ and $W_l(\boldsymbol{\beta})$ are given in Equations (1.5) and (2.6) respectively.

In practice, ρ is unknown and ignoring this within-subject correlation may result in a loss of efficiency in general problems. The within-subject correlations can be consistently estimated using the method of moments given in (1.7) and (1.8). We use the continuous differential SCAD penalty function with two unknown tuning parameters (δ, a), proposed by Fan and Li [2001] and defined in (1.10). Finding the profile EL function is a key step in applications of the subject-wise penalized EL likelihood ratio; it involves constrained minimization. However, in some situations, a solution may not exist. To avoid this problem, we introduce PAEL, obtained by adding a pseudo-observation to the data-set as discussed in Section 2.2.1.

4.3 Penalized Adjusted Empirical Likelihood for Longitudinal Data

The adjusted profile empirical log-likelihood ratio function is well defined after the addition of a pseudo-value $g_{k+1}(\boldsymbol{y};\boldsymbol{\beta},\hat{\rho}(\beta)) = -b_k\overline{g}_k(\boldsymbol{\beta},\hat{\rho}(\beta))$, as discussed in Section 2.2.1, where $\overline{g}_k(\boldsymbol{\beta},\hat{\rho}(\beta)) = \frac{1}{k}\sum_{i=1}^{k} g_i(\boldsymbol{\beta},\hat{\rho}(\beta))$ and $b_k = \log(k)/2$ is a positive constant. We now define the penalized subject-wise adjusted profile empirical log-likelihood ratio estimator of $\boldsymbol{\beta}$ as the minimizer of

$$\begin{aligned}\mathbf{L}_l^*(\boldsymbol{\beta}) &= W_l^*(\boldsymbol{\beta}) + (k+1)\sum_{j=1}^{p} p_\delta(|\beta_j|) \\ &= \sum_{i=1}^{k+1}\left[\log(1+\hat{\boldsymbol{\lambda}}^T(\beta)g_i(\boldsymbol{y};\boldsymbol{\beta},\hat{\rho}(\beta)))\right] + (k+1)\sum_{j=1}^{p} p_\delta(|\beta_j|) \qquad (4.3)\end{aligned}$$

with respect to $\boldsymbol{\beta}$, where $p_\delta(*)$ is the SCAD penalty function. This adjustment is useful because even for some undesirable values of $\boldsymbol{\beta}$ and the tuning parameters, the proposed algorithm guarantees a solution. By following Fan and Li [2001], we state the results showing that the PAEL estimates have oracle properties.

4.4 Oracle Properties

In this section, we present the oracle properties of the PAEL estimates.

For the ith subject, let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{it} \ldots, y_{im_i})^T$ be the response vector associated with $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im_i})^T$ the $m_i \times p$ matrix of covariates. We assume that all the subjects are independent and the repeated measurements y_{it} taken on each subject are correlated. Suppose $(\boldsymbol{y}_i, \boldsymbol{X}_i)$, $i = 1, 2, \ldots, k$, is a set of independent and identically distributed random vectors. We denote the true value of $\boldsymbol{\beta}_0$ by $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}_{10}^T, \boldsymbol{\beta}_{20}^T\right)^T$. The covariate matrix is partitioned into $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2})$ accordingly. Without loss of generality, we assume that $\boldsymbol{\beta}_{20} = \mathbf{0}$ and that the elements of $\boldsymbol{\beta}_{10}$ are all nonzero. We denote the dimension of $\boldsymbol{\beta}_{10}$ by s, where s may be fixed or growing with k.

Theorem 4.4.1 *Let $\hat{\boldsymbol{\beta}} = \left(\hat{\boldsymbol{\beta}}_1^T, \hat{\boldsymbol{\beta}}_2^T\right)^T$ be the minimizer of (4.3). Under regularity conditions A1-A5, as $k \to \infty$,*

(i) with probability tending to 1, $\hat{\boldsymbol{\beta}}_2 = \boldsymbol{0}$; and

(ii) $\sqrt{k}\left\{\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10} + (-\boldsymbol{\Delta})^{-1}p'_\delta(|\boldsymbol{\beta}_{10}|)\right\} \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Delta})$, *where* $\hat{\boldsymbol{\beta}}$ *is the PAEL estimate of* $\boldsymbol{\beta}$ *and*

$$\boldsymbol{\Delta} = \left[E_{\boldsymbol{\beta}_{10}}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T \left\{E_{\boldsymbol{\beta}_{10}}\left\{g(\boldsymbol{\beta}, \hat{\rho}(\beta))g^T(\boldsymbol{\beta}, \hat{\rho}(\beta))\right\}^{-1}\right\} E_{\boldsymbol{\beta}_{10}}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}\right]^{-1}.$$

Proof of Theorem 4.4.1(i):

The first result is corresponds to the Lemma 1 in Fan and Li [2001] which is about the estimator must have the sparsity property $\hat{\boldsymbol{\beta}}_2 = \mathbf{0}$, with probability tending to one. It is sufficient to show that for any $\boldsymbol{\beta}_1$ satisfying $\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{10} = O_p(k^{-1/2})$ and for any constant C some small $\epsilon_k = Ck^{-1/2}$ and $j = s+1, \ldots, p$,

$$\begin{aligned} \frac{\partial \mathbf{L}_l^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &< 0, \qquad 0 < \beta_j < \epsilon_k, \\ \frac{\partial \mathbf{L}_l^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &> 0, \qquad -\epsilon_k < \beta_j < 0. \end{aligned} \tag{4.4}$$

Assume that

$$\lim_{k\to\infty} \lim_{\boldsymbol{\beta}\to 0+} \left\{\frac{p'_{\delta_k}(\boldsymbol{\beta})}{\delta_k}\right\} > 0 \tag{4.5}$$

and by the condition on $p_{\delta_k}(|\boldsymbol{\beta}|)$, we are going to show that, uniformly in both $i = 1, 2, \ldots, k$ and $\boldsymbol{\beta}$,

$$\left|\frac{\partial W_l^*(\boldsymbol{\beta})}{\partial \beta_j}\right| = O_p(k^{2/3}).$$

Now consider

$$W_l^*(\beta_j) = \sum_{i=1}^{k+1} \log\left\{1 + \boldsymbol{\lambda}(\beta_j) g_i(\beta_j, \hat{\rho}(\beta))\right\}$$

where λ and g_i as functions of a specific component of β for simplicity. Note that

$$\frac{\partial W_l^*(\beta_j)}{\partial \beta_j} = \sum_{i=1}^{k+1} \frac{1}{1+\lambda(\beta_j)g_i(\beta_j, \hat{\rho}(\beta))} \left\{ \frac{\partial g_i(\beta_j, \hat{\rho}(\beta))}{\partial \beta_j} \right\} \lambda(\beta_j).$$

Since $\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{10} = O_p(k^{-1/2})$, then by Theorem 2.3.1, we have

$$\max_{1\le i\le k} \|g_i(\beta_j, \hat{\rho}(\beta))\| = O_p(k^{1/3}) \quad \text{and} \quad \|\lambda(\beta_j)\| = O_p(k^{-1/3}).$$

Hence,

$$\lambda(\beta_j)g_i(\beta_j, \hat{\rho}(\beta)) = o_p(1)$$

uniformly in both $i = 1, 2, \ldots, k$ and $\boldsymbol{\beta}$. Thus we have

$$\left| \frac{\partial W_l^*(\beta_j)}{\partial \beta_j} \right| \le \|\lambda(\beta_j)\| \sum_{i=1}^{k} \left\| \frac{\partial g_i(\beta_j, \hat{\rho}(\beta))}{\partial \beta_j}(1-b_k) \right\| [1 + o_p(1)],$$

where $g_{k+1}(\boldsymbol{\beta}, \hat{\rho}(\beta)) = -b_k \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))$ and b_k is a positive constant.

$$\left| \frac{\partial W_l^*(\beta_j)}{\partial \beta_j} \right| \le \|\lambda(\beta_j)\| \sum_{i=1}^{k} \left\| \frac{\partial g_i(\beta_j, \hat{\rho}(\beta))}{\partial \beta_j} \right\| [1 + o_p(1)]$$

$$= O_p(k^{-1/3})O_p(k)[1 + o_p(1)]$$

$$= O_p(k^{2/3}).$$

Thus, for every β_j, $j = s+1, \ldots, p$, it is true that

$$\frac{\partial \mathbf{L}^*(\beta)}{\partial \beta_j} = (k+1)\delta_k \left\{ -\delta_k^{-1} p'_{\delta_k}(|\beta_j|)\text{sgn}(\beta_j) + \delta_k^{-1} O_p(k^{-1/2}) \right\}.$$

By the assumption (4.5), $\sqrt{k}\delta_k \to \infty$, and $\delta_k \to 0$, the sign of the derivative is completely determined by that of β_j, hence the sparsity is proved. This completes the proof.

Second, we can prove the asymptotic normality of the PAEL estimate.

Proof of Theorem 4.4.1(ii):

Theorem 4.4.1(ii) shows that all nonzero PAEL estimator $\boldsymbol{\beta}_1$ is consistent and asymptotically normally distributed. Due to the sparsity property given in Theorem 4.4.1(i), it is seen that the PAEL estimator with proper tuning parameter δ_k minimizes $\mathbf{L}_l^*\{(\boldsymbol{\beta}_1, \mathbf{0})^T\}$ with respect to $\boldsymbol{\beta}_1$. Hence,

$$\frac{\partial \mathbf{L}_l^*(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})}{\partial \boldsymbol{\lambda}} = \mathbf{L}_{l,1,k+1}^*(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}) = 0, \frac{\partial \mathbf{L}_l^*(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})}{\partial \boldsymbol{\beta}} = \mathbf{L}_{l,2,k+1}^*(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}}) = 0$$

where

$$\mathbf{L}_{l,1,k+1}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{k} \sum_{i=1}^{k+1} \frac{g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}{1 + \boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}$$

and

$$\mathbf{L}_{l,2,k+1}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{k} \sum_{i=1}^{k+1} \frac{1}{1 + \boldsymbol{\lambda}^T(\beta) g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))} \left(\frac{\partial g_i(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}} \right)^T \boldsymbol{\lambda} + (k+1) p_\delta'(|\boldsymbol{\beta}_1|) \mathrm{sgn}(\boldsymbol{\beta}_1).$$

By using (2.19), $\mathbf{L}_{l,1,k+1}^*(\boldsymbol{\beta}, \boldsymbol{\lambda})$ can be written as

$$\mathbf{L}_{l,1,k+1}^*(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))(1 - b_k/k) - \boldsymbol{\lambda}^T(\beta) \boldsymbol{V}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))(1 + b_k^2/k) + o_p(k^{-1/3}).$$

The partial derivatives of $\mathbf{L}_{l,1,k+1}^*(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$ and $\mathbf{L}_{l,2,k+1}^*(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$ at $(\boldsymbol{\beta} = \boldsymbol{\beta}_0, \boldsymbol{\lambda} = \mathbf{0})$ are

$$\begin{aligned}
\frac{\mathbf{L}_{l,1,k+1}^*(\boldsymbol{\beta}_0, \mathbf{0})}{\partial \boldsymbol{\beta}} &= \frac{1}{k} \sum_{i=1}^{k} \frac{\partial g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}} \to -E\left\{ \frac{\partial g(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}} \right\} \qquad , \\
\frac{\mathbf{L}_{l,1,k+1}^*(\boldsymbol{\beta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}} &= \frac{1}{k} \sum_{i=1}^{k} g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) g_i^T(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) \to E\left\{ g(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) g^T(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) \right\}, \\
\frac{\mathbf{L}_{l,2,k+1}^*(\boldsymbol{\beta}_0, \mathbf{0})}{\partial \boldsymbol{\beta}} &= p_{\delta_k}''(|\boldsymbol{\beta}_0|), \\
\frac{\mathbf{L}_{l,2,k+1}^*(\boldsymbol{\beta}_0, \mathbf{0})}{\partial \boldsymbol{\lambda}} &= \frac{1}{k} \sum_{i=1}^{k+1} \left\{ \frac{\partial g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}} \right\}^T
\end{aligned}$$

$$= \frac{1}{k}\sum_{i=1}^{k}\left\{\frac{\partial g_i(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T \left[1 - \frac{b_k}{k}\right] \to E\left\{\frac{\partial g(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T,$$

where $g_{k+1}(\boldsymbol{\beta}, \hat{\rho}(\beta)) = -b_k \overline{g}_k(\boldsymbol{\beta}, \hat{\rho}(\beta))$ and b_k is a positive constant. Now by following Fan and Li [2001] in Theorem 2, expanding $\mathbf{L}^*_{l,1,k+1}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$ and $\mathbf{L}^*_{l,2,k+1}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\lambda}})$ at $(\boldsymbol{\beta}_0, \mathbf{0})$ under the conditions A2-A5, which leads to

$$\begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10} \end{bmatrix} = S_k^{-1} \begin{bmatrix} -\mathbf{L}^*_{l,1,k+1}(\boldsymbol{\beta}_0, \mathbf{0}) + o_p(\delta_k) \\ -\mathbf{L}^*_{l,2,k+1}(\boldsymbol{\beta}_0, \mathbf{0}) + o_p(\delta_k) \end{bmatrix}$$

where

$$S_k = \begin{bmatrix} -E\left\{g(\boldsymbol{\beta}_0) g^T(\boldsymbol{\beta}_0, \hat{\rho}(\beta))\right\} & E\left\{\frac{\partial g(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\} \\ E\left\{\frac{\partial g(\boldsymbol{\beta}_0, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T & p''_{\delta_k}(|\boldsymbol{\beta}_0|) \end{bmatrix}.$$

Since $\mathbf{L}^*_{l,1,k+1}(\boldsymbol{\beta}_0, \mathbf{0}) = \overline{g}_k(\boldsymbol{\beta}_0, \hat{\rho}(\beta)) = O_p(k^{-1/2})$, we know that $\delta_k = O_p(k^{-1/2})$, so, when $p''_{\delta_k}(|\boldsymbol{\beta}|) \to \mathbf{0}$ as $k \to \infty$, the limiting distribution of $\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}$ will be asymptotically normal, i.e.,

$$\sqrt{k}\left\{\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10} + S_k^{22}\mathbf{L}^*_{l,2,k+1}(\boldsymbol{\beta}_0, \mathbf{0})\right\} \xrightarrow{D} N(\mathbf{0}, \Delta),$$

where

$$\boldsymbol{\Delta} = \left[E_{\boldsymbol{\beta}_{10}}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}^T \left\{E_{\boldsymbol{\beta}_{10}}\left\{g(\boldsymbol{\beta}, \hat{\rho}(\beta)) g^T(\boldsymbol{\beta}, \hat{\rho}(\beta))\right\}^{-1}\right\} E_{\boldsymbol{\beta}_{10}}\left\{\frac{\partial g(\boldsymbol{\beta}, \hat{\rho}(\beta))}{\partial \boldsymbol{\beta}}\right\}\right]^{-1},$$

and $S_k^{22} = -\Delta^{-1}$ is the $(2,2)^{th}$ element of S_k^{-1} assuming $p''_{\delta_k}(|\boldsymbol{\beta}|) = 0$. This completes the proof.

In practice, we arrive the simplest model using variable selection methods and make inference from the resulting model. Due to randomness in the resulting model, these inferences always not guarantee the classical statistical theory provides for tests and confidence intervals when the model has been chosen as priori. To overcome this, researchers provided different approaches (see Berk, Brown, Buja, Zhang and Zhao [2013]), which focus on the simultaneous inference procedures. We will explore this post-selection inference in PEL as future work.

4.5 Algorithm

To implement our method, we need an efficient algorithm. We minimize the PEL given in (4.2) or the PAEL given in (4.3) with respect to $\boldsymbol{\beta}$ using a modified Newton–Raphson algorithm. At each Newton–Raphson iteration, we compute the correlation parameter $\hat{\rho}(\beta)$ and the Lagrange multiplier $\hat{\boldsymbol{\lambda}}(\beta)$ for an updated value of $\boldsymbol{\beta}$. The parameter $\hat{\rho}(\beta)$ can be estimated by the method of moments. The algorithm given in Section 4.5.1 can easily be extended to the PAEL, by the addition of a pseudo-value $g_{k+1}(\boldsymbol{\beta}) = -b_k \overline{g}_k(\boldsymbol{\beta})$, where $b_k = \log(k)/2$ is a positive constant.

4.5.1 Algorithm for Optimizing Penalized Empirical Likelihood

Let $\hat{\boldsymbol{\lambda}}(\beta)$ be the estimated values of $\boldsymbol{\lambda}$ for a given value of $\boldsymbol{\beta}$ (see Section 2.4.1) and $\hat{\rho}(\beta)$ the estimated values of ρ for a neighbourhood of $\boldsymbol{\beta}$. We minimize the PEL or PAEL over $\boldsymbol{\beta}$ using the modified Newton–Raphson algorithm. Note that the penalty function $p_\delta(|\beta_j|)$ is irregular at the origin and may not have a second derivative at some points. Therefore, special care is needed in the application of the Newton–Raphson algorithm. The penalty function is locally approximated, as discussed by Fan and Li [2001]. We assume that the profile empirical log-likelihood function is smooth with respect to $\boldsymbol{\beta}$ so that its first two partial derivatives are continuous. Thus, the term in the profile empirical log-likelihood ratio can be locally approximated via Taylor's expansion. Therefore, the minimization problem can be reduced to a quadratic minimization problem, and the Newton–Raphson algorithm can be used. The modified Newton–Raphson algorithm for estimating $\boldsymbol{\beta}$ uses a quadratic approximation of the profile empirical log-likelihood ratio function. The algorithm for optimizing the PEL is as follows:

1. Set $\boldsymbol{\beta} = \boldsymbol{\beta}^0$, $h = 0$, and $\epsilon = 1e^{-08}$.

2. Let $\hat{\boldsymbol{\lambda}}(\beta) = \boldsymbol{\lambda}(\beta)$ and $\hat{\rho}(\beta)$ be the estimated values of $\boldsymbol{\lambda}$ and ρ.

3. Compute the new estimate of $\boldsymbol{\beta}$ via

$$\boldsymbol{\beta}^{(h+1)} = \boldsymbol{\beta}^{(h)} - \left\{W_l^{\boldsymbol{\beta\beta}}(\boldsymbol{\beta}^h) + k\Sigma_\delta(\boldsymbol{\beta}^h)\right\}^{-1} \left\{W_l^{\boldsymbol{\beta}}(\boldsymbol{\beta}^h) + kU_\delta(\boldsymbol{\beta}^h)\right\} \tag{4.6}$$

   where $W_l(\boldsymbol{\beta})$ is the profile empirical log-likelihood ratio function defined in (2.6), with

$$W_l^{\boldsymbol{\beta}} = \frac{\partial W_l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}},\ W_l^{\boldsymbol{\beta\beta}} = \frac{\partial^2 W_l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T},$$
$$\Sigma_\delta(\boldsymbol{\beta}^h) = \text{diag}\left[\frac{p'_\delta(|\beta_1^h|)}{|\beta_1^h|}, \ldots, \frac{p'_\delta(|\beta_p^h|)}{|\beta_p^h|}\right], \text{ and } U_\delta(\boldsymbol{\beta}^h) = \Sigma_\delta(\boldsymbol{\beta}^h)\boldsymbol{\beta}^h.$$

   Note that to compute $W_l^{\boldsymbol{\beta}}$ and $W_l^{\boldsymbol{\beta\beta}}$, we need to estimate the Lagrange multipliers $\hat{\boldsymbol{\lambda}}(\beta)$ and $\hat{\rho}(\beta)$.

4. If $\min \left|\boldsymbol{\beta}^{(h+1)} - \boldsymbol{\beta}^{(h)}\right| < \epsilon$ stop the algorithm and report $\boldsymbol{\beta}^{(h+1)}$; otherwise set $h = h + 1$ and go to Step 3.

The simplified expressions for $W_l^{\boldsymbol{\beta}}$ and $W_l^{\boldsymbol{\beta\beta}}$ are given in Section 2.4.2.

4.5.2 Selection of Thresholding Parameters

The SCAD penalty function involves two unknown parameters, δ and a. In practice, we could search for the best pair (δ, a) over a two-dimensional grid using cross-validation (CV; Stone [1974]) or generalized cross-validation (GCV; Craven and Wahba [1979]). However, this is computationally expensive. From the Bayesian point of view, Fan and Li [2001] suggested using $a = 3.7$, and we use this value throughout our simulation studies. Let the profile EL ratio function evaluated at $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\lambda}}(\beta)$, and $\hat{\rho}(\beta)$ be

$$W_l(\hat{\boldsymbol{\beta}}) = \left\{ \sum_{i=1}^{k} \log(1 + \hat{\boldsymbol{\lambda}}(\beta)^T g_i(\hat{\boldsymbol{\beta}}, \hat{\rho}(\beta)) \right\}.$$

Then, we define the GCV criterion to be

$$\mathrm{GCV}(\delta) = \frac{W_l(\hat{\boldsymbol{\beta}})}{k\left[1 - e(\delta)/k\right]^2}, \tag{4.7}$$

where $e(\delta)$ is the generalized degrees of freedom given by

$$e(\delta) = \mathrm{tr}\left\{ \left[W_l^{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}) + \Sigma_\delta(\hat{\boldsymbol{\beta}}) \right]^{-1} W_l^{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}) \right\},$$

where $W_l^{\boldsymbol{\beta}\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}})$ is the second derivative of the profile EL ratio function with respect to $\boldsymbol{\beta}$ (see (2.25)) evaluated at $\hat{\boldsymbol{\beta}}$, and tr denotes the trace of a matrix. We choose the tuning parameters δ to minimize $\mathrm{GCV}(\delta)$.

4.6 Performance Analysis for Penalized Empirical Likelihood Variable Selection

To assess the performance of our variable selection method, we conducted a series of Monte-Carlo simulations for longitudinal count and continuous data. In the simulations we compare the PEL and PAEL with the PGEE under different working correlation structures. Our performance measures are (i) the median of the relative model error (MRME), (ii) the average number of estimated zero coefficients that are initially set to zero, (iii) and the average number of zero coefficients that are initially set to nonzero. We also compare the estimated values of the nonzero coefficients and the corresponding simulated standard errors.

Following Tibshirani [1996], we compare the MRME (Fan and Li [2001]) rather than the mean relative model error because of the instability of the best-subset variable selection. The model error for the GLM is defined by

$$\mathrm{ME}(\hat{\boldsymbol{\beta}}) = E_x \left\{ \mu(\boldsymbol{X}\boldsymbol{\beta}) - \mu(\boldsymbol{X}\hat{\boldsymbol{\beta}}) \right\}^2,$$

where $\mu(\boldsymbol{X}\boldsymbol{\beta}) = E(\boldsymbol{y}|\boldsymbol{X})$. The relative model error is

$$\mathrm{RME} = \frac{\mathrm{ME}}{\mathrm{ME}_{\mathrm{full}}}$$

where $\mathrm{ME}_{\mathrm{full}}$ is the model error calculated by fitting the data with the unpenalized full model.

4.6.1 Correlation Models for Stationary and Nonstationary Count Data

We consider the stationary and nonstationary correlation models for count data given in Sections 3.2 and 3.5. For the analysis, we consider the covariates $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{i5})$, where $x_{i1} \sim \text{Bernoulli}(0.5)$, x_{i2} to x_{i5} are generated from a multivariate normal distribution with mean zero, the correlation between x_{il} and x_{jl} is $0.5^{|i-j|}$, $l = 2, \ldots, 5$, $\boldsymbol{\beta} = (0.5, 0.5, 0.6, 0, 0)$, and $\rho = 0.5$. There are $m = 4$ time points and $k = 100$ subjects.

We simulated 1000 data sets from each of these models follow the AR(1), EQC, or MA(1) structure, and used penalized methods to estimate the parameters using different working correlation such as AR(1), EQC, and MA(1) as well as lag correlation. We compute the MRME values based on the PEL and PAEL and compare them with the PGEE. Table 4.1 gives the results for the independent and AR(1) models with stationary covariates; Table 4.2 gives the results for the EQC and MA(1) models with stationary covariates; and Table 4.3 gives the results for the independent, AR(1), and EQC models with nonstationary covariates. We also report the average number of zero coefficients, the estimated values of the nonzero coefficients, and the corresponding simulated standard errors in parentheses. The column labelled "Correct" is the average number of estimated zero coefficients that were initially set to zero, and the column labelled "Incorrect" is the average number of zero coefficients that were initially set to nonzero.

Tables 4.1 and 4.2 show that the MRMEs of the PEL and PAEL are almost the same as that of the PGEE. For instance, in the AR(1)/AR(1) case the MRMEs based on the PGEE, PEL, and PAEL are 61.33, 60.04, and 59.62, and the average numbers of correct zero coefficients are 1.80, 1.97, and 1.97 respectively. This shows that the "Correct" values for PEL and PAEL are close to the target of two, and the nonzero estimates are close to the true values in all cases. However, when there are time-dependent covariates in the model, the consistency of the GEE, especially with the working correlation approach, is not guaranteed (see Section 1.5). Table 4.3 shows that the PEL and PAEL can provide substantial efficiency gains over the PGEE for nonstationary covariates. Overall, the methods have similar performance when the covariates are stationary. However, our methods are superior when the covariates are nonstationary.

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
IND/IND $\rho = 0.50$	PGEE	63.90	1.73	0.00	0.5018	0.5010	0.5973
					(0.040)	(0.044)	(0.044)
	PEL	64.81	1.94	0.00	0.5015	0.5021	0.5969
					(0.042)	(0.044)	(0.043)
	PAEL	63.07	1.93	0.00	0.5014	0.5026	0.5965
					(0.042)	(0.044)	(0.043)
AR(1)/AR(1) $\rho = 0.50$	PGEE	61.33	1.80	0.00	0.4965	0.5018	0.5976
					(0.058)	(0.061)	(0.061)
	PEL	60.04	1.97	0.00	0.5006	0.5035	0.5953
					(0.059)	(0.060)	(0.061)
	PAEL	59.62	1.97	0.00	0.5001	0.5028	0.5961
					(0.059)	(0.060)	(0.060)
AR(1)/lag $\rho = 0.50$	PGEE	64.44	1.80	0.00	0.4981	0.5010	0.5976
					(0.058)	(0.062)	(0.062)
	PEL	60.06	1.98	0.00	0.5018	0.5025	0.5963
					(0.058)	(0.060)	(0.060)
	PAEL	60.06	1.98	0.00	0.5016	0.5024	0.5967
					(0.059)	(0.060)	(0.061)

Table 4.1: Performance measures for count data with stationary covariates for the independent and AR(1) models.

4.6.2 Misspecified Working Correlation Structure

In the above simulation studies we set the working correlation to either the true correlation or the lag correlation. As discussed in Section 1.4.1, if the working correlation is misspecified, we may lose the efficiency of the parameter estimates. To assess this loss in efficiency, we conduct a simulation study in which the repeated counts follow the AR(1), EQC, or MA(1) structure but we use different working correlation structures. This model misspecification and the corresponding correlation estimation are discussed by Sutradhar and Das (1999). Tables 4.4, 4.5, and 4.6 give the results for the AR(1), EQC, and MA(1) models respectively.

We see that the PEL and PAEL are superior to the PGEE for misspecified working

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
EQC/EQC $\rho = 0.50$	PGEE	66.00	1.83	0.00	0.5050	0.5007	0.5948
					(0.065)	(0.066)	(0.066)
	PEL	63.32	1.97	0.00	0.5068	0.5046	0.5924
					(0.064)	(0.065)	(0.065)
	PAEL	62.36	1.97	0.00	0.5069	0.5043	0.5930
					(0.064)	(0.065)	(0.065)
EQC/lag $\rho = 0.50$	PGEE	64.52	1.83	0.00	0.5048	0.5015	0.5953
					(0.064)	(0.064)	(0.066)
	PEL	66.68	1.98	0.00	0.5078	0.5043	0.5928
					(0.066)	(0.066)	(0.068)
	PAEL	65.36	1.98	0.00	0.5088	0.5045	0.5928
					(0.065)	(0.066)	(0.067)
MA(1)/MA(1) $\rho = 0.50$	PGEE	66.00	1.83	0.00	0.5004	0.4993	0.5981
					(0.052)	(0.053)	(0.052)
	PEL	62.52	1.96	0.00	0.5017	0.5024	0.5965
					(0.051)	(0.052)	(0.051)
	PAEL	61.80	1.97	0.00	0.5005	0.5031	0.5967
					(0.051)	(0.052)	(0.050)
MA(1)/lag $\rho = 0.50$	PGEE	65.20	1.75	0.00	0.5005	0.5004	0.5970
					(0.051)	(0.053)	(0.053)
	PEL	63.68	1.96	0.00	0.5012	0.5021	0.5958
					(0.051)	(0.052)	(0.050)
	PAEL	62.98	1.97	0.00	0.5017	0.5021	0.5967
					(0.051)	(0.052)	(0.050)

Table 4.2: Performance measures for count data with stationary covariates for the EQC and MA(1) models.

correlation structures. For example, in the AR(1)/EQC case the MRMEs based on the PGEE, PEL, and PAEL are 69.76, 64.63, and 65.47, and the average numbers of correct zero coefficients are 1.84, 1.98, and 1.98 respectively. In the AR(1)/Lag case the MRMEs based on the PGEE, PEL, and PAEL are 66.90, 60.50, and 60.74, and the average numbers of correct zero coefficients are 1.85, 1.97, and 1.98 respectively. The MRME of the PGEE with a lag correlation structure is smaller than that for

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
IND/IND $\rho = 0.50$	PGEE	73.88	1.68	0.00	0.4994	0.5003	0.5989
					(0.041)	(0.046)	(0.040)
	PEL	76.12	1.93	0.00	0.4982	0.5003	0.5997
					(0.042)	(0.047)	(0.042)
	PAEL	77.60	1.94	0.0	0.4978	0.5011	0.5992
					(0.042)	(0.047)	(0.043)
AR(1)/AR(1) $\rho = 0.50$	PGEE	74.35	1.82	0.00	0.4960	0.5070	0.5945
					(0.060)	(0.066)	(0.058)
	PEL	66.25	1.96	0.00	0.4981	0.5194	0.6085
					(0.059)	(0.063)	(0.055)
	PAEL	65.26	1.97	0.00	0.4978	0.5207	0.6079
					(0.059)	(0.063)	(0.055)
AR(1)/lag $\rho = 0.50$	PGEE	74.37	1.81	0.00	0.4956	0.5079	0.5949
					(0.056)	(0.063)	(0.055)
	PEL	74.29	1.97	0.00	0.4970	0.5077	0.5933
					(0.054)	(0.062)	(0.054)
	PAEL	74.51	1.97	0.00	0.4971	0.5080	0.5930
					(0.054)	(0.062)	(0.054)
EQC/EQC $\rho = 0.50$	PGEE	73.68	1.85	0.00	0.4978	0.5033	0.5957
					(0.065)	(0.076)	(0.068)
	PEL	60.14	1.97	0.00	0.5014	0.5256	0.6094
					(0.064)	(0.071)	(0.064)
	PAEL	59.87	1.97	0.00	0.5009	0.5266	0.6091
					(0.064)	(0.071)	(0.064)
EQC/lag $\rho = 0.50$	PGEE	74.10	1.70	0.00	0.4988	0.5020	0.5962
					(0.066)	(0.075)	(0.069)
	PEL	69.89	1.97	0.00	0.5025	0.5034	0.5963
					(0.065)	(0.073)	(0.066)
	PAEL	71.44	1.97	0.00	0.5026	0.5019	0.5984
					(0.065)	(0.074)	(0.066)

Table 4.3: Performance measures for count data with nonstationary covariates for the independent, AR(1), and EQC models.

the PGEE with a misspecified working correlation, supporting the findings of Section 1.4.1. The average numbers of correct zero coefficients for the PEL and PAEL are close to the target of two in all cases. This result shows that the PEL and PAEL outperform the PGEE since they are nonparametric methods based on a data-driven likelihood ratio function. It also shows that a working correlation based on the PGEE variable selection procedure is sensitive to the choice of covariance structure, leading to a loss of efficiency of the regression estimators.

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
AR(1)/EQC $\rho = 0.70$	PGEE	69.76	1.84	0.00	0.5030	0.5047	0.5935
					(0.067)	(0.073)	(0.072)
	PEL	64.63	1.98	0.00	0.5068	0.5086	0.5918
					(0.066)	(0.071)	(0.071)
	PAEL	65.47	1.98	0.00	0.5061	0.5089	0.5914
					(0.066)	(0.071)	(0.070)
AR(1)/lag $\rho = 0.70$	PGEE	66.90	1.85	0.00	0.5034	0.5052	0.5929
					(0.066)	(0.074)	(0.071)
	PEL	60.50	1.97	0.00	0.5083	0.5091	0.5889
					(0.065)	(0.070)	(0.068)
	PAEL	60.74	1.98	0.00	0.5077	0.5091	0.5896
					(0.065)	(0.070)	(0.069
AR(1)/MA(1) $\rho = 0.49$	PGEE	70.57	1.65	0.00	0.5011	0.5021	0.5986
					(0.060)	(0.062)	(0.062)
	PEL	64.01	1.98	0.00	0.5063	0.5930	0.5930
					(0.058)	(0.061)	(0.061)
	PAEL	63.42	1.97	0.00	0.5023	0.5068	0.5929
					(0.059)	(0.061)	(0.061)
AR(1)/lag $\rho = 0.49$	PGEE	67.64	1.79	0.00	0.5014	0.5029	0.5973
					(0.059)	(0.062)	(0.062)
	PEL	62.52	1.96	0.00	0.5017	0.5056	0.5932
					(0.058)	(0.061)	(0.061)
	PAEL	63.11	1.97	0.00	0.5026	0.5063	0.5920
					(0.058)	(0.061)	(0.061)

Table 4.4: Performance measures for count data with stationary covariates when the working correlation is misspecified for an AR(1) model.

4.6.3 Over-dispersed Stationary and Nonstationary Count Data

In this section, we consider the performance of our approach when the variance function is misspecified, in the context of stationary and nonstationary count data. We

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
EQC/AR(1) $\rho = 0.70$	PGEE	63.37	1.70	0.00	0.5025 (0.074)	0.5066 (0.076)	0.5914 (0.076)
	PEL	58.53	1.98	0.00	0.5098 (0.074)	0.5074 (0.074)	0.5918 (0.074)
	PAEL	57.82	1.98	0.00	0.5093 (0.074)	0.5090 (0.072)	0.5906 (0.073)
EQC/lag $\rho = 0.70$	PGEE	62.61	1.87	0.00	0.5026 (0.074)	0.5069 (0.076)	0.5916 (0.074)
	PEL	58.63	1.99	0.00	0.5098 (0.073)	0.5100 (0.073)	0.5887 (0.073)
	PAEL	59.65	1.98	0.00	0.5094 (0.073)	0.5102 (0.073)	0.5880 (0.073)
EQC/MA(1) $\rho = 0.49$	PGEE	75.50	1.59	0.00	0.4996 (0.065)	0.5023 (0.068)	0.5980 (0.073)
	PEL	63.75	1.98	0.00	0.5035 (0.064)	0.5044 (0.067)	0.5948 (0.069)
	PAEL	65.49	1.97	0.00	0.5020 (0.064)	0.5050 (0.067)	0.5950 (0.069)
EQC/lag $\rho = 0.49$	PGEE	66.40	1.82	0.00	0.5017 (0.065)	0.5046 (0.068)	0.5938 (0.070)
	PEL	63.93	1.97	0.00	0.5037 (0.064)	0.5072 (0.067)	0.5916 (0.069)
	PAEL	63.35	1.98	0.00	0.5028 (0.064)	0.5072 (0.066)	0.5911 (0.068)

Table 4.5: Performance measures for count data with stationary covariates when the working correlation is misspecified for the EQC model.

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
MA(1)/AR(1) $\rho = 0.67$	PGEE	69.39	1.78	0.00	0.5025 (0.052)	0.5016 (0.056)	0.5967 (0.055)
	PEL	64.76	1.96	0.00	0.5018 (0.051)	0.5013 (0.055)	0.5963 (0.054)
	PAEL	66.90	1.97	0.00	0.5012 (0.051)	0.5011 (0.055)	0.5965 (0.054)
MA(1)/lag $\rho = 0.67$	PGEE	69.03	1.76	0.00	0.5021 (0.053)	0.5022 (0.056)	0.5963 (0.055)
	PEL	64.99	1.97	0.00	0.5008 (0.052)	0.5031 (0.055)	0.5962 (0.054)
	PAEL	65.00	1.97	0.00	0.5013 (0.052)	0.5022 (0.055)	0.5969 (0.053)
MA(1)/EQC $\rho = 0.67$	PGEE	71.37	1.71	0.00	0.5006 (0.052)	0.5008 (0.056)	0.5974 (0.058)
	PEL	59.73	1.97	0.00	0.5010 (0.051)	0.5025 (0.055)	0.5955 (0.055)
	PAEL	60.18	1.98	0.00	0.5010 (0.051)	0.5024 (0.055)	0.5955 (0.055)

Table 4.6: Performance measures for count data with stationary covariates when the working correlation is misspecified for the MA(1) model.

generate over-dispersed count data with over-dispersion parameter $\omega = 1/4$ as discussed in Section 3.4. We simulated 1000 data sets with the parameter set used in Section 4.6.1. We report the MRME, the average number of zero coefficients, the estimated values of the nonzero coefficients, and the corresponding simulated standard errors in parentheses.

Table 4.7 gives the results for the independent and AR(1) models with stationary covariates; Table 4.8 gives the results for the EQC and MA(1) models with stationary covariates; and Table 4.9 gives the results for the independent, AR(1), and EQC models with nonstationary covariates.

The tables show that the MRMEs of the PEL and PAEL are smaller than the MRMEs of the PGEE in all cases. For instance, in the EQC/EQC case in Table 4.7 we see that the MRMEs based on the PGEE, PEL, and PAEL are 82.80, 68.08, and 67.14, and the average numbers of correct zero coefficients are 1.53, 1.98, and 1.98 respectively. This shows that the PEL and PAEL are superior to the PGEE, and the average numbers of correct zero coefficients for the PEL and PAEL are close to the target of two in all cases. The PGEE based on lag correlation has a similar pattern to the PGEE based on working correlation. When there is over-dispersion, the PEL and PAEL outperform the PGEE. Note that the PEL and PAEL did not model the over-dispersion. This shows that the EL approaches are robust to model misspecification.

4.6.4 Correlation Models for Continuous Data

In this section, we compare the performance of our PEL and PAEL approaches under stationary and nonstationary correlation models for continuous data. The correlated normal responses are generated from model (3.5) with $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)^T$ and $p = 8$. For the ith covariate $\boldsymbol{X}_i = (x_{i1}, \dots, x_{ip})$ are generated from the multivariate

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
	PGEE	84.94	1.00	0.00	0.4962	0.5035	0.5905
					(0.082)	(0.098)	(0.111)
IND/IND	PEL	59.85	1.97	0.00	0.5037	0.5062	0.5896
$\rho = 0.50$					(0.080)	(0.092)	(0.097)
	PAEL	59.24	1.98	0.00	0.5053	0.5068	0.5891
					(0.080)	(0.093)	(0.095)
	PGEE	73.07	1.47	0.00	0.5018	0.5121	0.5864
					(0.092)	(0.102)	(0.108)
AR(1)/AR(1)	PEL	63.32	1.98	0.00	0.5132	0.5161	0.5862
$\rho = 0.50$					(0.091)	(0.101)	(0.097)
	PAEL	64.26	1.98	0.00	0.5119	0.5159	0.5878
					(0.091)	(0.101)	(0.099)
	PGEE	69.04	1.52	0.00	0.5037	0.5110	0.5874
					(0.089)	(0.102)	(0.102)
AR(1)/lag	PEL	65.54	1.99	0.00	0.5115	0.5171	0.5882
$\rho = 0.50$					(0.086)	(0.097)	(0.099)
	PAEL	66.10	1.99	0.00	0.5107	0.5162	0.5893
					(0.086)	(0.098)	(0.100)

Table 4.7: Performance measures for over-dispersion count data with stationary covariates for the independent and AR(1) models.

normal distribution with mean 0 and an AR(1) covariance matrix with marginal variance 1 and auto-correlation coefficient 0.5. The random errors $(\epsilon_1, ..., \epsilon_4)^T$ are generated as in Section 3.7. We simulated 1000 data sets with $k = 50$ individuals from the models given in Section 3.7. We compute the MRME values based on the PEL and PAEL and compare them with the PGEE.

Table 4.10 gives the results for the independent and AR(1) models with stationary covariates; Table 4.11 gives the results for the EQC and MA(1) models with stationary covariates; Table 4.12 gives the results for the independent and AR(1) models with nonstationary covariates; and Table 4.13 gives the results for the EQC and MA(1) models with nonstationary covariates. We also report the average number of zero

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
EQC/EQC $\rho = 0.50$	PGEE	82.80	1.53	0.00	0.5117	0.5075	0.5832
					(0.093)	(0.109)	(0.116)
	PEL	68.08	1.98	0.00	0.5163	0.5168	0.5825
					(0.092)	(0.106)	(0.106)
	PAEL	67.14	1.98	0.00	0.5170	0.5168	0.5830
					(0.092)	(0.106)	(0.105)
EQC/lag $\rho = 0.50$	PGEE	81.79	1.51	0.00	0.5104	0.5082	0.5850
					(0.092)	(0.107)	(0.118)
	PEL	65.47	1.98	0.00	0.5161	0.5168	0.5838
					(0.091)	(0.106)	(0.106)
	PAEL	64.55	1.98	0.00	0.5151	0.5191	0.5822
					(0.091)	(0.106)	(0.106)
MA(1)/MA(1) $\rho = 0.50$	PGEE	73.78	1.31	0.00	0.5059	0.5071	0.5892
					(0.087)	(0.099)	(0.111)
	PEL	59.64	1.98	0.00	0.5101	0.5125	0.5882
					(0.086)	(0.091)	(0.101)
	PAEL	61.28	1.98	0.00	0.5092	0.5143	0.5875
					(0.086)	(0.092)	(0.103)
MA(1)/lag $\rho = 0.50$	PGEE	68.37	1.46	0.00	0.5054	0.5058	0.5892
					(0.086)	(0.097)	(0.106)
	PEL	58.86	1.97	0.00	0.5103	0.5096	0.5882
					(0.085)	(0.092)	(0.101)
	PAEL	59.05	1.97	0.00	0.5107	0.5115	0.5879
					(0.085)	(0.092)	(0.102)

Table 4.8: Performance measures for over-dispersion count data with stationary covariates for the EQC and MA(1) models.

coefficients, the estimated values of the nonzero coefficients, and the corresponding simulated standard errors in parentheses.

The tables show that the MRMEs of the PEL and PAEL are similar to those of the PGEE, but the average numbers of correct zero coefficients for the PEL and PAEL are close to the target value of five in all cases. For example, in the EQC/lag case in Table 4.10 we see that the MRMEs based on the PGEE, PEL, and PAEL are 32.34,

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
IND/IND $\rho = 0.50$	PGEE	96.20	0.82	0.00	0.4988	0.5106	0.5881
					(0.102)	(0.115)	(0.119)
	PEL	62.92	1.96	0.01	0.5036	0.5087	0.5940
					(0.097)	(0.112)	(0.102)
	PAEL	62.31	1.96	0.01	0.5021	0.4817	0.6272
					(0.101)	(0.113)	(0.102)
AR(1)/AR(1) $\rho = 0.50$	PGEE	85.19	1.28	0.00	0.4970	0.5068	0.5907
					(0.102)	(0.115)	(0.121)
	PEL	72.04	1.93	0.00	0.4701	0.5055	0.5694
					(0.101)	(0.113)	(0.119)
	PAEL	70.25	1.93	0.00	0.4704	0.5047	0.5677
					(0.101)	(0.114)	(0.118)
AR(1)/lag $\rho = 0.50$	PGEE	85.55	1.27	0.00	0.4985	0.5067	0.5908
					(0.103)	(0.116)	(0.121)
	PEL	67.09	1.97	0.01	0.4968	0.5152	0.5812
					(0.102)	(0.112)	(0.108)
	PAEL	68.44	1.98	0.01	0.4997	0.5126	0.5805
					(0.102)	(0.112)	(0.109)
EQC/EQC $\rho = 0.50$	PGEE	77.92	1.37	0.00	0.5010	0.4999	0.5893
					(0.109)	(0.116)	(0.115)
	PEL	71.79	1.88	0.00	0.4689	0.5006	0.5824
					(0.106)	(0.114)	(0.112)
	PAEL	69.91	1.86	0.00	0.4722	0.4992	0.5824
					(0.106)	(0.114)	(0.112)
EQC/lag $\rho = 0.50$	PGEE	77.47	1.19	0.00	0.4910	0.5051	0.5884
					(0.109)	(0.118)	(0.121)
	PEL	71.71	1.97	0.03	0.5042	0.5102	0.5908
					(0.126)	(0.116)	(0.118)
	PAEL	71.96	1.98	0.03	0.5055	0.5084	0.5922
					(0.106)	(0.117)	(0.119)

Table 4.9: Performance measures for over-dispersion count data with nonstationary covariates for independent, AR(1), and EQC models.

31.99, and 31.92, and the average numbers of correct zero coefficients are 4.59, 4.96, and 4.97 respectively. This clearly indicates that our methods perform as well as the PGEE when the model assumptions are correct. However, the results in Tables 4.12 and 4.13 show that the PEL and PAEL are superior to the PGEE. For instance, in the EQC/EQC case in Table 4.12 the MRMEs based on the PGEE, PEL, and PAEL are 41.04, 34.65, and 34.13, and the average numbers of correct zero coefficients are 3.97, 4.83, and 4.85 respectively. This shows that the PGEE estimates are not necessarily consistent when there are time-dependent covariates in the model; see Section 1.5.

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_5$
IND/IND $\rho = 0.50$	PGEE	35.79	4.24	0.00	2.998 (0.091)	1.501 (0.088)	2.001 (0.075)
	PEL	33.40	4.85	0.00	3.000 (0.088)	1.500 (0.087)	1.996 (0.073)
	PAEL	33.40	4.87	0.00	3.000 (0.088)	1.500 (0.087)	1.996 (0.074)
AR(1)/AR(1) $\rho = 0.50$	PGEE	34.53	4.54	0.00	2.997 (0.120)	1.504 (0.118)	2.004 (0.106)
	PEL	34.45	4.94	0.00	2.999 (0.121)	1.504 (0.119)	2.001 (0.107)
	PAEL	34.21	4.96	0.00	2.999 (0.121)	1.503 (0.119)	2.002 (0.107)
AR(1)/lag $\rho = 0.50$	PGEE	34.50	4.49	0.00	2.997 (0.117)	1.501 (0.118)	2.002 (0.107)
	PEL	34.64	4.94	0.00	2.999 (0.121)	1.500 (0.121)	1.999 0.107)
	PAEL	34.86	4.95	0.00	3.001 (0.122)	1.500 (0.122)	2.000 (0.107)

Table 4.10: Performance measures for continuous data with stationary covariates for the independent and AR(1) models.

4.6.5 Misspecified Correlation Models for Continuous Data

We now consider the performance of our method when the model is misspecified. We generate a class of stationary and nonstationary correlation models for longitudinal misspecified continuous data from (3.5), where the correlated random errors $(\epsilon_1, ..., \epsilon_4)^T$ are generated from the $\chi^2(1) - 1$ distribution instead of the normal distribution, as in Section 3.7. We simulated 1000 data sets with the parameter set of Section 4.6.4 for the models discussed in Section 3.7. However, we construct the

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_5$
EQC/EQC $\rho = 0.50$	PGEE	32.99	4.60	0.00	3.000 (0.132)	1.495 (0.131)	2.004 (0.117)
	PEL	33.23	4.96	0.00	2.998 (0.134)	1.500 (0.133)	2.002 (0.118)
	PAEL	33.35	4.97	0.00	2.999 (0.135)	1.499 (0.133)	2.002 (0.118)
EQC/lag $\rho = 0.50$	PGEE	32.34	4.59	0.00	3.002 (0.134)	1.495 (0.136)	2.002 (0.119)
	PEL	31.99	4.96	0.00	2.998 (0.133)	1.503 (0.135)	2.000 (0.118)
	PAEL	31.92	4.97	0.00	2.999 (0.133)	1.502 (0.136)	2.001 (0.118)
MA(1)/MA(1) $\rho = 0.50$	PGEE	35.60	4.46	0.00	3.004 (0.109)	1.497 (0.110)	2.001 (0.096)
	PEL	35.02	4.90	0.00	3.001 (0.108)	1.499 (0.110)	2.001 (0.095)
	PAEL	34.86	4.91	0.00	3.001 (0.109)	1.498 (0.111)	2.000 (0.096)
MA(1)/lag $\rho = 0.50$	PGEE	34.82	4.46	0.00	3.005 (0.110)	1.494 (0.110)	2.003 (0.094)
	PEL	34.66	4.92	0.00	3.001 (0.111)	1.500 (0.111)	1.999 (0.095)
	PAEL	34.77	4.96	0.00	3.001 (0.110)	1.500 (0.111)	2.000 (0.096)

Table 4.11: Performance measures for continuous data with stationary covariates for the EQC and MA(1) models.

PGEE, PEL, and PAEL under the assumption that there is no misspecification. We report the MRME, the average number of zero coefficients, the estimated values of the nonzero coefficients, and the corresponding simulated standard errors in parentheses. Table 4.14 gives the results for the independent and AR(1) models with stationary covariates; Table 4.15 gives the results for the EQC and MA(1) models with stationary covariates; Table 4.16 gives the results for the independent and AR(1) models with

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_5$
	PGEE	39.34	4.28	0.00	3.001	1.497	2.001
					(0.087)	(0.086)	(0.076)
IND/IND	PEL	36.46	4.90	0.00	2.999	1.498	2.003
$\rho = 0.50$					(0.086)	(0.085)	(0.074)
	PAEL	36.61	4.91	0.00	2.998	1.498	2.003
				(0.085)	(0.085)	(0.074)	
	PGEE	38.34	3.99	0.00	3.000	1.501	1.999
					(0.069)	(0.067)	(0.061)
AR(1)/AR(1)	PEL	33.62	4.79	0.00	3.002	1.499	2.000
$\rho = 0.50$					(0.068)	(0.066)	(0.060)
	PAEL	33.52	4.84	0.00	3.001	1.498	2.000
					(0.068)	(0.066)	(0.060)
	PGEE	38.38	3.91	0.00	3.001	1.497	2.001
					(0.069)	(0.069)	(0.062)
AR(1)/lag	PEL	33.35	4.81	0.00	3.001	1.498	1.998
$\rho = 0.50$					(0.068)	(0.068)	(0.061)
	PAEL	32.66	4.83	0.00	3.001	1.498	2.000
					(0.068)	(0.068)	(0.061)

Table 4.12: Performance measures for continuous data with nonstationary covariates for the independent and AR(1) models.

nonstationary covariates; and Table 4.17 gives the results for the EQC and MA(1) models with nonstationary covariates.

he results show that the MRMEs of the PEL and PAEL are smaller than that of the PGEE in all cases. For instance, in the AR(1)/AR(1) case in Table 4.16 we see that the MRMEs based on the PGEE, PEL, and PAEL are 46.43, 30.74, and 30.44, and the average numbers of correct zero coefficients are 3.03, 4.90, and 4.93 respectively. This shows that the PEL and PAEL outperform the PGEE, and the average numbers of correct zero coefficients for the PEL and PAEL are close to the target of five in all cases. The PGEE based on lag correlation has a similar pattern to the PGEE based on working correlation. This clearly indicates that our method is

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_5$
EQC/EQC $\rho = 0.50$	PGEE	41.04	3.97	0.00	2.998	1.499	2.002
					(0.067)	(0.068)	(0.058)
	PEL	34.65	4.83	0.00	2.997	1.500	2.001
					(0.066)	(0.066)	(0.057)
	PAEL	34.13	4.85	0.00	2.998	1.499	2.001
					(0.066)	(0.066)	(0.057)
EQC/lag $\rho = 0.50$	PGEE	41.22	3.91	0.00	3.001	1.497	1.999
					(0.068)	(0.067)	(0.059)
	PEL	33.74	4.81	0.00	2.998	1.499	2.002
					(0.067)	(0.066)	(0.058)
	PAEL	34.36	4.84	0.00	2.998	1.498	2.000
					(0.067)	(0.066)	(0.058)
MA(1)/MA(1) $\rho = 0.50$	PGEE	39.94	4.09	0.00	3.002	1.501	2.005
					(0.079)	(0.080)	(0.070)
	PEL	35.48	4.82	0.00	3.000	1.502	2.002
					(0.076)	(0.077)	(0.069)
	PAEL	35.79	4.85	0.00	3.002	1.502	2.003
					(0.076)	(0.077)	(0.069)
MA(1)/lag $\rho = 0.50$	PGEE	45.22	3.69	0.00	2.995	1.502	1.996
					(0.062)	(0.062)	(0.057)
	PEL	34.55	4.74	0.00	2.997	1.496	2.000
					(0.060)	(0.061)	(0.055)
	PAEL	32.94	4.78	0.00	2.999	0.1496	1.999
					(0.060)	(0.061)	(0.055)

Table 4.13: Performance measures for continuous data with nonstationary covariates for the EQC and MA(1) models.

superior to the PGEE when the model assumptions are incorrect, and the PEL and PAEL are robust to model misspecification.

4.7 Summary

Our performance analysis shows that our PEL and PAEL have consistent performance when the model assumptions are correct for count and continuous responses with stationary covariates. However, when the model is misspecified our PEL and PAEL are superior to the PGEE. This shows that the PEL and PAEL are robust to model misspecification since they are nonparametric methods based on a data-driven likelihood ratio function. When there are time-dependent covariates in the model, the consistency of the PGEE is not guaranteed. Our results show that the PEL and PAEL can provide substantial efficiency gains over the PGEE for nonstationary covariates.

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_5$
IND/IND $\rho = 0.50$	PGEE	41.93	3.74	0.00	2.995 (0.114)	1.506 (0.120)	2.000 (0.106)
	PEL	30.24	4.91	0.00	2.997 (0.112)	1.504 (0.117)	1.998 (0.100)
	PAEL	30.06	4.94	0.00	2.997 (0.112)	1.505 (0.117)	1.998 (0.100)
AR(1)/AR(1) $\rho = 0.50$	PGEE	37.66	3.95	0.00	2.991 (0.186)	1.515 (0.187)	1.995 (0.169)
	PEL	32.69	4.98	0.00	2.990 (0.185)	1.502 (0.185)	2.005 (0.156)
	PAEL	32.69	4.99	0.00	2.990 (0.185)	1.502 (0.185)	2.004 (0.155)
AR(1)/lag $\rho = 0.50$	PGEE	51.37	2.99	0.00	2.986 (0.181)	1.497 (0.193)	2.004 (0.174)
	PEL	31.97	4.88	0.00	2.997 (0.180)	1.501 (0.181)	1.995 (0.153)
	PAEL	32.16	4.89	0.00	2.990 (0.180)	1.502 (0.182)	2.004 (0.155)

Table 4.14: Performance measures for misspecified continuous data with stationary covariates for the independent and AR(1) models.

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_5$
EQC/EQC $\rho = 0.50$	PGEE	37.09	3.76	0.00	3.001	1.502	2.002
					(0.170)	(0.169)	(0.148)
	PEL	33.86	4.98	0.00	3.004	1.496	2.000
					(0.168)	(0.168)	(0.147)
	PAEL	33.79	4.98	0.00	3.003	1.498	2.002
					(0.168)	(0.168)	(0.146)
EQC/lag $\rho = 0.50$	PGEE	39.17	3.74	0.00	3.001	1.501	2.000
					(0.171)	(0.171)	(0.149)
	PEL	33.58	4.97	0.00	3.002	1.501	1.999
					(0.170)	(0.170)	(0.148)
	PAEL	33.71	4.98	0.00	3.003	1.500	2.000
					(0.170)	(0.170)	(0.148)
MA(1)/MA(1) $\rho = 0.50$	PGEE	39.28	3.77	0.00	3.009	1.501	2.001
					(0.165)	(0.175)	(0.147)
	PEL	33.40	4.99	0.00	2.993	1.512	2.003
					(0.161)	(0.174)	(0.143)
	PAEL	33.56	4.98	0.00	2.993	1.513	2.003
					(0.161)	(0.174)	(0.143)
MA(1)/lag $\rho = 0.50$	PGEE	39.42	3.81	0.00	3.008	1.502	2.006
					(0.167)	(0.175)	(0.146)
	PEL	34.97	4.97	0.00	2.996	1.510	2.005
					(0.165)	(0.174)	(0.142)
	PAEL	34.97	4.97	0.00	2.996	1.510	2.005
					(0.165)	(0.174)	(0.142)

Table 4.15: Performance measures for misspecified continuous data with stationary covariates for the EQC and MA(1) models.

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_5$
	PGEE	35.89	3.63	0.00	3.003	1.497	1.993
					(0.113)	(0.114)	(0.103)
IND/IND	PEL	29.35	4.93	0.00	3.006	1.501	1.998
$\rho = 0.50$					(0.107)	(0.111)	(0.099)
	PAEL	28.79	4.96	0.00	3.006	1.501	1.998
					(0.108)	(0.113)	(0.099)
	PGEE	46.43	3.03	0.00	3.006	1.499	1.997
					(0.111)	(0.113)	(0.103)
AR(1)/AR(1)	PEL	30.74	4.90	0.00	3.007	1.498	1.998
$\rho = 0.50$					(0.107)	(0.111)	(0.091)
	PAEL	30.44	4.93	0.00	3.007	1.497	1.998
					(0.107)	(0.112)	(0.092)
	PGEE	51.16	2.78	0.00	3.002	1.502	1.997
					(0.114)	(0.117)	(0.112)
AR(1)/lag	PEL	30.74	4.91	0.00	3.005	1.499	2.001
$\rho = 0.50$					(0.113)	(0.113)	(0.095)
	PAEL	30.36	4.94	0.00	3.007	1.498	2.002
					(0.113)	(0.113)	(0.094)

Table 4.16: Performance measures for misspecified continuous data with nonstationary covariates for the independent and AR(1) models.

True model/ Working correlation	Method	MRME%	Avg. no. of zero coefficients		Estimates of nonzero coefficients		
			Correct	Incorrect	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_5$
	PGEE	44.80	3.14	0.00	3.007	1.495	1.998
					(0.124)	(0.122)	(0.118)
EQC/EQC	PEL	31.59	4.93	0.00	3.005	1.497	2.002
$\rho = 0.50$					(0.121)	(0.121)	(0.109)
	PAEL	31.72	4.95	0.00	3.004	1.498	2.001
					(0.121)	(0.122)	(0.109)
	PGEE	47.14	3.08	0.00	3.009	1.498	1.997
					(0.129)	(0.129)	(0.122)
EQC/lag	PEL	30.57	4.95	0.00	2.999	1.499	1.999
$\rho = 0.50$					(0.126)	(0.128)	(0.113)
	PAEL	30.37	4.97	0.00	2.999	1.498	1.999
					(0.126)	(0.128)	(0.116)
	PGEE	42.23	3.35	0.00	3.001	1.500	2.001
					(0.110)	(0.115)	(0.108)
MA(1)/MA(1)	PEL	29.89	4.92	0.00	3.005	1.493	2.000
$\rho = 0.50$					(0.109)	(0.113)	(0.097)
	PAEL	30.28	4.93	0.00	3.006	1.493	2.001
					(0.109)	(0.113)	(0.097)
	PGEE	49.90	2.91	0.00	3.006	1.500	1.998
					(0.115)	(0.116)	(0.098)
MA(1)/lag	PEL	28.84	4.93	0.00	3.003	1.497	2.002
$\rho = 0.50$					(0.111)	(0.110)	(0.097)
	PAEL	28.67	4.95	0.00	3.004	1.497	2.002
					(0.110)	(0.111)	(0.097)

Table 4.17: Performance measures for misspecified continuous data with nonstationary covariates for the EQC and MA(1) models.

Chapter 5

Applications

In this chapter, we illustrate the applicability of our proposed method to two real-world examples.

5.1 Health Care Utilization Study

We consider longitudinal health care utilization data (Sutradhar [2003]) that was collected by Eastern Health, St. John's, Newfoundland, Canada. These longitudinal count data contain complete records for $k = 144$ individuals for the $m = 4$ years from 1985 to 1988. The response of interest was the number of visits to a physician by each individual during a given year. Information on four covariates, namely, gender, number of chronic conditions, education level, and age, was recorded for each individual. Background information allows us to assume that the response variable, marginally, follows the Poisson distribution, and the repeated counts over the four years will be longitudinally correlated. Since the data indicate over-dispersion, we consider a negative binomial model with two variance functions

$$\mathrm{var}(y) = \mu + \alpha\mu$$

and

$$\text{var}(y) = \mu + \alpha\mu^2.$$

Thus, the variance function is different from that of the Poisson model, $\text{var}(y) = \mu$. To confirm the over-dispersion, we test $H_0 : \alpha = 0$ against $H_a : \alpha > 0$ using the likelihood ratio test. The result confirms the presence of over-dispersion in both variance function models.

Parameter	Estimate	95% Confidence Interval	
		GEE	EL
Working Correlation: AR(1)			
Gender effect ($\hat{\beta}_1$)	-0.1929	(-0.313, -0.073)	**(-0.421, 0.020)**
Chronic effect ($\hat{\beta}_2$)	0.1668	(0.177, 0.216)	(0.094, 0.241)
Education effect ($\hat{\beta}_3$)	-0.4738	(-0.624, -0.324)	(-0.768, -0.180)
Age effect ($\hat{\beta}_4$)	0.0308	(0.029, 0.033)	(0.028, 0.033)
Working Correlation: EQC			
Gender effect ($\hat{\beta}_1$)	-0.1772	(-0.306, -0.048)	**(-0.407, 0.034)**
Chronic effect ($\hat{\beta}_2$)	0.1681	(0.115, 0.222)	(0.095, 0.237)
Education effect ($\hat{\beta}_3$)	-0.4354	(-0.597, -0.274)	(-0.726, -0.146)
Age effect ($\hat{\beta}_4$)	0.0302	(0.028, 0.033)	(0.027, 0.033)
Working Correlation: MA(1)			
Gender effect ($\hat{\beta}_1$)	-0.1922	(-0.299, -0.086)	**(-0.421, 0.021)**
Chronic effect ($\hat{\beta}_2$)	0.1669	(0.123, 0.211)	(0.094, 0.241)
Education effect ($\hat{\beta}_3$)	-0.4720	(-0.605, -0.339)	(-0.766, -0.179)
Age effect ($\hat{\beta}_4$)	0.0308	(0.029, 0.033)	(0.028, 0.033)
Working Correlation: Lag			
Gender effect ($\hat{\beta}_1$)	-0.1819	(-0.311, -0.053)	**(-0.411, 0.029)**
Chronic effect ($\hat{\beta}_2$)	0.1677	(0.114, 0.221)	(0.095, 0.238)
Education effect ($\hat{\beta}_3$)	0.4469	(-0.608, -0.286)	(-0.738, -0.156)
Age effect ($\hat{\beta}_4$)	0.0304	(0.028, 0.033)	(0.027, 0.033)

Table 5.1: Regression estimates for health care utilization count data.

Our analysis used the GEE with a working correlation matrix (AR(1), EQC, MA(1), or lag correlation) and our EL approach. Table 5.1 gives the regression parameter estimates and 95% CIs. The gender covariate was coded as 1 for male and 0 for female. Under the AR(1) structure, the estimate of its regression coefficient is $\hat{\beta}_1 = -0.1929$, suggesting that females make more visits to physicians. The GEE CI indicates that this variable is significant, but the EL CI does not. The estimated values $\hat{\beta}_2 = 0.1668$ and $\hat{\beta}_4 = 0.0308$ suggest that individuals with chronic diseases and older individuals pay more visits to physicians, as expected. The corresponding CIs show that both variables are significant. The education covariate was coded as 1 for less than high school and 0 for higher education. The value $\hat{\beta}_3 = -0.4738$ indicates that educated individuals pay more visits to physicians, showing that they are more concerned about their health or they can afford it. The corresponding CIs show that this variable is significant. Table 5.1 shows that different working correlations lead to slightly different parameter estimates, but the overall conclusion remains the same. Since the data indicate over-dispersion, the GEE-based approach may be inefficient, as shown in our performance analysis. We conclude that the EL approach is more appropriate for this data set, and the significant variables identified by this approach are more reliable.

5.1.1 Penalized Variable Selection for Health Care Utilization Data

To examine whether there are any interaction effects between the covariates, we included some two factor interactions and employed penalized variable selection to identify important covariates. The goal is to take the longitudinal correlations into account. Tables 5.2 and 5.3 summarize the results. The PGEE results indicate that CHRONIC, EDUCATION, AGE, GENDER*CHRONIC, and GENDER*AGE are significant. Under AR(1) and MA(1), the interaction GENDER*EDUCATION is also important. The PEL approach selects CHRONIC, EDUCATION, AGE, GENDER*CHRONIC, and GENDER*AGE. Since the data indicate over-dispersion, the EL approach may be more appropriate. Note that the PEL approach selected the simplest model.

Variable	Penalized Estimates			
	PGEE(AR(1))	PGEE(EQC)	PGEE(MA(1))	PGEE(Lag)
GENDER	**0.000**	**0.000**	**0.000**	**0.000**
CHRONIC	0.105	0.103	0.104	0.104
EDUCATION	-0.492	-0.432	-0.489	-0.443
AGE	0.033	0.032	0.033	0.033
GENDER*CHRONIC	0.143	0.144	0.144	0.143
GENDER*EDUCATION	0.053	**0.000**	0.050	**0.000**
GENDER*AGE	-0.009	-0.009	-0.009	-0.009

Table 5.2: PGEE regression estimates for health care utilization data under different working correlation structures.

Variable	Penalized Estimates			
	PEL(AR(1))	PEL(EQC)	PEL(MA(1))	PEL(Lag)
GENDER	**0.000**	**0.000**	**0.000**	**0.000**
CHRONIC	0.037	0.034	0.037	0.036
EDUCATION	-0.471	-0.468	-0.476	-0.466
AGE	0.035	0.035	0.035	0.035
GENDER*CHRONIC	0.005	0.005	0.005	0.005
GENDER*EDUCATION	**0.000**	**0.000**	**0.000**	**0.000**
GENDER*AGE	-0.005	-0.006	-0.005	-0.006

Table 5.3: PEL regression estimates for health care utilization data under different working correlation structures.

5.2 Longitudinal CD4 Cell Counts of HIV Seroconverters

This data set contains 2376 observations of the CD4 cell counts of $k = 369$ men infected with the HIV virus (Zeger and Diggle [1994]). The goal of our analysis is to estimate the average evolution over time of the CD4 counts by considering the effects of AGE, SMOKE (smoking status measured by packs of cigarettes per day), DRUG (yes = 1; no = 0), SEXP (number of sex partners), DEPRESSION (measured by the CESD scale) and YEAR (time since seroconversion). To examine whether there are any interaction effects between the covariates, we included all the two-factor interactions in our model.

Figure 5.1 shows the subject-specific evolution over time of the CD4 cell counts with and without drug use respectively. Figure 5.2 shows the evolution of the square root of the cell counts with and without drug use respectively. The cell counts are right-skewed, so a count model is not appropriate. Therefore, we will work with the square root of the counts. Tables 5.4 to 5.6 summarize the analysis for the AR(1), EQC, and lag working correlations. The GEE indicates that SMOKE, DRUG, SEXP, AGE.SEXP, SMOKE.DRUG, SMOKE.SEXP, and DRUG.SEXP are significant. Under EQC, AGE.SMOKE and AGE.DRUG are also significant. The EL selects SMOKE, DRUG, SEXP, and DRUG.SEX. Under EQC and lag AGE.SEXP is also significant. The GEE approach is sensitive to the choice of correlation structure. In this real data set, the true correlation structure is unknown, so the lag correlation approach is appropriate since it can accommodate all three correlation structures. The Shapiro–Wilk test shows that the CD4 cell counts are not normally distributed. The GEE-based method is therefore not appropriate. We therefore, conclude that the EL is a better choice.

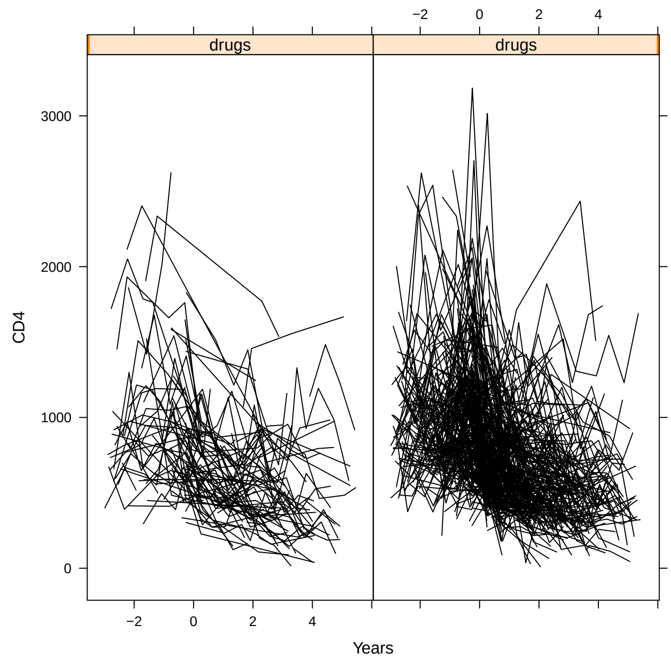

Figure 5.1: Evolution of CD4 cell count measurements with and without drug use.

Variable	Method	
	GEE	EL
INTERCEPT	25.37 (25.25, 25.49)	25.37 (24.97, 25.77)
AGE	**-0.001 (-0.014, 0.012)**	**-0.001 (-0.060, 0.060)**
SMOKE	0.938 (0.864, 1.012)	0.938 (0.669, 1.211)
DRUG	0.716 (0.597, 0.834)	0.716 (0.316, 1.115)
SEXP	0.390 (0.365, 0.414)	0.390 (0.306, 0.471)
AGE*SMOKE	**-0.001 (-0.007, 0.004)**	**-0.001 (-0.032, 0.031)**
AGE*DRUG	**0.001 (-0.013, 0.013)**	**0.001 (-0.069, 0.071)**
AGE*SEXP	0.008 (0.006, 0.009)	**0.008 (-0.003, 0.016)**
SMOKE*DRUG	-0.242 (-0.315, -0.169)	**-0.242 (-0.500, 0.023)**
SMOKE*SEXP	0.041 (0.033, 0.049)	**0.041 (-0.005, 0.089)**
DRUG*SEXP	-0.270 (-0.295, -0.245)	-0.270 (-0.358, -0.183)

Table 5.4: Estimated coefficients for CD4 data set using AR(1) working correlation.

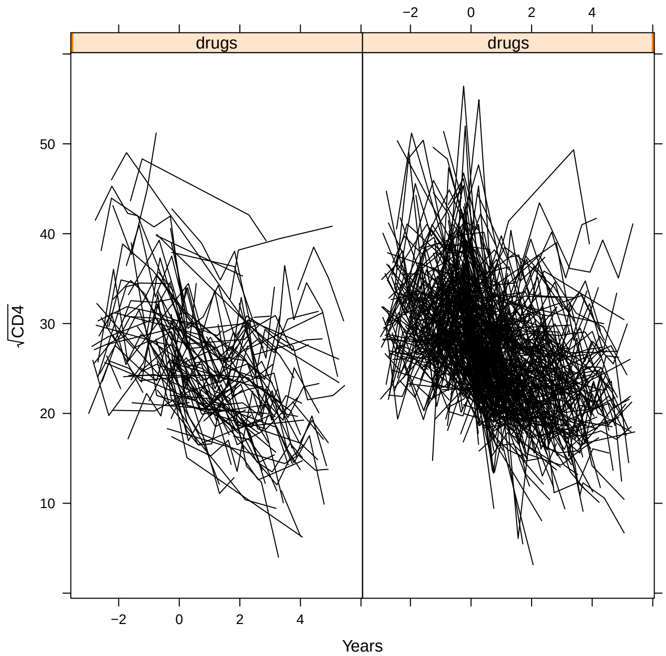

Figure 5.2: Evolution of square root of CD4 cell count measurements with and without drug use.

Variable	Method	
	GEE	EL
INTERCEPT	25.10 (24.96, 25.25)	25.10 (24.70, 25.50)
AGE	**-0.023 (-0.037, 0.008)**	**-0.023 (-0.085, 0.043)**
SMOKE	1.241 (1.161, 1.322)	1.241 (0.972, 1.515)
DRUG	1.132 (1.006, 1.257)	1.132 (0.732, 1.532)
SEXP	0.545 (0.521, 0.569)	0.545 (0.459, 0.633)
AGE*SMOKE	-0.011 (-0.016, -0.006)	**-0.011 (-0.044, 0.023)**
AGE*DRUG	0.036 (0.022, 0.050)	**0.036 (-0.031, 0.106)**
AGE*SEXP	0.017 (0.015, 0.018)	0.017 (0.007, 0.027)
SMOKE*DRUG	-0.398 (-0.477, -0.319)	**-0.398 (-0.650, 0.131)**
SMOKE*SEXP	0.038 (0.030, 0.045)	**0.038 (-0.010, 0.091)**
DRUG*SEXP	-0.184 (-0.209, -0.159)	-0.184 (-0.274, -0.091)

Table 5.5: Estimated coefficients for CD4 data set using EQC working correlation.

5.2.1 Penalized Variable Selection for CD4 Cell Counts of HIV Seroconverters

Since there are many regression parameters, we perform penalized variable selection for simultaneous estimation and variable selection. Table 5.7 gives the estimates

Variable	Method	
	GEE	EL
INTERCEPT	25.35 (25.22, 25.46)	25.35 (25.34, 25.36)
AGE	**-0.001 (-0.015, 0.012)**	**-0.001 (-0.061, 0.059)**
SMOKE	0.942 (0.867, 1.016)	0.942 (0.673, 1.215)
DRUG	0.727 (0.608, 0.845)	0.727 (0.327, 1.127)
SEXP	0.389 (0.364, 0.414)	0.389 (0.305, 0.470)
AGE*SMOKE	**-0.002 (-0.007, 0.003)**	**-0.002 (-0.032, 0.030)**
AGE*DRUG	**0.002 (-0.011, 0.015)**	**0.002 (-0.067, 0.072)**
AGE*SEXP	0.007 (0.005, 0.009)	0.007 (-0.003, 0.016)
SMOKE*DRUG	-0.233 (-0.306,-0.160)	**-0.233 (-0.490, 0.032)**
SMOKE*SEXP	0.042 (0.035, 0.050)	**0.042 (-0.004, 0.090)**
DRUG*SEXP	-0.268 (-0.356,-0.182)	-0.268 (-0.356,-0.182)

Table 5.6: Estimated coefficients for CD4 data set using lag working correlation.

of the regression coefficients. The PGEE indicates that SMOKE, DRUG, SEXP, SMOKE.DRUG, SMOKE.SEXP, and DRUG.SEXP are significant. The PEL indicates that SMOKE, DRUG, SEXP, and DRUG.SEX are significant. Under EQC SMOKE.DRUG and SMOKE.SEXP are also selected, whereas under lag SMOKE.DRUG is selected. PEL with AR(1) selected the simplest model.

Variable	AR(1)		EQC		Lag	
	PGEE	PEL	PGEE	PEL	PGEE	PEL
INTERCEPT	25.37	25.30	25.11	25.01	25.34	25.37
AGE	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
SMOKE	0.936	0.751	1.149	1.123	0.941	0.854
DRUG	0.715	0.685	1.101	1.123	0.729	0.673
SEXP	0.390	0.351	0.532	0.504	0.389	0.409
AGE*SMOKE	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
AGE*DRUG	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
AGE*SEXP	**0.000**	**0.000**	**0.000**	0.016	**0.000**	**0.000**
SMOKE*DRUG	-0.241	**0.000**	-0.292	-0.281	-0.232	-0.120
SMOKE*SEXP	0.041	**0.000**	0.034	0.020	0.042	**0.000**
DRUG*SEXP	-0.269	-0.178	-0.165	-0.110	-0.268	-0.257

Table 5.7: Penalized variable selection for CD4 cell count data under different working correlation structures.

Chapter 6

Summary and Future Work

In this chapter, we summarize our contributions, and discuss future work.

6.1 Summary

Longitudinal data modelling through GEEs assumes a working model for the within-subject correlation of the responses. When the working correlation is incorrectly specified, the GEE estimator is not necessarily consistent and may lose substantial estimation efficiency. To improve the efficiency, we can use a stationary lag correlation structure instead of the working correlation matrix. We also noticed that, to avoid losing efficiency, the first two moments of longitudinal responses needs to be correctly modelled. Any misspecification can cause estimates based on marginal models to be inefficient and misleading conclusions. Another problem with statistical inference such as confidence region construction and hypothesis testing are based on asymptotic normality, which may not hold since the finite-sample distribution may not be symmetric.

Taking these issues into account, we have proposed an EL-based longitudinal modelling based on a data-driven likelihood ratio approach sharing many of the properties of the parametric likelihood. We do not need to specify the complete parametric distribution to perform the inference. We can therefore use likelihood methods without assuming that the data come from a known family of distributions. We defined the subject-wise profile EL based on a set of GEEs. The estimation and confidence region construction using the EL approach proposed, which has advantages over other methods such as those based on normal approximations. A major advantage of EL is that involves no prior assumptions about the shape of an EL-based confidence region, which is determined automatically by the data. The construction of the confidence region based on the EL method does not involve any variance estimation.

We derived the asymptotic properties of the parameter estimates and developed an algorithm. Our performance analysis showed that our method for longitudinal count and continuous responses is comparable to the GEE when the model assumptions are satisfied. For instance, when the working correlation is correctly specified, the coverage probabilities of the intervals based on the EL, EEL, and AEL are similar to those of the GEE. CIs based on the regular EL have slight undercoverage compared with those of the GEE; the coverage probabilities are substantially improved with the EEL and AEL. Moreover, these methods are consistently more accurate than the regular EL. When the working correlation is misspecified, the coverage probabilities of the intervals based on the EL, EEL, and AEL are shown to be equally efficient to the GEE estimator with stationary lag correlation structure. Also the results shows that when the working correlation is misspecified, the GEE estimator with stationary lag correlation structure, EL, EEL, and AEL outperform the GEE with incorrect working correlation structure. When the model is misspecified such as marginal variance our method outperforms the GEE. This result shows that EL methods are robust to model misspecification. Moreover, the EL-based CI has a data-driven shape, whereas the GEE-based CI, based on the normal approximation, is symmetric.

We then extended PEL variable selection to high-dimensional longitudinal data with many covariates. We proposed simultaneous estimation and variable selection based on the subject-wise profile EL. This approach is possible with a proper choice of the tuning parameters. Under some regularity conditions, we proved that the PAEL estimators possess the oracle property. We also discussed the asymptotic properties of our method. Our algorithm produced accurate estimates of the regression parameters. Simulation studies showed that the PEL and PAEL for correlated count and continuous data are comparable with the PGEE when the model assumptions are correct and are superior when the model is misspecified. Moreover, when there are time-dependent covariates, the PGEE is not guaranteed to be consistent. Our results show that when there are nonstationary covariates our approaches can provide substantial efficiency gains over the PGEE. We applied our method to two real-world examples.

6.2 Future Work

Longitudinal and survival data are often associated in clinical trials and other medical and reliability studies. For example, in AIDS clinical trials, we repeatedly measure the number of CD4 cells per cubic ml of blood over time for each subject, and we may also be interested in the time to an event, such as death or disease progression (survival). In such situations, the longitudinal model may be the primary focus. Alternatively, the focus may be the survival data, especially when the time-dependent covariates are missing at failure times or there are measurement errors. For example, in HIV studies, the CD4 cell count per cubic ml of blood, which is measured repeatedly on the same individual throughout the study period, may be difficult to measure accurately, possibly because of machine imprecision. If we treat these mis-measured values as true values, we may draw incorrect conclusions. In particular, in regression models, if the covariates are measured with errors but considered accurate, the inference will be misleading, and a significant covariate may be found to be nonsignificant. Hence, measurement errors in the covariates must be taken into account. Two models are often assumed to be linked through shared parameters or shared unobserved variables. Our primary interest is in the survival model with measurement errors in the time-dependent covariates. The unobserved true values of the time-dependent covariates are the responses of the longitudinal model, so the two models share the same unobserved variables.

There are several methods for analyzing such data separately, including LMMs for longitudinal data and Weibull or semiparametric (Cox) proportional hazard (PH) models for time-to-event data. However, separate analyses may produce inefficient and biased results when the longitudinal variable is correlated with the survival event (time to event). In such situations, we need joint models (Tsiatis, DeGruttola and Wulfsohn [1995]; Wulfsohn and Tsiatis [1997]). The two common approaches are naive two-stage methods and likelihood methods.

The naive method uses one model to estimate the shared parameters or variables and then performs inference based on the other model using the estimated shared parameters or variables as if they were observed data. To apply this method to our situation, in the first step, we would estimate the true values of the time-dependent covariates by fitting the longitudinal model. In the second step, we would replace the unobserved variables by their estimated values and then perform inference based on the survival model.

The joint modelling of longitudinal and survival processes is based on the joint likelihood of all the longitudinal and survival data. The classical LMM can be used to model the time-dependent covariates. Maximum likelihood estimates of all the parameters can be obtained simultaneously by maximizing the joint likelihood. The maximum likelihood estimates are consistent, asymptotically efficient, and normal under the usual regularity conditions. We have discussed the disadvantages of using LMMs in a longitudinal context: they are suitable only for continuous responses. Discrete responses do not necessarily follow normal distributions. Moreover, the joint likelihood for longitudinal and survival data typically involves a high-dimensional and intractable integral. The process is computationally intensive and there may be convergence problems. Rizopoulos, Verbeke and Lesaffre [2009] use Laplace approximations for joint models, which can be especially useful for high-dimensional random effects. For maximization, the EM algorithm has been used. However, a serious drawback of the EM algorithm is its linear convergence rate, which results in slow convergence, especially close to the maximum. To overcome these difficulties, we propose an EL-based two-stage joint modelling of longitudinal and survival data.

In the above approaches, a well-defined parametric model is crucial. However, the parametric model is often not well defined, limiting the application of these methods. Another problem with parametric likelihood inference is the risk of model misspecification. Such misspecification can cause likelihood-based estimates to be inefficient. The confidence regions and hypothesis tests are based on asymptotic normality. The resulting CIs are symmetrical about the estimates, which may not be accurate since the finite-sample distribution may not be symmetric. This encourages us to investigate nonparametric likelihood. Nadarajah [2011] proposed the PEL variable selection for the Cox PH model, and we extended this approach to survival analysis.

6.2.1 Empirical-Likelihood-Based Two-Stage Joint Modelling

The observed covariate value for individual i at time t_{ij} is denoted $z_{ij} = z_i(t_{ij})$ ($i = 1, 2, ..., k,\ \ j = 1, 2, ..., m_i$). It is measured with measurement error, and the corresponding unobserved true value of the covariate is denoted z^*_{ij}. We consider the following Cox model with time-dependent and time-independent covariates:

$$h_i(t) = h_0(t)\exp(\boldsymbol{z}^*_i(t)\beta_1 + \boldsymbol{X}_i\boldsymbol{\beta}_2), \quad i = 1, 2, ..., k \tag{6.1}$$

where $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2)$. For the inference we must know the value of the time-dependent covariate $\boldsymbol{z}_i(t)$ at every event time t_i for all the individuals. Measurement errors in covariates are a form of missing data and are common in practice. The population parameters $\boldsymbol{\beta}$ in the survival model (6.1) are our main interest, and a longitudinal model is needed to address the measurement errors.

- In the first step, we estimate the true values of the covariates based on the EL approach discussed in Chapter 2 by minimizing the profile empirical log-likelihood ratio function of (2.6) with respect to $\boldsymbol{\beta}$, ignoring the survival model. We denote the predicted true values of the covariates by $\hat{\boldsymbol{z}}^*_i$.

- In the second step, we extend the EL-based Cox PH model (6.1) to perform inference. The profile empirical log-likelihood ratio estimator of $\boldsymbol{\beta}$ for the Cox PH model is the minimizer with respect to $\boldsymbol{\beta}$.

We will investigate the asymptotic properties of this approach. We will also investigate the use of PEL two-stage variable selection for longitudinal and survival data. The two-stage method has two limitations. First, it models each process separately, which may lead to biased estimates. For example, the longitudinal covariate data may be truncated by the event, so estimation based only on the observed covariate data may produce biased results. The bias may depend on the strength of the association between the longitudinal and survival processes or on the magnitude of the measurement errors in the covariates. Second, the information from the longitudinal process is not linked with that from the survival process to produce more efficient estimates. We plan to investigate these limitations.

Bibliography

Akaike, H. [1973]. Information theory and an extension of the maximum likelihood principle, *in* B. N. Petrove and F. Csaki (eds), *Second Symposium of Information Theory*, Akademiai Kiado, Budapest, pp. 267–282.

Akaike, H. [1974]. A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19**: 716–723.

Antoniadis, A. and Fan, J. [2001]. Regularization of wavelets approximations, *Journal of the American Statistical Association* **96**: 939–967.

Berk, R., Brown, L., Buja, A., Zhang, K. and Zhao, L. [2013]. Valid post-selection inference, *The Annals of Statistics* **41**(2): 802–837.

Breslow, E. N. and Clayton, G. D. [1993]. Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* **88**(421): 9–25.

Cantoni, E., Flemming, J. M. and Ronchetti, E. [2005]. Variable selection for marginal longitudinal generalized linear models, *Biometrics* **61**: 507–514.

Chen, J., Sitter, R. R. and Wu, C. [2002]. Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys, *Biometrika* **89**: 230–237.

Chen, J., Variyath, A. M. and Abraham, B. [2008]. Adjusted empirical likelihood and its properties, *Journal of Computational Graphics and statistics* **17**: 426–443.

Chen, S. X. [1994]. Empirical likelihood confidence intervals for linear regression coefficients, *Journal of Multivariate Analysis* **49**: 24–40.

Chen, S. X., Peng, L. and Qin, Y. L. [2009]. Effects of data dimension on empirical likelihood, *Biometrika* **96**: 711–722.

Chen, S. X. and Van Keilegom, I. [2009]. A review on empirical likelihood methods for regression (with discussion), *Test* **18**: 415–447.

Corcoran, S. A., Davison, A. C. and Spady, R. H. [1995]. Reliable inference from empirical likelihood, *Technical report*, Nuffield College, University of Oxford.

Craven, P. and Wahba, G. [1979]. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation, *Numerische Mathematika* **31**: 377–403.

Crowder, M. [1995]. On use of a working correlation matrix in using generalized linear models for repeated measures, *Biometrika* **82**: 407–410.

Diggle, P., Heagerty, P., Liang, K.-Y. and Zeger, S. [2002]. *The Analysis of Longitudinal Data, 2nd edition.*, Oxford Statistical Science, 2 edn, Oxford University Press, Oxford UK.

Dziak, J. J., Li, R. and Qu, A. [2009]. An overview on quadratic inference function approaches for longitudinal data, *in* J. Fan, J. S. Liu and X. Lin (eds), *New developments in biostatistics and bioinformatics*, Vol. 1 of *Frontiers of Statistics*, World scientific, pp. 49–72. Chapter 3.

Emond, M., Ritz, J. and Oakes, D. [1997]. Bias in gee estimates from misspecified models for longitudinal data, *Communications in Statistics* **26**: 15–32.

Fan, J. [1997]. Comments on “wavelets in statistics: A review” by a. antoniadis, *Journal of the Italian Statistical Association* **6**: 131–138.

Fan, J. and Li, R. [2001]. Variable selection via non concave penalized likelihood and its oracle properties, *Journal of the American Statistical Association* **96**: 1348–1360.

Fan, J. and Li, R. [2004]. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis, *Journal of the American Statistical Association* **99**: 710–723.

Fielding, A. H. and Bell, J. F. [2002]. A review of methods for the assessment of prediction erros in conservation presence/absence models, *Environmental Conservation* **24**: 38–49.

Fitzmaurice, G. [1995]. A caveat concerning independence estimating equations with multiple multivariate binary data, *Biometrics* **51**: 309–317.

Hall, P. and La Scala, B. [1990]. Methodology and algorithm of empirical likelihood, *International Statistical Review* **58**: 109–127.

Hjort, N. L., McKeague, I. and Van Keilegom, I. [2009]. Extending the scope of empirical likelihood, *Annals of Statistics* **37**: 1079–1111.

Hu, F.-C. [1993]. *A statistical methodology for analyzing the causal health effect of a timedependent exposure from longitudinal data*, Dissertation, Harvard School of Public Health, Department of Biostatistics, Boston, MA. Unpublished Sc.D.

Liang, K. Y. and Zeger, S. L. [1986]. Longitudinal data analysis using generalized linear models, *Biometrika* **73**: 13–22.

Lv, J. and Fan, Y. [2009]. A unified approach to model selection and sparse recovery using regularized least squares, *The Annals of Statistics* **37**: 3498–3528.

McCullagh, P. and Nelder, J. [1989]. *Generalized Linear Models*, Chapman and Hall, second edn, CRC press.

McKenzie, E. [1988]. Some arma models for dependent sequences of poisson counts, *Advances in Applied Probability* **20**: 822–835.

Nadarajah, T. [2011]. *Penalized empirical likelihood based variable selection*, Master's thesis, Memorial University of Newfoundland, Canada.

Nadarajah, T., Variyath, A. M. and Loredo-Osti, J. C. [2016]. Penalized generalized quasi-likelihood based variable selection for longitudinal data, *ISS-2015 Proceedings Volume on Advances in Parametric and Semiparametric Analysis of Multivariate, Time Series, Spatial-temporal, and Familial-longitudinal Data*, Springer Lecture Notes in Statistics.

Owen, A. B. [1988]. Empirical likelihood ratio confidence interval for a single functional, *Biometrika* **75**: 237–249.

Owen, A. B. [1990]. Empirical likelihood confidence regions, *The Annals of Statistics* **18**: 90–120.

Owen, A. B. [1991]. Empirical likelihood for linear models, *The Annals of Statistics* **19**: 1725–1747.

Owen, A. B. [2001]. *Empirical Likelihood*, Chapman and Hall, 2 edn, CRC press.

Pan, W. [2001]. Akaik's information criterion in generalized estimating equations, *Biometrics* **57**: 120–125.

Pan, W., Louis, T. and Connett, J. [2000]. A note on marginal linear regression with correlated response data, *American Statistician* **54**: 191–195.

Pepe, M. and Anderson, G. [1994]. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data, *Communications in Statistics, Part B-Simulation and Computation* **23**: 939–951.

Prentice, R. L. and Zhao, L. P. [1991]. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses, *Biometrics* **47**: 825–839.

Qin, J. and Lawless, J. [1994]. Empirical likelihood and general estimating equations, *Annals of Statistics* **22**: 300–325.

Qu, A., Lindsay, B. G. and Li, B. [2000]. Improving estimating equations using quadratic inference functions, *Biometrika* **87**: 823–836.

Rizopoulos, D., Verbeke, G. and Lesaffre, E. [2009]. Fully exponential laplace approximations for the joint modelling of survival and longitudinal data, *Journal of the Royal Statistical Society Series B* **71**(3): 637–654.

Schwarz, G. [1978]. Estimating the dimension of a model, *The Annals of Statistics* **6**: 461–464.

Stone, M. [1974]. Cross-validatory choice and assessment of statistical prediction, *Journal of the Royal Statistical Society* **36**(2): 111–147.

Sutradhar, B. C. [2003]. An overview on regression models for discrete longitudinal responses, *Statistical Science* **18**: 377–393.

Sutradhar, B. C. [2011]. *Dynamic Mixed Models for Familial Longitudinal Data*, Springer Series in Statistics, Springer.

Sutradhar, B. C. and Das, K. [1999]. On the efficiency of regression estimators in generalized linear models for longitudinal data, *Biometrika* **86**: 459–465.

Sutradhar, B. C. and Kovacevic, M. [2000]. Analysing ordinal longitudinal survey data: Generalized estimating equations approach, *Biometrika* **87**: 837–848.

Tang, C. Y. and Leng, C. [2010]. Penalized high dimensional empirical likelihood., *Biometrika* **97**: 905–920.

Tang, C. Y. and Leng, C. [2012]. Penalized empirical likelihood and growing dimensional general estimating equations, *Biometrika* **99**(3): 703–716.

Tibshirani, R. [1996]. Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society* **58**(1): 267–288.

Tsao, M. [2004]. Bounds on coverage probabilities of the empirical likelihood ratio confidencedence regions, *The Annals of Statistics* **32**: 1215–1221.

Tsao, M. [2013]. Extending the empirical likelihood by domain expansion, *Canadian Journal of Statistics* **41**: 257–274.

Tsao, M. and Wu, F. [2013]. Empirical likelihood on the full parameter space, *Revised for the Annals of Statistics* **41**(4): 2176–2196.

Tsiatis, A., DeGruttola, V. and Wulfsohn, M. [1995]. Modeling the relationship of survival to longitudinal data measured with error: applications to survival and cd4 counts in patients with aids., *Journal of the American Statistical Association* **90**: 27–37.

Variyath, A. M. [2006]. *Variable selection in generalized linear models by empirical likelihood*, PhD thesis, University of Waterloo, Canada.

Variyath, A. M., Chen, J. and Abraham, B. [2010]. Empirical likelihood based variable selection, *Journal of Statistical Planning and Inference* **140**: 971–981.

Wang, H. and Leng, C. [2007]. Unified lasso estimation via least squares approximation, *Journal of American Statistical Association* **102**: 1039–1048.

Wang, L., Li, H. and Huang, J. [2008]. Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements, *Journal of the American Statistical Association* **103**: 1556–1569.

Wang, L. and Qu, A. [2009]. Consistent model selection and data-driven smooth tests for longitudinal data in the estimating equations approach, *Journal of the Royal Statistical Society, Series* **71**(1): 177–190.

Wang, L., Zhou, J. and Qu, A. [2012]. Penalized generalized estimating equations for high-dimensional longitudinal data analysis, *Biometrics* **68**: 353–360.

Wang, S., Qian, L. and Carroll, J. [2010]. Generalized empirical likelihood methods for analyzing longitudinal data, *Biometrics* **97**(1): 79–93.

Wedderburn, R. [1974]. Quasi-likelihood functions, generalized linear models, and the gauss-newton method, *Biometrika* **61**(3): 439–447.

Wilks, S. S. [1938]. The large-sample distribution of the likelihood ratio for testing composite hypotheses, *The Annals of Mathematical Statistics* **9**: 60–62.

Wulfsohn, M. and Tsiatis, A. [1997]. A joint model for survival and longitudinal data measured with error, *Biometrics* **53**(1): 330–339.

Xiao, N., Zhang, D. and Zhang, H. H. [2009]. Variable selection for semiparametric mixed models in longitudinal studies, *Biometrics* **66**: 79–88.

Xu, P., W.-P., Wang, Y. and Zhu, L. [2010]. A gee based shrinkage estimation for the generalized linear model in longitudinal data analysis, *Technical report*, Department of Mathematics, Hong Kong Baptist University, Hong Kong.

Xue, L. G. and Zhu, L. X. [2007]. Empirical likelihood semiparametric regression analysis for longitudinal data, *Biometrika* **94**: 921–937.

Xue, L., Qu, A. and Zhou, J. [2010]. Consistent model selection for marginal generalized additive model for correlated data, *Journal of the American Statistical Association* **105**: 1518–1530.

You, J., Chen, G. and Zhou, Y. [2006]. Block empirical likelihood for longitudinal partially linear regression models, *Canadian Journal of Statistics* **34**: 79–96.

Zeger, L. and Diggle, J. [1994]. Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters, *Biometrics* **50**(3): 689–699.

Zhang, H. H. and Lu, W. [2007]. Adaptive lasso for cox's proportional hazards model, *Biometrika* **94**: 691–703.

Zou, H. [2006]. The adaptive lasso and its oracle properties, *Journal of the American Statistical Association* **101**(476): 1418–1429.