9-2013

Shrinkage empirical likelihood estimator in longitudinal analysis with time-dependent covariates: Application to modeling the health of Filipino children

Denis H. Y. LEUNG
Singapore Management University, denisleung@smu.edu.sg

Dylan SMALL
University of Pennsylvania

Jing QIN
National Institutes of Health

Min ZHU
Queensland University of Technology

Published in Biometrics, 2013, 69, (3), 624-632.
http://doi.org/10.1111/biom.12039

Shrinkage empirical likelihood estimator in longitudinal analysis with time-dependent covariates– application to modeling the health of Filipino children

Denis Heng-Yan Leung[1], Dylan Small[2], Jing Qin[3] and Min Zhu[4]

[1]School of Economics, Singapore Management University, Singapore

[2]Department of Statistics, The Wharton School, University of Pennsylvania, USA

[3]National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA

[4] School of Economics and Finance, Queensland University of Technology, Australia

SUMMARY: The method of generalized estimating equations (GEE) is a popular tool for analysing longitudinal (panel) data. Often, the covariates collected are time-dependent in nature, *e.g.*, age, relapse status, monthly income. When using GEE to analyse longitudinal data with time-dependent covariates, crucial assumptions about the covariates are necessary for valid inferences to be drawn. When those assumptions do not hold or cannot be verified, Pepe and Anderson (1994) advocated using an independence working correlation assumption in the GEE model as a robust approach. However, using GEE with the independence correlation assumption may lead to significant efficiency loss (Fitzmaurice, 1995). In this paper, we propose a method that extracts additional information from the estimating equations that are excluded by the independence assumption. The method always includes the estimating equations under the independence assumption and the contribution from the remaining estimating equations is weighted according to the likelihood of each equation being a consistent estimating equation and the information it carries. We apply the method to a longitudinal study of the health of a group of Filipino children.

KEY WORDS: Estimating functions; empirical likelihood; generalized estimating equations; longitudinal data.

1. Introduction

A popular method for analysing longitudinal data is the generalized estimation equation (GEE, Liang and Zeger, 1986). A GEE model is defined by marginal mean and intra-observation correlation structures. When the covariates are time-invariant, a well known robustness property of the GEE is that estimates of the mean parameters remain consistent even if the (intra-observation) correlation structure is misspecified. However, it is common that in a longitudinal study some of the covariates may vary over time. Under such a situation, Pepe and Anderson (1994) found that the robustness property of the GEE no longer holds because some of the estimating functions generated by the longitudinal data are no longer unbiased under an arbitrary correlation structure. They showed that only the independence correlation structure guarantees consistency in such situations. The use of an independence assumption may lead to substantial efficiency loss (*e.g.*, Fitzmaurice, 1995).

The problem noted by Fitzmaurice (1995) is due to the fact that, under the independence correlation structure, only a subset of all the unbiased estimating functions is used. To solve this problem, Pan and Connett (2002) suggested using correlation structures other than the independence correlation structure. They proposed choosing, among a number of commonly used correlation structures, the one that minimizes the predictive mean squared error of the model. Their method requires estimating the predictive mean squared error using resampling methods. Lai and Small (2007) identified three types of time-dependent covariates and classified the estimating functions according to each of these types of time-dependent covariates. For data analysis, they suggested a hypothesis testing procedure to decide between the different types of time-dependent covariates and then used a generalized method of moments (GMM, Hansen, 1982) to combine the estimating functions.

The problem of selecting the appropriate estimating functions for parameter estimation falls under a large body of works called the moment selection problem. That problem can be broadly classified into one of two types. The first type begins with a pool of unbiased estimating functions and the goal is to select those that are the most informative; the second type allows some of candidate estimating functions to be biased and the goal is to identify those that are unbiased and most informative. Here, an informative estimating function is an unbiased estimating function whose value is sensitive to the values of the parameters and hence its inclusion will help to identify the true values of the parameters. On the other hand, an un-informative estimating function does not respond to changes in the values of the parameters and it generates noise rather than signal about the parameters. In our context, the estimating functions are generated from solving the GEE; in other contexts, estimating functions can also be generated as a means to parameter estimation, such as the case in instrumental variable analyses (*e.g.*, Angrist and Krueger, 1992).

Within the first type of moment selection problems, a common strategy is to select estimating functions by minimizing some criterion; see, for example, Kolaczyk (1995), Donald and Newey (2001), Okui (2009) and Wang and Qu (2009). For the second type, a common approach is to use a test to identify from a pool of candidate estimating functions those that are likely to be unbiased, followed by estimation using the supposedly unbiased estimating functions. Previous works include Eichenbaum, Hansen, and Singleton (1988), Gallant, Hsieh, and Tauchen (1997), Andrews (1999) and Lai, Small, and Liu (2008).

In this paper, we consider a different approach for analysing longitudinal data with time varying covariates. Our method separates the estimating functions into two groups. One group is always used. In the context of this paper, this group corresponds to all the estimating functions under the independence correlation assumption. In other situations, this group may also include estimating functions that are known to be unbiased *a priori*. The second group are those that may improve the asymptotic efficiency of the parameter estimates. This group includes all other estimating functions, some of which may be informative, but some may be un-informative or may even be biased. This group must be handled delicately as inclusion of un-informative or biased estimating functions may hurt performance (see, *e.g.*, Newey and Smith, 2004). Unlike previous selection methods for identifying the un-informative and biased estimating functions, we create shrinkage parameters that appropriately shrink the estimating functions in this group, according to the likelihood of each being a biased, un-informative or informative estimating equation. In this way, our method solves the high dimensional problem of exhaustively finding the "best" subset of estimating functions from all the candidate estimating functions. Our method is related to the shrinkage estimator of Okui (2009) but it differs from Okui (2009) in two ways. First, Okui (2009) assumed the pool of estimating functions are all unbiased. His shrinkage estimator is a weighted sum of two estimators, one obtained using all the estimating functions that are known to be informative and another obtained from the remaining estimating functions. Second, his estimator uses the same amount of shrinkage for all the remaining estimating functions. In this paper, we do not assume the pool of estimating functions to be all unbiased and we allow different shrinkage parameters for different estimating functions.

We apply our method to use anthropometric measures (body mass index) to predict future morbidity in children in a rural area in the Philippines using a longitudinal data set collected by the International Food Policy Research Institute (IFPRI). It is estimated that about two-thirds of child deaths around the world are directly or indirectly associated with nutritional deficiencies (Caballero, 2002). Nutrition related mortality and morbidity is especially prevalent in developing countries (United Nations Children's Fund, 2007). Anthropometry is an inexpensive and non-invasive measure of nutrition status of an individual or a population. Anthropometric examination is widely used for identifying children at risk of illness or death and for prioritizing areas to be targeted for government programs (Martorell and Habicht, 1986, Zerfas et al., 1986, World Health Organization, 1995). Anthropometric measures are time-dependent covariates and the outcome from a previous period may affect future values of the covariates, which in turn may affect future outcomes. The case in point is a sick child may not eat well which affects his/her nutrition status, which puts the child at a higher risk for future morbidity (Scrimshaw and SanGiovanni, 1997). Because anthropometric measures are time-dependent covariates that may have feedbacks with the outcome, some of the estimating equations generated by longitudinal data may not be unbiased. We give weights to the estimating equations according to the likelihood of each estimating equation being unbiased and the information it carries to obtain efficient estimates of the effect of anthropometric measures at a given time point on morbidity four months later.

The rest of this paper is organized as follows. Section 2 considers longitudinal analysis with time-dependent covariates. In Section 3, we describe our shrinkage method. We used simulations to study the method's finite sample behavior and the results are given in Section 4. In Section 5, we illustrate our method using the IFPRI dataset.

2. Estimation with time-dependent covariates

Consider a longitudinal study where there are n observations each of which is measured at T time points. In practice, it is possible that not all observations are measured at all time points. But for ease of illustration, we assume herein observations are measured at all T time points. Let $\mathbf{y}_i = (y_{i1}, ..., y_{iT})^\tau$ denote the outcome for the i-th observation and let $\mathbf{X}_i = (\mathbf{x}_{i1}, ..., , \mathbf{x}_{iT})^\tau$ be an associated pT matrix of covariate values, where $\mathbf{x}_{it} = (x_{it}^1, ..., x_{it}^p)^\tau$ is a p-vector of covariates observed at time t. Assume for the moment that the covariates are time-independent, so $\mathbf{x}_{i1} = ... = \mathbf{x}_{iT}$. Let the marginal mean outcome at the t-th time point for the i-th observation be

$$E(y_{it}|\mathbf{x}_{it}) \equiv \mu_{it}(\boldsymbol{\beta}) = g(\boldsymbol{\beta}^\tau \mathbf{x}_{it}),$$

where g is a known function and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\tau$ is a p-vector of unknown parameters. Under this formulation, we assume the parameters for all n observations are identical. For conciseness, we suppress the explicit association of μ_{it} with $\boldsymbol{\beta}$ if there is no risk of confusion.

Following Liang and Zeger (1986), a GEE can be used to estimate the vector of regression parameters, $\boldsymbol{\beta}$, by solving a set of p equations,

$$\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i}{\partial \beta_j} \mathbf{V}_i^{-1} \{\mathbf{y}_i - \boldsymbol{\mu}_i\} \equiv \sum_{i=1}^{n} \sum_{s=1}^{T} \sum_{t=1}^{T} \frac{\partial \mu_{is}}{\partial \beta_j} \bar{v}_i^{st} \{y_{it} - \mu_{it}\} = 0, \quad j = 1, ..., p \tag{1}$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, ..., \mu_{iT})^\tau$ and $\bar{v}_i^{st}, s, t = 1, ..., T$ is the (s,t)-th entry of $\mathbf{V}_i^{-1}$, where $\mathbf{V}_i$ is an estimate of the covariance matrix of $\mathbf{y}_i$. The matrix $\mathbf{V}_i$ can be characterized using $\phi \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$, where $\mathbf{A}_i$ is a diagonal matrix representing the variances of y_{it}, $\mathbf{R}_i(\boldsymbol{\alpha})$ is a symmetric positive definite matrix of correlations depending on a vector of unknown parameters $\boldsymbol{\alpha}$ and ϕ is a scale parameter used to model over- or under-dispersion. Apart from some limited cases, $\boldsymbol{\alpha}$ is a function of $\boldsymbol{\beta}$ and ϕ, and ϕ is a function of $\boldsymbol{\beta}$. When the covariate is time-invariant, $\mu_{is} = \mu_{it}, \forall s, t$; therefore, $\partial \mu_{is}/\partial \beta_j \{y_{it} - \mu_{it}\} = \partial \mu_{it}/\partial \beta_j \{y_{it} - \mu_{it}\}$ has expectation zero under the true value of $\boldsymbol{\beta}$, ie., unbiased. Consequently, for time-invariant covariates, consistency of $\boldsymbol{\beta}$ is not dependent on the correct specification of $\mathbf{V}_i$.

When covariates are time-dependent so that $\mathbf{x}_{is} \neq \mathbf{x}_{it}$, for some s, t, then $\mu_{is} \neq \mu_{it}$ for some s, t and consequently, $\partial \mu_{is}/\partial \beta_j \{y_{it} - \mu_{it}\}$ may not be unbiased for all s, t. Pepe and Anderson (1994) showed that a sufficient condition for the GEE analysis to remain unbiased is $\mathrm{E}(y_{it}|\mathbf{X}_i) = \mathrm{E}(y_{it}|\mathbf{x}_{it})$. This condition requires the mean outcome given the covariates at any time to be the same as that on all past, present and future covariate values. This condition is unlikely to be satisfied in many situations. For example, if the outcome is cancer incidence and the covariates are exposures of risk factors such as radiation, smoking, etc., then $\mathrm{E}(y_{it}|\mathbf{x}_{it})$ captures the association of the current exposure and cancer risk, but $\mathrm{E}(y_{it}|\mathbf{X}_i)$ includes, among others, the association between functions of exposure history and the disease.

So even though we may have a model for the outcome and current exposure, the fact that there is influence from historical exposure could render the GEE analysis biased.

Since the j-th equation in (1) can be viewed as a linear combination of $\partial\mu_{is}/\partial\beta_j\{y_{it}-\mu_{it}\}$ with coefficients $\bar{v}_i^{st}$, any potentially biased $\partial\mu_{is}/\partial\beta_j\{y_{it}-\mu_{it}\}$ may be removed by setting its corresponding coefficient to zero. Pepe and Anderson (1994) used this idea and suggested using an independence assumption for the correlation structure for $\mathbf{R}_i(\boldsymbol{\alpha})$. Their assumption implies $\bar{v}_i^{st}=0, s\neq t$ and the GEE solves a set of p equations

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial\mu_{it}}{\partial\beta_j}\bar{v}_i^{tt}\{y_{it}-\mu_{it}\}=0,\quad j=1,...,p. \tag{2}$$

A GEE with independence correlation assumption may lose efficiency because information is discarded through the removal of a large number of estimating functions, ie., $\partial\mu_{is}/\partial\beta_j\{y_{it}-\mu_{it}\}, s\neq t$. Lai and Small (2007) argued that for certain types of covariates, some of these estimating functions could be utilized. They classified covariates into one of three types: Type I if $\mathrm{E}(\partial\mu_{is}/\partial\beta_j\{y_{it}-\mu_{it}\})=0, \forall s,t$, Type II if $\mathrm{E}(\partial\mu_{is}/\partial\beta_j\{y_{it}-\mu_{it}\})=0,$ for $s\geqslant t$ and Type III if $\mathrm{E}(\partial\mu_{is}/\partial\beta_j\{y_{it}-\mu_{it}\})\neq 0,$ for some $s\geqslant t$. The conditions under which a covariate belongs to these three types are given in Lai and Small (2007). If a covariate is of Type I or II, additional unbiased estimating functions could be used for efficiency gain.

3. Empirical likelihood with moment shrinkage

The method of Lai and Small (2007) requires testing each time-dependent covariate for Type I, II vs. III. As in any statistical test, there is the possibility of false positives, especially when the number of tests (covariates) is moderately large, which may be the case in many practical situations. Therefore, their method does not guarantee improved efficiency over a GEE with an independence correlation assumption. In this paper, we consider a method that has the following attributes: (1) improved efficiency over a GEE with an independence correlation assumption for certain types of time-dependent covariates, such as those defined by Lai and Small (2007) and (2) robustness against biased estimating functions that might have been included due to a falsely classified time-dependent covariate.

We first define some notations. For a generic observation $(y_{\cdot t}, \mathbf{x}_{\cdot t})$ at time t, define $\mu_{\cdot t} \equiv g(\boldsymbol{\beta}^{\tau}\mathbf{x}_{\cdot t})$. Let $\mathbf{S}(\boldsymbol{\beta}) = \{\partial\mu_{\cdot s}/\partial\beta_j\{y_{\cdot t} - \mu_{\cdot t}\}, j = 1, ..., p, s, t = 1, ..., T\}$. Note that $\mathbf{S}(\boldsymbol{\beta})$ is a vector of pT^2 estimating functions. We divide the pT^2 estimating functions into two groups. The first group includes all estimating functions in $\mathbf{S}(\boldsymbol{\beta})$ that are known *a priori* to be unbiased. The second group consists of all other estimating functions in $\mathbf{S}(\boldsymbol{\beta})$. We call the first group the main estimating functions as all estimating functions in this group will be selected and the second group the auxiliary estimating functions. We denote the main and the auxiliary group as $\mathbf{S}^M(\boldsymbol{\beta})$ and $\mathbf{S}^A(\boldsymbol{\beta})$, respectively. In our context, $\mathbf{S}^M(\boldsymbol{\beta}) = \{\partial\mu_{\cdot t}/\partial\beta_j\{y_{\cdot t} - \mu_{\cdot t}\}, j = 1, ..., p, t = 1, ..., T\}$, ie., all the estimating functions that would be used by a GEE with independence assumption. We introduce a vector, $\boldsymbol{\gamma}$, of shrinkage parameters with the same dimension as $\mathbf{S}^A(\boldsymbol{\beta})$; each element $\boldsymbol{\gamma}$ is a real number in $[0, 1]$ that depends on the data. We let $\mathbf{S}^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta}) = \boldsymbol{\gamma}^{\tau}\mathbf{S}^A(\boldsymbol{\beta})$. Note that the dimension of $\mathbf{S}^M(\boldsymbol{\beta})$ is pT and the dimension of $\mathbf{S}^A(\boldsymbol{\beta})$ is $p(T^2 - T)$, but the dimension of $\mathbf{S}^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})$ is one. So we have reduced the dimension of $\mathbf{S}^A(\boldsymbol{\beta})$ through the use of $\boldsymbol{\gamma}$.

For a particular choice of $\boldsymbol{\gamma}$, we have estimating functions $\{\mathbf{S}^M(\boldsymbol{\beta}), \mathbf{S}^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\}$. We can use empirical likelihood (EL, Owen, 1988) to combine these estimating functions.

Let $r_1, ..., r_n$ be non-negative weights associated with the observations $\{(\mathbf{y}_i, \mathbf{X}_i)\}_{i=1}^n$. Given $\{\mathbf{S}^M(\boldsymbol{\beta}), \mathbf{S}^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\}$, an EL for the parameter $\boldsymbol{\beta}$, is

$$L^{\boldsymbol{\gamma}}(\boldsymbol{\beta}) = \max \prod_{i=1}^{n} r_i \tag{3}$$

subject to the constraints

$$0 \leqslant r_i \leqslant 1, i = 1, ..., n; \quad \sum_{i=1}^{n} r_i = 1; \quad \sum_{i=1}^{n} r_i\{\mathbf{S}_i^M(\boldsymbol{\beta}), \mathbf{S}_i^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\} = 0,$$

where $\mathbf{S}_i^M(\boldsymbol{\beta})$ is $\mathbf{S}^M(\boldsymbol{\beta})$ applied to $(\mathbf{y}_i, \mathbf{X}_i)$ and $\mathbf{S}_i^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})$ is similarly defined. By introducing Lagrange multipliers $\nu \equiv \nu(\boldsymbol{\beta})$ and $\boldsymbol{\lambda}^{\tau} \equiv \boldsymbol{\lambda}^{\tau}(\boldsymbol{\beta})$, where $\boldsymbol{\lambda}^{\tau}$ is a vector of dimension $pT + 1$, the log-EL can be written as

$$\sum_{i=1}^{n} \log r_i + \nu(1 - \sum_{i=1}^{n} r_i) - n\boldsymbol{\lambda}^{\tau} \sum_{i=1}^{n} r_i\{\mathbf{S}_i^M(\boldsymbol{\beta}), \mathbf{S}_i^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\}. \tag{4}$$

The values of $\{r_i\}_{i=1}^n$ can be profiled out by differentiating the log-EL with respect to r_i

$$\frac{1}{r_i} - \nu - n\boldsymbol{\lambda}^{\tau}\{\mathbf{S}_i^M(\boldsymbol{\beta}), \mathbf{S}_i^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\} = 0 \Rightarrow n - \nu = 0 \Rightarrow \nu = n. \tag{5}$$

Expression (5) implies the optimal values of $\{r_i\}_{i=1}^n$ are

$$r_i \;=\; \frac{1}{n}\frac{1}{1 + \boldsymbol{\lambda}^{\tau}\{\mathbf{S}_i^M(\boldsymbol{\beta}), \mathbf{S}_i^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\}}, \tag{6}$$

and the constraint $\sum_{i=1}^n r_i\{\mathbf{S}_i^M(\boldsymbol{\beta}), \mathbf{S}_i^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\} = 0$ implies $\boldsymbol{\lambda}^{\tau}$ satisfies the following equation

$$\sum_{i=1}^{n} \frac{\{\mathbf{S}_i^M(\boldsymbol{\beta}), \mathbf{S}_i^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\}}{1 + \boldsymbol{\lambda}^{\tau}\{\mathbf{S}_i^M(\boldsymbol{\beta}), \mathbf{S}_i^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\}} \;=\; 0. \tag{7}$$

Using (5) and (6), we now have a profile log-EL

$$\ell^{\boldsymbol{\gamma}}(\boldsymbol{\beta}) = -\sum_{i=1}^{n} \log\{1 + \boldsymbol{\lambda}^{\tau}\{\mathbf{S}_i^M(\boldsymbol{\beta}), \mathbf{S}_i^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\}\} - n\log(n). \tag{8}$$

Differentiating (8) with respect to $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ leads to the maximum EL estimates $(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}})$.

In practice, the set of estimating functions $\{\mathbf{S}^M(\boldsymbol{\beta}), \mathbf{S}^{A,\boldsymbol{\gamma}}(\boldsymbol{\beta})\}$ may include some estimating functions that are biased and therefore, using them may lead to a $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}} \to \boldsymbol{\beta}_{**} \neq \boldsymbol{\beta}_{*}$, the true value of $\boldsymbol{\beta}$. The purpose of the vector of shrinkage parameters, $\boldsymbol{\gamma}$, is to down weight those biased and un-informative estimating functions to arrive at a consistent and efficient estimator $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}$. We now describe how to choose $\boldsymbol{\gamma}$. Our strategy is to assign different shrinkage parameters to each estimating function in $\mathbf{S}^A$, depending on whether it is suspected to be biased or not. Of course, we do not know *a priori* which estimating function in $\mathbf{S}^A$ is biased. One option is to test each of the estimating functions in $\mathbf{S}^A$, but that would involve a large number of tests with accumulation of probability of a Type I error. As pointed out by Andrews (1999), in practice, it is only feasible to test groups of estimating functions. We follow this strategy and adopt the testing procedure of Lai and Small (2007) to test each covariate for Type I, II vs. III. In that procedure, all estimating functions associated with a Type I covariate are deemed unbiased, but for a Type II or III covariate, only a subset of the estimating functions are considered unbiased. Details of this procedure are given under *Supplementary Material*, which is available with this paper at the Biometrics website on Wiley Online Library. Without loss of generality, let the estimating functions in $\mathbf{S}^A$ be $\{S_k, k = 1, ..., p(T^2 - T)\}$ and the shrinkage parameters as $\boldsymbol{\gamma} = \{\gamma^k, k = 1, ...p(T^2 - T)\}$. Let $\mathbf{c}$ be a $p(T^2 - T)$ vector with the k-th element, c_k taking a value of 1 if S_k is deemed unbiased by the testing procedure of Lai and Small (2007) and is 0 otherwise. In fact, shrinkage only affects estimating functions with $c_k = 1$ but not those functions with $c_k = 0$. We define the shrinkage parameters as

$$\gamma^k = c_k \exp\left(-\gamma_0 \left|\sum_{i=1}^{n} S_{k,i}(\tilde{\boldsymbol{\beta}})\right|\right), \quad k = 1, ..., p(T^2 - T), \tag{9}$$

where $\tilde{\boldsymbol{\beta}}$ is the solution of (2) and γ_0 is a constant in $\mathbb{R}^+$ that depends on the data and γ_0 satisfies the following conditions: $\gamma_0/\sqrt{n} \to 0$ and $\gamma_0 \to \infty$ as $n \to \infty$. Since $\sum_{i=1}^{n} S_{k,i}(\tilde{\boldsymbol{\beta}}) \to 0$ as $n \to \infty$ for an unbiased estimating function, (9) give less weight to those estimating functions that are likely to be biased. By changing the value of γ_0, the relative weights of the estimating functions in $\mathbf{S}^A$ would change. The special of case of $\gamma_0 = 0$ implies all the estimating functions in $\mathbf{S}^A$ would be weighted equally.

Lai and Small (2007) showed that their test is asymptotically consistent in the sense that, in large samples, it correctly classifies each covariate as Type I, II vs. III. However, in finite samples, there is still a chance that the test may return a wrong classification. The shrinkage parameters help to shrink the influence of biased estimating functions associated with such erroneous results, as shall be seen in the simulation study.

The value of γ_0 must be determined to implement shrinkage. We choose γ_0 by minimizing the mean squared error of $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}$. Instead of using an analytic approximation of the mean squared error, which requires strict structural assumptions of the model, we use the leave-one-out cross-validation (Stone, 1974). Define $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)}$ as the EL estimate of $\boldsymbol{\beta}$ with the i-th observation $(\mathbf{y}_i, \mathbf{X}_i)$ removed from the dataset. Then $\boldsymbol{\gamma}$ can be chosen to minimize

$$\frac{1}{n}\sum_{i=1}^{n}\{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)} - \tilde{\boldsymbol{\beta}}\}^{\tau}\{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)} - \tilde{\boldsymbol{\beta}}\}. \tag{10}$$

Note that (10) is different from the usual leave-one-out estimator where $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}$ would have appeared instead of $\tilde{\boldsymbol{\beta}}$. However, in the current set-up, $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}$ may be a biased estimator of $\boldsymbol{\beta}$ for certain choices of $\boldsymbol{\gamma}$ whereas $\tilde{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$. We can write (10) as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)} - \hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}\}^{\tau}\{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)} - \hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}\} + \{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\beta}}\}^{\tau}\{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\beta}}\} + 2\{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)} - \hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}\}^{\tau}\{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\beta}}\} \\
= \;& \frac{1}{n}\sum_{i=1}^{n}\{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)} - \hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}\}^{\tau}\{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)} - \hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}\} + \{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\beta}}\}^{\tau}\{\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\beta}}\} + O_p\left(\frac{1}{n}\right).
\end{aligned} \tag{11}$$

Since $\tilde{\boldsymbol{\beta}}$ is a $\sqrt{n}$-consistent estimator for $\boldsymbol{\beta}$, therefore, the first term in (11) can be viewed as a variance estimate for $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}$ whereas the second term in (11) can be viewed as an estimate of the squared bias. Hence (10) is a proper estimate of the mean squared error.

To implement (10), we need to find $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)}$ for $i = 1, ..., n$. For any reasonably large sample size n, it is not practically feasible to find $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)}$ for $i = 1, ..., n$ since that requires repeating EL estimation n times, each based on a sample of $n-1$ observations. However, we can use a strategy similar to that proposed by Zhu, Ibrahim, Tang, and Zhang (2008) to obtain an approximation of $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)}$ based on $\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}$. Writing $Q^{\boldsymbol{\gamma}}_1(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \partial/\partial\boldsymbol{\beta}\ell^{\boldsymbol{\gamma}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} q^{\boldsymbol{\gamma}}_{1,i}(\boldsymbol{\beta})$ and $Q^{\boldsymbol{\gamma}}_2(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \partial/\partial\boldsymbol{\lambda}\ell^{\boldsymbol{\gamma}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} q^{\boldsymbol{\gamma}}_{2,i}(\boldsymbol{\beta}, \boldsymbol{\lambda})$, and $Q^{\boldsymbol{\gamma}}_{1,-i}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ and $Q^{\boldsymbol{\gamma}}_{2,-i}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ be their analogues with $(\mathbf{y}_i, \mathbf{X}_i)$ removed. By Taylor series expansion

$$\begin{aligned}
0 = Q^{\boldsymbol{\gamma}}_{1,-i}(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)}, \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}}_{(-i)}) \;&=\; Q^{\boldsymbol{\gamma}}_{1,-i}(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}}) + \{\frac{\partial}{\partial\boldsymbol{\beta}}Q^{\boldsymbol{\gamma}}_{1,-i}(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}})(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)} - \hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}) + \\
&\quad \frac{\partial}{\partial\boldsymbol{\lambda}}Q^{\boldsymbol{\gamma}}_{1,-i}(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}})(\hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}}_{(-i)} - \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}})\}\{1 + o_p(1)\} \\
0 = Q^{\boldsymbol{\gamma}}_{2,-i}(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)}, \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}}_{(-i)}) \;&=\; Q^{\boldsymbol{\gamma}}_{2,-i}(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}}) + \{\frac{\partial}{\partial\boldsymbol{\beta}}Q^{\boldsymbol{\gamma}}_{2,-i}(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}})(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}_{(-i)} - \hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}) + \\
&\quad \frac{\partial}{\partial\boldsymbol{\lambda}}Q^{\boldsymbol{\gamma}}_{2,-i}(\hat{\boldsymbol{\beta}}^{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}})(\hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}}_{(-i)} - \hat{\boldsymbol{\lambda}}^{\boldsymbol{\gamma}})\}\{1 + o_p(1)\},
\end{aligned}$$

which jointly imply that

$$\begin{pmatrix} \hat{\boldsymbol\beta}^{\boldsymbol\gamma}_{(-i)} - \hat{\boldsymbol\beta}^{\boldsymbol\gamma} \\ \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}_{(-i)} - \hat{\boldsymbol\lambda}^{\boldsymbol\gamma} \end{pmatrix} \approx -\begin{pmatrix} Q^{\boldsymbol\gamma}_{1,-i}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) \\ Q^{\boldsymbol\gamma}_{2,-i}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial\boldsymbol\beta} Q^{\boldsymbol\gamma}_{1,-i}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) & \frac{\partial}{\partial\boldsymbol\lambda} Q^{\boldsymbol\gamma}_{1,-i}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) \\ \frac{\partial}{\partial\boldsymbol\beta} Q^{\boldsymbol\gamma}_{2,-i}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) & \frac{\partial}{\partial\boldsymbol\lambda} Q^{\boldsymbol\gamma}_{2,-i}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) \end{pmatrix}^{-1},$$

$$\begin{pmatrix} \hat{\boldsymbol\beta}^{\boldsymbol\gamma}_{(-i)} \\ \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}_{(-i)} \end{pmatrix} \approx \begin{pmatrix} \hat{\boldsymbol\beta}^{\boldsymbol\gamma} \\ \hat{\boldsymbol\lambda}^{\boldsymbol\gamma} \end{pmatrix} - \begin{pmatrix} q^{\boldsymbol\gamma}_{1,i}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) \\ q^{\boldsymbol\gamma}_{2,i}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial\boldsymbol\beta} Q^{\boldsymbol\gamma}_{1}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) & \frac{\partial}{\partial\boldsymbol\lambda} Q^{\boldsymbol\gamma}_{1}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) \\ \frac{\partial}{\partial\boldsymbol\beta} Q^{\boldsymbol\gamma}_{2}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) & \frac{\partial}{\partial\boldsymbol\lambda} Q^{\boldsymbol\gamma}_{2}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) \end{pmatrix}^{-1} \quad (12)$$

where (12) is based on the fact that: $Q^{\boldsymbol\gamma}_1(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) = 0$, $Q^{\boldsymbol\gamma}_2(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) = 0$, and $\partial/\partial\boldsymbol\beta Q^{\boldsymbol\gamma}_1(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma}) = \partial/\partial\boldsymbol\beta Q^{\boldsymbol\gamma}_{1,-i}(\hat{\boldsymbol\beta}^{\boldsymbol\gamma}, \hat{\boldsymbol\lambda}^{\boldsymbol\gamma})\{1 + O_p(n^{-1})\}$, etc. Using (12), we can estimate $\hat{\boldsymbol\beta}^{\boldsymbol\gamma}_{(-i)}$ by a single evaluation of the EL based on all n observations.

4. Simulation Study

In this section, we report the results of a simulation study designed to examine the finite sample properties of the proposed method. We simulated data using four different models. Each of these models was chosen to highlight the strengths and weaknesses of the proposed method. One thousand simulation runs were generated under each model. Throughout the simulation study, we used two different sample sizes, $n = 100$ or 500 and we used $T = 5$ throughout. We assumed complete data at all time points within each observation. For each model, five estimators were compared:

1. GEE using an independence working correlation (GEE_{I}).
2. GMM assuming the estimating functions are Type II estimating functions, as defined in Lai and Small (2007) (GMM_{T2}).
3. GMM method of Lai and Small (2007) using a 5 % significance test for deciding between Type I, II and III estimating functions (GMM_{SEL}).
4. Combining the same set of estimating functions found in GMM_{SEL} using an empirical likelihood approach (EL_1) but without a vector of shrinkage parameters. We could interpret

EL_1 as the shrinkage estimator considered in this paper with the parameter γ_0 set to 0 (no shrinkage). The inclusion of this estimator is to show the role of the shrinkage parameters.

5. EL method proposed in this paper (EL_2) using the same set of estimating functions found in $\mathrm{GMM}_{\mathrm{SEL}}$, with the vector of shrinkage parameters $\boldsymbol{\gamma}$ chosen to minimize (10).

The four models we studied are given below. In each case, we are interested in the regression coefficients of $\mathrm{E}(y_{it}|\mathbf{x}_{it}) = \zeta_0 + \boldsymbol{\beta}^{\tau}\mathbf{x}_{it}$. Details of the derivations of the true values of the regression coefficients are given in the Supplementary Material. The first model is chosen to study the possible efficiency gains in including a Type II covariate; the last three models are chosen to evaluate the robustness of the methods.

Model 1: A Type II time-dependent covariate. The data generating process is

$$\begin{aligned} y_{it} &= \zeta_0 + \zeta_1 x_{it} + \zeta_2 x_{i,t-1} + b_i + \epsilon_{it} \\ x_{it} &= \rho x_{i,t-1} + e_{it} \end{aligned} \quad , \quad t = 1, ..., 5$$

where $b_i \sim N(0, \sigma_b^2 = 4)$, $\epsilon_{it} \sim N(0, \sigma_\epsilon^2 = 1)$, $e_{it} \sim N(0, \sigma_e^2 = 1)$ are mutually independent and $x_{i0} \sim N\left(0, \sigma_e^2/(1-\rho^2)\right)$ with $\rho = 0.5$, $\zeta_0 = 0$, $\zeta_1 = \zeta_2 = 1$. This model comes from Diggle et al. (2002). For this model, the outcome variable depends on the most recent two values of the covariate and the covariate itself depends on its previous measurement. In this model, the covariate is Type II because x_{it} is only dependent on $x_{is}, s < t$.

Model 1a: A Type III time-dependent covariate. We use the same data generating process as Model 1, except that

$$x_{it} = \rho x_{i,t-1} + \kappa b_i + e_{it}, \quad t = 1, ..., 5$$

where $\kappa = 0.05$. Since x_{it} and y_{it} share a common factor b_i, so there is a feedback effect of past outcome values on future covariate values and hence, the covariate is a Type III covariate. We could imagine this type of model to be useful for studying the effects of medication on chronic diseases, such as migrane or psychological disorder. The theory is medication would have an effect on the medical condition but medications are often given based on the medical condition (*e.g.*, Ten Have and Morabia, 2002).

Model 2: Three Type III time-dependent covariates. The data generating process is

$$\left.\begin{array}{l} y_{it} = \zeta_0 + \boldsymbol{\zeta}_1^\tau \mathbf{x}_{it} + \boldsymbol{\zeta}_2^\tau \mathbf{x}_{i,t-1} + b_i + \epsilon_{it} \\ \mathbf{x}_{it} = (x_{it}^1, x_{it}^2, x_{it}^3)^\tau; \quad x_{it}^j = \rho x_{i,t-1}^j + \kappa b_i + e_{it}^j \end{array}\right. , \quad t = 1, ..., 5$$

where $b_i \sim N(0, \sigma_b^2 = 4)$, $\epsilon_{it} \sim N(0, \sigma_\epsilon^2 = 1)$, $e_{it}^j \sim N(0, \sigma_e^2 = 1), j = 1, 2, 3$ are mutually independent and $\mathbf{x}_{i0} \sim MVN\,(0, \mathrm{diag}[(\sigma_b^2\kappa^2 + \sigma_e^2)/(1-\rho^2)]_{3\times 3})$, with $\rho = 0.5$, $\kappa = 0.05$, $\zeta_0 = 0$, $\boldsymbol{\zeta}_1 = \boldsymbol{\zeta}_2 = (1, 1, 1)^\tau$. In this model, $\mathbf{x}_{it} = (\overbrace{x_{it}^1, x_{it}^2, x_{it}^3}^{\text{Type III}})^\tau$ for the same reason as Model 1a.

Model 3: Three time-dependent covariates with two Type III and one Type II. The data generating process is

$$\left.\begin{array}{l} y_{it} = \boldsymbol{\theta}^\tau \mathbf{x}_{it} + \kappa y_{i,t-1} + \epsilon_{it} \\ \mathbf{x}_{it} = (x_{it}^1, x_{it}^2, x_{it}^3)^\tau; \quad x_{it}^j = \begin{cases} \psi y_{i,t-1} + e_{it}^j, & j = 1, 2 \\ e_{it}^j, & j = 3 \end{cases} \end{array}\right. , \quad t = 1, ..., 5$$

where $\epsilon_{it} \sim N(0, \sigma_\epsilon^2 = 1)$ and $e_{it}^j \sim N(0, \sigma_e^2 = 1), j = 1, 2, 3$ are mutually independent variables. The initial value $y_{i0} \sim (0, (\sigma_e^2 \sum_{j=1}^3 \theta_j^2 + \sigma_\epsilon^2)/\{1 - (\psi \sum_{j=1}^2 \theta_j + \kappa)^2\})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^\tau = (0.1, 0.1, 0.1)$, $\kappa = 0.5$ and $\psi = 0.6$. In this model, $\mathbf{x}_{it} = (\overbrace{x_{it}^1, x_{it}^2,}^{\text{Type III}} \overbrace{x_{it}^3}^{\text{Type II}})^\tau$ because x_{it}^j depends on $y_{i,t-2}^j$, for $j = 1, 2$ whereas $\mathbf{x}_{it}^3$ is independent of y_{is}^3. The model postulates that current outcome value may be affected by its value in the last period and also by the time-dependent covariates; in addition, some of the covariates are affected by past outcome values. An example of this type of model is found in the next section.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

The results of the simulation study are tabulated in Tables 1-4. In each table, we give the bias and the root mean squared error (RMSE) of each method in estimating the regression parameters. We only give results for the slope parameter in each model as the intercept is a time-independent covariate and is seldom of interest.

Table 1 corresponds to the situation where there is a single Type II time-dependent covariate, all estimators are approximately unbiased for both $n = 100$ and $n = 500$. In this situation, we expect GMM_{T2} to perform the best, as reflected in the results, which show that the relative efficiency of GMM_{T2} to GEE_{I} to be about 170% ($(0.1313/0.1006)^2 \times 100\%$ for $n = 100$ and $(0.538/0.410)^2 \times 100\%$ for $n = 500$). The performances of GMM_{SEL}, EL_1 and EL_2 are similar and they are all slightly less efficient than GMM_{T2} due to testing for the covariate (and shrinkage in the case of EL_2) but all three estimators are better than GEE_{I}. This set of simulations shows that using additional estimating functions in the presence of a Type II covariate can lead to efficiency gain over a GEE with independence assumption.

Table 2 corresponds to the situation with a single Type III time-dependent covariate, therefore bias results if the covariate is mistaken as a Type II covariate. This fact shows up in the results, where substantial bias is observed in GMM_{T2} for both $n = 100$ and $n = 500$. For GMM_{SEL}, the test sometimes could identify the covariate as a Type III covariate but sometimes, it misses that fact, and the results still shows substantial bias for $n = 100$. The power of the test improves with a larger sample size ($n = 500$) so the estimator becomes unbiased for the larger sample situation. Using empirical likelihood to combine the estimating functions (EL_1) leads to less bias than GMM_{SEL} for $n = 100$. The use of a vector of shrinkage parameters (EL_2) substantially reduces the bias from GMM_{SEL} or EL_1.

Table 3 corresponds to the situation with three Type III time-dependent covariates. The results show that GEE_{I} is the best, as expected, since no additional estimating functions can be added. There are significant biases in the parameter estimates by assuming the covariates as Type II, as seen in the results for $\mathrm{GMM}_{\mathrm{T2}}$. For $n = 100$, the ordering of the performances of $\mathrm{GMM}_{\mathrm{SEL}}$, EL_1 and EL_2 is the same as in Table 2. In addition, EL_2 now performs almost as well as $\mathrm{GEE}_{\mathrm{I}}$. For $n = 500$, all estimators except $\mathrm{GMM}_{\mathrm{T2}}$ are nearly unbiased. This model and the previous one are possible situations where the shrinkage parameter is useful.

Table 4 corresponds to the situation with two Type III and one Type II time-dependent covariates. Using $\mathrm{GMM}_{\mathrm{T2}}$, which incorrectly assumes all covariates are Type II, now incurs noticeable biases in estimating the parameters (β_1, β_2) that correspond to the Type III covariates. In addition, there is a spillover effect on the estimation of the parameter β_3 for the Type II covariate, such that while there is no significant bias, the RMSE using $\mathrm{GMM}_{\mathrm{T2}}$ is much bigger than that using $\mathrm{GEE}_{\mathrm{I}}$. For the other methods, the Type III covariate is correctly identified but now, there is no efficiency gain either in estimating the parameter associated with the Type II covariate.

We draw the following conclusions. Our method works best when there are Type III covariates and when the power of the test that identifies Type II vs III covariates is low. When the data consists of only Type II covariates, its performance is similar to other estimators that use a test, all are better than using a GEE with the independence assumption.

We also repeated the simulations considered here, but under situations where some of the data are missing. The trends of the simulation results are similar to those presented here and hence, to conserve space, we moved the results to the Supplementary Material.

5. Analysis of Filipino children's data

We apply the methods in the previous section to the data described in the Introduction. The data was collected in 1984-1985 by surveying 448 households living within a 20-mile radius. Data was collected at four survey time points, separated by four month intervals. For more details, see Bouis and Haddad (1990) and Bhargava (1994).

The anthropometric measure we use is body mass index (BMI), which equals weight (in kg) divided by height (in cm) squared. We seek to predict morbidity (the outcome) four months into the future based on BMI. We also use the predictors of age (in months) and gender. BMI, age and gender are widely considered as building blocks of anthropometric measures (Cogill, 2003). In addition, we use dummy variables for the round of the survey (to account for seasonality in morbidity) as a predictor. We fit the data using a linear model.

Although we have the history of BMI at multiple time points, we focus on the effect of BMI at a given time point because in many developing countries, a full history of BMI is not available and public health decisions must be made only based on the child's current BMI (Martorell and Habicht, 1986, Zerfas et al., 1986, World Health Organization, 1995).

Following Bhargava (1994), we focus on the youngest child (1-14 years) in each household and only consider those children who have complete data at all time points, resulting in 370 children with three observations of (BMI at time t, morbidity at time $t+$ 4 months) each. For the morbidity outcome, we use the empirical logistic transformation (Bhargava, 1994) of the proportion of time over the two weeks prior to the interview that the child was sick,

$$y_{it} = \log\left(\frac{\text{days over last two weeks prior to time } t \text{ child was sick} + 0.5}{\text{14.5 - days over last two weeks prior to time } t \text{ child was sick}}\right) \tag{13}$$

Gender is a time-independent covariate. Age is a Type II time-dependent covariate since the age of a child at any time t determines his/her age at other times. A similar argument holds for the survey round dummy variables. BMI is a time-dependent covariate but it can also be viewed as part of the outcome process. Two reasons for this interpretation are (a) if a child is sick, the child may not eat much and this could affect the child's height and weight in the future and (b) infections have generalized effects on nutrient metabolism and utilization (Martorell and Ho, 1984). However, both (a) and (b) are most relevant for diarrheal infections and the proportion of sick children (over a two week period) who had diarrheal infection was only 9%. Furthermore, because survey rounds were four months apart, the effect of a child's sickness in one round on the child's weight at the next round is likely to be small. The test of

Lai and Small (2007) does not reject BMI as a Type II time-dependent covariate ($p = 0.21$). The program codes for analysing the data are given in the Supplementary Material.

[Table 5 about here.]

Table 5 gives the parameter estimates and standard errors of the five methods in Section 4. The results from the different methods show a similar general trend. A higher BMI is associated with a lower morbidity risk. Older kids and boys are less likely to get sick, which may be explained by the fact that in many parts of Asia, boys are favoured over girls and therefore, boys may be receiving a better share of the food staple than girls, and older kids have higher immunity against diseases. All methods except $\mathrm{GMM}_{\mathrm{SEL}}$ show a trend of increasing morbidity risk over the survey rounds. Interestingly, the proportion of households with $< 80\%$ of required caloric intake also increases across these three rounds (25%, 40%, 47%, respectively). Similar trends are seen for households with inadequate Iron, Vitamins A and C across these rounds. Therefore, the trend could be related to nutrition. A caveat in this analysis is the lack of significance in most of the estimates, which may be due to the small sample size and the difficulty in accurately measuring food intake. In this example, $\mathrm{GEE}_{\mathrm{I}}$, $\mathrm{GMM}_{\mathrm{T2}}$ and $\mathrm{GMM}_{\mathrm{SEL}}$ have smaller standard error estimates than the EL methods.

6. Supplementary Material

The Supplementary Material referenced in Sections 3 to 5, is available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

We thank the referees and the Associate Editor for the detailed and insightful comments, that have led to a greatly improved version of this paper. Denis Leung, Dylan Small and Zhu Min were partially supported by the Research Center at Singapore Management University.

References

Andrews, D. W. K. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica* **67,** 543–564.

Angrist, J. and Krueger, A. B. (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *Journal of the American Statistical Association* **87,** 328–336.

Bhargava, A. (1994). Modelling the health of Filipino children. *Journal of the Royal Statistical Society, A* **157,** 417–432.

Bouis, H. and Haddad, L. (1990). Effects of agricultural commercialization on land tenure, household resource allocation, and nutrition in the Philippines. Research Report 79, International Food Policy Research Institute, Washington.

Caballero, B. (2002). Global patterns of child health: The role of nutrition. *Annals of Nutrition & Metabolism* **46(suppl 1),** 3–7.

Cogill, B. (2003). *Anthropometric Indicators Measurement Guide*. Academy for Educational Development, Washington, D.C.

Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford:OUP, second edition.

Donald, S. G. and Newey, W. K. (2001). Choosing the number of instruments. *Econometrica* **69,** 1161–1191.

Eichenbaum, M. S., Hansen, L. P., and Singleton, K. J. (1988). A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *Quarterly Journal of Economics* **103,** 51–78.

Fitzmaurice, G. M. (1995). A caveat concerning independence estimating equations with multivariate binary data. *Biometrics* **51,** 309–317.

Gallant, A. R., Hsieh, D., and Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics* **81,** 159–192.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50,** 1029–1054.

Kolaczyk, E. (1995). An information criterion for empirical likelihood with general estimation equations. Unpublished manuscript, Department of Statistics, University of Chicago.

Lai, T., Small, D., and Liu, J. (2008). Statistical inference in dynamic panel data models. *Journal of Statistical Planning and Inference* **138,** 2763–2776.

Lai, T. L. and Small, D. (2007). Marginal regression analysis of longitudinal data with time-dependent covariates: a generalized method-of-moments approach. *Journal of the Royal Statistical Society, B* **69,** 79–99.

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

Martorell, R. and Habicht, J. P. (1986). Growth in early childhood in developing countries. In Falkner, F. and Tanner, J. M., editors, *Human Growth: A Comprehensive Treatise*. Plenum Press: New York.

Martorell, R. and Ho, T. J. (1984). Malnutrition, morbidity and mortality. *Population and Development Review* **10,** 49–68.

Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72,** 219–255.

Okui, R. (2009). Instrumental variable estimation in the presence of many moment conditions. *Journal of Econometrics* Accepted for publication.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75,** 237–49.

Pan, W. and Connett, J. E. (2002). Selecting the working correlation structure in generalized estimating equations with application to the Lung Health Study. *Statistica Sinica* **12,** 475–490.

Pepe, M. S. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics, Simulations and Computation* **23,** 939–951.

Scrimshaw, N. S. and SanGiovanni, J. P. (1997). Synergism of nutrition, infection, and immunity: An overview. *American Journal of Clinical Nutrition* **66,** 464S–477S.

Stone, M. (1974). A unified approach to regression analysis under double-sampling designs. *Journal of the Royal Statistical Society, B* **36,** 111–147.

Ten Have, T. and Morabia, A. (2002). An assessment of non-randomized medical treatment of long-term schizophrenia relapse using bivariate binary-response transition models. *Biostatistics* **3,** 119–131.

United Nations Children's Fund (2007). State of the World's Children 2008. United Nations Children's Fund: New York.

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, B* **71,** 177–190.

World Health Organization (1995). Physical status: The use and interpretation of anthropometry. WHO Technical Report Series 854, World Health Organization: Geneva.

Zerfas, A., Jelliffe, D. B., and Jelliffe, E. F. P. (1986). Epidemiology and nutrition. In Falkner, F. and Tanner, J. M., editors, *Human Growth: A Comprehensive Treatise*. Plenum Press: New York.

Zhu, H. T., Ibrahim, J. G., Tang, N., and Zhang, H. (2008). Diagnostic measures for empirical likelihood of general estimating equations. *Biometrika* **95,** 489–507.

Table 1
Bias (RMSE) for five estimators in Model 1

		$n = 100$		$n = 500$	
Method	Parameter	Bias	RMSE	Bias	RMSE
GEE_{I}	β_1	-0.0061	0.1313	0.0013	0.0538
GMM_{T2}	β_1	-0.0143	0.1006	-0.0009	0.0410
GMM_{SEL}	β_1	-0.0161	0.1150	-0.0013	0.0438
EL_1	β_1	-0.0101	0.1107	0.0009	0.0455
EL_2	β_1	-0.0096	0.1185	0.0010	0.0464

GEE_{I}: GEE with independence assumption
GMM_{T2}: GMM assuming all covariates are of Type II
GMM_{SEL}: GMM with testing
EL_1: EL with no shrinkage
EL_2: EL with shrinkage

Table 2
Bias (RMSE) for five estimators in Model 1a

		$n = 100$		$n = 500$	
Method	Parameter	Bias	RMSE	Bias	RMSE
GEE_{I}	β_1	0.0076	0.1387	0.0082	0.0581
GMM_{T2}	β_1	-0.1648	0.2003	-0.1808	0.1862
GMM_{SEL}	β_1	-0.0837	0.1822	-0.0062	0.0772
EL_1	β_1	-0.0670	0.1689	-0.0012	0.0729
EL_2	β_1	-0.0307	0.1496	0.0028	0.0655

GEE_{I}: GEE with independence assumption
GMM_{T2}: GMM assuming all covariates are of Type II
GMM_{SEL}: GMM with testing
EL_1: EL with no shrinkage
EL_2: EL with shrinkage

Table 3
Bias (RMSE) for five estimators in Model 2

		$n = 100$		$n = 500$	
Method	Parameter	Bias	RMSE	Bias	RMSE
$\mathrm{GEE_I}$	β_1	0.0030	0.1439	0.0168	0.0688
	β_2	0.0160	0.1476	0.0177	0.0666
	β_3	0.0096	0.1441	0.0174	0.0661
$\mathrm{GMM_{T2}}$	β_1	-0.1418	0.2198	-0.1346	0.1473
	β_2	-0.1296	0.2126	-0.1349	0.1488
	β_3	-0.1358	0.2128	-0.1343	0.1483
$\mathrm{GMM_{SEL}}$	β_1	-0.0680	0.1842	-0.0058	0.0855
	β_2	-0.0613	0.1886	-0.0063	0.0838
	β_3	-0.0667	0.1811	-0.0049	0.0820
$\mathrm{EL_1}$	β_1	-0.0462	0.1681	0.0017	0.0836
	β_2	-0.0371	0.1688	0.0021	0.0803
	β_3	-0.0417	0.1626	0.0029	0.0796
$\mathrm{EL_2}$	β_1	-0.0185	0.1524	0.0074	0.0767
	β_2	-0.0081	0.1552	0.0078	0.0741
	β_3	-0.0138	0.1520	0.0084	0.0729

$\mathrm{GEE_I}$: GEE with independence assumption
$\mathrm{GMM_{T2}}$: GMM assuming all covariates are of Type II
$\mathrm{GMM_{SEL}}$: GMM with testing
$\mathrm{EL_1}$: EL with no shrinkage
$\mathrm{EL_2}$: EL with shrinkage

Table 4
Bias (RMSE) for five estimators in Model 3

		$n = 100$		$n = 500$	
Method	Parameter	Bias	RMSE	Bias	RMSE
$\mathrm{GEE_I}$	β_1	-0.0005	0.0792	-0.0021	0.0342
	β_2	-0.0070	0.0782	0.0007	0.0340
	β_3	0.0001	0.0928	-0.0003	0.0416
$\mathrm{GMM_{T2}}$	β_1	-0.0122	0.1527	-0.0305	0.0682
	β_2	-0.0430	0.1603	-0.0255	0.0645
	β_3	0.0087	0.2002	-0.0048	0.0741
$\mathrm{GMM_{SEL}}$	β_1	-0.0050	0.0871	-0.0034	0.0354
	β_2	-0.0108	0.0867	-0.0008	0.0350
	β_3	-0.0005	0.1058	-0.0004	0.0424
$\mathrm{EL_1}$	β_1	-0.0005	0.0794	-0.0020	0.0343
	β_2	-0.0076	0.0787	0.0007	0.0340
	β_3	0.0000	0.0928	0.0001	0.0414
$\mathrm{EL_2}$	β_1	-0.0006	0.0794	-0.0021	0.0342
	β_2	-0.0072	0.0786	0.0007	0.0340
	β_3	0.0001	0.0927	-0.0001	0.0414

$\mathrm{GEE_I}$: GEE with independence assumption
$\mathrm{GMM_{T2}}$: GMM assuming all covariates are of Type II
$\mathrm{GMM_{SEL}}$: GMM with testing
$\mathrm{EL_1}$: EL with no shrinkage
$\mathrm{EL_2}$: EL with shrinkage

Table 5
Estimated coefficients (standard errors) for the Filipino children's data using different estimators.

Variable	GEE_{I}	GMM_{T2}	GMM_{SEL}	EL_1	EL_2
Intercept	-0.9718	-0.7095	-0.3634	-0.9538	-1.0223
	(0.7437)	(0.6590)	(0.7019)	(0.7336)	(0.7483)
BMI	-0.0618	-0.0804	-0.0965	-0.0374	-0.0239
	(0.0436)	(0.0379)	(0.0412)	(0.0563)	(0.0584)
Age	-0.0125	-0.0124	-0.0135	-0.009	-0.0122
	(0.0032)	(0.0030)	(0.0031)	(0.0105)	(0.0085)
Gender	-0.2796	-0.3098	-0.2993	-0.3138	-0.2262
	(0.1106)	(0.1151)	(0.1161)	(0.1314)	(0.1822)
Survey Round 2	0.0243	0.0011	-0.0072	0.0776	0.0791
	(0.1314)	(0.1359)	(0.1360)	(0.1577)	(0.1666)
Survey Round 3	0.1452	0.1305	0.1011	0.1987	0.1192
	(0.1063)	(0.1026)	(0.1040)	(0.1194)	(0.1561)