Harvard University
Harvard University Biostatistics Working Paper Series

Year 2006 *Paper* 52

Estimation in Semiparametric Transition Measurement Error Models for Longitudinal Data

Wenqin Pan* Donglin Zeng†
Xihong Lin‡

*Duke University, wendy.pan@duke.edu
†University of North Carolina, dzeng@bios.unc.edu
‡Havard School of Public Health, xlin@hsph.harvard.edu

Estimation in semiparametric transition measurement error models for longitudinal data

BY WENQIN PAN

Department of Biostatistics and Bioinformatics, Duke University, Durham, 27705, USA
wendy.pan@duke.edu

AND DONGLIN ZENG

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA
dzeng@bios.unc.edu

AND XIHONG LIN

Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA
xlin@hsph.harvard.edu

August 21, 2006

SUMMARY

We consider semiparametric transition measurement error models for longitudinal data, where one covariate is measured with error and no distributional assumption is made for the underlying unobserved covariate. An estimating equation approach based on the pseudo conditional score method is proposed. We show the resulting estimators of the regression coefficients are consistent and asymptotic normal. We derive the semiparametric efficiency score and study the efficiency loss of the pseudo conditional score estimator. In the presence of validation data, we propose a one-step estimator that achieves the semiparametric efficient bound. Simulation studies are conducted to examine the small-sample performance of our estimator. A real data set is analyzed for illustration.

Some Key words: Asymptotic efficiency, Conditional score method, Functional modeling, Measurement Error, Longitudinal data, Semiparametric inference, Transition models.

1. INTRODUCTION

Longitudinal data are common in health sciences research, where repeated measures are obtained for each subject over time. Diggle, et al. (2002) provide a comprehensive overview of statistical methods for analyzing longitudinal data. One class of longitudinal models is the transitional model, where the conditional mean of an outcome at the current time point is modeled as a function of its values at the previous time points and covariates (Diggle, et al., 2002, Chapter 10). This model is useful when one is interested in studying the effects of covariates and the past responses on the current response or predicting the future response given the past history. The within-subject correlation is easily accounted for by conditioning on the past responses, and the model can be easily fit within the generalized linear model framework. Transition models have been studied in a number of literatures and applications (Young et al. 1999, Heagerty 2002, Have and Morabia 2002, Yu et al. 2003, Yang et al., Dunson 2003, Roy and Lin 2005).

Measurement error is a common problem in longitudinal data collection, due to reasons such as equipment limitation, longitudinal variation, or recall bias. Classical covariate measurement error examples include CD4 counts in AIDS studies (Tsiatis, Degruttola and Wulfsohn 1995), blood pressure and fat intake in nutritional studies (Carroll, Ruppert, and Stefanski 1995). In one study from the AIDS Costs and Services Utilization Survey (ACSUS) (Berk, Maffeo and Schur 1993) which consists 2487 subjects in 10 randomly selected U.S. cities with the highest AIDS rates, one main outcome was whether an interviewee had hospital admission (yes/no) during the past 3 months and a question of interest is to predict how CD4 count affects the risk of future hospitalization given subject's past history. A natural model for analyzing this data is to use transition model. However, CD4 count is known to contain measurement errors due to its substantial variability (Tsiatis et al. 1995) and another source of error in this study is due to the fact that CD4 count was not measured at the time of each interview but abstracted from each respondent's most recent medical record.

For independent data, a comprehensive review on measurement error methods is given in Fuller (1987) and Carroll, et al. (1995). It is well known in traditional regression settings that ignoring covariate measurement error would lead to attenuated regression coefficient estimators. For longitudinal data, Buonaccorsi, Demidenko and Tosteson (2000) and Wang, et al. (1998) among others considered measurement error in mixed effects models. Limited work has been done for modeling measurement error in transition models. Schmid, Segal and Rosner (1994) and Schmid (1996) studied measurement error in first-order autoregressive models for continuous longitudinal outcome. Pan, Lin and Zeng (2006) proposed maximum likelihood estimation in generalized transitional measurement error models by assuming the repeated measures of the unobserved covariate follows a parametric multivariate normal distribution with the first order auto-regressive or AR(1) correlation structure. Consistency of the maximum likelihood estimator requires that the normality assumption holds and the correlation structure of the repeated measures of the unobserved covariate is correctly specified. However, in reality, such a normality assumption is often too strong. See the histogram of CD4 count of the ACSUS study in Figure 1, which shows considerable non-normality even after a log transformation. Further, the correlation structure of the repeated measures of the unobserved covariate is difficult to be specified correctly. It is hence desirable to develop a semiparametric method which leave the distribution of the repeated measures of the unobserved covariate fully unspecified. We develop such a semiparametric method for transition measurement error models in this paper.

For independent data, estimation in measurement error models without specifying a distribution for the unobserved covariate has been considered by several authors, when validation data are available. Stefanski and Cook (1995) proposed the SIMEX method, which is simple to implement but the resulting estimator is often inconsistent. Carroll et al. (1991) discussed using the validation data to obtain a kernel estimator of the density for the error-prone covariate then plugging it into the score equation to produce a consistent regression coefficient estimator. Recently, Schafer (2001) considered using the EM algorithm to maximize the observed likelihood function by treating the distribution for the unobserved covariate as a discrete function on a finite set of points. However, neither of these approaches is applicable to longitudinal data. A major difficulty is that the unobserved covariate has repeated measures which are likely to be correlated. The kernel method of Carroll et al. (1991) requires large validation data due to the curse of dimensionality needed for constructing a multivariate kernel density estimator. For the same reason, the number of points chosen for estimating the multivariate distribution of Schafter (2001) has to be unrealistically large.

Instead of estimating the multivariate distribution of the repeated measures of the unobserved covariate, we propose two semiparametric methods in this paper. Our first approach is based on an estimating equation method by modifying the conditional score method, which was originally proposed for measurement error regression for independent data by Stefanski and Carroll (1987). However, its generalization to the transition model is not trivial for longitudinal data in the presence of repeated measures of the unobserved covariate. We next derive the semiparametric efficiency score and study the efficiency loss of the pseudo conditional score estimator. In the presence of validation data, we propose a one-step estimator and show it reaches the semiparametric efficiency bound.

The rest of the paper is structured as follows. In §2, we present the semiparametric transition measurement error model for longitudinal data. In §3, we study the asymptotic bias when the distribution of the unobserved covariate is misspecified. In §4, we derive the general conditional score estimating equation and study the theoretical properties of the conditional score estimator, and apply the approach to both the linear and logistic transition models, then illustrate the method using simulation studies and an analysis of the ACSUS data. In §5, we derive the semiparametric efficiency score, and study efficiency loss of the pseudo conditional score estimator. When validation data are available, we propose a one-step estimator that is shown to be semiparametric efficient. Some numerical results are provided. Discussions are given in Section 6.

2. SEMIPARAMETRIC TRANSITION MEASUREMENT ERROR MODEL FOR LONGITUDINAL DATA

Suppose we observe longitudinal data from n subjects, and each subject has m repeated measures over time. Let Y_{ij} be the response at time j ($j = 1, \cdots, m$) of subject i ($i = 1, \cdots, n$). Let W_{ij} be a scalar observed error-prone covariate, which measures the unobserved covariate X_{ij} with error. Let Z_{ij} be a vector of covariates that are accurately measured. The transition model assumes the conditional distribution of Y_{ij} given the history of the outcome Y and the history of the true covariates X and Z satisfies the (q, r)-order Markov property (Ch 10, Diggle et al., 2002) and belongs to the exponentially family. Specifically, we assume that Y_{ij} depends on the past history only via $Y_{i,j-1}, ..., Y_{i,j-q}$ and $X_{ij}, \cdots, X_{i,j-r+1}$, $Z_{ij}, \cdots, Z_{i,j-r+1}$ for $j > (r-1) \vee q$, where $(r-1) \vee q = \max(r-1, q)$. Furthermore, the conditional distribution of Y_{ij} follows the exponential family

$$f(Y_{ij}|\bullet) = \exp\left\{(Y_{ij}\eta_{ij} - b(\eta_{ij}))/a\phi + c(\bullet, \phi)\right\}, \tag{1}$$

where $\bullet = \{Y_{i,j-1}, \cdots, Y_{i,j-q}, X_{ij}, ..., X_{i,j-r+1}, Z_{ij}, ..., Z_{i,j-r+1}\}$, $f(\cdot)$ denotes a density function, a is a prespecified weight, ϕ is a scale parameter, and $b(\cdot)$ and $c(\cdot)$ are specific functions associated with exponential family. We assume a canonical generalized linear model (McCullagh and Nelder, 1989) for $\mu_{ij} = E(Y_{ij}|\bullet) = b'(\eta_{ij})$ as

$$g(\mu_{ij}) = \eta_{ij} = \beta_0 + \sum_{k=1}^{q} \alpha_k Y_{i,j-k} + \sum_{l=1}^{r} \{\beta_{xl} X_{i,j-l+1} + \beta_{zl} Z_{i,j-l+1}\}, \tag{2}$$

where $g(\cdot)$ is a canonical link function and satisfies $g^{-1}(\cdot) = b'(\cdot)$, β_0, α_k ($k = 1, ..., q$), $\beta_l = (\beta_{xl}, \beta_{zl})^T$ ($l = 1, ..., r$) are regression coefficients. Additionally, we treat $Y_{i1}, ..., Y_{i,(r-1)\vee q}$ as initial states of this transition and assume that their distribution does not depend on β's and α's.

We assume that the measurement error is additive as

$$W_{ij} = X_{ij} + U_{ij}, \tag{3}$$

where the measurement error U_{ij} are independent of the X_{ij} and are independent and identically distributed and follow $U_{ij} \sim N(0, \sigma_u^2)$ for a known variance σ_u^2. Pan, et al. (2006), in their maximum likelihood estimation approach, assumed a multivariate normal distribution for the unobserved covariate vector $\{X_{i1}, \cdots, X_{im}\}$ with an auto-regressive correlation structure. The consistency of their maximum likelihood estimator requires the normality assumption. In this paper, we leave the joint distribution of $\{X_{i1}, \cdots, X_{im}\}$ fully unspecified and proceed with semiparametric estimation.

We assume that measurement error is non-differential, i.e., for each subject i, conditional on his/her history of Y and the true covariates X, Z, $\{Y_{ij}\}$ and $\{W_{ij}\}$ are independent, i.e.,

$$f(Y_{ij}, W_{ij}|\bullet) = f(Y_{ij}|\bullet) f(W_{ij}|X_{ij}),$$

where $\bullet$ was defined in (1). This means conditional on the true unobserved covariate (X, Z), the observed covariate W does not contain additional information about Y. We further assume that conditional on the past history of (Y, X, Z), the covariates (X_{ij}, Z_{ij}) only depends on the past history of the covariates of (X, Z), i.e.,

$$f(X_{ij}, Z_{ij}|Y_{i,j-1}, ..., Y_{i1}, X_{i,j-1}, ..., X_{i1}, Z_{i,j-1}, ..., Z_{i1}) = f(X_{ij}, Z_{ij}|X_{i,j-1}, ..., X_{i1}, Z_{i,j-1}, ..., Z_{i1}).$$

It follows that the log-likelihood function for the observed data is given by

$$\sum_{i=1}^{n} \log \int \prod_{j=(r-1)\vee q+1}^{m} f(Y_{ij}|\bullet) f(W_{ij}|X_{ij}) f(X_{ij}, Z_{ij}|X_{i,-j}, Z_{i,-j}) dX_{i1} \cdots dX_{im}, \tag{4}$$

where $\bullet$ is the same as before, $f(Y_{ij}|\bullet)$ is given in (1) and $f(W_{ij}|X_{ij})$ is the normal density under model (3), $X_{i,-j} = (X_{i,j-1}, \cdots, X_{i1})^T$ and a similar definition of $Z_{i,-j}$.

3. ASYMPTOTIC BIAS ANALYSIS OF THE MAXIMUM LIKELIHOOD ESTIMATOR WHEN THE DISTRIBUTION OF X IS MISSPECIFIED

To reveal the importance of our interest in leaving the distribution of the unobserved covariate X unspecified, we first study the asymptotic bias in maximum likelihood estimator when the distribution of X is misspecified. To highlight the key issue, without loss of generality, we focus on the case of $q = 1$ and $r = 1$ in (2) and X being the only covariate in the regression; that is, we consider the following simple generalized linear transition model:

$$g(\mu_{ij,x}) = \beta_0 + X_{ij}\beta_x + Y_{ij-1}\alpha. \tag{5}$$

To study the asymptotic bias of the maximum likelihood estimator when the distribution of X is misspecified, we assume that the true model for X_{ij} follows a first-order Markov model

$$X_{ij} = \gamma_0 + X_{ij-1}\gamma_x + e_{xij}, \tag{6}$$

where e_{xij} are independent of the U_{ij} in the error model (3) and are independent $N(0, \sigma_x^2)$. Equivalently, under the general stationary assumption, the true X model can be rewritten as

$$X_i = 1_i \frac{\gamma_0}{1-\gamma_x} + e_{xi} = 1_i \mu_x + e_{xi},$$

where 1_i is an $m \times 1$ vector of ones, $\gamma_0/(1-\gamma_x) = \mu_x$ is the mean of X_i, and e_{xi} is an AR(1) Gaussian process with mean 0 and covariance matrix Σ_{xi}, whose (j,k)th element is $\sigma_x^2(1-\gamma_x^2)^{-1}\gamma_x^{|j-k|}$. In the following context, we name this model as the *AR(1)* model.

We study the asymptotic biases in maximum likelihood estimators when one misspecifies the X model as an independent model. That is, the incorrect X model used in the maximum likelihood estimation is no longer a first-order autoregressive model, but instead, a model given by

$$X_i = 1_i \tilde{\mu}_x + \tilde{e}_{xi}, \tag{7}$$

where $\tilde{\mathbf{e}}_{xi} \sim N(0, \tilde{\sigma}_x^2 \mathbf{I})$ and $\tilde{\mu}_x$ and $\tilde{\sigma}_x^2$ are two unknown parameters. Equivalently, one misspecifies the observations X_{ij} as generated from independent and identically distributed $N(\tilde{\mu}_x, \tilde{\sigma}_x^2)$. We name this model as the *independent* model.

Some more notation is as follows. Denote the asymptotic limits of the maximum likelihood estimators of $\theta_Y = (\beta_0, \beta_x, \alpha)^T$ and $\theta_X = (\tilde{\mu}_x, \tilde{\sigma}_x^2)^T$ based on the misspecified *independent* X model as $\theta_{Y,indep} = (\beta_{0,indep}, \beta_{x,indep}, \alpha_{indep})^T$ and $\theta_{X,indep} = (\mu_{x,indep}, \sigma^2_{x,indep})$. Furthermore, define the reliability coefficient by $\lambda = \mathrm{var}(X_{ij})/\{\mathrm{var}(X_{ij}) + \sigma_u^2\} = \sigma_x^2(1-\gamma_x^2)^{-1}/\{\sigma_x^2(1-\gamma_x^2)^{-1} + \sigma_u^2\}$. In the following subsections, we investigate the asymptotic biases of the maximum likelihood estimators for Gaussian outcomes and non-Gaussian outcomes separately.

3.1 Asymptotic Biases Under the Linear Transition Model for Gaussian Responses

In this section, we study the asymptotic biases of the maximum likelihood estimators under a misspecified X model when Y follows a linear transition model

$$Y_{ij} = \beta_0 + X_{ij}\beta_x + Y_{ij-1}\alpha + \epsilon_{ij}, \quad \epsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_y^2). \tag{8}$$

Under the AR(1) X model, the results of Theorem 1 of Pan et al. (2006) show that Y_{ij} given the observed data W_{ij}, Y_{ij-1} satisfies

$$E(Y_{ij}|W_{ij}, Y_{ij-1}) = \beta_0^* + \lambda^* \beta_x W_{ij} + (\alpha + \lambda^{**})Y_{ij-1},$$

where β_0^* is some constant, and

$$\lambda^* = \frac{var(X_2)var(Y_1) - cov^2(X_2, Y_1)}{\{var(X_2) + \sigma_u^2\}var(Y_1) - cov^2(X_2, Y_1)},$$

$$\lambda^{**} = \frac{\beta_x \sigma_u^2 cov(X_2, Y_1)}{\{var(X_2) + \sigma_u^2\} var(Y_1) - cov^2(X_2, Y_1)}.$$

Under the *independent* X model, one can easily show that the $Y_{ij}|W_{ij}, Y_{ij-1}$ model takes the form

$$E(Y_{ij}|W_{ij}, Y_{ij-1}) = \beta_{0,indep} + \lambda \beta_{x,indep} W_{ij} + \alpha_{indep} Y_{ij-1}.$$

Thus, we obtain the following result.

THEOREM 1 *Under the conditions that* $|\alpha| < 1$ *and* $|\gamma_x| < 1$, *we have*

$$\beta_{x,indep} = \frac{\lambda^*}{\lambda} \beta_x, \quad \alpha_{indexp} = \alpha + \lambda^{**}.$$

Furthermore, Theorem 2 in Pan et al. (2006) shows that $\lambda^* \leq \lambda$ and λ^{**} has the same sign as γ_x. As a result, we obtain $|\beta_{x,indep}| \leq |\beta_x|$; α_{indep} is greater than α when $\gamma_x > 0$, while less than α when $\gamma_x < 0$. It follows that the maximum likelihood estimator of β_x under the misspecified X model is still attenuated, but its bias is less than the corresponding naive estimate when measurement error is ignored by of replacing W by X in (8), since $\lambda < 1$. The maximum likelihood estimator of the coefficient of the historical response α under the misspecified *independent* X model is equal to its corresponding naive estimator when the measurement error is ignored. Clearly, if in the true model $\gamma_x = 0$, i.e., the *AR(1)* model is equivalent to the *independent* model, $\beta_{x,indep}$ and α_{indep} are consistent estimators of the true parameters β_x and α.

In Figures 2, we numerically evaluate the asymptotic relative biases in $\beta_{x,indep}$ and α_{indep} as a function of the measurement error variance σ_u^2. The parameter configurations are that $\beta_0 = -1, \beta_x = 1, \alpha = 0.5, \sigma^2 = 1$, and $\gamma_0 = 0.4, \gamma_x = 0.6, \sigma_x^2 = 0.5$. The relative bias is defined as the bias of a parameter divided by its true value. The figure clearly shows that the maximum likelihood estimate of β_x under the *independent* X model is attenuated. The maximum likelihood estimate of α under the *independent* X model is inflated. The biases become more severe as σ_u^2 increases.

3.2. Asymptotic biases in the generalized linear transition model for non-gaussian response

When the response Y is non-Gaussian, the bias analysis of the maximum likelihood estimator under the misspecified X model is much more complicated, since the variance structure of the outcome depends on the measure structure. Closed form expressions of $\beta_{x,indep}$ and α_{indep} are generally unavailable, and numerical calculations are hence needed. We first describe the general theoretical results under the generalized linear transition model (5), then show as an example the detailed numerical calculation results of the asymptotic bias analysis in the logistic transition model

$$logit\{P(Y_{ij}=1|X_{ij},Y_{ij-1})\}=\beta_0+X_{ij}\beta_x+Y_{ij-1}\alpha. \tag{9}$$

The maximum likelihood estimator $(\theta_{Y,indep},\theta_{X,indep})$ under the misspecified *independent* X model maximizes the log-likelihood

$$n^{-1}\sum_{i=1}^{n}\ell_{indep}(Y_i,W_i;\theta_{Y,indep},\theta_{X,indep}),$$

where $\ell_{indep}(Y_i,W_i;\theta_{Y,indep},\theta_{X,indep})$ is the log-likelihood function of the ith subject under the independent model (5),(3) and (7). Suppressing the subscript i, the asymptotic limit of the maximum likelihood estimate $(\theta_{Y,indep},\theta_{X,indep})$ maximizes the probability limit (as $n\longrightarrow\infty$) of the independent log-likelihood, which equals $E\{\ell_{indep}(Y,W;\theta_{Y,indep},\theta_{X,indep})\}$, where the expectation is taken with respect to (Y,W,X) under the true models (5),(3) and (6). To compute this expectation, since under the *independent* X model, $X_{ij}=(1-\lambda)\mu_{x,indep}+\lambda W_{ij}+e_{wij}$, where $e_{wij}\sim N(0,(1-\lambda)\sigma^2_{x,indep})$, we plug this expression into the generalized linear transition model (5) and obtain the following equation for the conditional mean of Y_i given W_i, Y_{ij-1}, and $\mathbf{e}_{wij}$ as

$$g(\mu_{ij,w})=\{\beta_{0,indep}+(1-\lambda)\mu_{x,indep}\beta_{x,indep}\}+W_{ij}\lambda\beta_{x,indep}+Y_{ij-1}\alpha_{indep}+\beta_{x,indep}e_{wij}. \tag{10}$$

Therefore, the joint log-likelihood function for Y_i and W_i under the misspecified independent X model is

$$\begin{aligned}\ell_{indep}(Y_i,W_i;\theta_{indep}) &= \log\int L_{indep}(Y_i|W_i,\mathbf{e}_{wi};\theta_{Y,indep},\theta_{X,indep})dF(\mathbf{e}_{wi})\\ &= \log\int L_{indep}(Y_i|W_i,\mathbf{e}_{wi};\theta_{Y,indep},\theta_{X,indep})\sqrt{(1-\lambda)\sigma^2_{x,indep}}d\Phi(e_{wi}),\end{aligned}$$

where $L_{indep}(Y_i|W_i,\mathbf{e}_{wi};\theta_{Y,indep},\theta_{X,indep})$ is the conditional density of Y_i given W_i and $\mathbf{e}_{wi}$ based on the model (10). Particularly, when Y_i is binary,

$$\begin{aligned}&L_{indep}(Y_i|W_i,\mathbf{e}_{wi};\theta_{Y,indep},\theta_{X,indep})\\ =&\prod_{j=1}^{m}\{[g^{-1}([\beta_{0,indep}+(1-\lambda)\mu_{x,indep}\beta_{x,indep}]+W_{ij}\lambda\beta_{x,indep}+Y_{ij-1}\alpha_{indep}+\beta_{x,indep}e_{wij})]^{Y_{ij}}\\ &\times[1-g^{-1}([\beta_{0,indep}+(1-\lambda)\mu_{x,indep}\beta_{x,indep}]+W_{ij}\lambda\beta_{x,indep}+Y_{ij-1}\alpha_{indep}+\beta_{x,indep}e_{wij})]^{1-Y_{ij}}\}\end{aligned}$$

Together with using Gauss-Hermite quadrature and Monte-Carlo simulations in calculating numerical integrations, the expectation $E\{\ell_{indep}(Y,W;\theta_{indep})\}$ is evaluated, which is a function of θ_{indep} and the true value of θ.

As an example, we perform the detailed numerical calculations for the asymptotic limit of the maximum likelihood estimator under the misspecified independent X model for binary outcomes using the logistic transition model (9). Figure 3 shows the asymptotic relative biases in $\beta_{x,indep}$ and α_{indep} a function of the measurement error variance σ_u^2. The parameter configurations are the same as those in the linear transition model case. A similar pattern to Figure 2 is observed and as σ_u^2 increases, the biases become larger.

4. THE PSEUDO CONDITIONAL SCORE METHOD

4.1 The pseudo conditional score estimating equation

Estimation by directly maximizing the likelihood function (4) requires a parametric specification of the density function of $\{X_{i1}, \cdots, X_{im}\}$ and high dimensional integration, and is subject to bias if the distribution of X is misspecified as shown in our asymptotic bias analysis. It is hence desirable to construct a more robust estimator that does not require specifying the distribution of X. We propose in this section a pseudo conditional score method.

Specifically, in a similar spirit of Stefanski and Carroll (1987), we pretend θ to be known but treat the X_{ij} as fixed parameters by writing X_{ij} as x_{ij}, and calculate sufficient statistics for $(x_{i1}, ..., x_{im})$, and construct score equations of model parameters of interest based on the conditional likelihood function of the observed data given the sufficient statistics. Unfortunately, due to the transition structure and the possibly nonlinear link function in (1), sufficient statistics for x_{ij} based on the distribution of $Y_i = (Y_{i1}, ..., Y_{im})$ and $W_i = (W_{i1}, ..., W_{im})$ do not exist except for the linear transition model with normal errors. This makes the task of directly adopting the conditional score method of Stefanski and Carroll (1987) to our setting difficult. However, we note that for each $j = (r-1) \vee q + 1, ..., m$, the conditional density of $(Y_{ij}, W_{ij}, ..., W_{i,j-r+1})$ given $(Y_{i,-j}, Z_{i,-j}, Z_{ij})$ and $(x_{ij}, x_{i,-j})$ is given by

$$\exp\Bigg[Y_{ij}(\beta_0 + \sum_{k=1}^{q}\alpha_k Y_{i,j-k} + \sum_{l=1}^{r}\{\beta_{xl}x_{i,j-l+1} + \beta_{zl}Z_{i,j-l+1}\})/a\phi$$
$$-b(\beta_0 + \sum_{k=1}^{q}\alpha_k Y_{i,j-k} + \sum_{l=1}^{r}\{\beta_{xl}x_{i,j-l+1} + \beta_{zl}Z_{i,j-l+1}\})/a\phi + c(Y_{i,-j}, x_{i,-j}, Z_{i,-j}, \phi)$$
$$-\sum_{l=1}^{r}(W_{i,j-l+1} - x_{i,j-l+1})^2/2\sigma_u^2 - r\log\sqrt{2\pi\sigma_u^2}\Bigg].$$

We immediately recognize that this conditional density still belongs to an exponential family and moreover, we find that the sufficient statistics for $x_{i,j-k+1}, k = 1, ..., r$ are

$$T_{i1}^{(j)} = \frac{\beta_{x1}}{a\phi}Y_{ij} + \frac{1}{\sigma_u^2}W_{ij},\quad T_{i2}^{(j)} = \frac{\beta_{x2}}{a\phi}Y_{ij} + \frac{1}{\sigma_u^2}W_{i,j-1},\quad ...,\quad T_{ir}^{(j)} = \frac{\beta_{xr}}{a\phi}Y_{ij} + \frac{1}{\sigma_u^2}W_{i,j-r+1}.$$

Therefore, the distribution of Y_{ij} given $Y_{i,-j}$, $(Z_{ij}, Z_{i,-j})$ and $(T_{i1}^{(j)}, ..., T_{ir}^{(j)})$ only depends on ϕ, β_0, $\alpha_k(k=1,...,q)$ and $\beta_l = (\beta_{xl}, \beta_{zl})^T(l=1,...,r)$. We abbreviate this distribution as $\tilde{f}(Y_{ij}|V_{ij}(\theta);\theta)$, where θ consists of all the regression parameters and $V_{ij}(\theta)$ denotes those sufficient statistics conditioned on. Under the special case when $r=1$, $\tilde{f}(Y_{ij}|V_{ij}(\theta);\theta)$ is the same as the conditional distribution of Y_{ij} given $Y_{i,-j}$, $(Z_{ij}, Z_{i,-j})$ and $T_{i1}^{(j)}$ only.

From the property

$$E_{\theta_0}\left\{\nabla_\theta \log \tilde{f}(Y_{ij}|V_{ij}(\theta_0);\theta)\Big|_{\theta=\theta_0}\right\} = E_{\theta_0}\left[E_{\theta_0}\left\{\nabla_\theta \log \tilde{f}(Y_{ij}|V_{ij}(\theta_0);\theta)|V_{ij}(\theta_0)\right\}\Big|_{\theta=\theta_0}\right] = 0$$

where ∇_θ denote the gradient with respect to θ, we can construct the following estimating equation

$$\sum_{i=1}^{n} \sum_{j=(r-1)\vee q+1}^{m} g(Y_{ij}|v_{ij} = V_{ij}(\theta);\theta) = 0, \qquad (11)$$

where $g(y_{ij}|v_{ij};\theta)$ denotes the gradient of $\log \tilde{f}(y_{ij}|v_{ij};\theta)$ with respect to θ. Note that calculations of this gradient is done by viewing v_{ij} as fixed, not a function of θ and then evaluating v_{ij} at $v_{ij} = V_{ij}(\theta)$. To distinguish (11) from the conditional score equation in Stefanski and Carroll (1987), we call our proposed estimating equation the *pseudo conditional score equation.*

The Newton-Raphson iteration can be used to solve the equation. The following theorem gives the asymptotic property of any consistent solution to (11).

THEOREM 2. *Let θ_0 denote the true value of θ . Assume that with probability 1, in a neighborhood of θ_0, $\nabla_\theta g\{Y_{ij}|V_{ij}(\theta);\theta\}$ is Lipschitz continuous with respect to θ and moreover,*

$$E_{\theta_0}\left[\sum_{j=(r-1)\vee q+1}^{m} \nabla_\theta g(Y_{ij}|V_{ij}(\theta);\theta)\big|_{\theta=\theta_0}\right] \text{ is non-singular.}$$

Then there exists a solution, $\widehat{\theta}_n$, to equation (11) and $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ converges in distribution to a normal distribution with mean zero and covariance

$$\begin{aligned}\Sigma(\theta_0) &= E_{\theta_0}\left[\sum_{j=(r-1)\vee q+1}^{m} \nabla_\theta g(Y_{ij}|V_{ij}(\theta);\theta)\big|_{\theta=\theta_0}\right]^{-1} \\ &\quad \times E_{\theta_0}\left[\left\{\sum_{j=(r-1)\vee q+1}^{m} g(Y_{ij}|V_{ij}(\theta_0);\theta_0)\right\}\left\{\sum_{j=(r-1)\vee q+1}^{m} g(Y_{ij}|V_{ij}(\theta_0);\theta_0)\right\}^T\right] \\ &\quad \times E_{\theta_0}\left[\sum_{j=(r-1)\vee q+1}^{m} \nabla_\theta g(Y_{ij}|V_{ij}(\theta);\theta)\big|_{\theta=\theta_0}\right]^{-1}.\end{aligned}$$

The proof is given in Appendix. A consistent estimator for Σ is

$$
\begin{aligned}
\widehat{\Sigma}_n &= n\left[\sum_{i=1}^{n}\sum_{j=(r-1)\vee q+1}^{m} \nabla_{\theta} g(Y_{ij}|V_{ij}(\theta);\theta)\big|_{\theta=\widehat{\theta}_n}\right]^{-1} \\
&\quad \times \left[\sum_{i=1}^{n}\left\{\sum_{j=(r-1)\vee q+1}^{m} g(Y_{ij}|V(\widehat{\theta}_n);\widehat{\theta}_n)\right\}\left\{\sum_{j=(r-1)\vee q+1}^{m} g(Y_{ij}|V(\widehat{\theta}_n);\widehat{\theta}_n)\right\}^{T}\right] \\
&\quad \times \left[\sum_{i=1}^{n}\sum_{j=(r-1)\vee q+1}^{m} \nabla_{\theta} g(Y_{ij}|V_{ij}(\theta);\theta)\big|_{\theta=\widehat{\theta}_n}\right]^{-1}.
\end{aligned}
$$

4.2 Numerical studies

We apply our proposed method to two special examples. In the first example, we consider a linear transition model with $r = 1$ and $q = 1$. Then it is easy to calculate that for $j \geq 2$, $\tilde{f}(Y_{ij}|V_{ij}(\theta);\theta)$, which is the conditional density of Y_{ij} given $T_{i1}^{(j)} = \beta_x Y_{ij}/\sigma_y^2 + W_{ij}/\sigma_u^2$ and $(Y_{i,j-1},...,Y_{i1})$ as well as $(Z_{ij},...,Z_{i1})$, is the same as the conditional density of Y_{ij} given $Q_{ij} = \beta_x(Y_{ij} - \beta_0 - \alpha Y_{i,j-1} - \beta_z^T Z_{ij})/\sigma_y^2 + W_{ij}/\sigma_u^2$ and $(Y_{i,j-1},...,Y_{i1})$ as well as $(Z_{ij},...,Z_{i1})$. Direct calculation gives that the logarithm of this conditional density is equal to

$$
-\log\sqrt{2\pi\sigma_y^{*2}} - (2\sigma_y^{*2})^{-1}(Y_{ij} - \beta_0 - \alpha Y_{i,j-1}\alpha - \beta_z^T Z_{ij} - Q_{ij}\beta_x^*)^2, \quad j = 2,...,m,
$$

where $\beta_x^* = \beta_x/(\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)$ and $\sigma_y^{*2} = (\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)^{-1}\sigma_y^2/\sigma_u^2$. After differentiating with respect to all the parameters then substituting the expression of Q_{ij}, we obtain that the following pseudo-conditional score equations

$$
\begin{aligned}
0 &= \sum_{i=1}^{n}\sum_{j=2}^{m}\begin{pmatrix}1\\ Y_{i,j-1}\\ Z_{ij}\end{pmatrix}\left\{Y_{ij} - \beta_0 - \alpha Y_{i,j-1} - \beta_z^T Z_{ij} - \beta_x W_{ij}\right\}, \\
0 &= \sum_{i=1}^{n}\sum_{j=2}^{m}\left\{(Y_{ij} - \beta_0 - \alpha Y_{i,j-1} - \beta_z^T Z_{ij})\beta_x + W_{ij}\sigma_y^2/\sigma_u^2\right\}(Y_{ij} - \beta_0 - \alpha Y_{i,j-1} - \beta_z^T Z_{ij} - \beta_x W_{ij}), \\
0 &= \sum_{i=1}^{n}\sum_{j=2}^{m}\left\{(Y_{ij} - \beta_0 - \alpha Y_{i,j-1} - \beta_z^T Z_{ij} - \beta_x W_{ij})^2 - (\beta_x^2\sigma_u^2 + \sigma_y^2)\right\}.
\end{aligned}
$$

In the second example, we consider a logistic transition model with $r = q = 1$, where Y_{ij} is a Bernoulli variable and follows the logistic regression model. The likelihood function yields that the sufficient statistics for x_{ij} is given by $T_{i1}^{(j)} = \beta_x Y_{ij} + W_{ij}/\sigma_u^2$, for $j = 2,...,m$. Thus, the logarithm of the conditional density $\tilde{f}(Y_{ij}|T_{i1}^{(j)};\theta)$ is obtained as

$$
-\frac{(T_{i1}^{(j)} - Y_{ij}\beta_x)^2\sigma_u^2}{2} + Y_{ij}(\beta_0 + \beta_z^T Z_{ij} + \alpha Y_{i,j-1})
$$

$$-\log\left[\exp\left\{-\frac{(T_{i1}^{(j)}-\beta_x)^2\sigma_u^2}{2}+(\beta_0+\beta_z^T Z_{ij}+\alpha Y_{i,j-1})\right\}+\exp\left\{-\frac{{T_{i1}^{(j)}}^2\sigma_u^2}{2}\right\}\right].$$

After differentiating the above function with respect to all the parameters then substituting the expression of $T_{i1}^{(j)}$, we obtain the following pseudo-conditional score equations

$$\begin{aligned} 0 &= \sum_{i=1}^n\sum_{j=2}^m \begin{pmatrix} 1 \\ Y_{i,j-1} \\ Z_{ij} \end{pmatrix} \left[Y_{ij}-\frac{1}{1+\exp\left\{(1/2-Y_{ij})\beta_x^2\sigma_u^2-\beta_x W_{ij}-(\beta_0+\beta_z^T Z_{ij}+\alpha Y_{i,j-1})\right\}}\right], \\ 0 &= \sum_{i=1}^n\sum_{j=2}^m \left[Y_{ij}W_{ij}-\frac{(Y_{ij}\beta_x+W_{ij}/\sigma_u^2-\beta_x)\sigma_u^2}{1+\exp\left\{(1/2-Y_{ij})\beta_x^2\sigma_u^2-W_{ij}\beta_x-(\beta_0+\beta_z^T Z_{ij}+\alpha Y_{i,j-1})\right\}}\right]. \end{aligned}$$

We implement these two set of equations in our simulation studies. Especially, in the first simulation, the longitudinal response Y_{ij} is generated from

$$Y_{ij}=-1+0.4Y_{i,j-1}+3X_{ij}+0.8Z_i+N(0,1), \quad i=1,...,n, j=2,...,5,$$

where Z_i is a Bernoulli variable with $P(Z_i=1)=0.5$ and X_{ij} follows another transition model

$$X_{ij}=0.5+0.8X_{i,j-1}+N(0,1), \quad i=1,...,n, j=2,...,5.$$

Moreover, we use $X_{i1}=0.25$ and $Y_{i1}=-5/12+5Z_i/3$ as values at time one. The measurement error distribution in (3) has a variance 0.5. In the second simulation, we generate binary response from a logistic transition model with mean

$$E[Y_{ij}|H_{ij}]=\frac{\exp\{-1+0.5Y_{i,j-1}+X_{ij}+0.8Z_{ij}\}}{1+\exp\{-1+0.5Y_{i,j-1}+X_{ij}+0.8Z_{ij}\}}, \quad i=1,...,n, j=2,...,5,$$

where Z_i is generated from a Bernoulli distribution with $P(Z_i=1)=0.5$ and X_{ij} follows

$$X_{ij}=0.4+0.5Z_i+0.6X_{i,j-1}+N(0,0.5) \quad i=1,...,n, j=2,...,5.$$

The measure error has variance 0.5. In both simulations, we solve the pseudo-conditional score equations to derive the estimators and estimate the asymptotic variance using the formula $\widehat{\Sigma}_n$. Table 1 reports the summary results from both simulation with sample sizes 100 or 200 after 1000 repetitions. Table 1 indicates that in small sample, the estimates have virtually no bias and the estimated standard errors agree well with the true standard errors.

In a third simulation study to examine the robustness of the proposed approach, we use the same setting as in the first simulation study except that X's error distribution is a mixture of $N(-3,1)$ and $N(3,1)$ with mixing probability 0.5. We then estimate the regression parameters either using the pseudo-conditional score approach or using the "maximum likelihood approach" assuming a misspecified normal error for X. The simulation results from $n=100$ and $n=200$ based on 1000 repetitions are reported in Table 2. The table shows that the estimates from the pseudo-conditional score approach have bias as small as 1% of the true values while the "maximum likelihood approach" produces bias as large as 15% of the true values.

As an example, we apply our method to analyze the ACSUS data. Specifically, we restricted our attention to 533 who completed the first year interviews. These interviews occurred every 3 months. The outcome of interest is whether they had hospital admission (yes/no) during the four interviews. One interest is to estimate the risk of CD4 count on the hospitalization given the past history. Thus, a natural model for analyzing this data is via the transition model while accounting for the measurement error in the CD4 count. Particularly, a logistic transition model is used to fit the data with covariate $W = \log(CD4/100)$, a transformation that reduces the marked skewness of CD4 count, and other covariates including patient's age category from 1 to 10, whether s/he used antiretroviral drug, whether s/he was HIV-symptomatic at the start of the study, patient's race and gender. Additionally, the past hospitalization history is also adjusted for in the analysis. The size of the measurement error for W, σ_u^2, is set to be 1/3 of the variance of baseline W and it is equal to 0.38. This value is also close to the estimated value 0.39 by Wulfsohn and Tsiatis (1995) using data from a clinical trial conducted by Burroughs-Wellcome.

To determine the transition orders, we first note that the first order autocorrelations among W's are all above 0.85; thus this suggests that only current CD4 count is sufficient to represent the previous CD4 history, i.e., $r = 1$. Since the total number of measurements per subject is 4, the maximal value of q can only be 3. We then fit the data with $q = 3$ while treating the outcomes at the first three interviews as initial states. The result shows that the coefficients for the second and third order terms are highly nonsignificant. Hence, our final model has transition order $q = 1$. The fitted result is given in Table 3 and it shows that there exists significant difference between females and males and even after adjusting for the previous hospitalization status, the effect of CD4 on the risk of hospitalization is still significant. The patients who had previous hospital admission history and who had lower CD4 counts would be more likely to be hospitalized in the future. We also fit the model by letting σ_u^2 be 0.18 which responds to the coefficient of variation being 50% in the baseline W and the findings as shown in Table 3 are similar.

5. SEMIPARAMETRIC EFFICIENT ESTIMATION

5.1. Asymptotic efficiency in pseudo conditional score estimation

The pseudo-conditional score equation approach relies on the conditional likelihood function, so it does not utilize the full data information; as the results, it may not give the efficient estimators. Thus, it is useful to know how much efficiency is lost when using such an approach. Since deriving the asymptotic efficiency bound for model (2) is generally difficult, we focus our discussion on the situation that Y_{ij} is a Gaussian outcome and $r = 1$ and $q = 1$ in (2). Additionally, we assume $\{Z_{ij}\}$ and $\{X_{ij}\}$'s are independent but we allow the repeated measures of X to be correlated and the repeated measures of Z to be correlated.

From the previous discussion, we have known that $Q_{ij} = \beta_x(Y_{ij} - \alpha Y_{i,j-1} - \beta_z Z_{ij})/\sigma_y^2 + W_{ij}/\sigma_u^2, j = 2, ..., m$ are sufficient statistics for $X_{ij}, j = 2, ..., m$. In fact, they are also complete sufficient statistics. Therefore, following Bickel et al. (1993, Chap 4, pp. 130), one can show that the efficient score function for $\theta = (\beta_0, \beta_z, \alpha, \beta_x, \sigma_y^2)$ is equal to

$$\dot{\ell}_\theta^*(Y_i, W_i, Z_i; \theta, G) = E[\dot{\ell}_\theta^c(Y_i, W_i, Z_i, X_i; \theta)|Y_i, W_i, Z_i] - E[\dot{\ell}_\theta^c(Y_i, W_i, X_i, Z_i; \theta)|Q_i, Z_i],$$

where $Y_i = (Y_{i2}, ..., Y_{im}), W_i = (W_{i2}, ..., W_{im}),\ Q_i = (Q_{i2}, ..., Q_{im})$ and $\dot{\ell}_\theta^c$ is the score function for θ with the complete data (Y, X, Z). Here, $G(\cdot)$ denotes the joint distribution of $(X_{i2}, ..., X_{im})$. Specifically, we obtain

$$\dot{\ell}_\theta^*(Y_i, W_i, Z_i; \theta, G) = \frac{1}{\sigma_y^2}\sum_{j=2}^m \begin{pmatrix} \widetilde{\epsilon}_{ij} - E[\widetilde{\epsilon}_{ij}|Q_{ij}] \\ Z_{ij}(\widetilde{\epsilon}_{ij} - E[\widetilde{\epsilon}_{ij}|Q_{ij}]) \\ Y_{i,j-1}\widetilde{\epsilon}_{ij} - E[Y_{j-1}\widetilde{\epsilon}_{ij}|Q_{ij}] - \beta_x(Y_{i,j-1} - E[Y_{i,j-1}|Q_{ij}])E[X_{ij}|Q_{ij}] \\ (\widetilde{\epsilon}_{ij} - E[\widetilde{\epsilon}_{ij}|Q_{ij}])E[X_{ij}|Q_{ij}] \\ (\widetilde{\epsilon}_{ij}^2 - E[\widetilde{\epsilon}_{ij}^2|Q_{ij}] - 2\beta_x(\widetilde{\epsilon}_{ij} - E[\widetilde{\epsilon}_{ij}|Q_{ij}])E[X_{ij}|Q_{ij}])/(2\sigma_y^2) \end{pmatrix}, \tag{12}$$

where $\widetilde{\epsilon}_{ij} = Y_{ij} - \beta_0 - Z_{ij}^T\beta_z - Y_{i,j-1}\alpha$. We can further explicitly calculate each term of $\dot{\ell}_\theta^*$ using the fact that $(\widetilde{\epsilon}_{i1}, ..., \widetilde{\epsilon}_{im})^T$ given Q_i follows a multivariate normal distribution with mean $\beta_x(\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)^{-1}Q_i$ and covariance $\sigma_y^2/\sigma_u^2(\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)^{-1}I_{m\times m}$. Especially, we have

$$\begin{aligned}
E[\widetilde{\epsilon}_{ij}|Q_i] &= \frac{\beta_x}{\beta_x^2/\sigma_y^2 + 1/\sigma_u^2}Q_{ij}, \\
E[\widetilde{\epsilon}_{ij}^2|Q_i] &= \frac{\sigma_y^2/\sigma_u^2}{\beta_x^2/\sigma_y^2 + 1/\sigma_u^2} + \left(\frac{\beta_x}{\beta_x^2/\sigma_y^2 + 1/\sigma_u^2}Q_{ij}\right)^2, \\
E[Y_{i,j-1}|Q_i] &= \sum_{k=1}^{j-1}\alpha^{j-1-k}\left(\beta_0 + \beta_z^T Z_k + \frac{\beta_x}{\beta_x^2/\sigma_y^2 + 1/\sigma_u^2}Q_{ik}\right) + \alpha^{j-2}Y_0, \\
E[Y_{i,j-1}\widetilde{\epsilon}_{ij}|Q_i] &= E[Y_{i,j-1}|Q_i]E[\widetilde{\epsilon}_{ij}|Q_i], \\
E[X_{ij}|Q_i] &= \frac{\int X_{ij}q(Q_i|X_i, \theta)dG(X_i)}{\int q(Q_i|X_i, \theta)dG(X_i)},
\end{aligned}$$

where $q(Q_i|X_i, \theta)$ is the conditional density of Q_i given X_i, also given by

$$q(Q_i|X_i, \theta) = \left\{\sqrt{2\pi(\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)}\right\}^{-m}\exp\left\{-\frac{\sum_{j=2}^m[Q_{ij} - (\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)X_{ij}]^2}{2(\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)}\right\}.$$

It follows that the semiparametric efficiency bound is given by $\Sigma_e = \{E[\dot{l}^*_{\theta}(Y_i, W_i, Z_i; \theta, g)^{\otimes 2}]\}^{-1}$. Then the efficiency loss in the pseudo-conditional score estimating equations can be evaluated by comparing Σ_e with Σ, where Σ is given in Theorem 1. Particularly, the explicit forms of Σ_e and Σ are given in Appendix A.2 when X_i follows an AR(1) model.

We utilize a concrete example to illustrate the efficiency loss. Suppose that (Y_i, W_i) follows

$$Y_{ij} = -1 + 0.5Y_{i,j-1} + X_{ij} + 0.6Z_i + N(0, 2),$$

$$W_{ij} = X_{ij} + N(0, 0.5),$$

where Z_i is a Bernoulli variable with $P(Z_i = 1) = 0.5$ and X is generated from the following transition model

$$X_{ij} = 0.4 + 0.5X_{i,j-1} + N(0, \sigma_x^2).$$

For different choices of $\sigma_x^2 = 0.3$ or 0.15 and different cluster size $m = 3$ or 4, we compute the asymptotic relative efficiency of the estimators for β_x, β_z, α in the pseudo-conditional score approach and compared with the semiparametric efficient bound. The results are presented in Table 4. The results in Table 4 show that using the pseudo conditional score method, almost no efficiency is lost in estimating β_z; however, the efficiency loss in the estimators of β_x and α varies for different choices of the cluster size and the error variation in X and such a loss can be as large as 20% in some scenarios.

5.2. Semiparametric efficient estimation with validation data

When a set of validation data for X, say $\widetilde{X}_1, ..., \widetilde{X}_N$, is available, we propose a one-step estimator to improve efficiency by taking advantage of the explicit expression of the efficient score function for θ. Especially, the new estimator for θ is given by

$$\widetilde{\theta}_n = \widehat{\theta}_n + \left\{ \frac{1}{n} \sum_{i=1}^{n} \dot{\ell}^*_{\theta}(Y_i, W_i, Z_i; \widehat{\theta}_n, \widehat{G}_n) \dot{\ell}^*_{\theta}(Y_i, W_i, Z_i; \widehat{\theta}_n, \widehat{G}_n)^T \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \dot{l}^*_{\theta}(Y_i, W_i, Z_i; \widehat{\theta}_n, \widehat{G}_n) \right\}, \quad (13)$$

where $\widehat{G}_n$ is the empirical distribution of X from the validation set and $\dot{\ell}^*_{\theta}(\cdot)$ is the efficient score function given in (12). The following theorem shows that the one-step estimator $\widetilde{\theta}_n$ from (13) attains the semiparametric efficiency bound and its asymptotic variance can be consistently estimated by

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \dot{\ell}^*_{\theta}(Y_i, W_i, Z_i; \widehat{\theta}_n, \widehat{G}_n) \dot{\ell}^*_{\theta}(Y_i, W_i, Z_i; \widehat{\theta}_n, \widehat{G}_n)^T \right\}^{-1}.$$

THEOREM 3. *Suppose $n, N \to \infty$. Then $\sqrt{n}(\widetilde{\theta}_n - \theta_0)$ converges in distribution to a normal distribution with mean zero with variance equal to $E\left[\dot{\ell}^*_{\theta}(Y_i, W_i, Z_i; \theta_0, G_0) \dot{\ell}^*_{\theta}(Y_i, W_i, Z_i; \theta_0, G_0)^T\right]^{-1}$, where G_0 is the true distribution of X.*

The proof of the theorem is given in Appendix. Particularly, when $N/n \to 0$, i.e., the information from the validation set is nuisance as compared to the full data information, then the semiparametric efficiency bound even with validation observations is still the same as $E[\dot{\ell}^*_\theta(Y_i, W_i, Z_i; \theta_0, G_0)$ $\dot{\ell}^*_\theta(Y_i, W_i, Z_i; \theta_0, G_0)^T]^{-1}$. Thus, Theorem 3 implies that when $N/n \to 0$, $\widetilde{\theta}_n$ attains the asymptotic efficiency bound.

We also conduct a simulation study to examine the performance of the one-step estimator. The simulation setting is the same as in the previous section and σ_x^2 is chosen to be 0.3 and 0.15 and the cluster size m is 3 or 4. Moreover, we let $X_{i1} = 0.8$ and $Y_{i1} = -0.4 + 2Z_i$. In order to compare the pseudo conditional score estimator and the one-step estimator, we generate $N = n/4$ observations of $X_i = (X_{i1}, ..., X_{im}), i = 1, ..., N$. Our results from 1000 repetitions are summarized into Table 5. We observe that both the pseudo-conditional score estimate and the one-step estimate perform well in the sample size 200 and 400 and the corresponding inference is accurate. The variance for the estimate of β_x increases significantly when σ_x^2 decreases from 0.3 to 0.15; however, the estimates for both β_z and α do not change much. Efficiency is not gained with the one-step procedure for estimating β_z, while efficiency is gained a very small fraction in estimating α. However, using the one-step estimate, the efficiency is gained in estimating β_x and such an efficiency gain vary from 5% to more than 20% when σ_x^2 decreases from 0.30 to 0.15. Additionally, the more validation data are used or the smaller cluster each subject has, such an efficiency gain is more significant. Therefore, our simulation results comply with the previous theoretical calculations in Table 4, where we indicate that the one-step procedure does not improve the efficiency in estimating β_z and most improve the estimation for β_x.

To understand why such an efficiency gain increases with the validation size N while decreases with the cluster size and the σ_x^2, we recall that in the one-step procedure, it is necessary to obtain an empirical estimate for $E[X_i|Q_i]$ using the validation data. Therefore, when the variance of X_i is smaller, the cluster size is smaller, or the validation size is larger, such an estimate will be more accurate in finite sample calculation then the one-step estimate's efficiency gain will be more likely to be observed. This conclusion has also been confirmed by our other simulations not reported here, where when the size of the validation data is small and the σ_x^2 is relatively large, we observe little efficiency gain using the one-step procedure.

6. DISCUSSION

We consider in this paper transition measurement error models for longitudinal data. We show that the maximum likelihood estimator is likely to be asymptotically biased when the distribution of the unobserved covariate is misspecified. We propose a pseudo conditional score approach that does not require specifying the distribution of the unobserved covariate. We investigate the efficiency loss of such estimators and propose a semiparametric efficient one-step estimator when a small set of validation data is available. Both numerical calculations and simulation studies show that the estimators using the pseudo conditional score equations perform well and subject to small loss of efficiency. The one-step estimator using the validation data may improve the efficiency.

We acknowledge that the one-step efficient estimation relies on the explicit formulation of the semiparametric efficient score function. However, this formulation does not exist for more complicated model such as logistic transition models. One possible approach is to maximize the observed likelihood function, where the unknown distribution of X is substituted with a discrete distribution on the observed validation observations. Such an approach generally requires a large size of validation data and computation can be expensive.

One important issue in fitting a transition model is the selection of transition orders of r and q. Currently there does not exist any literature on choosing r and q in our current semiparametric setting. However, order selection has been discussed in detail via either Akaike information criteria or Bayesian information criteria for parametric structural models in Pan et al. (2006). Thus, we suggest practical users to first select transition orders using structural models then obtain robust estimates using our semiparametric method.

Another important issue is to determine the size of measurement error, σ_u^2. When neither validation set nor prior knowledge is available, one possible strategy is to conduct sensitivity analysis and report the estimates and their variations under a reasonable range of measurement error sizes. Such analysis can be useful in practice.

APPENDIX

Proof of Theorem 2

From the condition and the inverse mapping theorem, the map

$$\theta \mapsto n^{-1} \sum_{i=1}^{n} \sum_{j=(r-1)\vee q+1}^{m} g(Y_{ij}|V_{ij}(\theta);\theta)$$

is invertible in a neighborhood of θ_0. Since n is large, 0 is in the image of the map, we conclude that there exists a solution $\widehat{\theta}_n$ to equation (11). The asymptotic normality follows from Theorem 5.41 (van der Vaart, 1998).

Calculation of Σ_e and Σ in Section 5.1

To facilitate the calculation, we let $C_0 = \beta_x^2\sigma_u^2 + \sigma_y^2$, $C_1 = \sigma_y^2/\sigma_u^2(\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)^{-1}$ and $C_2 = \beta_x(\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)^{-1}$. Also define $\Delta_{ij} = \widetilde{\epsilon}_{ij} - E[\widetilde{\epsilon}_{ij}|Q_i]$ and $\tau_{ij}(Q_i) = E[\widetilde{\epsilon}_{ij}|Q_i] - \beta_x E[X_{ij}|Q_i]$.

We first derive the expression of Σ_e. Using the new notation, we rewrite (12) as

$$
\begin{aligned}
& \dot{l}^*_\theta(Y_i, W_i, Z_i; \theta, g) \\
= \; & \frac{1}{\sigma_y^2}\begin{pmatrix}
\Delta_{i1} + \sum_{j=2}^{m-1}\Delta_{ij} + \Delta_{im} \\
Z_{i1}\Delta_{i1} + \sum_{j=2}^{m-1} Z_{ij}\Delta_{ij} + Z_{im}\Delta_{im} \\
A_1(Q_i)\Delta_{i1} + \sum_{j=2}^{m-1} A_j(Q_i)\Delta_{ij} + A_m(Q_i)\Delta_{im} + B(Q_i) + \sum_{j=2}^m\sum_{k=1}^{j-1}\alpha^{j-1-k}\Delta_{ik}\Delta_{ij} \\
E[X_{i1}|Q_i]\Delta_{i1} + \sum_{j=2}^{m-1} E[X_{ij}|Q_i]\Delta_{ij} + E[X_{im}|Q_i]\Delta_{im} \\
\frac{1}{\sigma_y^2}\tau_{i1}(Q_i)\Delta_{i1} + \sum_{j=2}^{m-1}\frac{1}{\sigma_y^2}\tau_{ij}(Q_i)\Delta_{ij} + \frac{1}{\sigma_y^2}\tau_{im}(Q_i)\Delta_{im} - \frac{1}{2\sigma_y^2}var(\widetilde{\epsilon}_{ij}|Q_i) + \frac{1}{2\sigma_y^2}\sum_{j=2}^m\Delta_{ij}^2.
\end{pmatrix}.
\end{aligned}
$$

where

$$
\begin{aligned}
A_1(Q_i) &= \sum_{k=2}^m \alpha^{k-2}\tau_{ik}(Q_i) + E[Y_{i0}|Q_i], \\
A_j(Q_i) &= \sum_{k=1}^{j-1}\alpha^{j-1-k}\beta_z(Z_{ik} - E[Z_{ik}|Q_i]) + \sum_{k=j+1}^m \alpha^{k-1-j}\tau_{ik}(Q_i) + E[Y_{i,j-1}|Q_i], \\
& \quad j = 2, ..., m-1, \\
A_m(Q_i) &= \sum_{k=1}^{m-1}\alpha^{m-k-1}\beta_z(Z_{ik} - E[Z_{ik}|Q_i]) + E[Y_{i,m-1}|Q_i], \\
B(Q_i) &= \sum_{j=2}^m\sum_{k=1}^{j-1}\alpha^{j-1-k}\beta_z(Z_{ik} - E[Z_{ik}|Q_i])\tau_{ij}(Q_i).
\end{aligned}
$$

Using the fact that $\Delta_{i1}, ..., \Delta_{im}$ are conditionally independent given Q_i and they follow normal distributions with mean zero and constant variance C_1, we obtain that Σ_e is equal to the inverse of

$$
\begin{pmatrix}
nC_1 & C_1\sum_{j=2}^m E[Z_{ij}] & C_1\sum_{j=2}^m E[Y_{i,j-1}] & C_1\sum_{j=2}^m E[X_{ij}] & 0 \\
C_1\sum_{j=2}^m E[Z_{ij}^T] & C_1\sum_{j=2}^m E[Z_{ij}Z_{ij}^T] & C_1\sum_{j=2}^m E[Z_{ij}Y_{i,j-1}] & C_1\sum_{j=2}^m E[Z_{ij}X_{ij}] & 0 \\
C_1\sum_{j=2}^m E[Y_{i,j-1}]C_1\sum_{j=2}^m & E[Y_{i,j-1}Z_{ij}^T] & \sigma_{33} & \sigma_{34} & \sigma_{35} \\
C_1\sum_{j=2}^m E[X_{ij}] & C_1\sum_{j=2}^m E[X_{ij}Z_{ij}^T] & \sigma_{34} & \sigma_{44} & \sigma_{45} \\
0 & 0 & \sigma_{35} & \sigma_{45} & \sigma_{45}
\end{pmatrix}/\sigma_y^4, \tag{A.1}
$$

where

$$
\begin{aligned}
\sigma_{33} &= \sum_{j=2}^m\sum_{k=1}^{j-1}\alpha^{2(j-1-k)}C_1^2 + E[A_1(Q_i)^2]C_1 + \sum_{j=2}^{n-1}E[A_j(Q_i)^2]C_1 + E[A_m(Q_i)^2]C_1 + E[B(Q_i)^2]; \\
\sigma_{34} &= E[A_1(Q_i)E[X_{i1}|Q_i]]C_1 + \sum_{j=2}^{m-1}E[A_j(Q_i)E[X_{ij}|Q_i]]C_1 + E[A_m(Q_i)E[X_{im}|Q_i]]C_1; \\
\sigma_{35} &= \frac{1}{\sigma_y^2}(E[A_1(Q_i)\tau_{i1}(Q_i)]C_1 + \sum_{j=2}^{m-1}E[A_j(Q_i)\tau_{ij}(Q_i)]C_1 + E[A_m(Q_i)\tau_{im}(Q_i)]C_1); \\
\sigma_{44} &= \sum_{j=2}^m E[E[X_{ij}|Q_i]^2]C_1;
\end{aligned}
$$

$$\sigma_{45} = \frac{1}{\sigma_y^2}\sum_{j=2}^{m} E[E[X_{ij}|Q_i]\tau_{ij}(Q_i)]C_1;$$

$$\sigma_{55} = \frac{1}{4\sigma_y^4}(2nC_1^2 + 4\sum_{j=2}^{m} E[\tau_{ij}(Q_i)^2]C_1).$$

To derive the expression of Σ, the asymptotic covariance matrix for the pseudo-conditional score estimator, we note that from Theorem 1, Σ is given by $\widetilde{\Sigma} = \widetilde{\Sigma}_2^{-1}\widetilde{\Sigma}_1(\widetilde{\Sigma}_2^{-1})^T$, in which

$$\widetilde{\Sigma}_2 = \sum_{j=2}^{m}\begin{pmatrix} 1 & E[Z_{ij}^T] & E[Y_{i,j-1}] & E[X_{ij}] & 0 \\ E[Z_{ij}] & E[Z_{ij}Z_{ij}^T] & E[Z_{ij}Y_{i,j-1}] & E[Z_{ij}X_{ij}] & 0 \\ E[Y_{i,j-1}] & E[Y_{i,j-1}Z_{ij}^T] & E[Y_{i,j-1}^2] & E[Y_{j-1}X_{ij}] & 0 \\ C_0/\sigma_u^2 E[X_{ij}] & C_0/\sigma_u^2 E[X_{ij}Z_{ij}^T] & C_0/\sigma_u^2 E[X_{ij}Y_{i,j-1}] & C_0/\sigma_u^2 E[X_{ij}^2] & \beta_x \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad (A.2)$$

and

$$\widetilde{\Sigma}_1 = C_0\sum_{j=2}^{m}\begin{pmatrix} 1 & E[Z_{ij}^T] & E[Y_{i,j-1}] & C_0/\sigma_u^2 E[X_{ij}] & 0 \\ E[Z_{ij}] & E[Z_{ij}Z_{ij}^T] & E[Z_{ij}Y_{i,j-1}] & C_0/\sigma_u^2 E[Z_{ij}X_{ij}] & 0 \\ E[Y_{i,j-1}] & E[Y_{i,j-1}Z_{ij}^T] & E[Y_{i,j-1}^2] & C_0/\sigma_u^2 E[Y_{i,j-1}X_{ij}] & 0 \\ C_0/\sigma_u^2 E[X_{ij}] & C_0/\sigma_u^2 E[X_{ij}Z_{ij}^T] & C_0/\sigma_u^2 E[X_{ij}Y_{i,j-1}] & C_0(\sigma_y^2/\sigma_u^2 + C_0/\sigma_u^4 E[X_{ij}^2]) & 0 \\ 0 & 0 & 0 & 0 & 2C_0 \end{pmatrix}. \qquad (A.3)$$

We can evaluate each term in the above expressions of (A.1), (A.2), and (A.3) when assuming
(M.1) (Y_i, W_i) follows $Y_{ij} = \beta_0 + \beta_z Z_{ij} + \beta_x X_{ij} + \alpha Y_{i,j-1} + \epsilon_{ij}$, $W_{ij} = X_{ij} + U_{ij}$;
(M.2). X is generated from the transition model $X_{ij} = \gamma_0 + \gamma_x X_{i,j-1} + \epsilon_{xij}$;
(M.3) $Z_{ij} = ... = Z_{i1}$ has mean m_z and variance v_z and it is independent of X;
(M.4) Y_0 has mean m_y and variance v_y and X_0 has mean m_x and variance v_x;
(M.5) $(\epsilon_{ij}, U_{ij}, \epsilon_{xij})$ are independently from normal distribution with mean zero and variance $\sigma_y^2, \sigma_u^2, \sigma_x^2$ respectively.

For example, in calculating σ_{kl} in the matrix (A.1), we need calculate $E[X_i|Q_i]$. We first notice that the joint density of (Q_i, X_i) is proportional to

$$\exp\{-\frac{(Q - (\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)X)^T(Q_i - (\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)X_i)}{2(\beta^2/\sigma_y^2 + 1/\sigma_u^2)} - \frac{(X_i - \mu_x)^T\Sigma_x^{-1}(X_i - \mu_x)}{2}\},$$

where $\mu_x = (E[X_{i1}], ..., E[X_{im}])'$ and Σ_x is the covariance matrix of X_i, i.e., its (k, l)-element is equal to $E[X_{ik}X_{il}] - E[X_{ik}]E[X_{il}]$ for $1 \leq k, l \leq m$. Hence, X_i given Q_i is a multivariate-normal distribution with mean $E[X_i|Q_i] = [\Sigma_x^{-1} + (\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)I_{m\times m}]^{-1}(\Sigma_x^{-1}\mu_x + Q_i)$. Moreover, since $E[\widetilde{\epsilon}_{ij}|Q_i] = C_2 Q_{ij}$ and $E[Y_{i,j-1}|Q_i] = \sum_{k=1}^{j-1}\alpha^{j-1-k}(\beta_0 + \beta_z m_z + C_2 Q_{ik}) + \alpha_{j-1}m_y$, each term in the expression of $\sigma_{33}, \sigma_{34}, \sigma_{35}, \sigma_{44}, \sigma_{45}$ and σ_{55} is simply the expectation of a quadratic function of Q_i. Thus, Σ_e can be calculated from the additional facts that

$$Q_i \sim \text{Multinormal}((\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)E[X_i], \quad (\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)I_{m\times m} + (\beta_x^2/\sigma_y^2 + 1/\sigma_u^2)^2 Cov(X_i))$$

and that

$$\begin{aligned}
E[X_{ij}] &= \gamma_x^j m_x + \gamma_0 \frac{1-\gamma_x^j}{1-\gamma_x} + m_z\gamma_z \frac{1-\gamma_x^j}{1-\gamma_x}, \\
E[X_{ij}X_{ik}] &= E[X_{ij}]E[X_{ik}] + \sigma_x^2 \gamma_x^{j-k} \frac{1-\gamma_x^{2k}}{1-\gamma_x^2} + \gamma_x^{2j} v_x + \gamma_z^2 v_z \gamma_x^{j-k} (\frac{1-\gamma_x^k}{1-\gamma_x})^2, \quad k \le j, \\
E[Y_{ij}] &= \alpha^j m_y + \beta_0 \frac{1-\alpha^j}{1-\alpha} + m_z\beta_z \frac{1-\alpha^j}{1-\alpha} + \beta_x \sum_{k=1}^{j} \alpha^{j-k} E[X_{ik}], \\
E[Y_{ij}^2] &= E[Y_{ij}]^2 + v_z\beta_z^2 (\frac{1-\alpha^j}{1-\alpha})^2 + \alpha^{2j} v_y + \sigma_y^2 \frac{1-\alpha^{2j}}{1-\alpha^2} \\
&\quad + \sum_{k=1}^{j}\sum_{k'=1}^{j} \alpha^{j-k}\alpha^{j-k'} (E[X_{ik}X_{ik'}] - E[X_{ik}]E[X_{ik'}]), \\
E[X_{ij}Y_{i,j-1}] &= E[X_{ij}](\beta_0 \frac{1-\alpha^{j-1}}{1-\alpha} + \beta_z m_z \frac{1-\alpha^{j-1}}{1-\alpha} + m_y \alpha^{j-1}) \\
&\quad + \sum_{k=1}^{j-1} \beta_x \alpha^{j-k-1} E[X_{ij}X_{ik}], \quad j \ge 2.
\end{aligned}$$

Similarly, Σ can be calculated using the above equalities.

Proof of Theorem 3

We prove the same results under an even more general setting: Suppose that n i.i.d observations, $O_1, ..., O_n$ are available but $X_1, ..., X_n$ are missing. Moreover, the following assumptions hold:

(C.1) The conditional density of O given X is given by $f(O|X;\theta)$ and X has a density $g(X)$; moreover, $f(O|X;\theta)$ are continuously twice differentiable with respect θ;

(C.2) Q is a function of O and θ and in addition, Q is sufficient statistics for x in the family $\{f(O|x;\theta)\}$ indexed by both x and θ;

(C.3) Q is also a complete statistics for x in the above family; that is, if $E[w(Q)|X] = 0, a.s.$, then $w(Q) = 0, a.s.$

(C.4) there exists a consistent estimator $\widehat{\theta}_n$ such that $|\widehat{\theta}_n - \theta_0| = O_p(n^{-1/2})$;

(C.5) the distribution of X is estimated by $\widehat{G}_n(x)$ and for some metric ρ and some function $G^*(x)$, $\rho(\widehat{G}_n, G^*) \to 0$ in probability.

From (C.1)-(C.5), using the result in Page 130-131 (BKRW, 1993), we immediately obtain that the efficient score function for θ is given by

$$\dot{l}_\theta^*(O;\theta,G) = E[\dot{l}_\theta^c(O,X;\theta)|O] - E[\dot{l}_\theta^c(O,X;\theta)|Q], \tag{A.4}$$

where the subscript θ means the derivative with respect to θ and $\dot{l}_\theta^c(O,X;\theta) = \nabla_\theta \log f(O,X;\theta)$.

Therefore, the efficient influence function for θ is given by

$$\widetilde{l}_{\theta}(O;\theta,G) = -\{E[\ddot{l}^{*}_{\theta\theta}(O;\theta,G)]\}^{-1}\dot{l}^{*}_{\theta}(O;\theta,G) = \{E[\dot{l}^{*}_{\theta}(O;\theta,G)^{\otimes 2}]\}^{-1}\dot{l}^{*}_{\theta}(O;\theta,G),$$

where $\dot{l}^{*}_{\theta\theta}(O;\theta,G)$ is the derivative of $\dot{l}^{*}_{\theta}(O;\theta,G)$ with respect to θ.

Following the description in Section 5.2, a one-step estimator is constructed as follows:

$$\widetilde{\theta}_n = \widehat{\theta}_n - \left\{\frac{1}{n}\sum_{i=1}^{n}\ddot{l}^{*}_{\theta\theta}(O_i;\widehat{\theta}_n,\widehat{G}_n)\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\dot{l}^{*}_{\theta}(O_i;\widehat{\theta}_n,\widehat{G}_n)\right\}.$$

Then the following property holds for this one-step estimator $\widetilde{\theta}_n$.

THEOREM A.1. *Let (θ_0, G_0) denote the true parameters and denote $E_{\theta,G}[w(O)]$ as the expectation of $w(O)$ when the parameters are (θ, G). In addition to (C.1)-(C.5), we suppose the following smoothness assumptions are also satisfied:*

(C.6). $\{\dot{l}^{*}_{\theta}(O;\theta,G) : |\theta-\theta_0| < \delta_0, \rho(G,G^*) < \delta_0\}$ is a Donsker class for a small δ_0, where ρ is a semi-metric defined for g.

(C.7). $E_{\theta_0,G_0}[\dot{l}^{*}_{\theta\theta}(O;\theta,G)]$ is continuous in (θ_0, G^*).

(C.8). $E_{\theta_0,G_0}[\dot{l}^{*}_{\theta\theta}(O;\theta_0,G^*)]$ is a non-singular matrix.

Then $\sqrt{n}(\widetilde{\theta}_n - \theta_0)$ weakly converges to a multivariate normal distribution with mean zero and covariance

$$\Sigma = \{E[\dot{l}^{*}_{\theta\theta}(O;\theta_0,G^*)]\}^{-1}E[\dot{l}^{*}_{\theta}(O;\theta_0,G^*)^{T}\dot{l}^{*}_{\theta}(O;\theta_0,G^*)]\{E[\dot{l}^{*}_{\theta\theta}(O;\theta_0,G^*)]\}^{-1}.$$

Furthermore, if $G^*(x) = G_0(x)$, then $\widehat{\theta}$ is an efficient estimator for θ; i.e., Σ is equal to the semiparametric efficiency bound.

PROOF. We use the notation $\mathbf{P}_n w(O) = \frac{1}{n}\sum_{i=1}^{n} w(O_i)$ and $\mathbf{P}w(O) = E_{\theta_0,G_0}[w(O)]$.

$$\begin{aligned}
&\sqrt{n}(\widetilde{\theta}_n - \theta_0)\\
=\;& \sqrt{n}(\widehat{\theta}_n - \theta_0) + \sqrt{n}(\mathbf{P}_n - \mathbf{P})\widetilde{l}_{\theta}(O;\widehat{\theta}_n,\widehat{G}_n) + \sqrt{n}\mathbf{P}\widetilde{l}_{\theta}(O;\widehat{\theta}_n,\widehat{G}_n)\\
=\;& \sqrt{n}(\widehat{\theta}_n - \theta_0) + \sqrt{n}(\mathbf{P}_n - \mathbf{P})\widetilde{l}_{\theta}(O;\widehat{\theta}_n,\widehat{G}_n)\\
&\quad -\sqrt{n}\{E[\dot{l}^{*}_{\theta\theta}(O;\widehat{\theta}_n,\widehat{G}_n)]\}^{-1}\mathbf{P}\dot{l}^{*}_{\theta}(O;\widehat{\theta}_n,\widehat{G}_n). \qquad (A.5)
\end{aligned}$$

By the assumption (C.6) and the Donsker theorem,

$$\sqrt{n}(\mathbf{P}_n - \mathbf{P})\widetilde{l}_{\theta}(O;\widehat{\theta}_n,\widehat{G}_n) = \sqrt{n}(\mathbf{P}_n - \mathbf{P})\widetilde{l}_{\theta}(O;\theta_0,G^*) + o_p(1). \qquad (A.6)$$

Moreover, since the density of O given Q is independent of X, $E_{\theta_0,G_0}[Q(O)|T] = E_{\theta_0,G}[Q(O)|T]$ for any integrable function $Q(O)$. Therefore,

$$\mathbf{P}\dot{l}^{*}_{\theta}(O;\theta_0,G) = E_{\theta_0,G_0}[E_{\theta_0,G}[\dot{l}^{c}_{\theta}(O,X;\theta)|O] - E_{\theta_0,G}[\dot{l}^{c}_{\theta}(O,X;\theta)|Q]]$$

$$\begin{aligned}
&= E_{\theta_0,G_0}[E_{\theta_0,G_0}[\{E_{\theta_0,G}[\dot{l}^c_\theta(O,X;\theta)|O] - E_{\theta_0,G}[\dot{l}^c_\theta(O,X;\theta)|Q]\}|Q]] \\
&= E_{\theta_0,G_0}[E_{\theta_0,G}[\{E_{\theta_0,G}[\dot{l}^c_\theta(O,X;\theta)|O] - E_{\theta_0,G}[\dot{l}^c_\theta(O,X;\theta)|Q]\}|Q]] \\
&= E_{\theta_0,G_0}[E_{\theta_0,G}[\dot{l}^c_\theta(O,X;\theta)|Q] - E_{\theta_0,G}[\dot{l}^c_\theta(O,X;\theta)|Q]] = 0.
\end{aligned}$$

In other words, no matter what G is, $\mathbf{P}\dot{l}^*_\theta(O;\theta_0,G)$ is always zero. Hence,

$$\mathbf{P}\dot{l}^*_\theta(O;\widehat{\theta}_n,\widehat{G}_n) = \mathbf{P}[\dot{l}^*_\theta(O;\widehat{\theta}_n,\widehat{G}_n) - \dot{l}^*_\theta(O;\theta_0,\widehat{G}_n)] = \mathbf{P}[\ddot{l}^*_{\theta\theta}(O;\theta_0,\widehat{G}_n)](\widehat{\theta}_n - \theta_0) + o_p(\frac{1}{\sqrt{n}}). \quad (A.7)$$

From (A.5), (A.6) and (A.7), we obtain that $\sqrt{n}(\widetilde{\theta}_n - \theta_0) = \sqrt{n}(\mathbf{P}_n - \mathbf{P})\widetilde{l}_\theta(O;\theta_0,G^*) + o_p(1)$. The first conclusion follows. The second conclusion is clear since when $G^* = G_0$, $\widetilde{l}_\theta(O;\theta_0,G^*)$ is the efficient influence function.

REMARK A.1. One consistent estimate for the asymptotic covariance Σ is

$$\left\{\frac{1}{n}\sum_{i=1}^n \ddot{l}^*_{\theta\theta}(O_i;\widetilde{\theta}_n,\widehat{G}_n)\right\}^{-1} \left\{\frac{1}{n}\sum_{i=1}^n \dot{l}^*_\theta(O_i;\widetilde{\theta}_n,\widehat{G}_n)^T \dot{l}^*_\theta(O_i;\widetilde{\theta}_n,\widehat{G}_n)\right\}$$
$$\times \left\{\frac{1}{n}\sum_{i=1}^n \ddot{l}^*_{\theta\theta}(O_i;\widetilde{\theta}_n,\widehat{G}_n)\right\}^{-1}.$$

REMARK A.2. In Theorem A.1, if $G^* = G_0$, i.e., $\widetilde{G}_n$ is consistent, one-step estimator can be generated using an alternative equation

$$\widetilde{\theta}_n = \widehat{\theta}_n + \left\{\frac{1}{n}\sum_{i=1}^n \dot{l}^*_\theta(O_i;\widehat{\theta}_n,\widehat{G}_n)\dot{l}^*_\theta(O_i;\widehat{\theta}_n,\widehat{G}_n)^T\right\}^{-1} \left\{\frac{1}{n}\sum_{i=1}^n \dot{l}^*_\theta(O_i;\widehat{\theta}_n,\widehat{G}_n)\right\}.$$

Following the same arguments in proving Theorem A.1, we can easily show $\widetilde{\theta}_n$ is semiparametric efficient.

REMARK A.3. In the application of Theorem A.1 to a linear transition model with validation data of X, we take $\widehat{G}_n$ as the empirical distribution induced by the validation data and the metric ρ is given by the weak convergence of the probability measures.

REFERENCES

Berk, M. L., Maffeo, C., and Schur, C. L. (1993), *Research Design and Analysis Objectives*, AIDS Cost and Services Utilization Survey Report No. 1, Rockville, MD: Agency for Health Care Policy and Research.

Bickel, P. J., Klaassen, C. A. I., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Estimation for Semi-parametric mModels.* John Hopkins University Press.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models*, London: Chapman and Hall.

Carroll, R. J., and Wand, M. P. (1991), Semiparametric estimation in logistic measurement error models, *JRSS B*. **53**, 573-385.

Cook, J. R., and Stefanski, L. A. (1994), Simulation-extrapolation estimation in parametric measurement error models, *JASA*. **89**, 1314-1328.

Dunson, D. B. (2003), Dynamic latent trait models for multidimensional longitudinal data, *JASA*, **98**, 555-563.

Fuller, W. A. (1987), *Measurement Error Models,* John Wiley & Sons, New York.

Have, T. R. 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.

Heagerty, P. J. (2002), Marginalized transition models and likelihood inference for longitudinal categorical data, *Biometrics*, 58, 342-351.

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

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, 2nd edition, London: Chapman and Hall.

Pan, W., Lin, X. and Zeng, D. (2006), Structural inference in transition measurement error models for longitudinal data. *Biometrics*, 62.

Roy, J. and Lin, X. (2005), Missing covariates in longitudinal data with informative dropouts: Bias analysis and inference, *Biometrics*, 61, 837-846

Schafer, D. W. (2001), Semiparametric maximum likelihood for measurement error model regression, *Biometrics*, 57, 53-61.

Spiegelman, D., Rosner, B., and Logan, R. (2000), Estimation and inference for logistic regression with covariates misclassification and measurement error in main study/validation study design, *JASA*, 95, 51-61.

Stefanski, L. A., and Carroll, R. J. (1987), Conditional scores and optimal scores for generalized linear measurement-error models, *Biometrika*, 74, 703-716.

Stefanski, L. A. and Cook, J. R. (1995), Simulation-extrapolation: The measurement error jackknife, *JASA*, 90, 1247-1256.

Sepanski, J. H., Knickerbocker, R., and Carroll, R. J. (1994), A semiparametric correction for attenuation, *JASA*. 89, 1366-1373.

Tsiatis, A. A., De Gruttola, V. and Wulfsohn, M. S. (1995), Modeling the relationship of survival to longitudinal data Measured with error applications to survival and CD4 counts in patients with AIDS, *JASA*, 90, 27-37.

Van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge University Press.

Wang, N., Lin, X., Gutierrez, R. G., and Carroll, R. J. (1998), Bias analysis and SIMEX approach in generalized linear mixed measurement error models, *JASA*, 93, 249-261.

Wulfsohn, M. S. and Tsiatis, A. A. (1997), A joint model for survival and longitudinal data measured with error, *Biometrics*, 53, 330-339.

Young, P. J., Weeden, S. and Kirwan, J. R. (1999), The analysis of a bivariate multi-state Markov transition model for rheumatoid arthritis with an incomplete disease history, *Statistics in Medicine*, 18, 1677-1690.

Yu, F., Morgenstern, H., Hurwitz, E. and Berlin, T.R. (2003), Use of a Markov transition model to analyze longitudinal low-back pain data, *Statistical Methods in Medical Research*, 12, 321-331.

Table 1: Simulation results for pseudo-conditional score equation approach from 1000 repetitions

Sample Size	Parameter	True Value	EST	ESE	SEE	CP	MSE
			linear transition model				
$n = 100$	β_x	3.0	3.023	0.217	0.225	0.94	0.051
	β_z	0.8	0.804	0.322	0.329	0.95	0.108
	α	0.4	0.396	0.039	0.039	0.94	0.0016
$n = 200$	β_x	3.0	3.017	0.152	0.150	0.95	0.023
	β_z	0.8	0.797	0.227	0.226	0.95	0.052
	α	0.4	0.397	0.027	0.027	0.95	0.0007
			logistic transition model				
$n = 100$	β_x	1.0	1.067	0.283	0.283	0.97	0.084
	β_z	0.8	0.796	0.384	0.398	0.95	0.158
	α	0.5	0.455	0.311	0.319	0.94	0.103
$n = 200$	β_x	1.0	1.024	0.185	0.186	0.96	0.035
	β_z	0.8	0.812	0.262	0.258	0.96	0.067
	α	0.5	0.481	0.216	0.214	0.95	0.046

Note: EST is the mean of the estimates; ESE is the mean of the estimated standard errors; SEE is the standard error of the estimators; MSE is the mean square error; CP denotes the coverage proportion of the 95% confidence intervals.

Table 2: Robustness analysis for pseudo-conditional score equation approach from 1000 repetitions

Sample Size	Parameter	True Value	EST	SEE	EST	SEE
			pseudo-conditional score		"MLE" approach	
$n = 100$	β_x	3.0	3.003	0.076	2.805	0.058
	β_z	0.8	0.801	0.309	0.739	0.221
	α	0.4	0.399	0.018	0.448	0.013
$n = 200$	β_x	3.0	3.005	0.050	2.806	0.040
	β_z	0.8	0.793	0.225	0.741	0.156
	α	0.4	0.399	0.012	0.448	0.009

Note: see Table 1. "MLE" assumes an AR(1) model for X and with normally distributed errors.

Table 3: Parameter estimates for the ACSUS study

	$\sigma_u^2 = 0.38$		$\sigma_u^2 = 0.18$	
Parameter	Estimate	Standard Error	Estimate	Standard Error
$\log(CD4/100)$ (β_x)	-0.460	0.072	-0.416	0.067
age	0.030	0.056	0.031	0.055
antireviral drug use	0.051	0.235	0.077	0.232
HIV symptomatic	0.086	0.191	0.069	0.188
race	0.208	0.214	0.209	0.211
sex (female vs. male)	0.621	0.243	0.577	0.239
previous hospitalization (α)	1.838	0.253	1.865	0.250

Table 4: Relative efficiency of pseudo-conditional score estimators

	Cluster Size	β_x	β_z	α
$\sigma_x^2 = 0.3$	$m = 4$	0.893	1.000	0.897
	$m = 3$	0.905	1.000	0.900
$\sigma_x^2 = 0.15$	$m = 4$	0.830	1.000	0.873
	$m = 3$	0.905	1.000	0.877

Table 5: Estimation from one-step procedure with $n/4$ validation Data

m	n	σ_x^2	θ	pseudo-conditional score approach					one-step procedure				
				EST	ESE	SEE	MSE	CP	EST	ESE	SEE	MSE	CP
4	200	0.30	β_x	1.018	0.159	0.164	0.0271	0.94	1.020	0.151	0.161	0.0263	0.93
			β_z	0.605	0.123	0.124	0.0155	0.95	0.605	0.124	0.125	0.0156	0.95
			α	0.495	0.040	0.041	0.0017	0.94	0.494	0.039	0.041	0.0017	0.94
		0.15	β_x	1.034	0.306	0.307	0.0952	0.96	1.043	0.260	0.291	0.0862	0.93
			β_z	0.609	0.127	0.125	0.0157	0.95	0.611	0.125	0.125	0.0157	0.95
			α	0.495	0.042	0.043	0.0018	0.95	0.494	0.041	0.042	0.0018	0.95
	400	0.30	β_x	1.001	0.110	0.110	0.0121	0.95	1.002	0.105	0.107	0.0115	0.94
			β_z	0.604	0.086	0.089	0.0079	0.95	0.604	0.087	0.088	0.0078	0.95
			α	0.497	0.028	0.029	0.0008	0.93	0.497	0.027	0.028	0.0008	0.94
		0.15	β_x	1.011	0.201	0.197	0.0389	0.96	1.015	0.179	0.184	0.0341	0.95
			β_z	0.608	0.087	0.086	0.0074	0.95	0.607	0.087	0.085	0.0073	0.96
			α	0.498	0.029	0.030	0.0009	0.95	0.496	0.029	0.029	0.0008	0.95
3	200	0.30	β_x	1.019	0.190	0.190	0.0368	0.95	1.025	0.180	0.186	0.0351	0.94
			β_z	0.608	0.144	0.146	0.0212	0.94	0.609	0.146	0.146	0.0213	0.95
			α	0.491	0.050	0.053	0.0029	0.93	0.490	0.050	0.052	0.0028	0.93
		0.15	β_x	1.054	0.387	0.382	0.1490	0.97	1.047	0.320	0.348	0.1233	0.95
			β_z	0.608	0.150	0.144	0.0209	0.96	0.609	0.148	0.143	0.0206	0.96
			α	0.494	0.054	0.052	0.0028	0.96	0.493	0.053	0.049	0.0025	0.96
	400	0.30	β_x	1.010	0.131	0.133	0.0177	0.95	1.010	0.125	0.128	0.0165	0.95
			β_z	0.603	0.101	0.100	0.0100	0.95	0.604	0.101	0.100	0.0100	0.95
			α	0.498	0.035	0.035	0.0013	0.94	0.497	0.035	0.035	0.0012	0.95
		0.15	β_x	1.034	0.249	0.250	0.0638	0.95	1.033	0.219	0.229	0.0536	0.94
			β_z	0.608	0.103	0.103	0.0106	0.96	0.609	0.103	0.102	0.0105	0.96
			α	0.496	0.037	0.036	0.0013	0.95	0.495	0.036	0.035	0.0012	0.96

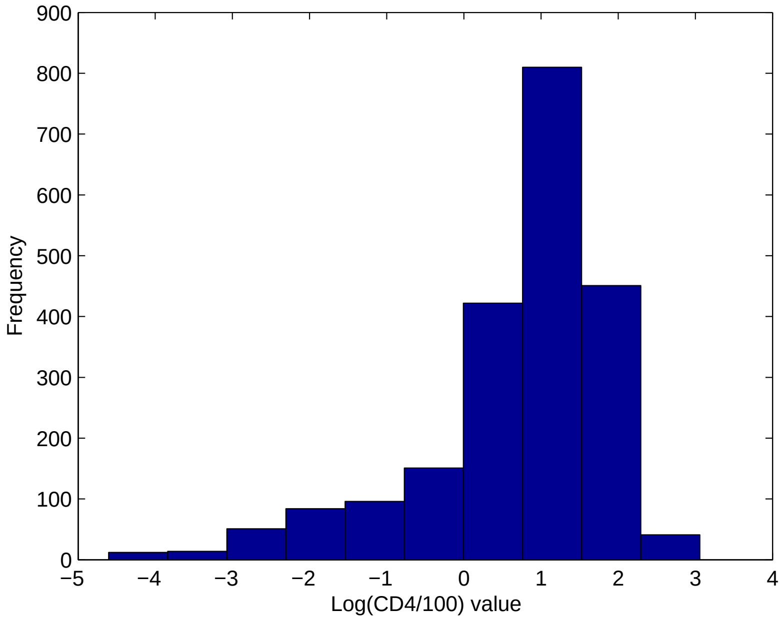

Figure 1: Histogram of the log-transformed CD4 count in the ACSUS data

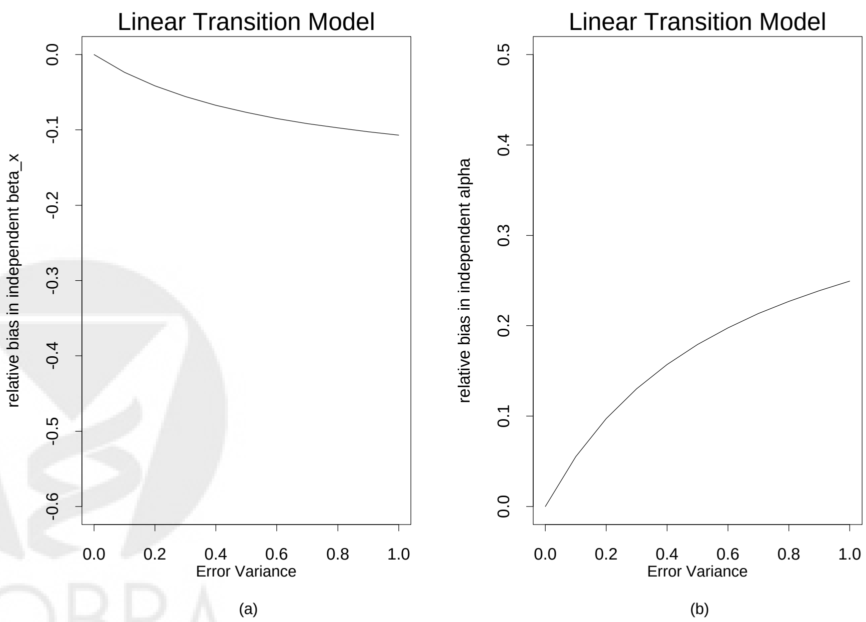

Figure 2: Asymptotic Relative Biases in independent "MLEs" of β_x and α in the Linear Transition Models for Gaussian Outcome, when AR(1) model for X is true. The true parameter values are $\beta_0 = -1, \beta_x = 1, \alpha = 0.5, \sigma^2 = 1$, and $\gamma_0 = 0.4, \gamma_x = 0.6, \sigma_x^2 = 0.5$. The two plots correspond to (a) asymptotic relative bias in $\beta_{x,indep}$; (b) asymptotic relative bias in α_{indep}.

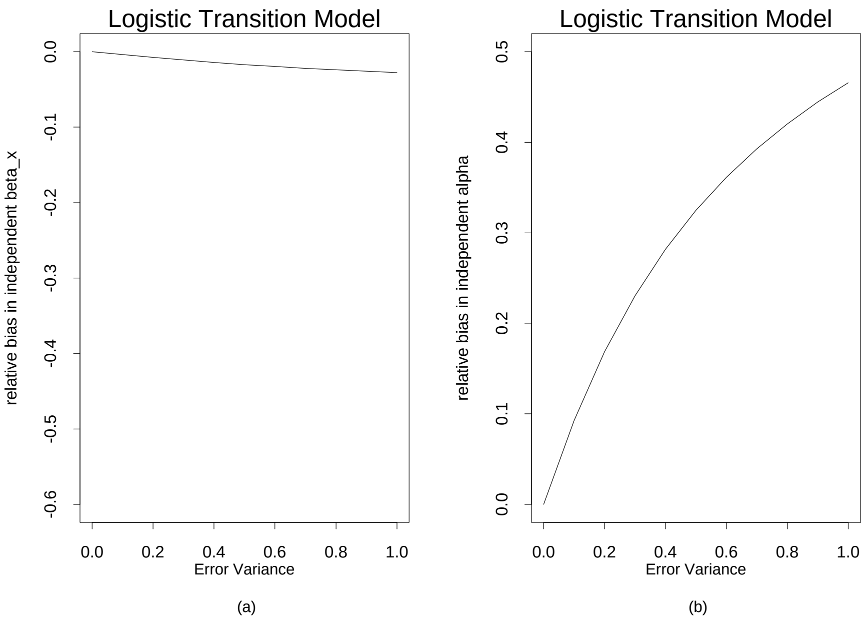

Figure 3: Asymptotic Relative Biases in the independent "MLEs" of β_x and α in the Generalized Linear Transition Models for non-Gaussian Outcome, when AR(1) model for X is true. The true parameter values are $\beta_0 = -1, \beta_x = 1, \alpha = 0.5$, and $\gamma_0 = 0.4, \gamma_x = 0.6, \sigma_x^2 = 0.5$. The two plots correspond to (a) asymptotic relative bias in $\beta_{x,indep}$; (b) asymptotic relative bias in α_{indep}.