Harvard University
Harvard University Biostatistics Working Paper Series

Year 2006 *Paper* 51

Structural Inference in Transition Measurement Error Models for Longitudinal Data

Wenqin Pan* Xihong Lin† Donglin Zeng‡

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

Structural Inference in Transition Measurement Error Models for Longitudinal Data

Wenqin Pan
Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27505, USA
wenqin.pan@duke.edu
Xihong Lin
Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA
xlin@umich.edu
and
Donglin Zeng
Department of Biostatistics, University of North Carolina, Chapel Hills, NC 27514, USA
dzeng@bios.unc.edu

Abstract

We propose a new class of models, transition measurement error models, to study the effects of covariates and the past responses on the current response in longitudinal studies when one of the covariates is measured with error. We show that the response variable conditional on the error prone covariate follows a complex transition mixed effects model. The naive model obtained by ignoring the measurement error correctly specifies the transition part of the model, but misspecifies the covariate effect structure and ignores the random effects. We next study the asymptotic bias in naive estimator obtained by ignoring the measurement error for both continuous and discrete outcomes. We show that the naive estimator of the regression coefficient of the error-prone covariate is attenuated; while the naive estimators of the regression coefficients of the past responses are generally inflated. We then develop a structural modeling approach for parameter estimation using the maximum likelihood estimation method. In view of the multi-dimensional integration required by full maximum likelihood estimation, an EM algorithm is developed to calculate maximum likelihood estimators, in which Monte-Carlo simulations are used to evaluate the conditional expectations in the E-step. We evaluate the performance of the proposed method through a simulation study, and apply it to a longitudinal social support study for elderly women with heart disease.

Key Words: Asymptotic bias; EM algorithm; Maximum likelihood estimator; Measurement error; Structural modeling; Transitional 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. 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.

A common problem in longitudinal studies is the presence of covariate measurement error. For example, it is well known that covariates such as CD4 counts (Tsiatis, Degruttola, and Wulfsohn, 1995) and blood pressure and nutrient intake (Carroll, Ruppert, and Stefanski, 1995) are often measured with error. In Section 6, we consider a longitudinal study of elderly women with heart disease. One of the study objectives was to investigate the effect of social support on the health outcomes. However, the social support level was estimated using the average score of several questions concerning social support in a simple questionnaire and hence measured the true social support level with considerable error.

There is an extensive literature on measurement error for independent data (Fuller, 1987; Carroll, *et al.*, 1995). For longitudinal data, Tosteson, Buonaccorsi, Demidenko (1998) and Buonaccorsi, Demidenko and Tosteson (2000), and Wang, et al (1998) considered modeling measurement error in linear and nonlinear 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. It should be noted that the results in classical generalized linear models with covariate measurement error (Carroll, *et al.*, 1995) are not applicable in generalized transition models, since (1) the past response is included as a covariate in transition models, and

(2) responses and unobserved covariates and their observed error prone values are measured repeatedly over time and are likely to be correlated.

We develop in this paper a new class of models, transition measurement error models, for continuous and discrete outcomes, to study the effects of covariates and the past responses on the current response in longitudinal studies when one of the covariates is measured with error (Section 2). We show in Section 3 that the response variable conditional on the error prone covariate follows a complex transition mixed effects model. The naive model obtained by ignoring the measurement error correctly specifies the transition part of the model, but misspecifies the covariate effect structure and ignores the random effects. We next perform an asymptotic bias analysis in Section 3 and show that ignoring the measurement error results in the regression coefficients of covariates attenuated and the regression coefficients of the past responses inflated. In contrast to the results in generalized linear mixed measurement error models (Wang, *et al.*, 1998), the biases in naive estimators do not depend on the cluster size. We develop a structural modeling approach for inference in Section 4 by accounting for measurement error by assuming the unobserved covariate follows a transition model. We study the finite sample performance of the proposed method in a simulation study in Section 5, and apply the method to the longitudinal study of elderly women with heart disease in Section 6, followed by discussions in Section 7.

2 The General Transition Measurement Error Model

Suppose the data are obtained from m subjects over time from a longitudinal study. Denote by Y_{ij} the outcome variable of the ith subject ($i = 1, ..., m$) at the jth time point ($j = 1, ..., n_i$), X_{ij} the unobserved true covariate (for simplicity, X_{ij} is assumed to be a scalar), W_{ij} the observed error-prone measure of X_{ij}, and $\mathbf{Z}_{ij}(p \times 1)$ the other covariates which are measured precisely. We assume a q-order generalized linear transition model for the outcome variable Y_{ij}, where the conditional distribution of Y_{ij} given its history $\{Y_{ij-1}, ..., Y_{i1}\}$ and the covariate history $\{X_{ij-1}, ..., X_{i1}\}$ and $\{\mathbf{Z}_{ij-1}, ..., \mathbf{Z}_{i1}\}$ is assumed to depend on on the prior q observations of the outcome $\mathbf{H}_{y,ij} = (Y_{ij-1}, ..., Y_{ij-q})^T$ ($q \leq n_i$) and current values of the covariates $\{X_{ij}, \mathbf{Z}_{ij}\}$. The transition model of the outcome Y_{ij} can be written as

$$g(\mu_{ij,x}) = \beta_0 + X_{ij}\beta_x + \mathbf{Z}_{ij}^T\boldsymbol{\beta}_z + \mathbf{H}_{y,ij}^T\boldsymbol{\alpha}, \tag{1}$$

where $\mu_{ij,x}$ is the conditional mean of Y_{ij} given X_{ij}, $\mathbf{Z}_{ij}$, and $\mathbf{H}_{y,ij}$, $g(\cdot)$ is a monotonic differentiable link function, $\beta_0, \beta_x, \boldsymbol{\beta}_z(p \times 1)$, and $\boldsymbol{\alpha}(q \times 1)$ are unknown regression coefficients.

Define $\mathbf{Y}_i = (Y_{i1}, ..., Y_{in_i})^T$, $\mathbf{X}_i$, $\mathbf{Z}_i$, $\mathbf{W}_i$ similarly, and $\boldsymbol{\theta}_Y = (\beta_0, \beta_x, \boldsymbol{\beta}_z^T, \boldsymbol{\alpha}^T)^T$. Then the joint log likelihood of $\mathbf{Y}_i$ given $\{\mathbf{X}_i, \mathbf{Z}_i\}$ for the ith subject is

$$\ell_i(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{Z}_i; \boldsymbol{\theta}_Y) = \sum_{j=q+1}^{n_i} \ell_{ij}(Y_{ij}|X_{ij}, \mathbf{Z}_{ij}, \mathbf{H}_{y,ij}) + \ell_i(Y_{i1}, ..., Y_{iq}|\mathbf{X}_i, \mathbf{Z}_i), \tag{2}$$

where $\ell_{ij}(Y_{ij}|X_{ij}, \mathbf{Z}_{ij}, \mathbf{H}_{y,ij})$ belongs to the exponential family distribution (McCullagh and Nelder, 1989) with mean $\mu_{ij,x}$ and variance $\phi a_{ij}^{-1} v(u)$, a_{ij} is a pre-specified weight, ϕ is a scale parameter, $v(\cdot)$ is a variance function, and $\ell_i(Y_{i1}, ..., Y_{iq}|\mathbf{X}_i, \mathbf{Z}_i)$ is assumed free of $\boldsymbol{\theta}_Y$.

The observed error prone covariate W_{ij} is assumed to be related to the true unobserved X_{ij} through an additive measurement error model,

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

where the measurement error U_{ij} are independent of X_{ij} and independently follow $N(0, \sigma_u^2)$.

The structural transition measurement error model is completed by specifying a distribution for the unobserved covariate X_{ij}. In the classical measurement error literature, it is common to assume the X_{ij} to be independent (Carroll, *et al.*, 1995). However, for longitudinal data, the X_{ij} observed from the same subject are likely to be correlated. Hence, paralleling the transition model for $\mathbf{Y}_i$, we consider a r-order linear transition model for the unobserved X_{ij} as

$$X_{ij} = \gamma_0 + \mathbf{Z}_{ij}^T\boldsymbol{\gamma}_z + \mathbf{H}_{x,ij}^T\boldsymbol{\gamma}_x + e_{x,ij}, \tag{4}$$

where $\mathbf{H}_{x,ij} = (X_{ij-1}, ..., X_{ij-r})^T$, $\boldsymbol{\theta}_X = (\gamma_0, \boldsymbol{\gamma}_z^T, \boldsymbol{\gamma}_x^T)^T$ is an unknown parameter vector, and the $e_{x,ij}$ are independent of U_{ij} and independently follow $N(0, \sigma_x^2)$.

Assuming the measurement error is non-differential, i.e., $L_i(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{Z}_i, \mathbf{W}_i) = L_i(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{Z}_i)$, the joint likelihood of the observed data $(\mathbf{Y}_i, \mathbf{W}_i|\mathbf{Z}_i)$ for the ith subject is

$$L_i(\mathbf{Y}_i, \mathbf{W}_i|\mathbf{Z}_i) = \int L_i(\mathbf{Y}_i|\mathbf{X}_i, \mathbf{Z}_i) L_i(\mathbf{W}_i|\mathbf{X}_i, \mathbf{Z}_i) L_i(\mathbf{X}_i|\mathbf{Z}_i) d\mathbf{X}_i, \tag{5}$$

which often does not have a closed form expression and involves n_i dimensional integration except for Gaussian outcomes.

3 Asymptotic Bias Analysis in Naive Estimators

It is of interest to understand how the transition model (1) can be misspecified if the measurement error is ignored and the asymptotic biases in naive estimators obtained by ignoring the measurement error. To understand the fundamental issues, for simplicity, we focus in our asymptotic bias analysis on the first-order transition models for both the outcome variable Y_{ij} and the unobserved covariate X_{ij}, i.e., set $q = r = 1$, $\mathbf{H}_{y,ij} = Y_{ij-1}$ and $\mathbf{H}_{x,ij} = X_{ij-1}$.

3.1 Misspecification of the naive model

The naive model is defined by ignoring the measurement error by simply replacing the unobserved true covariate X_{ij} with its error-prone value W_{ij} in model (1) as

$$g(\mu_{ij,w}) = \beta_0 + W_{ij}\beta_x + \mathbf{Z}_{ij}^T\boldsymbol{\beta}_z + Y_{ij-1}\alpha. \qquad (6)$$

To examine how the naive model (6) misspecifies the true $(\mathbf{Y}_i|\mathbf{W}_i, \mathbf{Z}_i)$ model, we first derive the true $(\mathbf{Y}_i|\mathbf{W}_i, \mathbf{Z}_i)$ model under the transition models (1), (3), (4), then compare it with (6).

For simplicity, in this investigation, we assume no covariates $\mathbf{Z}_i$ in the $\mathbf{X}$ model (4), i.e.,

$$X_{ij} = \gamma_0 + X_{ij-1}\gamma_x + e_{x,ij}, \qquad (7)$$

which can be rewritten as $\mathbf{X}_i = \mathbf{1}_i\gamma_0/(1-\gamma_x) + \mathbf{e}_{xi}$, where $\mathbf{1}_i$ is an $n_i \times 1$ vector of ones and $\mathbf{e}_{xi}$ is an AR(1) Gaussian process with mean 0 and covariance matrix $\mathbf{\Sigma}_{xi}$, whose (j,k)th element is $\sigma_x^2(1-\gamma_x^2)^{-1}\gamma_x^{|j-k|}$. Denote by $\mathbf{I}_i$ an $n_i \times n_i$ identity matrix and by $\mathbf{\Lambda}_i = \mathbf{\Sigma}_{xi}\{\mathbf{\Sigma}_{xi} + \sigma_u^2\mathbf{I}_i\}^{-1}$ the reliability matrix, where $\mathbf{\Sigma}_{xi}$ and $\mathbf{\Lambda}_i$ depend on i only through their dimensions n_i. Since $\mathbf{X}_i$ and $\mathbf{W}_i$ are jointly normally distributed and independent of $\mathbf{Z}_i$, one can show that $\mathbf{X}_i$ given $(\mathbf{W}_i, \mathbf{Z}_i)$ is also normally distributed with the conditional mean $E(\mathbf{X}_i|\mathbf{W}_i, \mathbf{Z}_i) = \gamma_0(1 -$

$\gamma_x)^{-1}(\mathbf{I}_i - \mathbf{\Lambda}_i)\mathbf{1}_i + \mathbf{\Lambda}_i\mathbf{W}_i$ and the conditional covariance $cov(\mathbf{X}_i|\mathbf{W}_i, \mathbf{Z}_i) = (\mathbf{I}_i - \mathbf{\Lambda}_i)\mathbf{\Sigma}_{xi}$, i.e.,

$$\mathbf{X}_i = (\mathbf{I}_i - \mathbf{\Lambda}_i)(\mathbf{1}_i\frac{\gamma_0}{1-\gamma_x}) + \mathbf{\Lambda}_i\mathbf{W}_i + \mathbf{e}^*_{xi}, \tag{8}$$

where $\mathbf{e}^*_{xi} = \mathbf{X}_i - E(\mathbf{X}_i|\mathbf{W}_i, \mathbf{Z}_i) = (\mathbf{I}_i - \mathbf{\Lambda}_i)\mathbf{e}_{xi} - \mathbf{\Lambda}_i\mathbf{U}_i$ follows $N\{0, (\mathbf{I}_i - \mathbf{\Lambda}_i)\mathbf{\Sigma}_{xi}\}$ and is independent of $(\mathbf{W}_i, \mathbf{Z}_i)$.

Plugging (8) into (1), some calculations show that the observed data $(\mathbf{Y}_i|\mathbf{W}_i, \mathbf{Z}_i)$ no longer follow a transition model, but follow the complex random effects transition model as

$$g(\mu_{ij,w}) = (\beta_0 + \gamma_{0j}\beta_x) + \mathbf{W}_i^T\boldsymbol{\alpha}_{wj}\beta_x + \mathbf{Z}_{ij}^T\boldsymbol{\beta}_z + Y_{ij-1}\alpha + e^*_{xij}\beta_x, \tag{9}$$

where $\mu_{ij,w} = E(Y_{ij}|W_{ij}, \mathbf{Z}_{ij}, Y_{ij-1})$, γ_{0j} is the jth element of the vector $(\mathbf{I}_i - \mathbf{\Lambda}_i)(\mathbf{1}_i\frac{\gamma_0}{1-\gamma_x})$, $\boldsymbol{\alpha}_{wj}$ is the transpose of the jth row of $\mathbf{\Lambda}_i$, and the random effect e^*_{xij} is the jth element of $\mathbf{e}^*_{xi}$ and is induced by measurement error in X. A comparison of the naive model (6) with the true $(\mathbf{Y}_i|\mathbf{W}_i, \mathbf{Z}_i)$ model (9) shows that the naive model correctly specifies the structure of the transitional part, but ignores the random effect e^*_{xij} and misspecifies the covariate structure by assuming it only depends on the current value W_{ij} instead of the whole vector $\mathbf{W}_i$. Hence ignoring the measurement error could result in biased estimates of $\boldsymbol{\beta} = (\beta_0, \beta_x, \boldsymbol{\beta}_z^T)^T$ and α.

It is of substantial interest to understand the direction and the magnitude of such biases. We investigate the asymptotic bias of the regression coefficients β_x and α and the effect of cluster size on the asymptotic bias when the measurement error is ignored. To illustrate the fundamental impact of the measurement error, we assume the same cluster size $n_i = n$ and no covariates $\mathbf{Z}_i$. Specifically the transition model in the asymptotic bias analysis is

$$g(\mu_{ij,x}) = \beta_0 + X_{ij}\beta_x + Y_{ij-1}\alpha, \tag{10}$$

and the unobserved covariate X_{ij} follows the linear transitional model (7).

The naive model simply replaces X_{ij} in (10) with W_{ij} as

$$g(\mu_{ij,w}) = \beta_{0,naive} + W_{ij}\beta_{x,naive} + Y_{ij-1}\alpha_{naive}. \tag{11}$$

The naive estimators are the MLEs under model (11). We assume in our asymptotic investigation the cluster size n is fixed and the number of clusters (subjects) $m \to \infty$, and investigate the asymptotic limits of the naive estimators as $m \to \infty$. We first investigate the effect of cluster size n on the asymptotic biases in naive estimators (Section 3.2), then study such asymptotic biases for Gaussian outcomes (Section 3.3) and for non-Gaussian outcomes (Section 3.4).

3.2 Cluster size effect on the asymptotic bias in naive estimator

We show in Theorem 1 that under some general assumptions, the biases in naive estimators do not depend on the cluster size n. This result differs from that under generalized linear mixed models with covariate measurement error, where the bias increases with the cluster size (Wang, *et al.*, 1998). The proof of Theorem 1 is given in Appendix A.1.

Theorem 1 *Suppose the subject index i. Suppose that the repeated measures $(X_j, Y_j)(j = 1, ..., n)$ are observed from a stationary two-dimensional first-order Markov process. Then the asymptotic biases in naive estimators obtained by ignoring the measurement error via fitting the naive model (11) does not depend on the cluster size n $(n \geq 2)$.*

Theorem 1 suggests that we can simply restrict our asymptotic bias analysis to an arbitrary fixed cluster size n $(n \geq 2)$. We can easily extend the results to the case where (X_j, Y_j) is a q-order stationary process and show that the asymptotic biases in naive estimators are free of the cluster size $n \geq q + 1$. Finally, we emphasize that the stationary process assumption is essential for Theorem 1. If this assumption is violated or the transition model is misspecified, the result in Theorem 1 may not be true. For instance, in one numerical study where data were generated from a non-stationary transition model, the biases in the regression coefficients varied with maximum bias 10% when cluster sizes changed from 5 to 10.

3.3 Asymptotic biases in naive estimators under the linear transition model for Gaussian outcomes

In this section, we study the asymptotic biases in naive estimators of β_x and α under the linear transition model for normally distributed outcomes Y_{ij} and assuming an identity link in (10. Denote the asymptotic limits of the naive estimators of the regression coefficients

$\boldsymbol{\theta}_Y = (\beta_0, \beta_x, \alpha)^T$ by $\boldsymbol{\theta}_{Y,naive} = (\beta_{0,naive}, \beta_{x,naive}, \alpha_{naive})^T$ as $m \to \infty$. Suppress the subject index i and denote by $U_{naive}(\mathbf{Y}, \mathbf{W}; \boldsymbol{\theta}_{Y,naive})$ the score function of the naive model (11). The asymptotic limit of the naive estimator $\boldsymbol{\theta}_{Y,naive}$ solves $E\{U_{naive}(\mathbf{Y}, \mathbf{W}; \boldsymbol{\theta}_{Y,naive})\} = 0$, where the expectation is taken under the true linear model (10) with $g(\cdot) = 1$, (3), (7) and is a function of the true parameter vector $(\boldsymbol{\theta}_Y^T, \sigma^2, \boldsymbol{\theta}_X^T, \sigma_x^2)^T$.

The results in Theorem 1 show the asymptotic biases of the naive estimators do not depend on the cluster size n. Without loss of generality, we set the cluster size $n = 2$ in our asymptotic bias calculations. It can be easily shown that $\boldsymbol{\theta}_{Y,naive}$ satisfies

$$\begin{aligned} E(Y_2 - \beta_{0,naive} - W_2\beta_{x,naive} - Y_1\alpha_{naive}) &= 0 \\ E\{W_2(Y_2 - \beta_{0,naive} - W_2\beta_{x,naive} - Y_1\alpha_{naive})\} &= 0 \\ E\{Y_1(Y_2 - \beta_{0,naive} - W_2\beta_{x,naive} - Y_1\alpha_{naive})\} &= 0. \end{aligned}$$

Some calculations give

$$\begin{bmatrix} 1 & E(X_2) & E(Y_1) \\ E(X_2) & E(X_2^2) + \sigma_u^2 & E(X_2Y_1) \\ E(Y_1) & E(X_2Y_1) & E(Y_1^2) \end{bmatrix} \begin{pmatrix} \beta_0 - \beta_{0,naive} \\ \beta_x - \beta_{x,naive} \\ \alpha - \alpha_{naive} \end{pmatrix} = \begin{pmatrix} 0 \\ \beta_x\sigma_u^2 \\ 0 \end{pmatrix}.$$

It follows that

$$\beta_{x,naive} = \lambda^*\beta_x, \qquad \alpha_{naive} = \alpha + \lambda^{**}, \tag{12}$$

where

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

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

and the expressions of $var(X_2), var(Y_1), cov(X_2, Y_1)$ are given in Appendix A.2.

It can be shown that λ^* and λ^{**} are bounded, and their properties are given in Theorem 2. The proof of Theorem 2 is given in Appendix A.2.

Theorem 2 *Under the stationary condition that* $|\alpha| < 1$ *and* $|\gamma_x| < 1$, *then*

(1) λ^* satisfies

$$\frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2} \le \lambda^* \le \frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2(1-\gamma_x^2)}, \tag{13}$$

(2) λ^{**} has the same sign with γ_x, where γ_x is defined in the $\mathbf{X}$ model (7).

Note that the assumption $|\alpha| < 1$ and $|\gamma_x| < 1$ in Theorem 2 guarantees both the $\mathbf{Y}$ and $\mathbf{X}$ processes being stationary. Theorem 2 shows that if the measurement error is ignored, the naive estimator of the regression coefficient β_x of the covariate X_{ij} is attenuated, while the naive estimator of the regression coefficient α of the past response is inflated if $\gamma_x > 0$ and underestimated if $\gamma_x < 0$. Since the X_{ij} are often positively correlated within each subject, $\gamma_x > 0$ in practice, the naive estimator of α is often inflated in practice. Denote by $\sigma_X^2 = \sigma_x^2/(1-\gamma_x^2)$ the marginal variance of X_{ij} under the $\mathbf{X}$ model (7). The right hand side of (13) can be written as $\sigma_X^2/(\sigma_X^2+\sigma_u^2)$, which is the traditional attenuation factor in a standard linear regression measurement error model (Carroll, et al, 1995). It follows that the attenuation of the naive estimator of the regression coefficient β_x in a linear transition model is usually more severe compared to that in a standard linear regression measurement error model.

In Figure 1, we numerically evaluate the asymptotic relative biases in $\beta_{x,naive}$ and α_{naive} as a function of the measurement error variance σ_u^2 under the linear transition measurement error model. The parameter configurations 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 relative bias is defined as the bias of a parameter estimator divided by its true value. Figure 1 clearly shows that the naive estimator of β_x is attenuated, while the naive estimator of α is inflated. The biases become more severe as σ_u^2 increases.

3.4 Asymptotic bias in naive estimator under the generalized linear transition model for non-Gaussian outcomes

When the response $\mathbf{Y}_i$ is non-Gaussian, the bias analysis is much more complicated and closed form expressions of the asymptotic limits of the naive estimators are usually unavailable. Numerical calculations are hence needed. We first describe the general theoretical results under the generalized linear transition model (10). We next illustrate as an example the detailed numerical calculations of the asymptotic biases in the logistic transition model in (10) with $g(\cdot$ being the logit link.

The naive estimator $\widehat{\boldsymbol{\theta}}_{Y,naive}$ maximizes the naive log likelihood $m^{-1}\sum_{i=1}^{m}\ell_{naive}(\mathbf{Y}_i,\mathbf{W}_i;\boldsymbol{\theta}_{Y,naive})$, where $\ell_{naive}(\mathbf{Y}_i,\mathbf{W}_i;\boldsymbol{\theta}_{Y,naive})$ is the log likelihood function of the ith subject under the naive model (11), and is given in (2) with X_{ij} replaced by W_{ij}. Suppressing the subscript i, the asymptotic limit of the naive estimator $\boldsymbol{\theta}_{Y,naive}$ maximizes the asymptotic limit (as $m\to\infty$) of the naive log likelihood, which is its expectation $E\{\ell_{naive}(\mathbf{Y},\mathbf{W};\boldsymbol{\theta}_{Y,naive})\}$, where the expectation is taken with respect to $(\mathbf{Y},\mathbf{W},\mathbf{X})$ under the true model (10), (3), (7). We have

$$E_{(Y,W,X)}\{\ell_{naive}(\mathbf{Y},\mathbf{W};\boldsymbol{\theta}_{Y,naive})\}=E_X(E_Y[E_U\{\ell_{naive}(\mathbf{Y},\mathbf{W}=\mathbf{X}+\mathbf{U};\boldsymbol{\theta}_{Y,naive})|\mathbf{X}\}]),\qquad(14)$$

which is a function of $\boldsymbol{\theta}_{Y,naive}$ and the true value $(\boldsymbol{\theta}_Y,\boldsymbol{\theta}_X,\sigma_x^2)$. Hence $\boldsymbol{\theta}_{Y,naive}$ maximizes (14) with respect to $\boldsymbol{\theta}_{Y,naive}$ as a function of the true value $(\boldsymbol{\theta}_Y^T,\boldsymbol{\theta}_X^T,\sigma_x^2)^T$. In evaluation of the logistic transition measurement error model, the conditional expectation for binary outcome $\mathbf{Y}$ is simply a discrete summation, and the expectation for $\mathbf{X}$ can be calculated using Gauss-Hermite quadrature or Monte-Carlo method.

As an example, we numerically calculate the asymptotic biases of $\beta_{x,naive}$ and α_{naive} in the logistic transition model. The results are presented in Figure 2, where the relative asymptotic biases of $\beta_{x,naive}$ and α_{naive} are plotted against the measurement error variance σ_u^2. The parameter configurations are the same as those in the linear transition model case. Similar to the results in the linear transition model, the naive estimator of β_x is attenuated, while the naive estimator of α is inflated. As σ_u^2 increases, the biases become larger.

4 Estimation in Generalized Linear Transition Measurement Error Models

The asymptotic bias analysis results in Section 3 show that the naive estimator ignoring the measurement error is asymptotically biased. Valid statistical inference hence requires properly accounting for the measurement error. We propose in this section maximum likelihood estimation by jointly modeling the outcome using the transition model (1) and the measurement error using the structural model (3) and (4). In view of the multi-dimensional integration required by maximizing the likelihood (5), we develop an EM algorithm to calculate the MLEs. We first discuss in Section 4.1 the EM algorithm for linear transition measurement error models for Gaussian outcomes, and extend in Section 4.2 the results to generalized linear transition measurement error models for non-Gaussian outcomes.

4.1 EM algorithm for linear transition measurement error models

In this section, we derive an EM algorithm to compute the maximum likelihood estimator when the response Y_{ij} is normal and follows a linear transition model

$$Y_{ij} = \beta_0 + X_{ij}\beta_x + \mathbf{Z}_{ij}^T\boldsymbol{\beta}_z + \mathbf{H}_{y,ij}^T\boldsymbol{\alpha} + \epsilon_{ij}, \tag{15}$$

where the ϵ_{ij} are independent and follow $N(0, \sigma^2)$. We assume the error-prone covariate W_{ij} follows (3) and the unobserved covariate X_{ij} follows the linear transition model (4).

The complete data are $(\mathbf{Y}, \mathbf{W}, \mathbf{X}, \mathbf{Z})$ and the observed data are $(\mathbf{Y}, \mathbf{W}, \mathbf{Z})$. It is easy to write out the complete data log likelihood function $\ell(\mathbf{Y}, \mathbf{W}, \mathbf{X}|\mathbf{Z})$ as

$$\begin{aligned}
\ell(\mathbf{Y}, \mathbf{W}, \mathbf{X}|\mathbf{Z}) &= \sum_{i=1}^{m}\Big\{ -\frac{n_i - q}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{Y}_i^{(R)} - \mathbf{C}_{Y_i}\boldsymbol{\theta}_Y)^T(\mathbf{Y}_i^{(R)} - \mathbf{C}_{Y_i}\boldsymbol{\theta}_Y) \\
&\quad -\frac{n_i}{2}\ln(2\pi\sigma_u^2) - \frac{1}{2\sigma_u^2}(\mathbf{W}_i - \mathbf{X}_i)^T(\mathbf{W}_i - \mathbf{X}_i) \\
&\quad -\frac{n_i - r}{2}\log(2\pi\sigma_x^2) - \frac{1}{2\sigma_x^2}(\mathbf{X}_i^{(R)} - \mathbf{C}_{X_i}\boldsymbol{\theta}_X)^T(\mathbf{X}_i^{(R)} - \mathbf{C}_{X_i}\boldsymbol{\theta}_X)\Big\},
\end{aligned}$$

where $\mathbf{Y}_i^{(R)} = (Y_{i,q+1}, ..., Y_{in_i})^T$, and $\mathbf{1}_i^{(Y)}$, $\mathbf{X}_i^{(Y)}$ and $\mathbf{Z}_i^{(Y)}$ are defined similarly, $\mathbf{Y}_i^{(H)}$ denotes the history outcomes $\mathbf{Y}_i^{(H)} = (\mathbf{H}_{y,i(q+1)}, ..., \mathbf{H}_{y,in_i})^T$, and $\mathbf{X}_i^{(H)}$ is defined similarly, $\mathbf{C}_{Y_i} = \{\mathbf{1}_i^{(Y)}, \mathbf{X}_i^{(Y)}, \mathbf{Z}_i^{(Y)}, \mathbf{Y}_i^{(H)}\}$ and $\boldsymbol{\theta}_Y = (\beta_0, \beta_x, \boldsymbol{\beta}_z^T, \boldsymbol{\alpha}^T)^T$ denote the design matrix and the regression coefficient vector in the response model (15), and $\mathbf{X}_i^{(R)} = (X_{i,r+1}, ..., X_{in_i})^T$, and $\mathbf{1}_i^{(X)}$ and $\mathbf{Z}_i^{(X)}$ are defined similarly, and $\mathbf{C}_{X_i} = \{\mathbf{1}_i^{(X)}, \mathbf{Z}_i^{(X)}, \mathbf{X}_i^{(H)}\}$ and $\boldsymbol{\theta}_X = (\gamma_0, \boldsymbol{\gamma}_z^T, \boldsymbol{\gamma}_x^T)^T$ denote the design matrix and the regression coefficient vector in the $\mathbf{X}$ model (4).

Denote by $\boldsymbol{\theta} = (\boldsymbol{\theta}_Y^T, \sigma^2, \boldsymbol{\theta}_X^T, \sigma_x^2)^T$ and its MLE by $\hat{\boldsymbol{\theta}}$. Assume the measurement error variance σ_u^2 is known. Let $\hat{\boldsymbol{\theta}}^{(k)}$ be the estimator of $\boldsymbol{\theta}$ at the kth iteration, the M-step updates $\hat{\boldsymbol{\theta}}^{(k)}$ by

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_Y^{(k+1)} &= \left[\sum_{i=1}^m E_X\left\{\mathbf{C}_{Y_i}^T\mathbf{C}_{Y_i}\Big|\,\mathbf{Y},\mathbf{W},\mathbf{Z};\hat{\boldsymbol{\theta}}^{(k)}\right\}\right]^{-1}\left[\sum_{i=1}^m E_X\left\{\mathbf{C}_{Y_i}^T\Big|\,\mathbf{Y},\mathbf{W},\mathbf{Z};\hat{\boldsymbol{\theta}}^{(k)}\right\}\mathbf{Y}_i^{(R)}\right];\\
\widehat{\sigma^2}^{(k+1)} &= \left\{\sum_{i=1}^m(n_i-q)\right\}^{-1}\sum_{i=1}^m E_X\left[\left\{\mathbf{Y}_i^{(R)}-\mathbf{C}_{Y_i}\hat{\boldsymbol{\theta}}_Y^{(k+1)}\right\}^T\left\{\mathbf{Y}_i^{(R)}-\mathbf{C}_{Y_i}\hat{\boldsymbol{\theta}}_Y^{(k+1)}\right\}\Big|\,\mathbf{Y},\mathbf{W},\mathbf{Z};\hat{\boldsymbol{\theta}}^{(k)}\right];\\
\hat{\boldsymbol{\theta}}_X^{(k+1)} &= \left[\sum_{i=1}^m E_X\left\{\mathbf{C}_{X_i}^T\mathbf{C}_{X_i}\Big|\,\mathbf{Y},\mathbf{W},\mathbf{Z};\hat{\boldsymbol{\theta}}^{(k)}\right\}\right]^{-1}\left[\sum_{i=1}^m E_X\left\{\mathbf{C}_{X_i}^T\Big|\,\mathbf{Y},\mathbf{W},\mathbf{Z};\hat{\boldsymbol{\theta}}^{(k)}\right\}\mathbf{X}_i^{(R)}\right];\\
\widehat{\sigma_x^2}^{(k+1)} &= \left\{\sum_{i=1}^m(n_i-r)\right\}^{-1}\sum_{i=1}^m E_X\left[\left\{\mathbf{X}_i^{(R)}-\mathbf{C}_{X_i}\hat{\boldsymbol{\theta}}_X^{(k+1)}\right\}^T\left\{\mathbf{X}_i^{(R)}-\mathbf{C}_{X_i}\hat{\boldsymbol{\theta}}_X^{(k+1)}\right\}\Big|\,\mathbf{Y},\mathbf{W},\mathbf{Z};\hat{\boldsymbol{\theta}}^{(k)}\right].
\end{aligned}$$

The conditional expectations in the above expressions are calculated at the E-step. Specifically, at the E-step, we calculate the conditional expectation of $\mathbf{X}_i^{(Y)}, \mathbf{X}_i^{(Y)^T}\mathbf{X}_i^{(Y)}, \mathbf{X}_i^{(H)}$, $\mathbf{X}_i^{(H)^T}\mathbf{X}_i^{(H)}, \mathbf{X}_i^{(R)}, \mathbf{X}_i^{(R)^T}\mathbf{X}_i^{(R)}, \mathbf{X}_i^{(H)^T}\mathbf{X}_i^{(R)}$ given the observed data $(\mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i), i = 1, ..., m$. Since the joint distribution of $(\mathbf{Y}_i, \mathbf{W}_i)$ is multivariate normal, these conditional expectations are easy to calculate and have closed forms, Their detailed expressions are given in Appendix A.3.

The covariance of the MLE $\widehat{\boldsymbol{\theta}}$ is calculated using the observed information

$$\mathbf{J}_{obs}(\hat{\boldsymbol{\theta}}) = -\left.\frac{\partial^2\ell(\mathbf{Y},\mathbf{W}|\mathbf{Z})}{\partial\boldsymbol{\theta}\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = E_X\left[\left\{\mathbf{J}_c(\hat{\boldsymbol{\theta}})-\mathbf{U}_c(\hat{\boldsymbol{\theta}})\mathbf{U}_c(\hat{\boldsymbol{\theta}})^T\right\}\Big|\,\mathbf{Y},\mathbf{W},\mathbf{Z}\right], \tag{16}$$

where $\mathbf{J}_c(\hat{\boldsymbol{\theta}})$ and $\mathbf{U}_c(\hat{\boldsymbol{\theta}})$ denote the observed information matrix and the score function of the complete data $(\mathbf{Y}, \mathbf{W}, \mathbf{X}, \mathbf{Z})$ (Louis, 1982). Since $(\mathbf{Y}, \mathbf{W}, \mathbf{X}|\mathbf{Z})$ is multivariate normal, (16) has a closed form and can be easily calculated.

4.2 The EM algorithm for generalized linear transition measurement error models

In this section, we discuss the EM algorithm for the generalized linear transition measurement error model (1) for non-Gaussian outcomes. We still assume $\mathbf{W}$ and $\mathbf{X}$ follow (3) and (4). Examination of equation (5) suggests that the likelihood function $L(\mathbf{Y}, \mathbf{W}|\mathbf{Z})$ does not have a closed form anymore and requires numerical integration. One can easily see that the estimators of $\boldsymbol{\theta}_X$ and σ_x^2 in the M-step can be updated in the same way as in Section 4.1. Estimation of $\boldsymbol{\theta}_Y$ in the M-step needs to be modified and requires solving the conditional score equation

$$\sum_{i=1}^{m} E_{X_i}\left[\mathbf{C}_{Y_i}^T \boldsymbol{\Delta}_i \mathbf{V}_i^{-1}\left\{\mathbf{Y}_i^{(R)} - \boldsymbol{\mu}_{i,x}\right\}|\mathbf{Y},\mathbf{W},\mathbf{Z};\hat{\boldsymbol{\theta}}^{(k)}\right] = 0, \tag{17}$$

where $\mathbf{C}_{Y_i}$ is defined at the beginning of Section 4, $\boldsymbol{\mu}_{i,x} = E\left\{\mathbf{Y}_i^{(R)}|\mathbf{X}_i^{(Y)},\mathbf{Z}_i^{(Y)},\mathbf{Y}_i^{(H)}\right\}$ with the jth component $\mu_{ij,x}$ given in (1), $\mathbf{V}_i = \text{diag}\{var(Y_{ij}|X_{ij},\mathbf{Z}_{ij},\mathbf{H}_{y,ij})\}$, and $\boldsymbol{\Delta}_i = \text{diag}\{\mu_{ij,x}^{(1)}(\cdot)\}$ and $\mu_{ij,x}^{(1)}(\cdot)$ denotes the first derivative of $\mu_{ij,x}$. The Fisher scoring method is used to solve (17).

At the E-step, one needs to calculate the conditional expectations of the form $E_{X_i}\{T_i(\mathbf{Y}_i,\mathbf{W}_i,\mathbf{X}_i,\mathbf{Z}_i;\boldsymbol{\theta})\}$ for some function $T(\cdot)$, which can be written as

$$\frac{\int T_i(\mathbf{Y}_i,\mathbf{W}_i,\mathbf{X}_i,\mathbf{Z}_i;\boldsymbol{\theta})f(\mathbf{Y}_i|\mathbf{X}_i,\mathbf{Z}_i;\hat{\boldsymbol{\theta}}^{(k)})f(\mathbf{W}_i|\mathbf{X}_i;\hat{\boldsymbol{\theta}}^{(k)})f(\mathbf{X}_i|\mathbf{Z}_i;\hat{\boldsymbol{\theta}}^{(k)})d\mathbf{X}_i}{\int f(\mathbf{Y}_i|\mathbf{X}_i,\mathbf{Z}_i;\boldsymbol{\theta}^{(k)})f(\mathbf{W}_i|\mathbf{X}_i;\hat{\boldsymbol{\theta}}^{(k)})f(\mathbf{X}_i|\mathbf{Z}_i;\hat{\boldsymbol{\theta}}^{(k)})d\mathbf{X}_i},$$

where $f(\cdot)$ denotes a density function. Both the numerator and the denominator can be evaluated using Monte Carlo simulations by generating $\mathbf{X}_i$ from the conditional distribution $\mathbf{X}_i|(\mathbf{W}_i,\mathbf{Z}_i)$, which is multivariate normal.

We found that calculations of the covariance of $\hat{\boldsymbol{\theta}}$ using (16) are not stable for non-Gaussian cases. In particular, the information matrix calculated using (16) may not be positive definite. We hence estimate the observed information using $\mathbf{J}_{obs}(\hat{\boldsymbol{\theta}}) = m^{-1}\sum_{i=1}^m \mathbf{U}_{i,obs}(\hat{\boldsymbol{\theta}})\mathbf{U}_{i,obs}(\hat{\boldsymbol{\theta}})^T$, where $\mathbf{U}_{i,obs}(\hat{\boldsymbol{\theta}})$ is the score of the observed data $(\mathbf{Y}_i,\mathbf{W}_i,\mathbf{Z}_i)$ for the ith subject.

5 Simulation Study

We conducted two simulation studies to evaluate the finite sample performance of the MLEs, one for normal outcomes under linear transition measurement error models (15) and one for binary outcomes under logistic transition measurement error models. We compared the MLEs with the estimates obtained using the true $\mathbf{X}$ and the naive estimates obtained by ignoring the the measurement error with $\mathbf{X}$ replaced by $\mathbf{W}$. We set the number of subjects $m = 100$ and the cluster size $n = 5$. The covariate Z_{ij} was assumed to be a treatment indicator, with half subjects being treated and half subjects not being treated. The **W** and **X** were generated under (3) and (4). For simplicity, to evaluate the performance of the MLE estimates, we assumed in the simulation studies first-order transition models for both the outcome variable Y_{ij} and the unobserved covariate X_{ij}, i.e. $q = r = 1$.

The true parameter values were: (1) $\beta_0 = -1, \beta_x = 1, \beta_z = 0.8, \alpha = 0.5$, ($\sigma^2 = 1$ for the normal case); (2) $\gamma_0 = 0.4, \gamma_z = 0.5, \gamma_x = 0.6, \sigma_x^2 = 0.5$; (3) $\sigma_u^2 = 0.5$ (assumed to be known). A total of 1000 replications were used for the normal case; and 500 replications for the binary case. The results are presented in Table 1. "True X" indicates that the results using the true covariate **X** (unobserved in practice) by fitting model (1). "Naive" corresponds to the results when the measurement error is ignored by fitting (1) with **X** replaced by **W**; "MLE" corresponds to the MLEs using the observed data **W** calculated using the EM algorithm in Section 4 to account for the measurement error. Programming was done in SAS/IML.

Our simulation results show that the estimates using the unobserved "true" **X** performs the best both in terms of biases and standard errors. The MLEs using the correct measurement error structure perform very well and have little bias, while the naive estimators are biased. In particular, the coefficient β_x of the unobserved covariate X_{ij} is attenuated; while the coefficient α of the past response Y_{ij-1} is inflated when measurement error is ignored. This result is consistent with our asymptotic bias analysis result in Section 3. We also noticed the trade-off between the bias and the variance. The MLEs effectively correct the biases in naive estimators, but have larger SEs. Using mean square errors (MSEs) as a measure of overall performance, the MLEs perform substantially better than the naive estimates and have substantially smaller MSEs. The estimated SEs and empirical SEs agree well.

The orders q and r in both the transition model of Y_{ij} and the transition model of X_{ij} are important, and their misspecification could induce considerable bias in the parameter estimates. To see this, we conducted an additional simulation study, where the outcome **Y** and the true covariate **X** were generated from a third-order and a second-order linear transition model respectively ($q = 3, r = 2$). The parameter setting was the same as in the previous simulation, except for the coefficients of the historical observations: $\alpha_1 = 0.4, \alpha_2 = 0.6, \alpha_3 = -0.5, \gamma_{x1} = 0.1, \gamma_{x2} = 0.5$. The sample size was 100 or 200. In the simulation study, we misspecified the transition models by assuming $q = r = 1$ and ran 1000 simulations. The average estimates of the regression coefficients of X_{ij} and Z_{ij} in the outcome model had relative biases as much as 15% and 20% respectively.

These results show that selection of the orders of the transition models is important in practice. We considered the AIC and the BIC. To examine different model selection methods to choose (q, r), for each of the above simulation data set, we further analyzed the data assuming different order transition models by varying (q, r) between 0 and 3 and maximized the AIC and the BIC to select the best model. The AIC is the twice log-likelihood function minus twice the number of parameters, and the BIC is the difference between twice the log-likelihood function and log(sample size) times the number of parameters. The AIC and BIC model selection results are summarized in Table 2. We considered the sample size $n = 100$ and $n = 200$. Our results show that the BIC works much better than the AIC and has a very high chance (93.4%) of choosing the correct model in 1000 simulations. We hence will use the BIC in the data example.

6 Application to the Women Take Pride Study

We apply the proposed transition measurement error model to the analysis of the "Women Take PRIDE" study (Janevic, *et al.*, 2002). The study involved 570 women who were aged 60 or older and had cardiac disease. One study objective was to explore the effects of social support on health outcomes. Since the social support level was measured using the average of a few questions in the questionnaire, it was subject to considerable measurement error.

After a telephone baseline interview, the participants were randomly assigned to either a control group or a 4-week intervention group designed to improve self-care ability. Subsequent telephone interviews were conducted at 4, 12, and 18 months. The social support level was calculated using the average of the scores of several questions concerning social support in the questionnaire and hence measured the true social support level with considerable error. The social support questions were asked at each followup time. The range of the social support level is continuous from 1 to 5 with a higher level indicating less support. A log-transformation was performed to make the normality assumption more plausible.

Due to the repeated measure nature of the data, examination of (4) and (3) shows that we can estimate σ_u^2 by fitting a linear mixed model for $\mathbf{W}$ on $\mathbf{Z}$ with the random effect following the AR(1) correlation structure. The estimated residual variance thus estimates the measurement error variance. For our data, σ_u^2 was estimated as 0.053, which was about one-fourth of the variation of $\mathbf{W}$ and indicated moderate measurement error.

Note that one could treat measurement error in social support as a misclassification problem (Espeland et al., 1987). However, since social support is an average of multiple questions, it is continuous. Treating it as a misclassified covariate, it would have many categories and model parameters would be difficult to interpret. Further, by treating it as a continuous variable, one could also estimate σ_u^2 from the data.

The health outcome is symptom bothersomeness. This score assesses the bothersomeness of 14 symptoms common to patients with heart disease. For each symptom, bothersomeness is assessed with the question "How much would you say this symptom bothers you?" with a response scale from 1="not at all" to 5="a lot". A symptom bothersomeness score was then calculated by summing the responses of the 14 symptoms. The resulting possible total score ranged from 0 to 70, with higher scores indicating greater symptom bothersomeness. A log transformation was performed to make the normality assumption more plausible.

We fit the data using different orders of the proposed transition measurement error models. The covariates $\mathbf{Z}$ included an intervention indicator, age, race(white/nonwhite), and education level(lower than high school, high school, higher than high school). In view of our simulation results, the BIC was used to choose the best model. The BIC chose the first-order transition models for both the outcome Y and the error-prone covariate X. The results obtained from the best model are presented in Table 3. The results show that women with more social support have less symptom bothersomeness. The intervention reduces symptom bothersomeness. None of the demographics has significant effects. A comparison of the naive estimates with the MLEs shows that the naive estimate of the social support effect is severely attenuated, while the naive estimate of the past outcome effect is inflated. These results are consistent with our theoretical findings in Section 3. The naive estimates and the MLEs of the coefficients of the other covariates were similar, indicating the measurement error in social support had little effects on the coefficients of the other covariates. This might be due to the small residual variance of e^*_{xi} in equation (9).

To examine the effect of σ^2_u, we refit the model to the data by assuming $\sigma^2_u = 0.025$ and$\sigma^2_u = 0.1$. The BIC still chose the first order transition models as the best model and the estimates of the coefficients are similar to what is presented in Table 3.

7 Discussion

This paper focuses on structural modelling for transition model with measurement errors. In the measurement error literature, an alternative method is called functional modelling. Structural modeling assumes a distribution for the unobserved covariate **X** and functional modeling does not assume a distribution for **X**. In principle, compared with the functional modeling approach, the structural modeling approach is likely to yield more efficient estimators when the **X** model is correctly specified but less robust estimators when the **X** model is misspecified. We focus in this paper on the structural modeling approach by specifying a linear transition model for **X** and the consistency of the MLEs requires the **X** model is correctly specified. We have also developed a functional approach for transition measurement error models, e.g., using SIMEX (Carroll, *et al.*, 1995). The results will be reported elsewhere. Although we consider in this paper a scalar unobserved covariate X_{ij}, our results can be easily extended to vector unobserved covariates $\mathbf{X}_{ij}$ in higher-order transition models.

We use the AIC and BIC to choose appropriate orders in the transition models. Our finite sample study indicates that the BIC outperforms the AIC and the BIC shows superior performance in our model. Other model selection methods might be used, such the FIC (focused information criterion) recently proposed by (Hjort and Claeskens, 2003) and the cross-validation method. For future research, it would be interesting to compare the performance of all these methods when applied to our model. It should be noted that when the number of repeated measures is small, there might not be sufficient information in the data to estimate the orders of transition models. For high order transition models, high-dimensional numerical methods will have to be used for the calculations in the E-step and computation complexity will increase dramatically for non-Gaussian data.

ACKNOWLEDGMENT

Lin's research was supported by National Cancer Institute grant CA–76404 and National Heart, Lung and Blood Institute grant HL–58611. The authors thank Noreen Clark and Nancy Janz for providing us the Women Take Pride data, and the associate editor and the referees for very helpful comments that have considerably strengthened the paper.

REFERENCES

Buonaccorsi, J., Demidenko, E., and Tosteson, T. (2000). Estimation in Longitudinal Random Effects Models with Measurement Error *Statistica Sinica* **10**, 885-903.

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

Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*, 2nd edition. Oxford.

Espeland, M. A. and Hui, S. L. (1987). A General Approach to Analyzing Epidemiologic Data That Contains Misclassification Errors. *Biometrics* **43**, 1001-1012.

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

Janevic, M. R., Janz, N. K., Dodge, J. A., Wang, Y., Lin, X., and Clark, N. M. (2004). Longitudinal effects of social support on the health and functioning of older women with heart disease. *The International Journal of Aging & Human Development*. Accepted.

Hjort, N. L. and Claeskens, G. (2003) Frequent model average estimators (with discussions). *Journal of American Statistical Association* , **98**, 879-945.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* **44**, 226-233.

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

Schmid, C. H. (1996). An EM Algorithm Fitting First-Order Conditional Autoregressive Models to Longitudinal Data. *Journal of American Statistical Association* **91**, 1322-1330.

Schmid, C. H., Segal, M. R., and Rosner B. (1994). Incorporating Measurement Error in the Estimation of Autoregressive Models for Longitudinal Data. *Journal of Statistical Planning and Inference* **42**, 1-18.

Tosteson, T. D., Buonaccorsi, J. P., and Demidenko, E. (1998). Covariate Measurement Error and the Estimation of Random Effect Parameters in a Mixed Model for Longitudinal Data. *Statistics in Medicine* **17**, 1959-1971.

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. *Journal of American Statistical Association* **90**, 27-37.

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

Appendix

A.1 Proof of Theorem 1

Under the transition structure of the naive model, the expectation $E\{\ell_{naive}(\mathbf{Y}, \mathbf{W}; \boldsymbol{\theta}_{Y,naive})\}$ is

$$\begin{aligned} E\{\ell_{naive}(\mathbf{Y}, \mathbf{W}; \boldsymbol{\theta}_{Y,naive})\} &= \sum_{j=2}^{n} E\{\ell_{naive}(Y_j|W_j, Y_{j-1}; \boldsymbol{\theta}_{Y,naive})\} \\ &= \sum_{j=2}^{n} E_X\{E_{Y,U}\{\ell_{naive}(Y_j|W_i = X_j + U_j, Y_{j-1}; \boldsymbol{\theta}_{Y,naive})]\} \\ &= \sum_{j=2}^{n} \int \ell_{naive}(Y_j|W_j = X_j + U_j, Y_{j-1}; \boldsymbol{\theta}_{Y,naive}) \times \\ & \quad f(Y_j, X_j, Y_{j-1}) f(U_j) dY_j dX_j dY_{j-1} dU_j, \end{aligned}$$

where $f(Y_j, X_j, Y_{j-1})$ is the density of (Y_j, X_j, Y_{j-1}) under the true model (10) and (7), $f(U_j)$ is the density of the measurement error following $N(0, \sigma_u^2)$. Since both U_j and (X_j, Y_j) are both stationary processes, $f(Y_j, X_j, Y_{j-1})$ and $f(U_j)$ do not depend on j. It follows that

$$E\{\ell_{naive}(\mathbf{Y}, \mathbf{W}; \boldsymbol{\theta}_{Y,naive})\} = (n-1)E\{\ell_{naive}(Y_j|W_j = X_j + U_j, Y_{j-1}; \boldsymbol{\theta}_{Y,naive})\}.$$

Hence, the asymptotic limit of the naive estimator $\boldsymbol{\theta}_{Y,naive}$, which maximizes $E\{\ell_{naive}(\mathbf{Y}, \mathbf{W}; \boldsymbol{\theta}_{Y,naive})\}$, does not depend on cluster size n ($n \geq 2$).

A.2 Proof of Theorem 2

Suppress subscript i. We first note the inequality $(a+c)/(b+c) \geq a/b$ for $c > 0, b \geq a > 0$. Let $a = var(X_2)var(Y_1) - cov^2(X_2, Y_1)$, $b = \{var(X_2) + \sigma_u^2\}var(Y_1) - cov^2(X_2, Y_1)$ and $c = cov^2(X_2, Y_1)$. Noting all three quantities are positive, simple calculations give

$$0 \leq \lambda^* \leq \frac{var(X_2)}{var(X_2) + \sigma_u^2} \leq 1.$$

Some calculations show that

$$var(X_j) = \frac{\sigma_x^2}{1-\gamma_x^2}, \quad var(Y_j) = \frac{\sigma^2 + \beta_x^2 \frac{\sigma_x^2}{1-\gamma_x^2} + \frac{2\alpha\beta_x^2\gamma_x\sigma_x^2}{(1-\alpha\gamma_x)(1-\gamma_x^2)}}{1-\alpha^2}, \quad cov(X_j, Y_{j-1}) = \frac{\beta_x\gamma_x\sigma_x^2}{(1-\alpha\gamma_x)(1-\gamma_x^2)}.$$

It follows that λ^* can be written as

$$\lambda^* = \frac{\sigma_x^2 var(Y_1) + \frac{\sigma_x^2\gamma_x^2}{1-\gamma_x^2} var(Y_1) - cov^2(X_2, Y_1)}{(\sigma_x^2+\sigma_u^2) var(Y_1) + \frac{\sigma_x^2\gamma_x^2}{1-\gamma_x^2} var(Y_1) - cov^2(X_2, Y_1)}.$$

Let

$$c = \frac{\sigma_x^2\gamma_x^2}{1-\gamma_x^2} var(Y_1) - cov^2(X_2, Y_1) = \frac{\sigma_x^2\gamma_x^2(\sigma^2 + \frac{\alpha^2\beta_x^2\sigma_x^2}{(1-\alpha\gamma_x)^2})}{(1-\alpha^2)(1-\gamma_x^2)} \geq 0,$$

and a and b be the numerator and the denominator of λ^*, application of the (a, b, c) inequality at the beginning of this section gives $\lambda^* \geq \sigma_x^2(\sigma_x^2+\sigma_u^2)^{-1}$. Using the above expression of $\text{cov}(X_2, Y_1)$, one can easily see that $\lambda^{**}\gamma_x \geq 0$. Hence λ^{**} has the same sign with γ_x.

A.3 The E-step for calculating the MLEs in linear transition measurement error models

The conditional expectations of $\mathbf{X}_i^{(Y)}, {\mathbf{X}_i^{(Y)}}^T\mathbf{X}_i^{(Y)}, \mathbf{X}_i^{(H)}, {\mathbf{X}_i^{(H)}}^T\mathbf{X}_i^{(H)}, \mathbf{X}_i^{(R)}, {\mathbf{X}_i^{(R)}}^T\mathbf{X}_i^{(R)}, {\mathbf{X}_i^{(H)}}^T\mathbf{X}_i^{(R)}$ given the observed data $(\mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i)$ $(i = 1, ..., m)$ need to be calculated in the E-step. Note that $\mathbf{X}_i^{(Y)} = \mathbf{L}_i^{(Y)}\mathbf{X}_i$, $\mathbf{X}_i^{(R)} = \mathbf{L}_i^{(R)}\mathbf{X}_i$, where $\mathbf{L}_i^{(Y)} = \begin{pmatrix} \mathbf{0}_{(n_i-q)\times q} & \mathbf{I}_{(n_i-q)} \end{pmatrix}$ and $\mathbf{L}_i^{(R)} = \begin{pmatrix} \mathbf{0}_{(n_i-r)\times r} & \mathbf{I}_{(n_i-r)} \end{pmatrix}$. Let $\mathbf{Q}_i^{(s)} = \begin{pmatrix} \mathbf{0}_{r\times(s-1)} & \tilde{\mathbf{I}}_r & \mathbf{0}_{r\times(n_i-r-s+1)} \end{pmatrix}$ for $s = r+1, ..., n_i$, where $\tilde{\mathbf{I}}_r$ is an $r \times r$ matrix whose $(l, r-l+1)$th elements $(l = 1, .., r)$ are 1's but 0's elsewhere, then

$$\mathbf{X}_i^{(H)} = (\mathbf{H}_{x,i(r+1)}, ..., \mathbf{H}_{x,in_i}) = \begin{pmatrix} (\mathbf{Q}_i^{(r+1)}\mathbf{X}_i)^T \\ \vdots \\ (\mathbf{Q}_i^{(n_i)}\mathbf{X}_i)^T \end{pmatrix}.$$

Therefore, the components concerning $\mathbf{X}_i$ in $\ell(\mathbf{Y}_i, \mathbf{W}_i, \mathbf{X}_i|\mathbf{Z}_i)$ are given by

$$-\frac{1}{2\sigma^2}(\mathbf{Y}_i^{(R)} - \mathbf{1}_i^{(Y)}\beta_0 - \mathbf{Z}_i^{(Y)}\boldsymbol{\beta}_z - \mathbf{Y}_i^{(H)}\boldsymbol{\alpha} - \beta_x\mathbf{L}_i^{(Y)}\mathbf{X}_i)^T(\mathbf{Y}_i^{(R)} - \mathbf{1}_i^{(Y)}\beta_0 - \mathbf{Z}_i^{(Y)}\boldsymbol{\beta}_z - \mathbf{Y}_i^{(H)}\boldsymbol{\alpha} - \beta_x\mathbf{L}_i^{(Y)}\mathbf{X}_i)$$
$$-\frac{1}{2\sigma_u^2}(\mathbf{W}_i - \mathbf{X}_i)^T(\mathbf{W}_i - \mathbf{X}_i)$$
$$-\frac{1}{2\sigma_x^2}(\mathbf{L}_i^{(R)}\mathbf{X}_i - \mathbf{1}_i^{(X)}\gamma_0 - \mathbf{Z}_i^{(X)}\boldsymbol{\gamma}_z - \boldsymbol{\gamma}_x^T\mathbf{Q}_i\mathbf{X}_i)^T(\mathbf{L}_i^{(R)}\mathbf{X}_i - \mathbf{1}_i^{(X)}\gamma_0 - \mathbf{Z}_i^{(X)}\boldsymbol{\gamma}_z - \boldsymbol{\gamma}_x^T\mathbf{Q}_i\mathbf{X}_i),$$

where $\mathbf{Q}_i = ({\mathbf{Q}_i^{(r+1)}}^T, \cdots, {\mathbf{Q}_i^{(n_i)}}^T)^T$. Thus, it is easy to see that given $(\mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i)$, $\mathbf{X}_i$ follows a multivariate-normal distribution with mean $-\mathbf{A}_i^{-1}\mathbf{b}_i$ and covariance matrix $\mathbf{A}_i^{-1}$, where

$$\mathbf{A}_i((\mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i); \boldsymbol{\theta}) = \Bigg\{\frac{\beta_x^2}{\sigma^2}{\mathbf{L}_i^{(Y)}}^T\mathbf{L}_i^{(Y)} + \frac{1}{\sigma_u^2}\mathbf{I}_{(n_i)}$$

$$
\begin{aligned}
& +\frac{1}{\sigma_x^2}\Big(\mathbf{L}_i^{(R)} - \begin{pmatrix} \boldsymbol{\gamma}_x^T \mathbf{Q}_i^{(r+1)} \\ \vdots \\ \boldsymbol{\gamma}_x^T \mathbf{Q}_i^{(n_i)} \end{pmatrix}\Big)^T \Big(\mathbf{L}_i^{(R)} - \begin{pmatrix} \boldsymbol{\gamma}_x^T \mathbf{Q}_i^{(r+1)} \\ \vdots \\ \boldsymbol{\gamma}_x^T \mathbf{Q}_i^{(n_i)} \end{pmatrix}\Big)\Big\}, \\
\mathbf{b}_i(\mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i; \boldsymbol{\theta}) \quad = \quad & -\frac{1}{\sigma^2}\beta_x \mathbf{L}_i^{(Y)^T}(\mathbf{Y}_i^{(R)} - \mathbf{1}_i^{(Y)}\beta_0 - \mathbf{Z}_i^{(Y)}\boldsymbol{\beta}_z - \mathbf{Y}_i^{(H)}\boldsymbol{\alpha}) - \frac{1}{\sigma_u^2}\mathbf{W}_i \\
& -\frac{1}{\sigma_x^2}(\mathbf{L}_i^{(R)} - \begin{pmatrix} \boldsymbol{\gamma}_x^T \mathbf{Q}_i^{(r+1)} \\ \vdots \\ \boldsymbol{\gamma}_x^T \mathbf{Q}_i^{(n_i)} \end{pmatrix})^T(\gamma_0 + \mathbf{Z}_i^{(X)}\boldsymbol{\gamma}_z).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
E\left\{\mathbf{X}_i^{(Y)} | \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\} &= -\left\{\mathbf{A}_i^{(k)}\right\}^{-1} \mathbf{b}_i^{(k)} \mathbf{L}_i^{(Y)^T}, \\
E\left\{\mathbf{X}_i^{(Y)^T}\mathbf{X}_i^{(Y)} | \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\} &= \operatorname{trace}\left\{\mathbf{L}_i^{(Y)}\left\{\mathbf{A}_i^{(k)}\right\}^{-1}\mathbf{L}_i^{(Y)^T}\right\} \\
& \quad + E\left\{\mathbf{X}_i^{(Y)} | \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\}^T E\left\{\mathbf{X}_i^{(Y)} | \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\}, \\
E\left\{\mathbf{X}_i^{(R)} | \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\} &= -\left\{\mathbf{A}_i^{(k)}\right\}^{-1} \mathbf{b}_i^{(k)} \mathbf{L}_i^{(R)^T}, \\
E\left\{\mathbf{X}_i^{(R)^T}\mathbf{X}_i^{(R)} \middle| \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\} &= \operatorname{trace}\left\{\mathbf{L}_i^{(R)}\left\{\mathbf{A}_i^{(k)}\right\}^{-1}\mathbf{L}_i^{(R)^T}\right\} \\
& \quad + E\left\{\mathbf{X}_i^{(R)} | \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\}^T E\left\{\mathbf{X}_i^{(R)} | \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\}, \\
E\left\{\mathbf{X}_i^{(H)} \middle| \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\} &= -\begin{pmatrix} (\mathbf{Q}_i^{(r+1)}\left\{\mathbf{A}_i^{(k)}\right\}^{-1}\mathbf{b}_i^{(k)})^T \\ \vdots \\ (\mathbf{Q}_i^{(n_i)}\left\{\mathbf{A}_i^{(k)}\right\}^{-1}\mathbf{b}_i^{(k)})^T \end{pmatrix}, \\
E\left\{\mathbf{X}_i^{(H)^T}\mathbf{X}_i^{(H)} \middle| \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\} &= \sum_{s=r+1}^{n_i} \mathbf{Q}_i^{(s)}\left\{\mathbf{A}_i^{(k)}\right\}^{-1}\mathbf{Q}_i^{(s)^T} \\
& \quad + E\left\{\mathbf{X}_i^{(H)} \middle| \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\}^T E\left\{\mathbf{X}_i^{(H)} \middle| \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\}, \\
E\left\{\mathbf{X}_i^{(H)^T}\mathbf{X}_i^{(R)} \middle| \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\} &= \sum_{s=r+1}^{n_i} \mathbf{Q}_i^{(s)}\left\{\mathbf{A}_i^{(k)}\right\}^{-1}\mathbf{L}_{i,s-r}^{(R)\;T} \\
& \quad + E\left\{\mathbf{X}_i^{(H)} \middle| \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\}^T E\left\{\mathbf{X}_i^{(R)} \middle| \mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)}\right\},
\end{aligned}
$$

where $\mathbf{A}_i^{(k)} = \mathbf{A}_i(\mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)})$, $\mathbf{b}_i^{(k)} = \mathbf{b}_i(\mathbf{Y}_i, \mathbf{W}_i, \mathbf{Z}_i, \boldsymbol{\theta}^{(k)})$, and $\mathbf{L}_{i,s-r}^{(R)}$ is the $(s-r)$th row of $\mathbf{L}_i^{(R)}$.

Table 1: Simulation results of the linear and logistic transition measurement error models based on 1000 replicates (linear) and 500 replicates (logistic) with the number of subjects $m = 100$ and cluster size $n = 5$, measurement error variance $\sigma_u^2 = 0.5$

Model	Parameter	True Value	Method	Mean	Est. SE	Emp. SE	MSE
Linear	β_x	1	True X	1.000	0.063	0.062	0.004
			Naive	0.559	0.053	0.054	0.198
			MLE	0.959	0.099	0.096	0.011
	β_z	0.8	True X	0.817	0.153	0.153	0.024
			Naive	1.006	0.172	0.167	0.071
			MLE	0.782	0.187	0.181	0.033
	α	0.5	True X	0.496	0.030	0.030	0.001
			Naive	0.584	0.033	0.031	0.008
			MLE	0.517	0.038	0.036	0.002
Logistic	β_x	1	True X	1.027	0.177	0.178	0.032
			Naive	0.579	0.126	0.123	0.193
			MLE	1.022	0.263	0.239	0.058
	β_z	0.8	True X	0.799	0.334	0.331	0.110
			Naive	1.199	0.315	0.326	0.265
			MLE	0.803	0.395	0.380	0.144
	α	0.5	True X	0.466	0.289	0.280	0.080
			Naive	0.602	0.279	0.271	0.084
			MLE	0.495	0.322	0.290	0.084

Est. SE: Estimated SE; Emp. SE: Empirical SE.

Table 2: Frequencies of Selecting (q, r)-models from 1000 Samples

Sample size	Method	Frequency table				
$n = 100$	AIC		$r = 1$	$r = 2$	$r = 3$	$r = 4$
		$q = 1$	0	0	0	0
		$q = 2$	0	0	0	0
		$q = 3$	0	645	134	58
		$q = 4$	0	130	20	13
	BIC		$r = 1$	$r = 2$	$r = 3$	$r = 4$
		$q = 1$	0	0	0	0
		$q = 2$	0	0	0	0
		$q = 3$	0	934	39	1
		$q = 4$	0	25	1	0
$n = 200$	AIC		$r = 1$	$r = 2$	$r = 3$	$r = 4$
		$q = 1$	0	0	0	0
		$q = 2$	0	0	0	0
		$q = 3$	0	672	115	65
		$q = 4$	0	119	15	14
	BIC		$r = 1$	$r = 2$	$r = 3$	$r = 4$
		$q = 1$	0	0	0	0
		$q = 2$	0	0	0	0
		$q = 3$	0	945	29	5
		$q = 4$	0	19	2	0

Table 3: Parameter estimates for the Women take PRIDE study under a first-order linear transition model for X and Y with $\sigma_u^2 = 0.053$

	MLE		Naive	
Parameter	Estimate	SE	Estimate	SE
Social support (β_x)	0.213	0.063	0.147	0.052
Intervention	-0.08	0.044	-0.085	0.044
Age	0.004	0.0040	0.004	0.0039
Race	0.006	0.073	-0.006	0.073
Education (HS vs. <HS)	-0.01	0.058	-0.005	0.058
Education (>HS vs. <HS)	-0.09	0.059	-0.083	0.059
Past symptom bothersome (α)	0.552	0.031	0.559	0.030

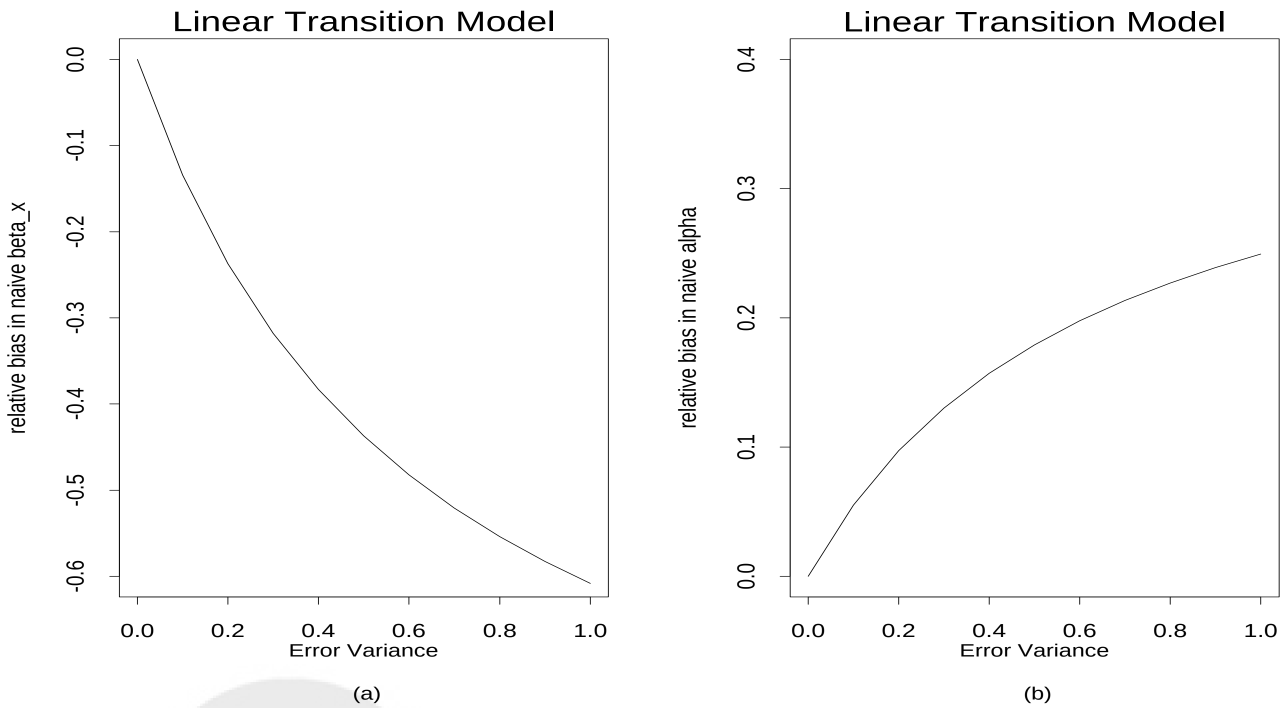

Figure 1: Asymptotic relative biases in naive estimators of β_x and α by ignoring measurement error in the first order linear transition measurement error model for Gaussian outcomes. The true parameter values are $\beta_0 = -1, \beta_x = 1, \alpha = 0.5, \sigma^2 = 1$, and $\gamma_0 = 0.4, \gamma_x = 0.6, \sigma_x^2 = 0.5$. The two plots correspond to (a) the relative bias in $\beta_{x,naive}$; (b) the relative bias in α_{naive}.

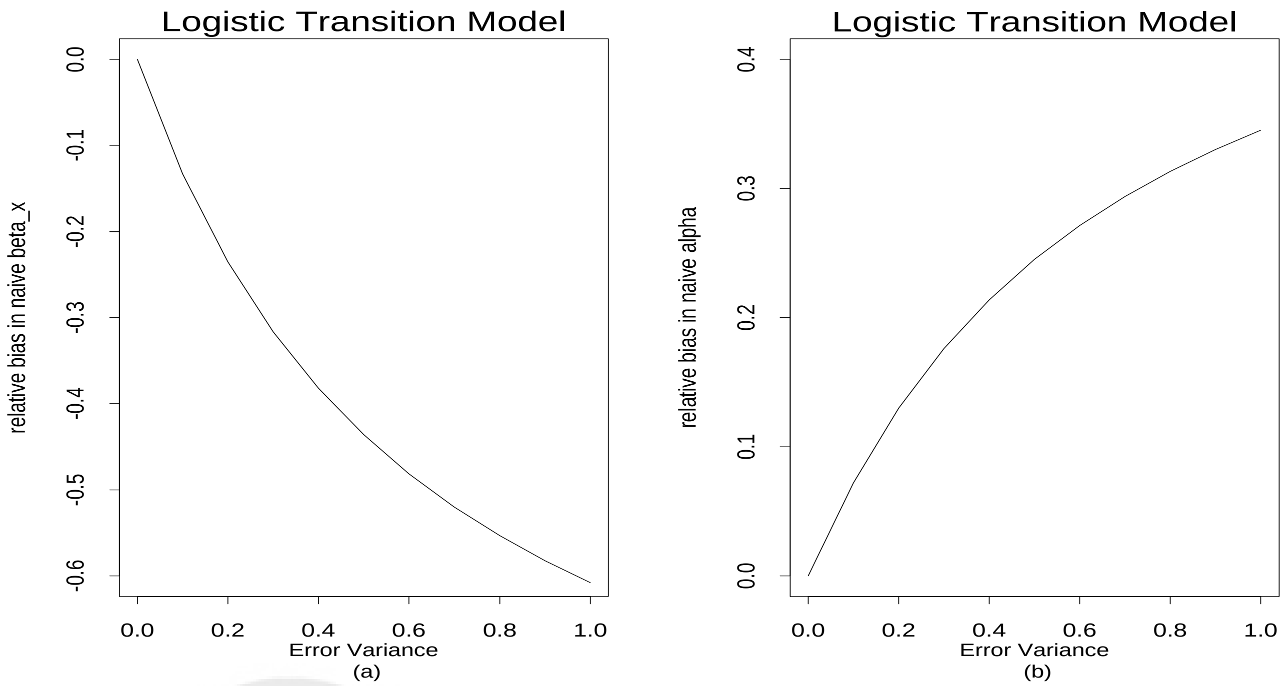

Figure 2: Asymptotic relative biases in naive estimates of β_x and α by ignoring measurement error in the first-order logistic transition measurement error model for binary outcomes. The true parameter values are $\beta_0 = -1, \beta_x = 1, \alpha = 0.5$, and $\gamma_0 = 0.4, \gamma_x = 0.6, \sigma_x^2 = 0.5$. The two plots correspond to (a) relative bias in $\beta_{x,naive}$; (b) relative bias in α_{naive}.