

# **Estimation of Effects of Endogenous Time-Varying Covariates: A Comparison Of Multilevel Linear Modeling and Generalized Estimating Equations**

Research Report

**Ward B. Eiling (9294163)**

Supervisors: Ellen Hamaker and Jeroen Mulder

*Master's degree in Methodology and Statistics for the Behavioural,  
Biomedical and Social Sciences*

*Utrecht University*

December 10, 2024

Word count: 2500/2500

FETC-approved: 24-2003

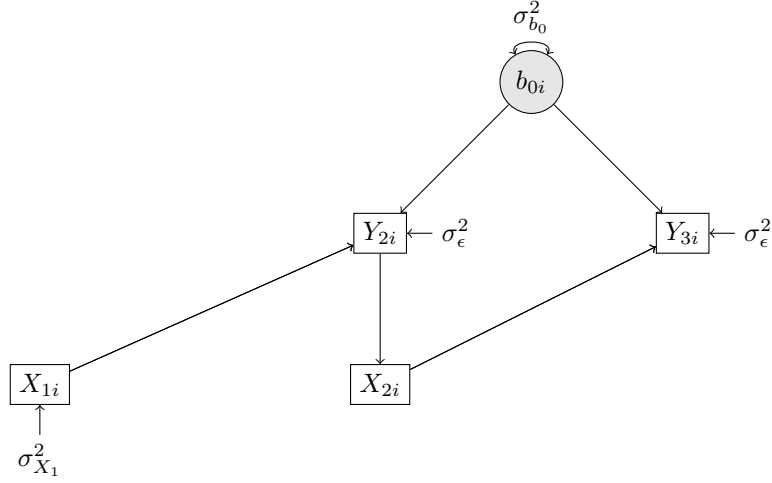
*Candidate journal: Psychological Methods*

# 1 Introduction

Across a wide range of disciplines, researchers analyze clustered longitudinal, observational data to investigate prospective causal relationships between variables. When analyzing such data, the psychological sciences most commonly resort to the multilevel linear model (MLM, [McNeish et al., 2017](#)), which—in the context of longitudinal data analysis—separates observed variance into stable between-person differences and within-person fluctuations ([Hamaker & Muthén, 2020](#)). Conversely, other fields, such as biostatistics and econometrics often favour generalized estimating equations (GEE) for the analysis of longitudinal data ([McNeish et al., 2017](#)). Despite some cross-disciplinary efforts to compare these methods ([McNeish et al., 2017](#); [Muth et al., 2016](#); [Yan et al., 2013](#)), their scarcity may leave researchers with limited guidance in choosing the most suitable approach for their application.

A recent study by Qian et al. ([2020](#)) highlighted an issue present in both methods—except for GEE with working independence—where controlling for *time-varying endogenous covariates* may lead to biased causal estimates. A time-varying covariate is *endogenous* if it is directly or indirectly influenced by prior treatment or outcome, meaning its value may be determined by earlier stages of the process ([Qian et al., 2020](#)). Figure 1 showcases a time-varying endogenous covariate  $X_{it}$  in the context of a simple multilevel linear model with a random intercept  $b_{0i}$ . As a result of including these covariates in these models, ordinary interpretations of the coefficients are no longer valid ([Qian et al., 2020, p. 3](#)). According to Diggle ([2002](#)), this issue not only pertains GEE and MLM, but *all* longitudinal data analysis methods.

Figure 1: Multilevel Linear Model with Time-Varying Endogenous Covariate  $X_{it}$ .



*Note.* Adapted from Section 2.2 of Qian et al. (2020).

However, due to a divide between the disciplines that employ these methods, such critiques of the MLM appear to have largely failed to reach the applied researcher in psychology. One specific reason might be that the technical jargon in other disciplines makes it difficult for researchers to recognize when and how these issues emerge[1]. Therefore, this report aims to understand and explain the issue of including endogenous covariates in analyses involving GEE and MLM in a psychological context. To achieve this aim, the current investigation employs (a) graphical tools such as the directed acyclic graph (DAG) and path diagram to assess potentially relevant assumptions, as well as (b) data simulations with additional scenarios to pinpoint the issue. Accordingly, the following research questions will be addressed:

- (1) When does the inclusion of endogenous variables in multilevel linear models result in biased estimates of the treatment effect?
- (2) When does the inclusion of endogenous covariates in multilevel linear models result in a discrepancy between conditional and marginal interpretations of the treatment effect?

## 2 Methods

To obtain a better understanding of the issue exposed by Qian et al. (2020), two methods were employed. First, graphical methods were used provide insight into the presence and extent of bias with potential violation of assumptions: (a) path diagrams were used to evaluate the conditional independence assumption and (b) directed acyclic graphs (DAGs) were used to evaluate the backdoor criterion (Pearl, 1988, 2009). Second, a simulation study was performed to reproduce the results for the generative models (GMs) from Qian et al. (2020) and to further isolate the issue using additional GMs. In this simulation, bias in the treatment effect (RQ 1) was assessed with analytical multilevel models. The discrepancy between conditional and marginal interpretations of the treatment effect (RQ 2) was assessed with GEE with working independence.

### 2.1 Data Generation

In the simulation Qian et al. (2020) considered three generative models (GMs), all of which have an endogenous time-varying covariate. In GM1 and GM2, the endogenous covariate  $X_{it}$  equals the previous outcome  $Y_{it}$  plus some random noise, so the *conditional independence* assumption is valid. In GM3, the endogenous covariate depends directly on  $b_{i0}$ , violating the assumption. To isolate the issue in GM3, we consider two variations on this model: GM3A, where the random slope  $b_{i2}$  for the treatment  $A_{it}$  is removed; GM3B, where the interaction term  $\beta_1 A_{it} X_{it}$  is removed. Note that the conditional independence assumption is violated in either of these variations. The details of the generative models are described below. We follow the notation of Qian et al. (2020) to allow for direct comparison, but rewrite the equations into within- and between-person models (see Raudenbush & Bryk, 2002).

### 2.1.1 Generative Model 1

In GM1, we considered a simple case with only a random intercept and a random slope for  $X_{it}$ . The outcome is generated according to the following repeated-observations or within-person model (level 1):

$$Y_{it+1} = \pi_{0i} + \pi_{1i}X_{it} + \pi_{2i}A_{it} + \pi_{3i}A_{it}X_{it} + \epsilon_{it+1}$$

with the person-level or between-person model (level 2):

$$\pi_{0i} = \alpha_0 + b_{i0}, \quad b_{i0} \sim \mathcal{N}(0, \sigma_{b0}^2),$$

$$\pi_{1i} = \alpha_1,$$

$$\pi_{2i} = \beta_0 + b_{i2}, \quad b_{i2} \sim \mathcal{N}(0, \sigma_{b2}^2),$$

$$\pi_{3i} = \beta_1.$$

By substitution, we get the single equation model:

$$\begin{aligned} Y_{it+1} &= \pi_{0i} + \pi_{1i}X_{it} + \pi_{2i}A_{it} + \pi_{3i}A_{it}X_{it} + \epsilon_{it+1} \\ &= (\alpha_0 + b_{i0}) + \alpha_1X_{it} + (\beta_0 + b_{i2})A_{it} + \beta_1A_{it}X_{it} + \epsilon_{it+1} \\ &= \alpha_0 + \alpha_1X_{it} + b_{i0} + A_{it}(\beta_0 + \beta_1X_{it} + b_{i2}) + \epsilon_{it+1}. \end{aligned}$$

The random effects  $b_{i0} \sim \mathcal{N}(0, \sigma_{b0}^2)$  and  $b_{i2} \sim \mathcal{N}(0, \sigma_{b2}^2)$  are independent of each other.

The covariate is generated as  $X_{i1} \sim \mathcal{N}(0, 1)$ , and for  $t \geq 2$ ,

$$X_{it} = Y_{it} + \mathcal{N}(0, 1).$$

The randomization probability  $p_t = P(A_{it} = 1 \mid H_{it})$  is constant at  $1/2$ . Thus,  $A_{it} \sim \text{Bernoulli}(0.5)$  for  $i = 1, \dots, N$  and  $t = 1, \dots, T$ . The exogenous noise is  $\epsilon_{it+1} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ .

Figure 2a shows the path diagram for the first couple observations of GM1.

### 2.1.2 Generative Model 2

In GM2, we considered the case with a random intercept and random slopes for (1) covariate  $X_{it}$ , (2) treatment  $A_{it}$ , and (3) the interaction between  $A_{it}$  and  $X_{it}$ ; and with a time-varying randomization probability for treatment. The outcome is generated according to the same repeated-observations model presented in GM1. However, the person-level model is different:

$$\pi_{0i} = \alpha_0 + b_{i0}, \quad b_{i0} \sim \mathcal{N}(0, \sigma_{b0}^2),$$

$$\pi_{1i} = \alpha_1 + b_{i1}, \quad b_{i1} \sim \mathcal{N}(0, \sigma_{b1}^2),$$

$$\pi_{2i} = \beta_0 + b_{i2}, \quad b_{i2} \sim \mathcal{N}(0, \sigma_{b2}^2),$$

$$\pi_{3i} = \beta_1 + b_{i3}, \quad b_{i3} \sim \mathcal{N}(0, \sigma_{b3}^2).$$

By substitution, we get the single equation model:

$$\begin{aligned}
Y_{it+1} &= \pi_{0i} + \pi_{1i}X_{it} + \pi_{2i}A_{it} + \pi_{3i}A_{it}X_{it} + \epsilon_{it+1} \\
&= (\alpha_0 + b_{i0}) + (\alpha_1 + b_{i1})X_{it} + (\beta_0 + b_{i2})A_{it} + (\beta_1 + b_{i3})A_{it}X_{it} + \epsilon_{it+1} \\
&= \alpha_0 + \alpha_1X_{it} + b_{i0} + b_{i1}X_{it} + A_{it}(\beta_0 + \beta_1X_{it} + b_{i2} + b_{i3}X_{it}) + \epsilon_{it+1}.
\end{aligned}$$

The random effects  $b_{ij} \sim \mathcal{N}(0, \sigma_{bj}^2)$ , for  $j = 0, 1, 2, 3$ , are independent of each other.

The covariate is generated as  $X_{i1} \sim \mathcal{N}(0, 1)$ , and for  $t \geq 2$ ,

$$X_{it} = Y_{it} + \mathcal{N}(0, 1).$$

The randomization probability depends on  $X_{it}$ :

$$p_t = P(A_{it} = 1 \mid H_{it}) = \begin{cases} 0.7 & \text{if } X_{it} > -1.27, \\ 0.3 & \text{if } X_{it} \leq -1.27, \end{cases}$$

where the cutoff  $-1.27$  was chosen so that  $p_t$  equals 0.7 or 0.3 for about half of the time. In other words, if the value of the covariate for any given person and time point is above the cutoff, the probability of receiving the treatment  $p_t$  is 0.7; otherwise, it is 0.3. Accordingly,  $A_{it} \sim \text{Bernoulli}(p_t)$  for  $i = 1, \dots, N$  and  $t = 1, \dots, T$ . The exogenous noise is  $\epsilon_{it+1} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ .

Figure 2b shows the path diagram for the first couple observations of GM2.

### 2.1.3 Generative Model 3

GM3 is the same as GM1, except that the covariate  $X_{it}$  depends directly on  $b_{i0}$ :

$$X_{i1} \sim \mathcal{N}(b_{i0}, 1), \quad X_{it} = Y_{it} + \mathcal{N}(b_{i0}, 1) \text{ for } t \geq 2.$$

Figure 2c shows the path diagram for the first couple observations of GM3.

#### 2.1.4 Generative Model 3A

GM3A is the same as GM3, except that the random slope  $b_{i2}$  for the treatment  $A_{it}$  is removed (see Figure 2d). The single equation model then becomes:

$$Y_{it+1} = \alpha_0 + \alpha_1 X_{it} + b_{i0} + A_{it}(\beta_0 + \beta_1 X_{it}) + \epsilon_{it+1}.$$

#### 2.1.5 Generative Model 3B

GM3B is the same as GM3, except that the interaction term  $\beta_1 A_{it} X_{it}$  is removed (see Figure 2e). The single equation model then becomes:

$$Y_{it+1} = \alpha_0 + \alpha_1 X_{it} + b_{i0} + A_{it}(\beta_0 + b_{i2}) + \epsilon_{it+1}.$$

#### 2.1.6 Parameter Values

The following parameter values were adapted from Qian et al. (2020):

$$\alpha_0 = -2, \quad \alpha_1 = -0.3, \quad \beta_0 = 1, \quad \beta_1 = 0.3,$$

$$\sigma_{b0}^2 = 4, \quad \sigma_{b1}^2 = \frac{1}{4}, \quad \sigma_{b2}^2 = 1, \quad \sigma_{b3}^2 = \frac{1}{4}, \quad \sigma_{\epsilon}^2 = 1.$$

## 2.2 Conditional Independence and Path Diagrams

Qian et al. (2020) proposes the use of the conditional independence assumption to identify whether bias may occur, which is given by:



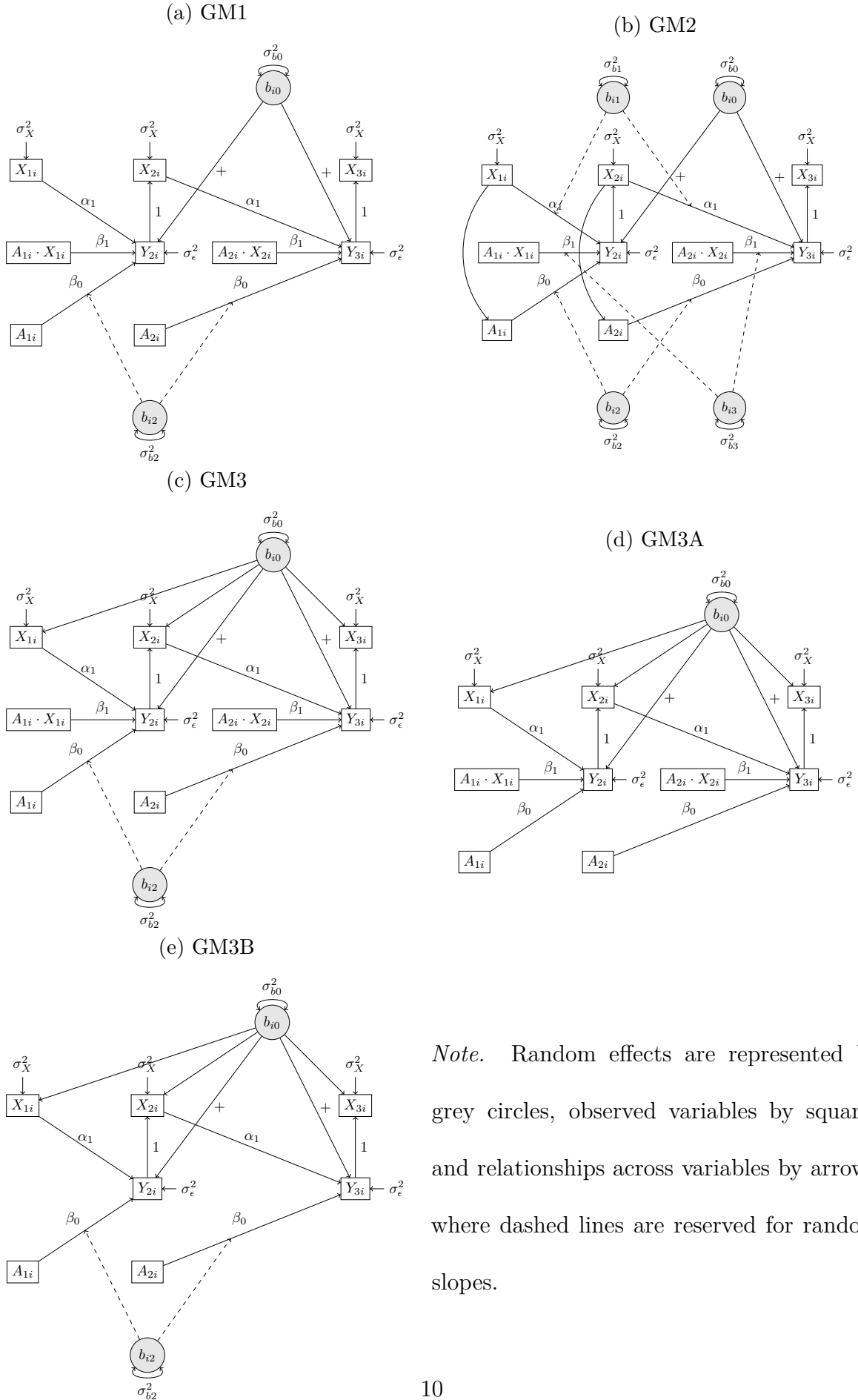
$$X_{it} \perp (b_{i0}, b_{i1}) \mid H_{it-1}, A_{it-1}, Y_{it}.$$

where  $H_{it-1}$  refers to the history of the set of covariates, which in this case are all observations of covariate  $X_{it}$  prior to the current timepoint  $t$ . This allows  $X_{it}$  to be endogenous, but the endogenous covariate  $X_{it}$  can only depend on the random effects through variables observed prior to  $X_{it}$ . If the only endogenous covariates are functions of prior treatments and prior outcomes, then the assumption automatically holds.

To make the application of the assumption more insightful, we accompany the equations of the GMs with path diagrams of the first three observations  $t$  for each generative model (see Figure 2). The path diagrams of the three data generating models shows the discrepancies between the different generative models—especially concerning the interaction effects—more clearly than DAGS.

When inspecting Figure 2a and Figure 2b, we may notice that  $X_{it}$  becomes independent of the random effects after conditioning on  $Y_{it}$ . On the other hand, we can see that this assumption is violated only in GM3/3A/3B, as  $X_{it}$  depends directly on  $b_{i0}$  and can thus not be made independent of the random effects by conditioning on prior variables such as  $Y_{it}$  (see Figure 2c, Figure 2d and Figure 2e). Thus, all things considered, we would expect biased estimates of the treatment effect for GM3/3A/3B but not for GM1/2.

Figure 2: Path Diagrams for Generative Models 1, 2, 3, 3A, 3B ( $t = 1, 2, 3$ )



*Note.* Random effects are represented by grey circles, observed variables by squares and relationships across variables by arrows, where dashed lines are reserved for random slopes.

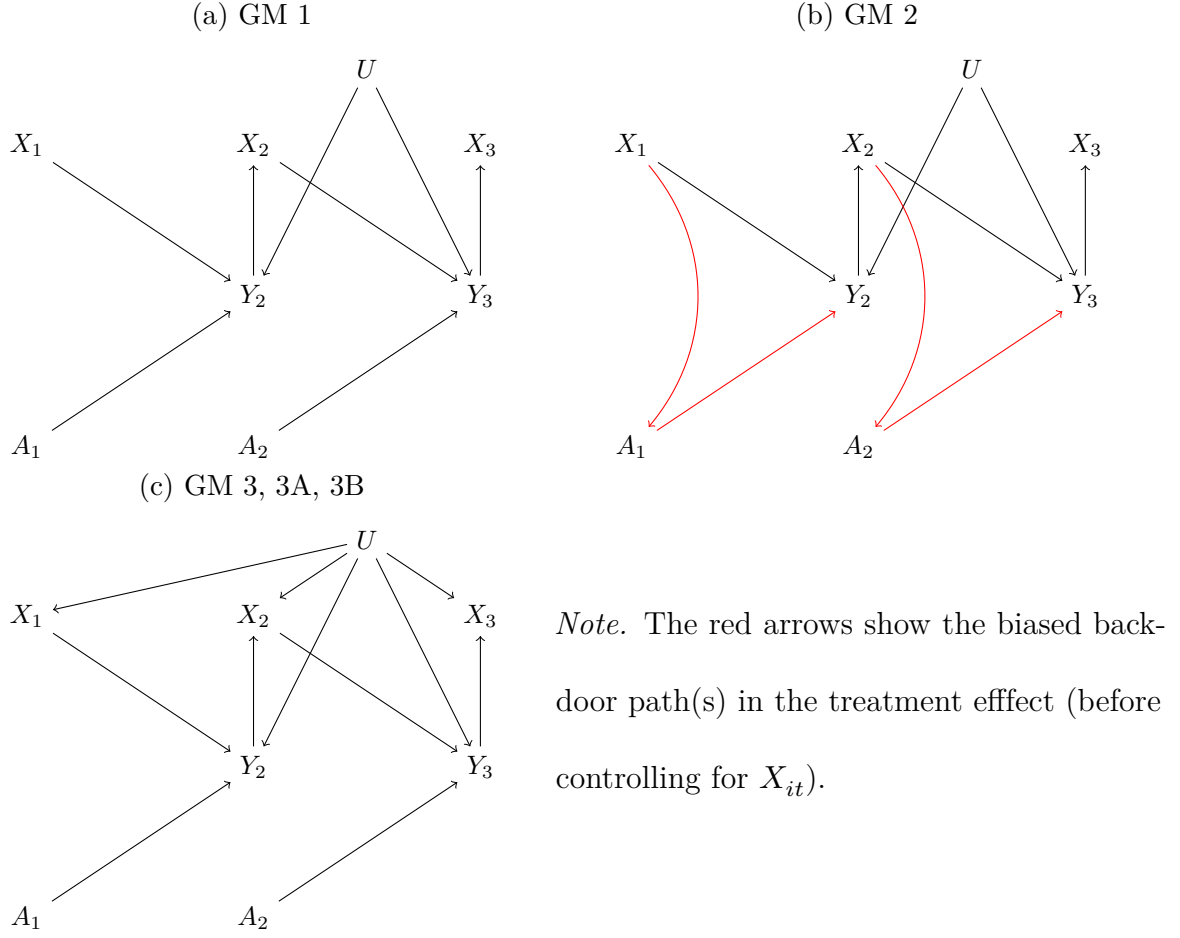
## 2.3 Backdoor Criterion and DAGs

According to the backdoor criterion (Pearl, 1988, 2009), a requirement for causal identification, causal effects can be identified by blocking non-causal paths through conditioning on intermediate variables (e.g., controlling or matching). If any non-causal paths cannot be blocked due to omitted variables or measurement error, treatment and outcome remain linked via backdoor paths, leading to biased estimates of the treatment effect (Kim & Steiner, 2021).

DAGs are a useful tool for representing causal relationships between variables and to evaluate the assumptions needed for causal identification (see Elwert & Winship, 2014 for a psychological example). We formulated the DAGs for the first three observations of each generative model, where the random disturbance  $b_{0i}$  was represented by the node  $U$  (e.g., Kim & Steiner, 2021, see Figure 3).

When applying Pearl’s backdoor criterion to GM1/3/3A/3B, it may be observed that there exists no backdoor path in the treatment effect  $A_{it} \rightarrow Y_{it+1}$ , as  $A_{it}$  does not have any parents. While we need not control for covariate  $X_{it}$  to obtain an unbiased total effect, doing so should not introduce bias. On the other hand, in GM2, there is a backdoor path in the treatment effect:  $A_{it} \leftarrow X_{it} \rightarrow Y_{it+1}$  (see Figure 3b). More specifically,  $X_{it}$  is a confounder in the relationship between  $A_{it}$  and  $Y_{it+1}$ . However, controlling for  $X_{it}$  blocks this backdoor path, making the treatment effect unbiased. In other words, the history of covariate  $X_{it}$  is a sufficient adjustment set for the treatment effect. All things considered, according to the backdoor criterion, controlling for the covariate  $X_{it}$  should not result in biased estimates of the treatment effect for any of the generative models.

Figure 3: DAGs for Generative Models 1, 2, 3, 3A, 3B ( $t = 1, 2, 3$ )



## 2.4 Data Analysis

We evaluated the performance of the models across a total of 30 different settings, each replicated 1,000 times, by systematically varying the following factors:

- **Generative Models (GM):** 1, 2, 3, 3A, 3B
- **Number of timepoints (T):** 10, 30
- **Sample size (N):** 30, 100, 200

All data generation and estimation was performed in R, version 4.4.2 ([Team, 2024](#)).

After the generation of data generation for any given setting, several models were fit. To

fit the standard MLM, the `lmer` function from the R-package `lme4` (Bates et al., 2015) was employed with restricted maximum likelihood estimation. For the MLM, the analytical models were equivalent to each of the respective data-generating models. To fit the GEE with the independent working correlation structure, the `geeglm` function from the R-package `geepack` (Halekoh et al., 2006) was employed with the identity link function. Since the random effects are not explicitly modelled in GEE, the analytical GEE models simply contain only the fixed effects of the generative model at hand.

### 3 Results

Table 1 presents the simulation results for each of the generative and analytical models. The estimates for the analytical MLM may be interpreted in terms of bias, where given the value of the treatment effect  $\beta_0 = 1$ , absolute bias of .05 would imply 5% relative bias. Here we find the greatest absolute bias of .02 – .06 for GM3,  $\leq .015$  for GM1/2,  $\leq .010$  for GM3B and ,  $\leq .005$  for GM3A. While the bias found for the original GMs 1, 2 and 3 was slightly larger here compared to Qian et al. (2020), the overall pattern remained the same. To conclude, once we remove either the dependency of the random intercept with the covariate (GM1), the random slope  $b_{i2}$  (GM3A) or the interaction  $\beta_1$  (GM3B) from GM3, the bias disappears or becomes very small. The bias in GM3 decreases as the number of timepoints  $T$  increases from 10 to 30. Note that the MLM model fitting success rates are particularly poor for GM2, where in the worst case, only 87 of the 1000 models were fitted.

For the GEE with independence, the values refer to the difference between the estimated marginal effect—which should be unbiased under endogenous covariates (see Pepe & Anderson, 1994)—and the specified conditional effect. Here we find that there is a enor-

mous difference between these effects for GM2, which increases along with an increase in  $T$  and  $N$ , up to a difference of more than 6,000. This is followed by a difference of around .07 – .09 for GM1, .02 – .04 for GM3,  $\leq .015$  for GM3B and close to zero ( $\leq .005$ ) for GM3A. The GEE models fitted successfully for all settings.

## 4 Discussion

This report employed both graphical methods and data simulations to understand and explain the issue of endogenous covariates. Now we first discuss the expected results based on the backdoor criterion Pearl (2009) and the conditional independence assumption (Qian et al., 2020), whereafter we discuss the findings relating to the two research questions.

Using the conditional independence assumption of Qian et al. (2020), we would expect, based on the path diagrams, that the treatment effect would be biased for GM3, 3A and 3B. On the other hand, the backdoor criterion suggested the absence of bias for all generative models. While Qian et al. (2020) show that GM3 is the only model with bias in the treatment effect, the backdoor criterion failed to identify this bias, as there is no backdoor path in the treatment effect. This may be explained by the fact that the classic DAG does not impose restrictions based on (a) the random slopes and (b) interaction effects.

The first research question—pertaining to the extent of treatment effect bias in MLM estimates of generative model that were nested in GM3—was investigated using the analytical multilevel model. First, we reproduced the findings by Qian et al. (2020) who found consistent estimators for GM1 and GM2 and inconsistent ones for GM3. Using additional generative models, we found that bias became indiscernable when removing from GM3 either the dependency between the random intercept and covariate (GM1),

Table 1: Simulation results for Generative Models 1, 2, 3, 3A and 3B over 1000 replications

GM	T	N	MLM			GEE-IND		
			Bias	SD	SR	Difference	SD	SR
1	10	30	0.000	0.238	0.998	0.071	0.296	1
		100	-0.012	0.129	1.000	0.074	0.169	1
		200	0.003	0.093	0.999	0.085	0.116	1
	30	30	-0.001	0.203	0.998	0.085	0.224	1
		100	-0.007	0.107	0.996	0.083	0.123	1
		200	0.001	0.079	0.996	0.094	0.088	1
2	10	30	0.011	0.282	0.925	0.306	1.630	1
		100	0.005	0.147	0.881	0.565	1.836	1
		200	0.008	0.103	0.844	0.935	1.887	1
	30	30	0.000	0.220	0.603	182.565	4751.387	1
		100	-0.014	0.114	0.247	-356.412	39799.388	1
		200	-0.013	0.087	0.087	6319.792	136201.790	1
3	10	30	-0.052	0.245	0.999	0.020	0.249	1
		100	-0.064	0.134	1.000	0.024	0.141	1
		200	-0.051	0.096	1.000	0.035	0.097	1
	30	30	-0.024	0.206	0.997	0.030	0.208	1
		100	-0.030	0.108	0.996	0.027	0.112	1
		200	-0.023	0.080	0.997	0.037	0.081	1
3A	10	30	0.000	0.126	1.000	-0.004	0.157	1
		100	0.004	0.073	1.000	0.001	0.090	1
		200	0.002	0.048	1.000	-0.001	0.062	1
	30	30	-0.001	0.071	1.000	-0.003	0.090	1
		100	0.000	0.040	1.000	-0.001	0.051	1
		200	0.000	0.028	1.000	0.000	0.036	1
3B	10	30	0.001	0.217	0.999	-0.013	0.241	1
		100	-0.008	0.121	1.000	-0.008	0.138	1
		200	0.005	0.087	1.000	0.003	0.097	1
	30	30	0.000	0.193	1.000	-0.004	0.200	1
		100	-0.008	0.103	0.997	-0.007	0.108	1
		200	0.001	0.075	0.999	0.001	0.080	1

*Note.* SR: model fitting success rate. Bias:  $\hat{\beta}_{0,MLM} - \beta_{0,MLM}$ . Difference:  $\hat{\beta}_{0,GEE} - \beta_{0,MLM}$ . SD: standard deviation of estimates across replications.

the random slope for treatment (GM3A) or the interaction effect (GM3B). This finding is in sharp contrast to the suggestion of the conditional independence assumption that the treatment effect would be biased for GM3, 3A and 3B.

The second research question—related to the discrepancy between marginal and conditional interpretations of the treatment effect—was assessed with analytical GEE with working independence. Here we found extreme differences between the estimated marginal and specified conditional effect for GM2, suggesting that the marginal interpretation breaks down the most for this generative model<sup>1</sup>. Hence, for this GM, a false interpretation of the MLM parameters as marginal, would potentially have great inferential consequences. For GM1 and GM3, there smaller but still noticeable differences between the marginal and conditional effect. This suggests that the marginal interpretation of the treatment effect may be recovered for GM3, 3A and 3B, but not for GM2. Especially for GM3A, this difference was practically indiscernable, suggesting that the marginal interpretation of the treatment effect may be recovered. Conversely, Qian et al. (2020) notes that if the random effect in the model does not interact with the treatment variable, the interaction recovers its marginal interpretation but the treatment effect does not (p. 382). This difference in conclusions may be explained through the difference in approach: while Qian et al. (2020) provides an analytical answer, the current study provides approximations through simulations.

For the GM2 setting of Qian et al. (2020), we found several issues, which were most pronounced for  $T = 30$ . First, we noticed extreme model fitting issues for the MLM, due to, among other things, a lack of convergence and singularity. It should be noted that unlike the script used here, Qian et al. (2020) deals only with errors of the `lmer()`

---

<sup>1</sup>Note, however, that this generative model may not be plausible given the extreme spread across the covariate and treatment variables.



function, but not with warnings (e.g., pertaining to non-convergence) in their script. This discrepancy may explain the slightly different estimates of MLM bias for GM2. Second, we found extremely large GEE estimates of the treatment effect. This may be explained by the fact that the values of the covariate and outcome were also extremely great, often exceeding a million. All things considered, this suggests that GM2 may be a poorly specified model.

The current research report leaves several avenues unexplored. First, it is unclear whether the simulation findings pertaining the generative models in Qian et al. (2020) and here generalize to other generative models. For instance, we found here that removal of a random slope or interaction from GM3 got rid of most if not all of the treatment effect bias. Thus, it is important to establish how this generalizes, so that practical recommendations can be formulated. This is particularly important, since while violations of model assumptions are never desired, the robustness against and the practical implications of a violation is what matters. Second, it is unclear how exactly the divide between the literatures pertaining to the focus of the MLM on different centering methods and within- and between-person interpretations and the focus of the GEE on marginal and conditional interpretations may be bridged. Consequently, future research could assess the implications of centering methods in MLMs on the extent to which the marginal interpretation of MLM breaks down. Third, we found that the classical DAG may not be sufficient to identify bias in the treatment effect for GM3, especially due to its lack of specification of interaction effects. Concerns regarding the use of Pearl’s backdoor criterion in situations with interaction effects have been voiced by several people (see Weinberg (2007); Attia et al. (2022)). Future research could explore to what extent proposed extensions of the DAG may be useful in identifying bias in the treatment effect for GM3. Finally, it

may be interesting to investigate the implications of endogenous covariates in MLMs for other types of longitudinal data analysis methods, such as dynamic structural equation modelling (DSEM; a widely used framework in the social sciences based on MLM).

Third, since the issue extends to all longitudinal data analysis methods according to Diggle ([2002](#)), in future research it may be interesting to investigate the implications of endogenous covariates in MLMs for other types of longitudinal data analysis methods, such as dynamic structural equation modelling (DSEM; a widely used framework in the social sciences based on MLM).

## 5 References

- Attia, J., Holliday, E., & Oldmeadow, C. (2022). A proposal for capturing interaction and effect modification using DAGs. *International Journal of Epidemiology*, 51(4), 1047–1053. <https://doi.org/10.1093/ije/dyac126>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 148. <https://doi.org/10.18637/jss.v067.i01>
- Diggle, P. (2002). *Analysis of Longitudinal Data*. OUP Oxford.
- Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, 40, 31–53. <https://doi.org/10.1146/annurev-soc-071913-043455>
- Halekoh, U., Højsgaard, S., & Yan, J. (2006). The r package geepack for generalized estimating equations. *Journal of Statistical Software*, 15/2, 111.
- Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods*, 25(3), 365–379. <https://doi.org/10.1037/met0000239>
- Kim, Y., & Steiner, P. M. (2021). Causal graphical views of fixed effects and random effects models. *British Journal of Mathematical and Statistical Psychology*, 74(2), 165–183. <https://doi.org/10.1111/bmsp.12217>
- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22(1), 114–140. <https://doi.org/10.1037/met0000078>
- Muth, C., Bales, K. L., Hinde, K., Maninger, N., Mendoza, S. P., & Ferrer, E. (2016). Alternative Models for Small Samples in Psychological Research: Applying Linear

- Mixed Effects Models and Generalized Estimating Equations to Repeated Measures Data. *Educational and Psychological Measurement*, 76(1), 64–87. <https://doi.org/10.1177/0013164415580432>
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
- Pepe, M. S., & Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics - Simulation and Computation*, 23(4), 939–951. <https://doi.org/10.1080/03610919408813210>
- Qian, T., Klasnja, P., & Murphy, S. A. (2020). Linear mixed models with endogenous covariates: Modeling sequential treatment effects with application to a mobile health study. *Statistical Science : A Review Journal of the Institute of Mathematical Statistics*, 35(3), 375–390. <https://doi.org/10.1214/19-sts720>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). SAGE.
- Team, R. C. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Weinberg, C. R. (2007). Commentary: Can DAGs clarify effect modification? *Epidemiology*, 18(5), 569–572. <https://www.jstor.org/stable/20486428>
- Yan, J., Aseltine, R. H., & Harel, O. (2013). Comparing Regression Coefficients Between Nested Linear Models for Clustered Data With Generalized Estimating Equations. *Journal of Educational and Behavioral Statistics*, 38(2), 172–189. <https://doi.org/10.1177/0013164413500000>

3102/1076998611432175