Treatment Effect Bias in Multilevel Linear Models under Time-Varying Endogeneity: A New Look at Qian et al. (2020)

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 20, 2024

Word count: 3331/2500

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, psychological researchers most commonly use the multilevel linear model[1] (MLM, Bauer & Sterba, 2011), which—in the context of longitudinal data analysis—partitions observed variance into stable between-person differences and within-person fluctuations (Hamaker & Muthén, 2020). Research questions explored with the MLM commonly lead to the availability of time invariant and/or time-varying covariates, the latter measured repeatedly over time. The inclusion of covariates is a common strategy to improve parameter precision (Boruvka et al., 2018) and address bias introduced by (time-varying) confounders (Daniel et al., 2013; Robins et al., 2000; Wodtke, 2020). Nevertheless, this approach is not universally beneficial, as conditioning on endogenous covariates—those influenced by (prior) treatment/exposure or outcome—can create challenges for standard methods like MLMs, which implicitly assume the exogeneity of covariates (Erler et al., 2019).

Originating from the biostatistics literature and dating back to the work of Pepe & Anderson (1994), this assumption has been shown to be non-trivial when endogenous covariates vary over time. In fact, their inclusion in longitudinal studies can lead to biased treatment effect estimates, an issue that, despite its significance, has received limited attention in psychological research. Building on this foundation, a recent paper by Qian et al. (2020) examined the suitability of MLM for estimating the causal effect of a time-varying exposure or treatment. Specifically, they focused on settings where the exposure is randomly assigned at each occasion within individuals. Such randomized exposures may include, for example, prompts delivered through push notifications to remind participants of cognitive or mindfulness-based strategies (Nahum-Shani et al., 2021; Walton et al., 2018). While random assignment with a constant

[1]The MLM is known by various names in different substantive fields, including: linear mixed model, hierarchical linear model, random-effect model and mixed-effects model.

probability might seem sufficient to identify (the presence and absence of) causal effects, Qian et al. (2020) showed that model fitting issues and parameter bias can arise when a *time-varying endogenous covariate* is present.

However, due to a divide between the disciplines that employ the MLM, such critiques 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. This report aims to explore why Qian et al. (2020) observed biased estimates of the treatment effect in certain data-generating mechanisms containing endogenous covariates, while not for others. Additionally, it seeks to explain this issue to an audience of psychologists. The study will first employ graphical diagrams to assess two criteria across various scenarios involving an endogenous time-varying covariate and randomized treatment: (a) path diagrams to evaluate the conditional independence assumption introduced by Qian et al. (2020) and (b) directed acyclic graphs (DAGs) to assess the backdoor criterion (Pearl, 1988, 2009). Subsequently, data simulations based on Qian et al. (2020)'s original scenarios, along with additional ones, will be performed to reproduce and isolate the underlying issue and evaluate whether these criteria can effectively detect bias in the treatment effect. The following research question will be addressed: *When does the inclusion of endogenous variables in multilevel linear models result in biased estimates of the treatment effect?*

2 Methods

In this section, we first formulate four generative models (GMs) that incorporate a time-varying endogenous covariate and a randomized treatment. We then outline the methodology used to investigate treatment effect bias across different settings.

2.1 Data Generation

We consider two GMs from Qian et al. (2020), one (GM A) being a special case of the general model (GM G) where bias was detected. To further isolate the source of bias, we introduce two additional special cases, labeled GM B and C. We first describe the general generative model (GM G) in detail, and then proceed to its three special cases: GM A, B, and C.

2.1.1 General Model: Generative Model G

Following the original notation of Qian et al. (2020), the outcome of GM G was generated as:

$$Y_{it+1} = \alpha_0 + \alpha_1 X_{it} + b_{i0} + A_{it}(\beta_0 + \beta_1 X_{it} + b_{i2}) + \epsilon_{it+1} \tag{1}$$

where Y_{it+1} is the outcome for person i at time $t+1$, X_{it} is the covariate for person i at time t, A_{it} is the treatment for person i at time t, b_{i0} is the random intercept, b_{i2} is the random slope for the treatment, and ϵ_{it+1} is the error term. Alternatively, the model can be rewritten in the multilevel notation of Raudenbush & Bryk (2002), with at the within-person level (level 1):

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

and at the between-person level (level 2):

$$\begin{aligned} \pi_{0i} &= \alpha_0 + b_{i0}, \quad \text{where} \quad b_{i0} \sim \mathcal{N}(0, \sigma^2_{b0}), \\ \pi_{1i} &= \alpha_1, \\ \pi_{2i} &= \beta_0 + b_{i2}, \quad \text{where} \quad b_{i2} \sim \mathcal{N}(0, \sigma^2_{b2}), \\ \pi_{3i} &= \beta_1. \end{aligned}$$

The parameters $\alpha_0 = -2$, $\alpha_1 = -0.3$, $\beta_0 = 1$, and $\beta_1 = 0.3$ are fixed effects that are constant

across individuals, while b_{i0} and b_{i2} are independent random effects that capture individual-specific deviations from population parameters. The presence of the interaction term β_1 implies treatment heterogeneity: the effect of the treatment A_{it} on the outcome depends on the value of the covariate X_{it}. The random intercept b_{i0} and random slope b_{i2} are assumed to be normally distributed with mean zero and variance $\sigma^2_{b0} = 4$ and $\sigma^2_{b2} = 1$, respectively. b_{i0} represents deviations from the population intercept α_0, and b_{i2} represents deviations from the population slope β_0. The exogenous noise ϵ_{it+1} in the outcome is normally distributed with mean zero and variance $\sigma^2_\epsilon = 1$.

The covariate is generated as:

$$X_{it} = \begin{cases} b_{i0} + \epsilon_{X_{it}} & \text{if } t = 1, \\ b_{i0} + Y_{it} + \epsilon_{X_{it}} & \text{if } t \geq 2, \end{cases} \quad \text{where} \quad \epsilon_{X_{it}} \sim \mathcal{N}(0, 1)$$

The treatment randomization probability is constant at $p_t = 0.5$, so $A_{it} \sim \text{Bernoulli}(0.5)$ for all i and t. In other words, for every given person i and every timepoint t, the probability that treatment is assigned is equivalent to a fair coinflip. Relationships between the (observed and latent) variables are illustrated in Figure 1a.

2.1.2 Special Cases: Generative Model A, B and C

We consider three special cases of GM G, namely GM A, B and C. The relation of each special case to GM G is summarized in Table 1. The specifics of each special case are described below.

Table 1: Summary of Differences Between Generative Models

Generative Model	Name in Qian et al. (2020)	dependency b_{i0} and X_{it}	random slope treatment b_{i2}	interaction β_1
G(eneral)	3	✓	✓	✓
A	1	×	✓	✓

Generative Model	Name in Qian et al. (2020)	dependency b_{i0} and X_{it}	random slope treatment b_{i2}	interaction β_1
B	NA	✓	×	✓
C	NA	✓	✓	×

GM A is a special case of GM G, where the effect of the random intercept b_{i0} on the covariate X_{it} is set to zero. This results in a model where the covariate X_{it} is not directly determined by the random intercept b_{i0} (see Figure 1b). Instead, the endogenous covariate X_{it} equals the previous outcome Y_{it} plus some random noise:

$$X_{it} = \begin{cases} \epsilon_{X_{it}} & \text{if } t = 1, \\ Y_{it} + \epsilon_{X_{it}} & \text{if } t \geq 2, \end{cases} \quad \text{where} \quad \epsilon_{X_{it}} \sim \mathcal{N}(0, 1)$$

GM B is a special case of GM G in which the random slope b_{i2} was removed (see Figure 1c) by setting the random slope variance σ^2_{b2} to zero. While the within-person model is the same as GM G, there is a slight alteration in the between-person model:

$$\pi_{2i} = \beta_0.$$

The composite model then becomes:

$$Y_{it+1} = (\alpha_0 + b_{i0}) + \alpha_1 X_{it} + \beta_0 A_{it} + \beta_1 A_{it} X_{it} + \epsilon_{it+1} \tag{2}$$

GM C is a special case of GM G, where the fixed interaction parameter β_1 is set to zero, which implies the removal of the interaction term $\beta_1 A_{it} X_{it}$ (see Figure 1d). This, in turn, removed π_{3i}, thereby creating a discrepancy in within-person model of GM C and GM G:

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

Nevertheless, the between-person model of π_{0i}, π_{1i} and π_{2i} remains the same as GM G. The composite model then becomes:

$$Y_{it+1} = \alpha_0 + \alpha_1 X_{it} + b_{i0} + A_{it}(\beta_0 + b_{i2}) + \epsilon_{it+1}. \tag{3}$$

2.2 Data Analysis

All data generation and estimation was performed in `R`, version 4.4.2 (Team, 2024). Choices in the simulation-setup were guided by the need for consistency with the methodology of Qian et al. (2020). After the generation of data for each setting, analytical MLMs were fit to the data[2]. Specifically, the MLM from Equation 1 was fit for GM G and A, the MLM from Equation 2 for GM B, and the MLM from Equation 3 for GM C. To fit the MLM, the `lmer` function from the R-package `lme4` (Bates et al., 2015) was employed with restricted maximum likelihood estimation.

In the simulation study, we evaluated the bias of the analytical models across different settings by systematically varying the following 3 factors: Generative Model (GM) with the levels G, A, B, and C; number of timepoints (T) with the levels 10 and 30; and sample size (N) with the levels 30, 100, and 200. By varying these factors, we created a total of 24 unique settings, each of which was replicated 10,000 times.

The bias was calculated as the difference between the estimated treatment effect and the true treatment effect, averaged across replications. The uncertainty in the bias estimates for each setting was quantified using the standard deviation and the Monte Carlo standard error (see Morris et al., 2019).

[2]Contrary to the data generating models, the analytical models do not model the covariate X_{it} and the treatment A_{it}.

3 Results

This section begins by constructing predictions about treatment effect bias for each GM, guided by the conditional independence assumption and the backdoor criterion. These predictions are then compared with the simulation study results, which present the bias across the different GMs.

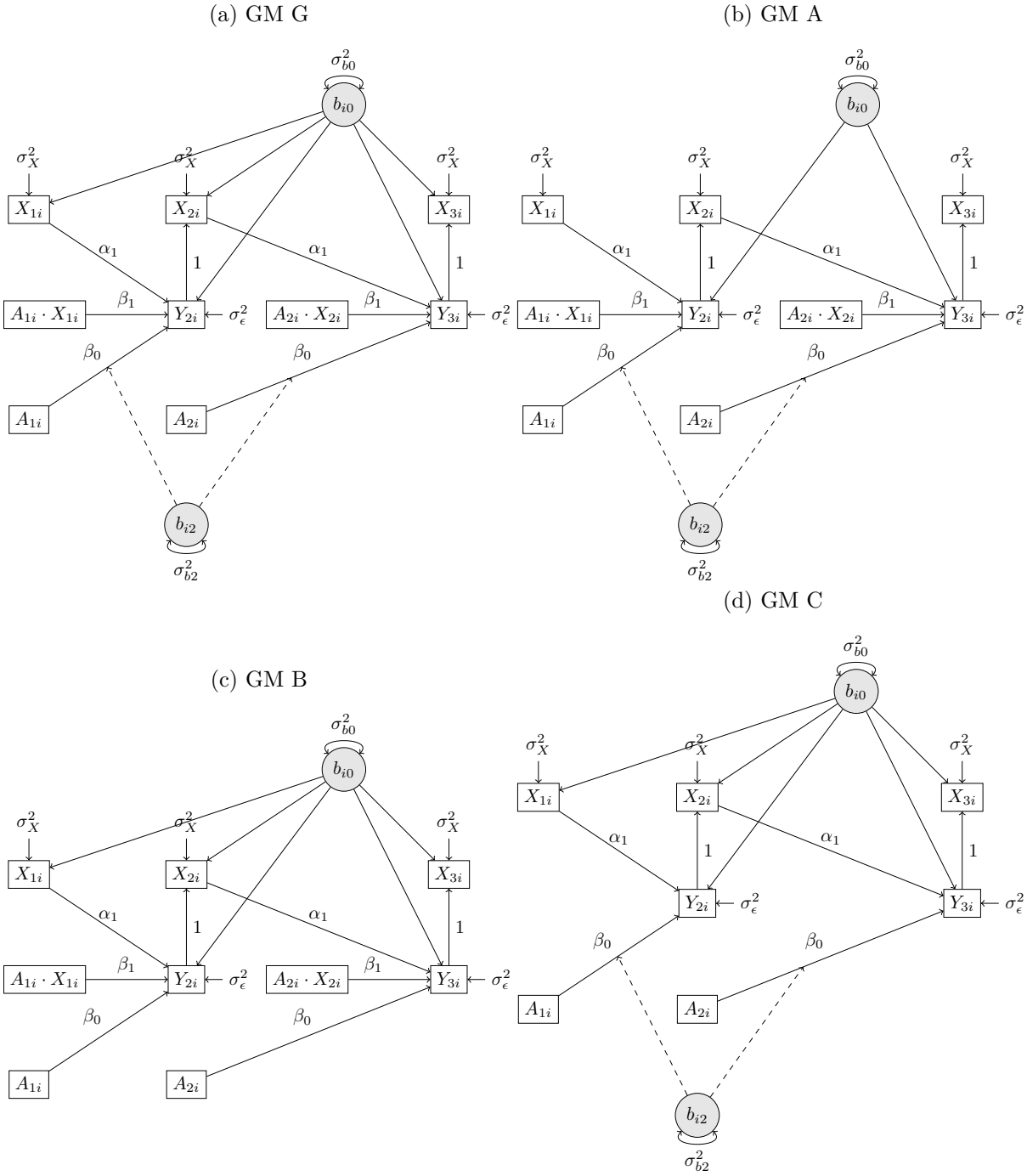
3.1 Conditional Independence and Path Diagrams

The first criterion for evaluating the presence of bias in treatment effect estimates is the *conditional independence assumption*, introduced by Qian et al. (2020) and based on the work of Sitlani et al. (2012). According to Qian et al. (2020), this assumption should identify whether estimators of the treatment effect are consistent and unbiased under randomized treatment assignment. The conditional independence assumption states that the covariate at time t (X_{it}) should be independent of the individual's random effects (intercept b_{i0} and slope(s) b_{i1}) once we account for their history of covariates up to timepoint $t-1$ (H_{it-1}), previous treatment (A_{it-1}), and prior outcome (Y_{it}).

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

This assumption allows for X_{it} to be influenced by earlier variables (e.g., outcomes or treatments) but not directly by unobserved individual characteristics (i.e., random effects). However, as Qian et al. (2020) highlights, ensuring this assumption holds requires careful consideration of theory and domain knowledge. To clarify the application of the conditional independence assumption, we pair the equations of the generative models (GMs) with path diagrams (Duncan, 1966; Wright, 1934) illustrating the first three timepoints (t) for each model (see Figure 1). We can observe that all models contain a time-varying *endogenous* covariate X_{it}, which is determined by previous outcome Y_{it}.

Figure 1: Path Diagrams for Generative Models G, A, B and C (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.

In GM G, the covariate X_{it} is directly influenced by unobserved individual factors (represented by the random effects, b_{i0}). Consequently, conditioning on prior variables, such as the

outcome at the previous timepoint Y_{it}, does not fully block or eliminate the influence of these unobserved factors. As a result, X_{it} remains dependent on the random effects, violating the assumption. This violation of the conditional independence assumption aligns with the biased estimates of the treatment effect observed in GM G, as identified by Qian et al. (2020).

In contrast, GM A, a special case of GM G where no bias was found by Qian et al. (2020), removes the direct link between X_{it} and the random effects b_{i0}. In this case, X_{it} is simply the previous outcome Y_{it} plus some random noise. While there remains an indirect connection between X_{it} and b_{i0} through Y_{it}, conditioning on Y_{it} effectively "breaks the link" between X_{it} and the random effects, satisfying the conditional independence assumption.

For GM B and GM C, the direct link between the random effects and X_{it} remains, as in GM G. As a result, these models also violate the conditional independence assumption, suggesting the presence of bias in treatment effect estimates.

In summary, GM G, B, and C violate the conditional independence assumption, which suggests that we would expect biased treatment effect estimates for these models. In contrast, GM A satisfies the assumption, supporting unbiased estimates of the treatment effect.

3.2 Backdoor Criterion and DAGs

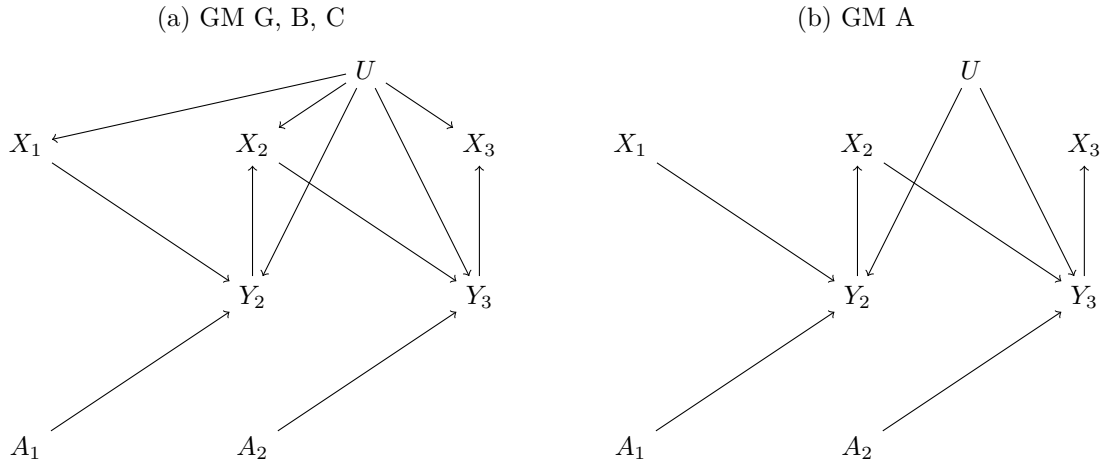
The second criterion for evaluating bias in treatment effect estimates is the *backdoor criterion* (Pearl, 1988, 2009). When estimating the treatment effect, the backdoor criterion can help us decide which variables to control for to ensure that we do not obtain a biased estimate of the treatment effect.

To detect backdoor paths, directed acyclic graphs (DAGs) (Pearl, 1995, 2009) are invaluable tools. DAGs generalize conventional path diagrams (Duncan, 1966; Wright, 1934) within a fully nonparametric framework. Unlike conventional linear path diagrams and those based on structural equation modelling (SEM), DAGs make no assumptions about distributional properties (e.g., multivariate normality) or functional forms (e.g., linearity). They encode qualitative causal

assumptions about the data-generating process, where arrows indicate direct causal effects that may vary across individuals (effect heterogeneity) or depend on other variables (effect interaction or modification) (Elwert & Winship, 2014). Notably, random slopes from random-effects models and interaction effects are not explicitly represented in DAGs, which precluded their use for evaluating the conditional independence assumption.

Using the direct causal effects specified in each generative model (GM), we can formulate DAGs for the first three observations, representing the random disturbance b_{0i} as the node U (e.g., Kim & Steiner, 2021, see Figure 2). These diagrams confirm that random slopes and fixed interaction effects are absent. Indeed, this absence explains why the DAGs for GMs G, B, and C are equivalent.

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



Note. The node U represents the random disturbance b_{0i} in the generative models.

We now apply the backdoor criterion to these DAGs to assess potential violations. Application of the backdoor criterion begins by identifying all paths that connect the treatment A_t to the outcome Y_{t+1}, excluding the direct treatment effect itself. Next, we determine whether these paths can transmit association by applying the d-separation rules[3]. A path that contains an arrow pointing to the treatment A_t is considered a *backdoor path*.

[3]See (Hayduk et al., 2003) for an accessible introduction into d-separation.

Examining the DAGs presented in Figure 2, we observe that none of the GMs contains any path connecting A_t to Y_{t+1} other than the direct treatment effect itself. This confirms that there are no backdoor paths between A_t and Y_{t+1}, as A_t lacks any parent nodes. Indeed, the random assignment of the treatment ensures, by design, the absence of backdoor paths and any shared common cause with the outcome. Consequently, the covariate X_t does not need to be controlled for to obtain an unbiased estimate of the total effect.

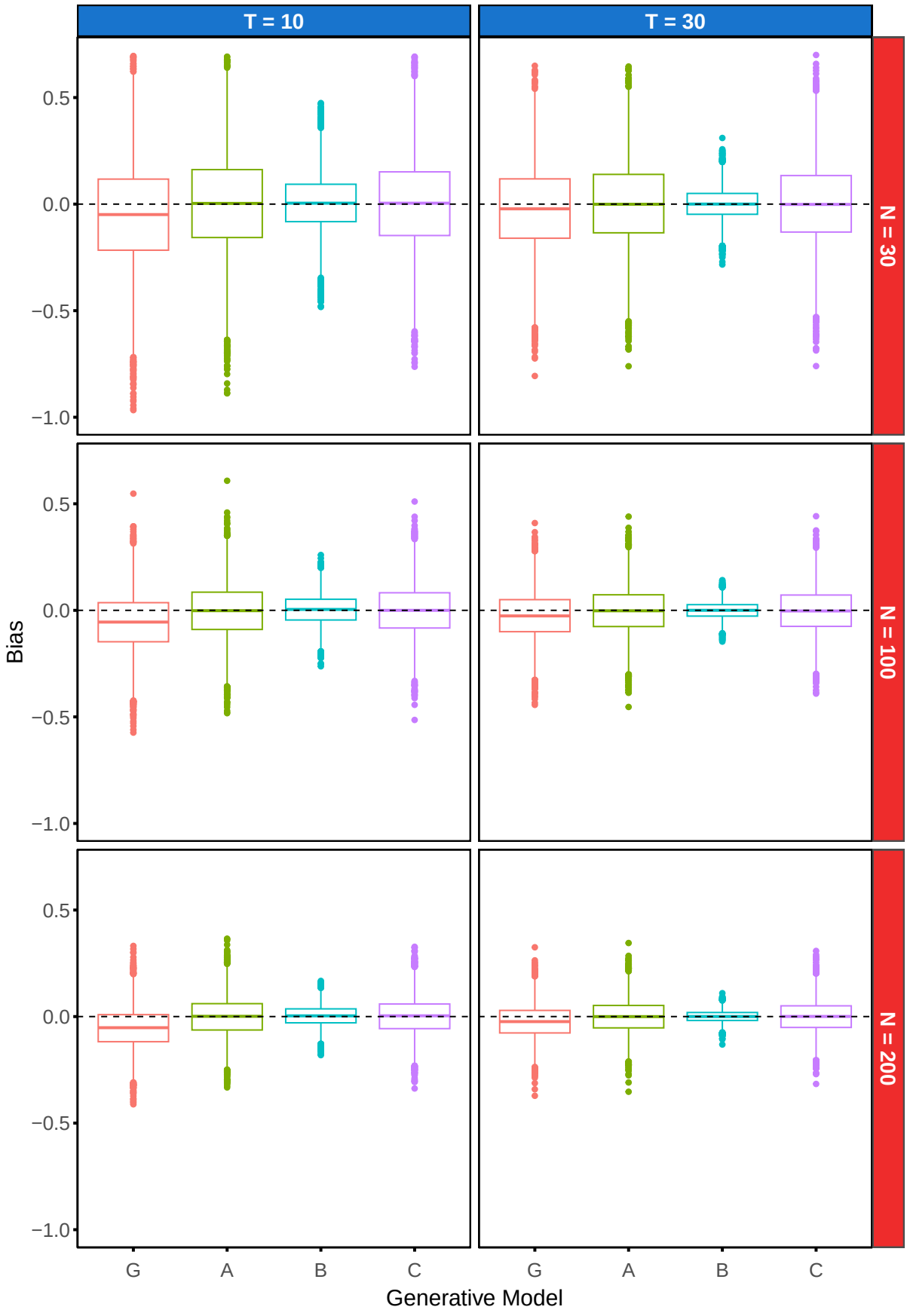
Additionally, the lack of any pathways connecting A_t to Y_{t+1} other than the direct treatment effect ensures that including X_t does not introduce identification issues. Its inclusion may increase the statistical power to detect the treatment effect. Therefore, according to the backdoor criterion, the inclusion of the time-varying covariate X_{it} should not lead to biased estimates of the treatment effect in any of the generative models.

3.3 Simulation Study

Figure 3 present the simulation results for each of the generative models[4], where bias refers to the difference between the mean of the estimated parameter values $\bar{\beta}_0$ and the prespecified treatment effect β_0. As $\beta_0 = 1$, an absolute bias of 0.05 implies a 5% relative bias.

[4]The supplemental table with additional information can be found at the OSF repository: https://osf.io/8xawt/?view_only=aad6a13b7a4a4d36aed76ed8aac584c4

Figure 3: Estimation bias for the fixed treatment effect β_0 of each generative model for different combinations of sample size N and number of timepoints T over 10,000 simulation replications



In this reproduction of Qian et al. (2020), the overall pattern was consistent with the original study: we observed substantial absolute bias ranging from 0.023 (2.3%) to 0.056 (5.6%) for the most general generative model (GM G) and no discernible bias ($\leq$ 0.002 or 0.2%) for GM A. These results align with expectations based on the conditional independence assumption, but contradict the backdoor criterion, which predicted the absence of bias for all the GMs. As shown in Figure 3, the magnitude of the bias in GM G decreased as the number of time points (T) increased. While a larger sample size (N) reduced the variability of the bias across replications, it did not reduce the bias itself. This indicates the bias is systematic, and the estimates are inconsistent, meaning they do not converge to the true value as N increases.

Similar to GM A, we observed no discernible bias for GM B ($\leq$ 0.005 or 0.5%) and GM C ($\leq$ 0.003 or 0.3%). These findings align with the backdoor criterion's prediction of no bias but contradict the expectations based on the conditional independence assumption, which suggests the presence of bias. Additionally, GM B showed much smaller variability across simulation replications compared to all other GMs. In contrast, the remaining models exhibited comparable levels of variability (see Figure 3).

In summary, these findings suggest that if the underlying GM did not include the direct dependency of the random intercept on the covariate (GM A), the random slope b_{i2} (GM B), or the interaction term β_1 (GM C), as in GM G, the bias disappears. However, neither the backdoor criterion nor the conditional independence assumption provided consistent predictions of treatment effect bias across all models.

4 Discussion

4.1 Main Findings

In this research report, we evaluated several GMs to investigate when endogenous time-varying covariates bias treatment effect estimates in MLMs under randomized treatment. We also as-

sessed the ability of the conditional independence assumption and the backdoor criterion to predict this bias. Consistent with Qian et al. (2020) and the conditional independence assumption, we observed biased and inconsistent estimates in the most general model (GM G), but not in a special case (GM A), that did not contain a direct effect of the random intercept on the covariate. Similarly, no discernible bias was found for the two additional special cases of GM G—where the random slope for treatment (GM B) or the interaction term between treatment and covariate (GM C) were present—despite violations of the conditional independence assumption. These findings suggest that, at least in this instance, the dependency between the covariate and treatment, the random slope for treatment, and the interaction effect are all essential for bias to occur.

One possibility is that heterogeneity in the treatment effect—both explained by the covariate and influenced by the random slope—is necessary for bias to occur. Both factors appear to be essential in driving the bias, rather than either factor alone. This aligns with @qian2020's assertion that "applying linear mixed models is problematic because potential moderators of the treatment effect are frequently endogenous" (p. 375). If treatment effect moderation by a covariate is indeed a prerequisite for bias, it could explain the absence of bias in GM C, where no interaction term was included. However this would not explain why the bias disappears in the model without the random slope (GM B). More generally, note of caution is due here since the current study only considered a specific set of GMs. Further research is needed to determine whether these findings generalize or are specific to the evaluated GMs, thereby informing practical recommendations for using MLMs with endogenous time-varying covariates.

Regarding the backdoor criterion and DAGs (Pearl, 1988, 2009), our results suggest that the classical non-parametric DAG may be insufficient to identify bias in GM G. While DAGs account for direct causal effects, they do not impose restrictions on random slopes or interaction effects[5], which are central to the conditional independence assumption. Similar concerns regard-

[5]Note that the term "effect modification", while often used interchangeably with "interaction", has a distinct definition in the counterfactual framework (VanderWeele, 2009).

ing the use of the DAG with Pearl's backdoor criterion in situations with interaction effects have been raised (Attia et al., 2022; Weinberg, 2007). Future research could explore to what extent proposed extensions of the DAG—that incorporate interaction effects—may allow the backdoor criterion to identify bias in the treatment effect estimates.

Another avenue for future investigation is the role of centering approaches (see Hamaker & Muthén (2020) for an overview). Namely, Antonakis et al. (2021) notes that the assumption of uncorrelatedness between the random effects and level 1 covariates can be relaxed by using Mundlak's contextual model[6] (Mundlak, 1978): adding the cluster means of each covariate as predictor of the random intercept. Such an approach of explicitly modeling the source of endogeneity, as advocated by Bell & Jones (2015), may further clarify the treatment effect bias in GM G.

Finally, Qian et al. (2020) only considered independent random effects, a restrictive assumption that may be violated in practice. Exploring correlated random effects using structural equation modeling frameworks (Rovine & Molenaar, 2000) could provide further insight.

4.2 Conclusion

This report is a first step towards understanding the implications of endogenous covariates in multilevel linear models. However, to recognize and understand completely when and why endogenous covariates may trouble an empirical investigation, further research is needed.

[6]This is referred to as the Correlated Random Effects (CRE) approach by Wooldridge (2002).

5 References

Antonakis, J., Bastardoz, N., & Rönkkö, M. (2021). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, *24*(2), 443–483. https://doi.org/10.1177/1094428119877457

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

Bauer, D. J., & Sterba, S. K. (2011). Fitting multilevel models with ordinal outcomes: Performance of alternative specifications and methods of estimation. *Psychological Methods*, *16*(4), 373–390. https://doi.org/10.1037/a0025813

Bell, A., & Jones, K. (2015). Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods*, *3*(1), 133–153. https://doi.org/10.1017/psrm.2014.7

Boruvka, A., Almirall, D., Witkiewitz, K., & Murphy, S. A. (2018). Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association*, *113*(523), 1112–1121. https://doi.org/10.1080/01621459.2017.1305274

Daniel, R. m., Cousens, S. n., De Stavola, B. l., Kenward, M. G., & Sterne, J. a. C. (2013). Methods for dealing with time-dependent confounding. *Statistics in Medicine*, *32*(9), 1584–1618. https://doi.org/10.1002/sim.5686

Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology*, *72*(1), 1–16. https://doi.org/10.1086/224256

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

Erler, N. S., Rizopoulos, D., Jaddoe, V. W., Franco, O. H., & Lesaffre, E. M. (2019). Bayesian imputation of time-varying covariates in linear mixed models. *Statistical Methods in Medical Research*, *28*(2), 555–568. https://doi.org/10.1177/0962280217730851

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

Hayduk, L., Cummings, G., Stratkotter, R., Nimmo, M., Grygoryev, K., Dosman, D., Gillespie, M., Pazderka-Robinson, H., & Boadu, K. (2003). Pearl's d-separation: One more step into causal thinking. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(2), 289–311. https://doi.org/10.1207/S15328007SEM1002_8

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

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://doi.org/10.1002/sim.8086

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, *46*(1), 69–85. https://doi.org/10.2307/1913646

Nahum-Shani, I., Potter, L. N., Lam, C. Y., Yap, J., Moreno, A., Stoffel, R., Wu, Z., Wan, N., Dempsey, W., Kumar, S., & al., et. (2021). The mobile assistance for regulating smoking (MARS) micro-randomized trial design protocol. *Contemporary Clinical Trials*, *110*, 106513. https://doi.org/10.1016/j.cct.2021.106513

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–688. https://doi.org/10.1093/biomet/82.4.669

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.

Robins, J. M., Hernán, M. Á., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550. https://journals.lww.com/epidem/fulltext/2000/09000/marginal_structural_models_and_causal_inference_in.11.aspx

Rovine, M. J., & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research*, *35*(1), 51–88. https://doi.org/10.1207/S15327906MBR3501_3

Sitlani, C. M., Heagerty, P. J., Blood, E. A., & Tosteson, T. D. (2012). Longitudinal structural mixed models for the analysis of surgical trials with noncompliance. *Statistics in Medicine*, *31*(16), 1738–1760. https://doi.org/10.1002/sim.4510

Team, R. C. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

VanderWeele, T. J. (2009). On the distinction between interaction and effect modification. *Epidemiology*, *20*(6), 863. https://doi.org/10.1097/EDE.0b013e3181ba333c

Walton, A., Nahum-Shani, I., Crosby, L., Klasnja, P., & Murphy, S. (2018). Optimizing digital

integrated care via micro-randomized trials. *Clinical Pharmacology & Therapeutics*, *104*(1), 53–58. https://doi.org/10.1002/cpt.1079

Weinberg, C. R. (2007). Commentary: Can DAGs clarify effect modification? *Epidemiology*, *18*(5), 569–572. https://www.jstor.org/stable/20486428

Wodtke, G. T. (2020). Regression-based adjustment for time-varying confounders. *Sociological Methods & Research*, *49*(4), 906–946. https://doi.org/10.1177/0049124118769087

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. MIT Press.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, *5*(3), 161–215. https://doi.org/10.1214/aoms/1177732676