Article

Organizational Research Methods
1-25
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1094428119842638
journals.sagepub.com/home/orm
SAGE

Challenges and Opportunities in the Estimation of Dynamic Models

Ran Xu[1], Richard P. DeShon[2], and Christopher R. Dishop[2]

Abstract

Interest in modeling longitudinal processes is increasing rapidly in organizational science. Organizational scholars often employ multilevel or hierarchical linear models (HLMs) to study such processes given that longitudinal data in organizational science typically consist of observations over a relatively small number of time intervals (*T*) nested within a relatively large number of units (*N*; e.g., people, teams, organizations). In this paper, we first distinguish *change* and *dynamics* as common research foci when modeling longitudinal processes and then demonstrate that a unique set of inferential hazards exists when investigating change or dynamics using multilevel models. Specifically, multilevel models that include one or more time-lagged values of the dependent variable as predictors often result in substantially biased estimates of the model parameters, inflated Type I error rates, and ultimately inaccurate inference. Using Monte Carlo simulations, we investigate the bias and Type I error rates for the standard centered/uncentered hierarchical linear model (HLM) and compare them with two alternative estimation methods: the Bollen and Brand structural equation modeling (SEM) approach and the Arrelano and Bond generalized method of moments using instrumental variables (GMM-IV) approach. We find that the commonly applied hierarchical linear model performs poorly, whereas the SEM and GMM-IV approaches generally perform well, with the SEM approach yielding slightly better performance in small samples with large autoregressive effects. We recommend the Bollen and Brand SEM approach for general use when studying change or dynamics in organizational science.

Keywords

longitudinal, change, dynamics, lagged response variable, hierarchical linear models, multilevel models, estimation, structural equation modeling, bias

[1]Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
[2]Michigan State University, East Lansing, MI, USA

Corresponding Author:
Ran Xu, Virginia Polytechnic Institute and State University, Blacksburg, 430 Northern Virginia Center, Virginia Tech, Falls Church, VA, 22043.
Email: ranxu@vt.edu

Organizational scholars are increasingly interested in change and how processes unfold over time. The role of time, for instance, is increasingly prevalent and explicit in theory development (Mitchell & James, 2001; Pitariu & Ployhart, 2010; Zaheer, Albert, & Zaheer, 1999), longitudinal data structures are now common and the number of time points included in these data sets appears to be increasing, and the use of time-oriented terms such as *change*, *process*, *mediation*, *dynamic*, *dynamics*, and *dynamical* is growing exponentially (DeShon, 2012). A wide variety of statistical and computational models are used to support inferences about the fundamental processes responsible for the trajectories observed in longitudinal data.

The current state of process-oriented investigations in organizational science is reminiscent of the early stages of the multilevel revolution before consensus emerged on how to best collect and model multilevel data structures. Just like the early multilevel literature, the still nascent process modeling literature likely contains a number of well-intentioned false starts and inferential errors resulting from the lack of a shared understanding of longitudinal modeling pitfalls and best practices. One such practice—the focus of the current presentation—is the use of multilevel or hierarchical linear models (HLMs) to capture patterns present in longitudinal data. Here we show that unless the number of observations over time is large, hierarchical linear models yield inconsistent estimates of key parameters for an important but distinct set of process inferences. We present and evaluate alternative modeling strategies that overcome the limitations of the standard hierarchical linear model.[1]

Longitudinal data structures in organizational science typically consist of observations on a relatively large number of units (e.g., people, teams, organizations), N, over a relatively small number of equally spaced time (T) intervals. In recent investigations, N is typically in the range of 50 to 500 and T is between 2 and 50, with the vast majority of studies having six or fewer measurements per unit (e.g., Chi, Chang, & Huang, 2015; Li, Burch, & Lee, 2017; Zhu, Wanberg, Harrison, & Diehn, 2016). Longitudinal data structures fall within the broader class of nested data structures, and so methods that account for the nonindependence of observations within the units (aka, clustered responses or unit effects) are needed to obtain accurate inferences. It is entirely sensible, then, that the standard HLM used to account for clustered responses in the multilevel literature would be used when modeling relationships found in longitudinal data. Unfortunately, a unique set of inferential hazards arise when fitting models to longitudinal data structures, and the use of HLMs is, for a given class of inference, a poor model choice. Specifically, HLMs are problematic when researchers are interested in change or dynamics.

In the following, we discuss change and dynamics as general classes of inference that researchers may be interested in when they explore patterns in longitudinal data. Then, we highlight problems associated with modeling change or dynamics within longitudinal data structures, investigate the extent to which several different modeling strategies—including the commonly used HLM, the Arrelano and Bond generalized method of moments using instrumental variables (GMM-IV) approach, and Bollen and Brand's structural equation modeling (SEM) approach—produce bias using Monte Carlo simulations and ultimately recommend Bollen and Brand's (2010) strategy for a number of discussed reasons.

Change and Dynamics

The notion of change plays a fundamental role in science, and it is a common focus in organizational research (Ployhart & Vandenberg, 2010). The appropriate conceptualization of change and best methods for modeling change have been debated for decades in the general psychological methods literature (e.g., Cronbach & Furby, 1970; Lord, 1967; Rogosa, 1988; Wainer, 1991). Given our somewhat unique methodological history dealing with difference scores, it is not surprising to find that we have strongly converged to models of change that include prior values of response variables as covariates predicting current or future values of the same response variable. In practice, the notion of change is often studied using only two or three waves of data, but more recent investigations tend to include observations taken at 5 to 10 time points.

Before presenting the full change model familiar to organizational researchers, we first unpack a number of simpler models on which it is based. The basic *autoregressive model* containing a single lag (i.e., $AR[1]$) may be represented mathematically as

$$y_t = b_0 + b_1 y_{(t-1)} + e_t, \tag{1}$$

where $t = 0, 1, 2, ...T$, y_t is the value of the variable at time point t, $y_{(t-1)}$ is the value of the variable at time point $t-1$, b_0 represents the y-intercept for $t = 0$ (i.e., the initial value), b_1 is the autoregression parameter relating current and prior values of the variable over time, and e_t is a random variable distributed as $N(0, \sigma_e)$. This short memory model allows immediately prior values of a variable to influence the current value of a variable. The existence of memory in this model makes it possible to represent phenomena that do not instantly dissipate, such as cognitive ability or organizational culture. If we have repeated measures on job satisfaction, for example, then this model represents how prior values of job satisfaction influence job satisfaction at the current moment. Additional lagged versions of the response variable may be included to represent processes and systems that have longer memories.

The aforementioned autoregressive model represents the trajectory of a single variable on only one unit (e.g., a person, team, or organization) over time. The autoregressive model can be generalized to cases with multiple units measured over multiple time points (i.e., longitudinal data or panel data rather than a single time-series) by simply adding a unit subscript:

$$y_{it} = b_0 + b_1 y_{i(t-1)} + e_{it}, \tag{2}$$

where $i = 1, 2, 3...N$ represent different units (e.g., different individuals when modeling job satisfaction over time) and all other terms are as defined in the previous paragraphs.

The models represented by Equations 1 and 2 express simple ideas of change in a response variable through time. The change model typically applied in organizational science extends on the aforementioned basic models by including a concurrent predictor. This model is shown in the first panel of Figure 1, where a response variable is predicted by both its prior values and the values of an additional, concurrent predictor. The full change model may be represented mathematically as,

$$y_{it} = b_{0_i} + b_1 y_{i(t-1)} + b_2 x_{it} + e_{it}, \tag{3}$$

where the only new term, x_{it} represents the value of the predictor for the i th individual at time t. As described previously, b_{0_i} represents a random unit intercept used to capture variance due to clustering of responses within units. When investigating change, $y_{i(t-1)}$ often remains uncentered either because it mathematically cannot be centered within units (i.e., $T = 2$) or there is a desire to keep the same scale for the response and the predictor variables.

Another concept that is closely related to change is dynamics. The conceptualization of dynamics is perhaps the key source of ambiguity in this relatively new literature. Different authors use the terms *dynamic*, *dynamics*, or *dynamical* to refer to a wide variety of concepts. Some use it to describe the concept of volatility where one or more variables change dramatically over time. Others use it to represent change or growth or even changes in the strength of relationships over time. Most of these uses are inconsistent with the way other scientific disciplines (e.g., physics, chemistry, biology, ecology) define the concept of dynamics. Here, we explicate the conceptualization of dynamics that dominates scientific thinking using words and visual and mathematical representations.

The concept of dynamics is rooted in Newton's development of calculus, which generated an entire discipline of applied mathematics focusing on the effect of various forces on the motion and

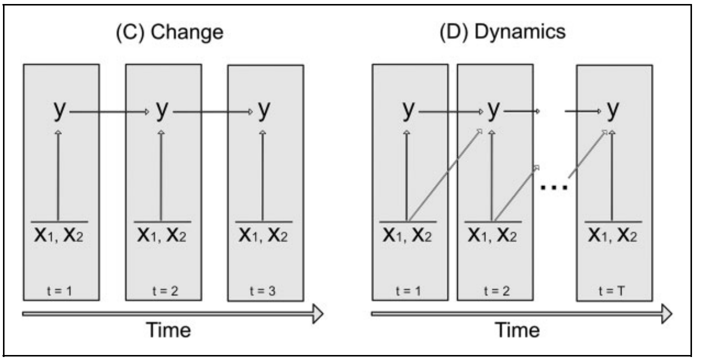

Figure 1. Common change and dynamic models for longitudinal data.

trajectories of objects. The notion of infinitesimally small changes in the position of an object (dx) over infinitesimally small time spans (dt) is the key insight underlying differential calculus and its innumerable applications. For our purposes, what matters is that conceptualizing dynamics and representing it explicitly in a model means that saying something about the future values (i.e., states) of a variable requires more or less knowledge about the past values of the variable. There are two broad classes of dynamic models: those that treat time as a continuous, differentiable variable and those that treat time as taking on discrete values. Dynamic equations based on continuous time are called *differential equations*, and those that treat time as discrete are called *difference equations*. Organizational scientists rarely work with deterministic, closed systems. Instead, we view systems as open and influenced in part by stochastic (i.e., random over time) processes. The stochastic versions of differential and difference equations are called, respectively, *stochastic differential equations* and *autoregressive processes*. We do commonly work with models that contain autoregressive components, and so, perhaps without awareness or intention, we have adopted an approach to the study of dynamics based on stochastic difference equations.

We now generalize the *change* equation presented previously and present the *dynamic* model as an autoregressive model with additional time-lagged predictors, which yields the well-known *autoregressive distributed lag* model (ARDL). Mathematically, an *ARDL* (*1*,*1*) model—meaning that there is a single lag in both the independent and dependent variables—is

$$y_{it} = b_{0i} + b_1 y_{i(t-1)} + b_2 x_{it} + b_3 x_{i(t-1)} + e_{it}. \tag{4}$$

where the only addition is the inclusion of $x_{i(t-1)}$ to represent the lagged independent variable. In this model, the current value of a dependent variable is related to the current value of an independent variable along with prior values of both the independent and dependent variables. The ARDL model is general enough to represent a wide range of organizational dynamics. Researchers can incorporate, for example, more variables to model systems with many influential predictors or longer lags to represent systems with different memory. The second panel in Figure 1 represents the ARDL model graphically, where a set of predictors has both concurrent and lagged (represented with the dotted arrow) influences on the response variable.

The autoregressive distributed lag model presented here is a specific instance of the more general class of linear, time-invariant dynamic models. High-degree polynomial terms may be included to represent complex curvilinear relations, but the ARDL model is linear in parameters. Also, the model parameters are constant over time. The levels of the variables may change in the ARDL model, but the parameters and their estimates are treated as constant over time, meaning that the strength of relationship for a given model component is considered stable over time. Interactions between predictors and an index of time may be included in the model to represent changes in the strength of relationship for a particular predictor, but the parameter estimate on the interaction is not free to vary over time. Time-varying models may be formed by simply including an index representing time as a subscript for any parameter in the model. Time-varying models are challenging to estimate unless the data contain a very large number of observations over time for each unit; we do not discuss them any further here.

Despite the slight differences we presented previously between change and dynamics, what matters is that both incorporate a lagged dependent variable or autoregressive term. We are not suggesting that models that include lagged versions of a predictor are fundamentally dynamic whereas models that include only concurrent predictors are fundamentally change. Rather, change and dynamics are related concepts in which the key factor—memory—is incorporated by modeling lagged versions of the *response* or *outcome* variable through time.

A number of researchers are exploring change and dynamics in empirical work through the use of HLMs. For example, Gabriel, Diefendorff, Chandler, Moran, and Greguras (2014) studied the relationships among work affect, job satisfaction, and perceptions of person-organization (P-O) and person-job fit. Two of their hypotheses include (a) perceived P-O fit relates positively to subsequent positive affect and (b) perceived P-O fit relates negatively with subsequent negative affect. They sampled 142 administrative assistants five times a day for 10 consecutive workdays and evaluated their hypotheses using HLM. The authors report (p. 401) that they used lagged variables such that the current (t) value of the dependent variable was predicted by the prior ($t - 1$) value of the independent variable while controlling for the prior ($t - 1$) value of the dependent variable. The predictors were within-person centered, but it is unclear whether the lagged dependent variables were centered or not. Similarly, Lanaj, Johnson, and Lee (2016) examined the relationship between transformational leadership and affect (in the leader). Two of their hypotheses include (a) transformational leadership will be associated with an increase in positive affect and (b) transformational leadership will be associated with a decrease in negative affect. They collected surveys from 50 managers twice a day over 15 workdays and evaluated their hypotheses using HLM. Values of positive and negative affect collected in the morning were included as covariates or predictors of positive and negative affect reported in the afternoon. Additional types of leader behaviors were included as predictors. The rationale provided for using prior affect as predictors of current or afternoon affect was that they were interested in examining daily changes in affect, thus they controlled for morning affect by entering it as an uncentered variable in the HLM regressions. Other researchers using HLMs to explore dynamic or change patterns include Judge, Simon, Hurst, and Kelley (2014); Johnson, Lanaj, and Barnes (2014); and D'Innocenzo, Luciano, Mathieu, Maynard, and Chen (2016). Table 1 abstracts some of the common research questions and inferences explored by researchers interested in change and dynamics.

Note that HLMs can be used to represent a variety of patterns present in longitudinal data. Researchers may also investigate purely concurrent relations between predictors and outcomes with models that do not include a lagged response variable (e.g., Barnes, Schaubroeck, Huth, & Ghumman, 2011) or growth in an outcome and its covariates (e.g., Li et al., 2017; Zhu et al., 2016). These other applications are not the emphasis of the current study because they do not incorporate lagged dependent variables (DVs) and place their inferences on different fundamental patterns (i.e., not change or dynamics).

Estimation Pitfalls for HLMs With Lagged Response Variables

There are two issues to highlight when researchers explore change or dynamic patterns in longitudinal data. First, models used to investigate change and dynamics typically include prior values of the response variable in the set of predictors. Although the inferential focus might differ slightly (see

Table 1. Common Change and Dynamic Inferences.

Inference	Example
1	There is memory/self-similarity/autoregression in Y.
2	X relates to change in Y.
3	X relates to subsequent change in Y.
4	Change in X relates to change in Y.
5	Change in X relates to subsequent change in Y.

Table 1), the underlying model is some variant of an autoregressive or dynamic model, and parameters are estimated using random coefficient modeling software (e.g., HLM) to account for nesting. This type of model is more and more common in organizational research. For example, a review of papers published in four organizational science journals (*Academy of Management Journal*, *Journal of Applied Psychology*, *Journal of Management*, and *Organizational Behavior and Human Decision Processes*) from 2014 to 2016 yielded 19 instances of such cases where lagged response variables were included in the model. Second, in organizational research, predictors are often within-unit centered when conducting HLM analyses. Within-unit centering is performed either when the focus is on within rather than between-unit relationships or when there are unobserved between-unit variables that are correlated with the predictor of interest. Within-group centering, however, is often not performed when prior values of the response variables are included as predictors in the model. Sometimes this is due to necessity: Within-unit centering of lagged response variables cannot be performed if only two waves of data are available. At other times, within-unit centering is not performed for a specific inferential reason, such as keeping the scale of the predictor and response variables consistent.

The aforementioned issues create a unique set of inferential hazards when researchers explore dynamic or change patterns in longitudinal data. Specifically, researchers must attend to dynamic bias, multi-unit dynamic bias, and dynamic heterogeneity bias. In the following, we explain each and discuss why HLM handles them poorly.

Dynamic Bias

To introduce the concept of dynamic bias, it is useful to remove the complexities that arise from nested observations found in longitudinal data and focus instead on a set of repeated observations for a single unit. In other words, we focus here on problems that arise when modeling relatively short time-series data (i.e., $T < 25$). The dynamic bias problem (Hurwicz, 1950) occurs whenever lagged response variables are included in a model as predictors and the number of repeated observations is relatively small. The dynamics present in these short times-series data can be represented using the simple autoregressive model,

$$y_t = b_0 + b_1 y_{(t-1)} + e_t. \tag{5}$$

where, for instance, y_t might be a person's current level of deep acting in an emotional labor study and $y_{(t-1)}$—immediately prior deep acting—is included as the sole predictor variable to represent the persistence of the person's deep acting over time. In this case, Hurwicz (1950) demonstrated that both ordinary least squares (OLS) and maximum likelihood (ML) estimates of the regression coefficient, b_1, are severely biased when the time-series is short. Later, Marriott and Pope (1954) proved that the bias is downward when $b_1 \in (0, 1)$ and the magnitude of bias is $1/T$ such that the bias decreases as the number of time points increases. In other words, dynamic bias has a substantial impact on parameter estimates in dynamic models when the number of time points, T, is small but

Table 2. Empirical Example of Dynamic Bias.

Parameter	$T = 4$	$T = 11$
$\hat{b}_1$	0.22	0.46
r_{y_{t-1},e_t}	–0.50	–0.31

the bias decreases as T gets large. The underlying cause of dynamic bias is due to violating the standard assumption that the errors are independent of the model predictors when including the lagged dependent variable as a predictor.

To be concrete, we provide a brief example. Consider the simple autoregressive data-generating process,

$$y_t = 1.0 + 0.8y_{t-1} + e_t, \tag{6}$$

where e_t is a white noise process distributed $N(0, 1)$. We illustrate the magnitude of bias present in OLS estimates of the regression coefficient, b_1; the magnitude of correlation between the model predictor, y_{t-1}; and the error, e_t, for short time-series of lengths $T = 4$ and $T = 11$. We sampled 1,000 data sets from this population for each time-series and fit an autoregressive model to the data using OLS estimation. For each fitted model, we recorded the estimate of b_1, and we computed and recorded the correlation between the lagged predictor and the model residuals, r_{y_{t-1},e_t}. The results of our simulations across the 1,000 samples are summarized in Table 2.

These results clearly demonstrate that the estimates for the dynamic regression coefficient, $\hat{b}_1$, are substantially biased downward from the true value of 0.80 as a result of the strong negative correlation between lagged values of the response variable, y_{t-1}, and the errors, e_t. It is also clear that the bias is more severe for shorter time-series due to the larger correlation between the model predictor and the error process. In the context of our example, this bias would mean that we would severely underestimate the persistence of deep acting over time, especially when T is small. The *dynamic bias* demonstrated here is a special case of sampling bias (Hurwicz, 1950; Marriott & Pope, 1954). As the number of repeated observations (T) increases, the bias decreases and both OLS and ML estimates of the dynamic coefficient(s) are consistent when T is large.

Multi-unit Dynamic Bias

We now return to the longitudinal data structures commonly found in organizational science. It is widely recognized that failing to account for clustering of observations within units can result in inaccurate inference, and HLMs are generally viewed as the solution to this problem. In the standard multilevel context, failing to account for nesting results in unbiased parameter estimates, but the associated standard errors are underestimated (Bryk & Raudenbush, 1992; Snijders & Bosker, 1999). This, unfortunately, is not the case when lagged response variables are included as predictors in HLM analyses. To show why this occurs, it is helpful to rewrite the standard HLM in a slightly different form. In its typical form, the standard HLM may be represented as

$$\textit{Level } 1 : y_{it} = b_{0_i} + b_1 x_{1_{it}} \ldots b_p x_{p_{it}} + e_{it}, \tag{7}$$

$$\textit{Level } 2 : b_{0_i} = b_0 + u_i, \tag{8}$$

where y_{it} is the response variable (e.g., job satisfaction) for unit i at time t, b_{0_i} is the regression intercept for the i^{th} unit assumed to be randomly distributed $N(b_0, \sigma_{b_0})$, $b_1 \ldots b_p$ are parameter estimates associated with each of the predictor variables (e.g., stress, organizational commitment, regulatory resources) for the i^{th} unit at time t, and e_{it} is the value of the error term for the i^{th} person at time t and is assumed to be distributed $N(0, \sigma_e)$. It is instructive to rewrite this standard HLM to include the grand intercept and the specific unit intercept effects as

$$y_{it} = b_0 + b_1 x_{1_{it}} \ldots b_p x_{p_{it}} + u_i + e_{it}, \tag{9}$$

where all terms are defined as previously described, b_0 is the grand intercept across units, and u_i represents the unit heterogeneity modeled as the intercept for the i^{th} unit that is assumed to vary randomly across units but is constant over time within each unit.

We first focus on the HLM model with predictors that are not within-group centered. To yield unbiased estimates, HLM requires that the u_i be uncorrelated with the predictors in the model (Wooldridge, 2010). This assumption is met in the standard multilevel context, but it must be violated when lagged response variables are included as predictors in the model. To be explicit, consider the previous model with one predictor and a lagged response variable as an additional predictor variable:

$$y_{it} = b_0 + b_1 x_{1_{it}} + b_2 y_{it-1} + u_i + e_{it}. \tag{10}$$

In this case, the unit effects, u_i, must correlate with the lagged dependent variable because the model requires that u_i has a direct effect on every y_{it}. If u_i is a component of y_{it}, then it must have also been a component of prior responses and so must be related to $y_{i(t-1)}$ (Nickell, 1981). Violating this independence assumption upwardly biases the coefficient for the lagged dependent variable and under most conditions, downwardly biases the coefficients for other predictor variables in the model. Moreover, this bias does not completely go away with large T or large N (see the next section for a simulation).

Dynamic Heterogeneity Bias

What happens when the predictor variables are within-group centered? Unfortunately, a different yet important problem arises under within-group centering. The standard HLM with a single group mean–centered predictor and no lagged response variables can be written as

$$y_{it} = b_0 + b_1 (x_{1_{it}} - \bar{x}_{1_i}) + u_i + e_{it}. \tag{11}$$

Within-group centering effectively removes the correlation between the random unit effects, u_i, and the transformed covariates $x_{1_{it}} - \bar{x}_{1_i}$ (Enders & Tofighi, 2007). This removes the multi-unit dynamic bias presented in the preceding text. However, within-group centering introduces a different assumption violation that results in a different source of bias we term *dynamic heterogeneity bias*.

To avoid introducing unnecessary complexity, we focus on an HLM model with random intercepts, a lagged response variable, and no additional covariates. This simple model may be represented as

$$y_{it} = b_0 + b_1 y_{i(t-1)} + u_i + e_{it}. \tag{12}$$

Now, if we within-group center the lagged response variable, the HLM becomes

$$y_{it} = b_0 + b_1 (y_{i(t-1)} - \bar{y}_{i\cdot}) + u_i + e_{it}. \tag{13}$$

In this case, $\bar{y}_{i\cdot}$ represents the average of $y_{i(t-1)}$ for each unit. The centered variable $(y_{i(t-1)} - \bar{y}_{i\cdot})$ will be uncorrelated with the random intercepts representing unit effects, u_i, hence eliminating this source of bias. However, $\bar{y}_{i\cdot}$ must now be correlated with the error term, e_{it}, because $\bar{y}_{i\cdot}$ is constructed from y_{it} across all time points (except the last time point), which have e_{it} as a component in each time period. Therefore, the within group–centered lagged response variable $(y_{i(t-1)} - \bar{y}_{i\cdot})$ will be

Table 3. Empirical Correlation for $N = 100$ and $T = 6$.

	$y_{i(t-1)}$	$y_{i_{within-centered}}$	u_i	e_{it}
$y_{i(t-1)}$	—	—	—	—
$y_{i_{within-centered}}$	0.15	—	—	—
u_i	0.91	0	—	—
e_{it}	0	–0.28	0	—

negatively correlated with e_{it}. This correlation causes bias even if the number of units, N, increases to infinity. The existence of this problem has been known in the econometrics literature since Nickell (1981), and it is commonly referred to as the Nickell bias. Nickell showed that the estimate of b_1 on the lagged response variable is downwardly biased as a function of $1/T$, irrespective of the sample size, and that the coefficients on any other predictors in the model are also biased.

As previously described, we use a simple simulation to demonstrate the effects of these two sources of bias that arise when including lagged response variables as predictors in HLMs. Assume the true data-generating process is

$$y_{it} = 1 + b_1 y_{i(t-1)} + u_i + e_{it}, \tag{14}$$

where $u_i = 0.1 * i$, $e_{it} \sim N(0,1)$, and b_1 range from 0.1 to 0.9. We estimate the model using HLM with both centered and uncentered versions of the lagged response variable using 100 units and six time points across 4,500 independent samples from the data-generating process. Table 3 presents the problematic correlations among the relevant terms, where $y_{i(t-1)}$ represents the lagged value of the response variable *without* within-group centering, $y_{i_{\text{within-centered}}}$ is the within group–centered lagged value of the response variable, and u_i and e_{it} are defined as previously described. The large positive correlation between the uncentered values of the lagged response variable and the random unit intercept effects results in *multi-unit dynamic bias*, and the substantial negative correlation between the centered values of the lagged response variable and the random errors is responsible for *dynamic heterogeneity bias*.

Figure 2 presents a graph of the bias in b_1 as a result of uncentered versus centered HLM models. HLMs that use an uncentered lagged response variable consistently overestimate the coefficient, whereas HLMs that use a group mean–centered lagged response variable consistently underestimate b_1.

Alternative Approaches

Fortunately, researchers are not limited to HLM; there are alternative methods for estimating models with lagged response variables that are designed to overcome the problems discussed previously. The first alternative is an instrumental variable-based approach common in economics, and the second is a relatively new SEM approach.

Estimation via GMM-IV

A number of estimation procedures have been developed to solve the problem of dynamic panel bias in the field of economics, including Anderson and Hsiao's (1982) instrumental variable (IV) approach, Kiviet's (1995) least squares dummy variable (LSDV) bias correction procedure, and Arellano and Bond's (1991) GMM-IV approach. Of these, the Arellano and Bond GMM-IV approach is, by far, the most commonly used method for estimating dynamic panel models, such as the ARDL model, in economics and political science.

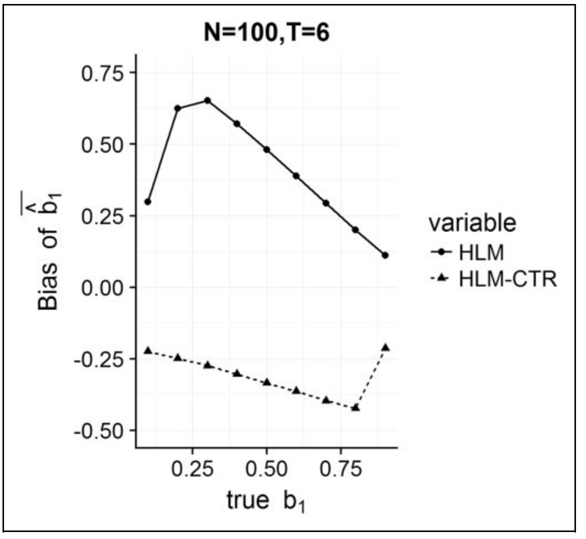

Figure 2. Bias in $\hat{b}_1$ as b_1 increases for $N=100$ and $T=6$.

We focus here on the logic of the GMM-IV approach rather than the computational details. Detailed descriptions on the method are found in Arellano and Bond's (1991) original paper and across numerous discussions and extensions of the method (e.g., Wansbeek & Bekker, 1996). In addition, several software packages (e.g., SAS, Stata, R) now include routines to implement the GMM approach and provide detailed descriptions of the actual computations (e.g., xtabond in Stata).

Arrelano and Bond's (1991) GMM-IV approach is an extension of Anderson and Hsiao's (1982) IV method. Following the previous discussion, we focus on a model with random intercepts, a lagged response variable, and no additional covariates. This simple model may be represented as

$$y_{it} = b_0 + b_1 y_{i(t-1)} + u_i + e_{it}. \tag{15}$$

The model can then be rewritten as a first-differenced model to remove unobserved heterogeneity:

$$y_{it} - y_{it-1} = b_0 + b_1(y_{i(t-1)} - y_{i(t-2)}) + (e_{it} - e_{it-1}). \tag{16}$$

In the first-differenced model, the correlation between the transformed lagged values of the dependent variable $(y_{i(t-1)} - y_{i(t-2)})$ used as predictors and the transformed errors $(e_{it} - e_{it-1})$ results in biased estimates even as $N \to \infty$. One approach to solve this problem is to use an instrumental variable that correlates with the term $(y_{i(t-1)} - y_{i(t-2)})$ but is uncorrelated with the errors $(e_{it} - e_{it-1})$. Under the sequential exogeneity assumption, stating that errors (i.e., shocks) in the future are independent of past values of the response variable (e.g., a random change in job satisfaction at a certain time point does not correlate with past values of job satisfaction), a natural instrument is provided by past values of the response variable (e.g., $y_{i(t-2)}$), which will correlate with the transformed lagged response variable $(y_{i(t-1)} - y_{i(t-2)})$ but will be uncorrelated with the transformed error term $(e_{it} - e_{it-1})$. Using job satisfaction as an example, this means we can use the past values of job satisfaction as instrumental variables for the immediate transformed lagged value of job satisfaction to achieve consistent estimation. The Anderson-Hsiao estimator uses only a single instrument, but the GMM-IV approach includes all possible past values of the response variable as instruments (i.e., $y_{i1}, y_{i2}, \ldots, y_{iT-2}$ as instruments for $(y_{i(T-1)} - y_{i(T-2)})$) and then uses a GMM estimator to obtain parameter estimates (Arellano & Bond, 1991).

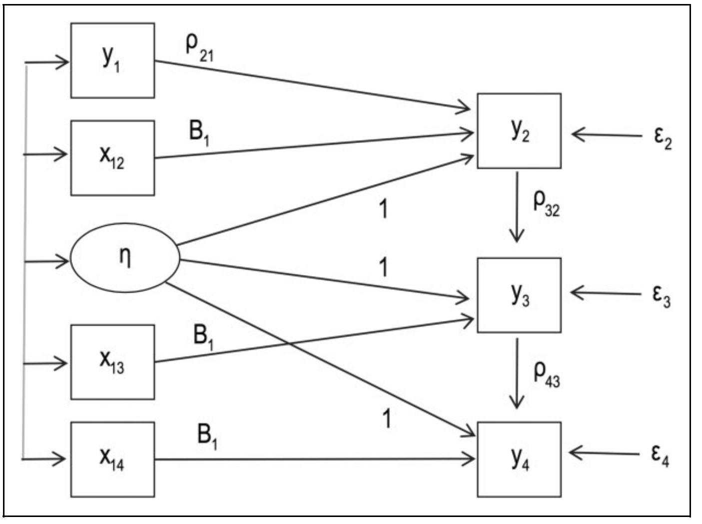

Figure 3. Bollen and Brand's (2010) structural equation modeling (SEM) approach to estimating the autoregressive distributed lag (ARDL) model parameters.

This GMM-IV approach is generally consistent and efficient as $N \to \infty$, but in empirical work, the optimal number of moment conditions that should be used for estimation is not clear (Judson & Owen, 1999; Kiviet, 1995; Wansbeek & Bekker, 1996). Further, simulations by Ziliak (1997) also show that there could be a downward bias in GMM estimates as the number of moment conditions increases. Perhaps more important for organizational scientists, the GMM-IV approach uses unfamiliar model formulations dealing with instrumental variables, and GMM estimation is not widely used in organizational research.

Estimation via SEM

In a little known paper, Bollen and Brand (2010) provided an approach to estimating parameters in dynamic models using the structural equation modeling approach familiar to most organizational scientists. Figure 3 provides a graphic depiction of their model. Here, Y represents the outcome of interest, and X is the contemporaneous exogenous variable (second subscripts indicate different time points). The path coefficients from the exogenous variable X to Y are constrained to be equal. The latent time-invariant variable, η_i, representing unobserved heterogeneity (or random intercept in HLM), has a path coefficient of 1 to the response variable Y at all time points and is allowed to correlate with both the exogenous variables and the response variable at the initial time point, Y_1. In practice, these restrictions can be loosened and evaluated, but we use this default model in all the evaluations presented in the following, except that we constrain the path coefficients from the lagged values of Y to Y to be equal as well. Using the job satisfaction example, this SEM approach models job satisfaction as a function of lagged job satisfaction and contemporaneous exogenous variables such as stress and organizational commitment. The model treats time-invariant between-unit unobserved heterogeneity in job satisfaction as a latent variable and allows it to correlate with both the lagged dependent variable as well as the time-varying covariates such as stress and organizational commitment.

In principle, the Bollen and Brand (2010) model should provide accurate estimates of change or dynamics since (a) it models unobserved heterogeneity as a distribution rather than modeling each individual specific effect separately, which avoids the biased estimation of each individual specific effect when T is small, commonly known as the incidental parameter problem (Lancaster, 2000), and (b) it allows correlation between unobserved effects and exogenous variables or lagged dependent variables, which is the correct specification of the underlying relationship. It also (c) conditions on the initial observation for the response variable, thereby avoiding the well-known initial conditions problem that occurs when using maximum likelihood estimation (Anderson & Hsiao, 1982; Wooldridge, 2010). Few simulation studies have been performed to evaluate the performance of this approach. In the following sections, therefore, we describe and implement a Monte Carlo simulation to examine the performance of this estimation method as compared to centered and uncentered HLM estimation and GMM-IV estimation.

Monte Carlo Study

We are primarily interested in how well the various estimation approaches proposed previously are able to recover the underlying data-generating mechanism when there are both unit effects (i.e., clustered responses) and a lagged response variable in the model. Specifically, we compare the performance of four approaches estimating panel and dynamic panel model parameters: (a) the random intercepts HLM approach using uncentered predictors, (b) the random intercepts HLM approach using within-unit centered predictors, (c) the Arrelano and Bond GMM-IV approach, and (d) Bollen and Brand's SEM approach. For each of these models, we generate simulated data across a wide range of conditions and then evaluate parameter estimate accuracy (i.e., bias) and the Type I error rates for the tests of significance.

Data Generation

Consistent with prior simulation research on this topic, we use a simple form of the autoregressive distributed lag model (ARDL) with one lagged dependent variable, one contemporaneous exogenous variable, and a time-invariant individual specific effect to generate the longitudinal data. The canonical data generating equation for our simulations is

$$y_{it} = \gamma y_{i(t-1)} + \beta x_{it} + u_i + e_{it}, \tag{17}$$

where y_{it} represents the response variable for the i^{th} unit at time t, x_{it} represents an exogenous variable assessed contemporaneously with y_{it} that is generated as a $N(0, \sigma_{x_{it}})$ white noise process yielding independent values with homogeneous variance across units. $y_{i(t-1)}$ is the lagged response variable for the i^{th} unit; u_i is a unit-specific parameter representing the combined effect of all unobserved, time-invariant variables responsible for between unit differences; and e_{it} is an independent, random sample from a $N(0, 1)$ distribution reflecting process error. For the simulations, the time-invariant unobserved variable u_i for each unit was drawn from a normal distribution $u_i \sim N(0, 1)$, and the initial condition or starting value was obtained as an independent, random sample from a $N(0, 1)$ distribution. Finally, note that the notation used for the ARDL parameters has been altered slightly to ease the presentation of the Monte Carlo simulation. Specifically, we use γ to represent the coefficient for the lagged response variable $y_{i(t-1)}$ and β to represent the coefficient for the exogenous variable x_{it}. The interpretation of these model parameters remains as described previously.

Simulation Design

The Monte Carlo simulations were designed to be sensitive to a number of factors. First, as shown by Clark and Linzer (2015), when the within-unit variance of the exogenous variable, $\sigma^2_{x_{it}}$, is substantial, the random effects approach performs as well as the within-unit centered approach. As a result, it is important to determine whether the within-unit variation of the exogenous variable influences bias in the parameter estimates and Type I error rates. Second, Kiviet (1995) demonstrated that the magnitude of the coefficients for both the lagged dependent variable and the exogenous variable impacts estimation bias. As a result, we need to systematically vary the coefficients of both the lagged dependent variable and the exogenous variable. Third, Nickell (1981) demonstrated that the within-unit centering approach generates consistent estimates as T approaches infinity, but N and T are typically small in the majority of organizational science data structures—namely, often the data reflect small T and larger N. Therefore, it is important to examine bias and Type I error rates under realistic conditions for N and T in longitudinal data.

To study these factors, we used the following two experiment conditions:

1. In the first set of simulations, we generated data using the aforementioned data-generating mechanism and varied the standard deviation of the exogenous variable, $\sigma_{x_{it}}$, from 1 to 9 while keeping β and γ fixed at 0.5. The rationale for this approach is based on the knowledge that in our data-generating model, the exogenous variable x correlates with the outcome of interest y but does not correlate with the unobserved heterogeneity term, u_i. Larger variances in x (as compared with the variance in unobserved heterogeneity) will result in larger within-unit variation in y (i.e., the part of y that is uncorrelated with unobserved heterogeneity, u_i) that in turn decreases the correlation between lagged y and unobserved heterogeneity. Varying the standard deviation of x allows us to evaluate the performance of the approaches under different levels of correlation between y and unobserved heterogeneity. This correlation is the key to the bias in most approaches. Furthermore, by varying the standard deviation of the exogenous variable $\sigma_{x_{it}}$ from 1 to 9, we can generate a wide range of values for the percentage of within-unit variance (roughly 20% to 80%) in the simulated data, which substantially overlaps with what is usually seen in empirical organizational research.
2. In the second sets of simulations, we generated data using the same data-generating equation as previously but set the standard deviation of x to a constant ($\sigma_{x_{it}} = 4$) and varied β and γ from 0.1 to 0.9 under the constraint that $\beta + \gamma = 1$. The choice of ($\sigma_{x_{it}} = 4$) generates reasonable amount of within-unit variation (roughly 50% within-unit variation when β and γ are 0.5). The constraint of $\beta + \gamma = 1$ follows Kiviet (1995), and it is needed for several reasons: (a) As γ increases and β decreases, the correlation between lagged y and unobserved heterogeneity increases. Therefore, varying the true coefficients makes it possible to examine the performance of the approaches under different levels of correlation between y and unobserved heterogeneity. (b) This constraint ensures that the data-generating process remains strictly stationary, and (c) the constraint implies that $\beta = 1 - \gamma$, so a change in γ only changes the dynamics of the relationship and does not confound the equilibrium relationship between y and x.

In each set of simulations, we examine two levels of N ($N = 150$ or $N = 250$) and T ($T = 6$ or $T = 12$) to represent real-world scenarios (in applied psychological research) where we have a relatively large number of units and a small number of time points. For example, these sample sizes and time points are consistent across research on mood and personality (Chi et al., 2015; Judge et al., 2014), turnover and performance (Call, Nyberg, Ployhart, & Weekley, 2015; Cao, Lemmon, Pan, Qian, & Tian, 2018), and other studies reported at the beginning of this article. In each simulation, we allow for at least 20 rounds of burn-in before we take the sample for estimatation to enable the trajectories to reach steady states. Both Nickell (1981) and Hurwicz (1950) showed that the estimate for the dynamic part of the model, γ, is less biased as T increases, mostly due to the decreased correlation between the lagged dependent variable and unobserved heterogeneity. However, data sets with large T are relatively uncommon in organization settings, and therefore, it is essential to

Table 4. Simulation Design.

	Parameters
Experiment 1	$\sigma_{x_{it}}$: 1 − 9; *N*: 150 or 250; *T*: 6 or 12; $\beta = 0.5$; $\gamma = 0.5$
Experiment 2	$\sigma_{x_{it}} = 4$; *N*: 150 or 250; *T*: 6 or 12; β:0.1 to 0.9; γ: 0.1 to 0.9; $\beta + \gamma = 1$

understand the robustness of the approaches to short time frames. Crossing these factors yields 36 ($4 \times 9 = 36$) distinct conditions in the first set of simulations and 36 ($4 \times 9 = 36$) distinct conditions in the second set of simulations. Details of the simulation parameters are summarized in Table 4. We performed 1,000 Monte Carlo simulations for each condition.

Analyses

We analyzed the simulated data using R and recorded the estimates for β, γ, and their respective standard errors. We estimated the uncentered and within-unit centered HLM models using restricted maximum likelihood estimation with the lme4 package. We estimated the GMM-IV model with all past values of y as instruments using generalized methods of moment estimation as implemented in the plm package. We estimated the structural equation model with maximum likelihood estimation as implemented in lavaan package. We assess the performance of each model on recovering β and γ in the data-generating model by calculating the mean bias and Type I error. Mean bias is calculated as the difference between the mean of the estimates and the true parameter value. Type I error is calculated as the percentage of tests that falsely rejects the null hypothesis under 5% significance level when the null hypothesis is true. Note that to calculate Type I error, we did not set the null hypothesis to be 0 but set it to equal to the true value of the parameter in the data-generating process. We calculate these measures for each of the 1,000 samples generated in each simulation condition.

Results

Study 1: Impact of σ_x

Bias. The impact of the standard deviation of the exogenous variable, σ_x, on the estimation of the exogenous regression parameter, β, and the autoregressive parameter, γ, is shown in Figures 4 and 5, respectively, for each of the four estimation methods. With respect to the regression parameter (β) for the exogenous variable (x), Figure 4 shows that as $\sigma_{x_{it}}$ increases, the estimates of the coefficient for the exogenous variable, β, yield little bias (smaller than 0.05), with the largest bias in the within-unit centered HLM approach. The bias decreases as σ_x increases in all four approaches. Additionally, bias decreases as T increases, but it is relatively insensitive to sample size, N. Note that the bias for β is relatively small as x is exogenous and does not directly correlate with the unobserved heterogeneity term u_i. However, the bias for β will be larger in situations where the x is also a function of u_i or the response variable y (e.g., cross-lagged models).

Figure 5 depicts the estimation bias for the dynamic coefficient γ (i.e., the regression coefficient for the lagged dependent variable Y_{t-1}) as the standard deviation of the exogenous variable (σ_x) increases. Generally, there is far more bias in estimating the dynamic coefficient (γ) than in estimating β. When the standard deviation of x is relatively small ($\sigma_x = 1 \sim 5$), there is substantial bias in estimates from all four approaches; uncentered HLM and within-unit centered HLM perform the worst by producing the largest upward bias and downward bias, respectively (mean biases are as large as 0.35 and –0.25, respectively, for the uncentered and within-unit centered HLM when $\sigma_x = 1$). In contrast, the SEM and GMM approaches produce much smaller bias in the same

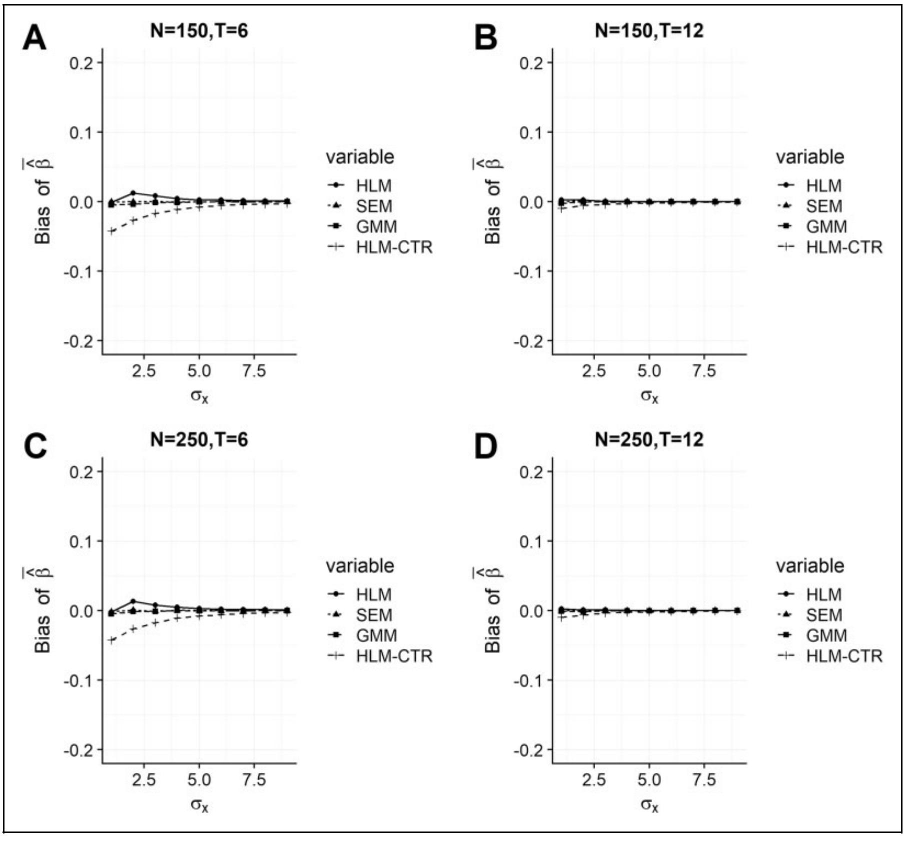

Figure 4. Bias in $\hat{\beta}$ as σ_x increases across patterns of N and T.

situation across all configurations of N and T. As σ_x increases, the bias decreases for estimates in all four approaches. As previously described, the magnitude of bias decreases substantially as T increases, especially for the uncentered and within-unit centered HLM. However, the magnitude of bias is not meaningfully affected by the number of units, N.

Overall, these results demonstrate that there is much more bias in estimating the dynamic coefficient (γ) than in estimating the coefficient for the exogenous variable (β). Uncentered and within-unit centered HLM perform worse than SEM and GMM approaches when estimating γ, especially when σ_x is relatively small. Furthermore, the magnitude of bias for estimates in all four approaches decreases as T increases but is relatively unaffected by changes in the number of units, N.

Type I Error Rates. The Type I error rates for the significance tests on the parameter estimates for β and γ are shown in Figures 6 and 7, respectively. The true values of β and γ were set to 0.5 in the data-generating process, and they represent the null hypothesis. With respect to the exogenous variable, as σ_x increases, the Type 1 error rates for the test of β are reasonably well maintained with respect to the nominal alpha of 0.05 for all combinations of N and T. The exception to this pattern occurs for the within unit–centered HLM when $T = 6$ and $1 \leq \sigma_x \leq 7$. Under these conditions, the Type I error rates for the within-unit centered HLM are unacceptably high, ranging from 0.15 to 0.40.

Type I error rates for the significance test of the dynamic parameter estimate, γ, as σ_x increases are shown in Figure 7. Both the uncentered and within unit–centered HLM approaches fail to

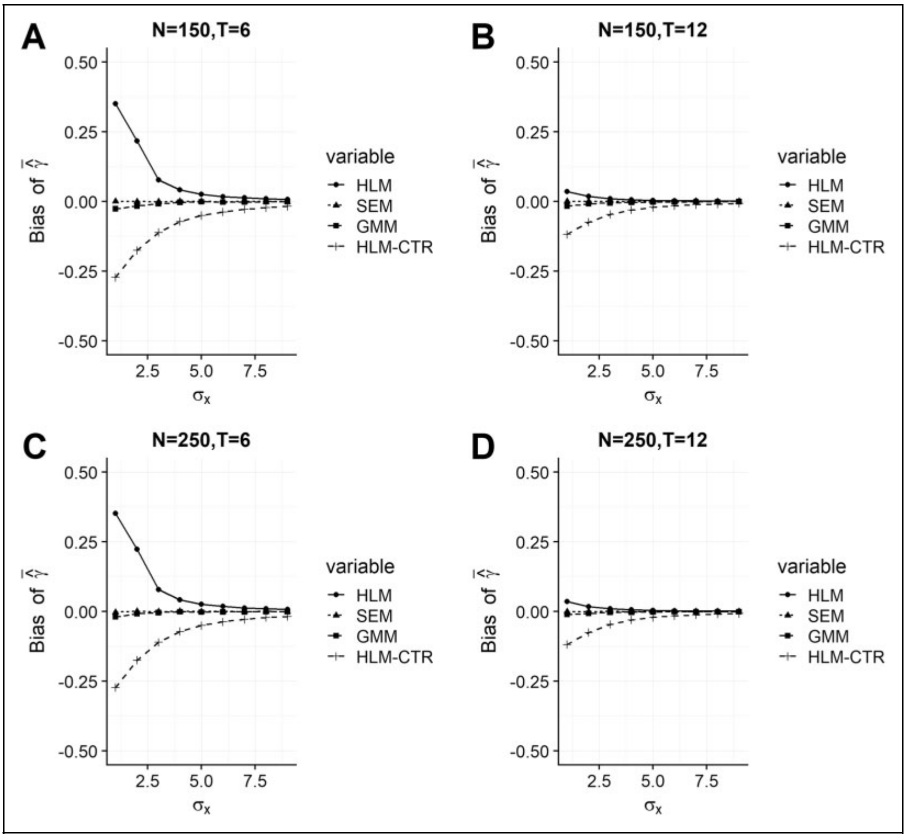

Figure 5. Bias in $\hat{\gamma}$ as σ_x increases across patterns of *N* and *T*.

maintain acceptable Type I error rates, and this result is particularly pronounced for small values of σ_x or when T is small. The significance tests for the GMM and SEM approaches generally maintain the nominal alpha level across all conditions.

Overall, the results depicted in Figures 6 and 7 indicate that Type I error rates can be problematic for testing the dynamic coefficient (γ) using the uncentered and within unit–centered HLM approaches and testing the coefficient of the exogenous variable (β) using the within unit–centered HLM approach, especially when the standard deviation of x is small or when T is small. Both SEM and GMM maintain good control of Type I error rates for the significance test on the coefficients of the exogenous variable (β) and the lagged dependent variable (γ). Generally, the Type I error rates did not vary much with the standard deviation of x, number of time points, or the number of units for the SEM and GMM approaches.

Study 2: Magnitude of dynamics

Bias. Bias in the estimation of γ for the lagged dependent variable y_{it-1} as the size of the true coefficient in the data-generating process increases is shown in Figure 8. Results show that when γ is small, all four approaches produce a minimum amount of bias. However, problems arise for the uncentered and within unit–centered HLM and GMM estimates as the magnitude of the dynamic coefficient, γ, increases. Interestingly, the uncentered HLM estimates are upwardly biased, while the

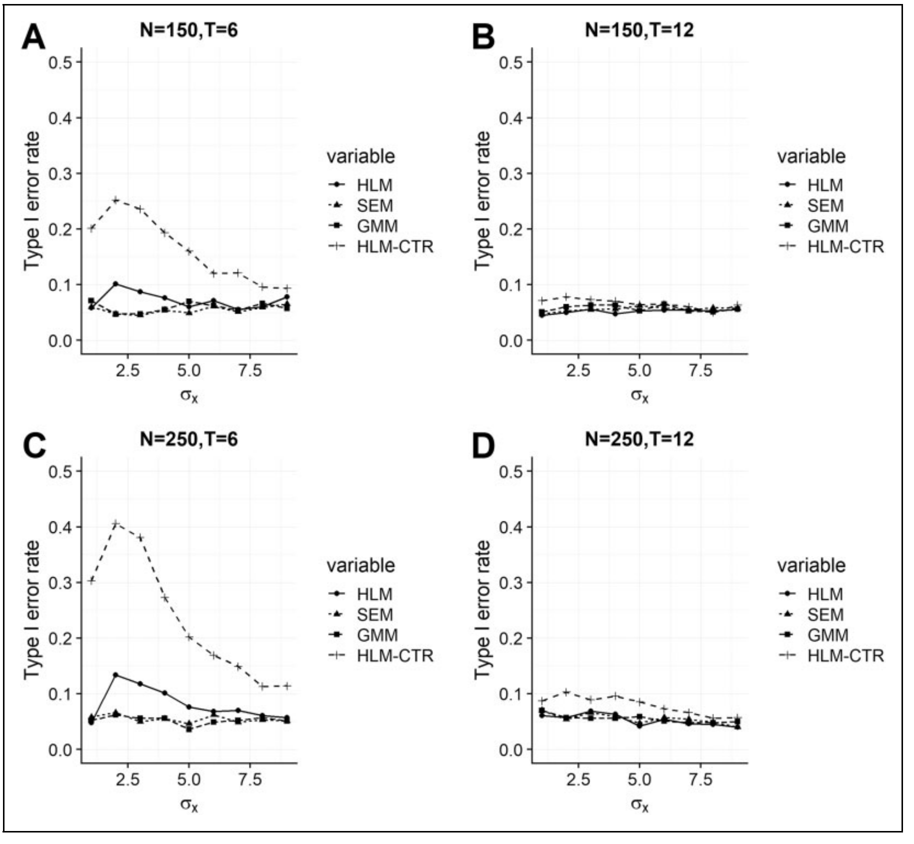

Figure 6. Type I error rate for $\hat{\beta}$ as σ_x increases across patterns of N and T.

within group–centered HLM and GMM-IV estimates are downwardly biased, consistent with Nickell (1981) and Ziliak (1997).

In addition, the size of T and N have different effects on the bias of the four approaches. Overall, the SEM approach yields the least biased estimates across all the conditions. There is a slight upward bias in the SEM estimates when the true coefficient of γ is relatively large, and the magnitude of the bias appears to be insensitive to both the number of units N and the number of time points, T.

The GMM estimates tend to underestimate the size of true γ, and this pattern is particularly pronounced for larger autoregressive effects when the number of time points, T, is small. Even in the best case scenario studied ($N = 250$, $T = 12$), the GMM estimates are still downwardly biased when the dynamics are strongest.

The within unit–centered HLM estimates have unacceptable levels of downward bias and perform worst when the dynamics in the true data-generating process are large. This is particularly problematic since the magnitude of the downward bias is 0.4 when the true γ is approximately 0.9. Interestingly, the magnitude of bias decreases with the number of time points, T, while the number of units, N, does not have much impact on the bias of within unit–centered HLM approach.

The uncentered HLM estimates are upwardly biased, and the approach performs worst when γ in the true data-generating process is large. This is particularly problematic since the upward bias is ≤ 0.25 when the true γ is approximately 0.7. As a result of this pattern of bias, a stationary dynamic process could easily be mistaken for a random walk or an exponential growth. Similarly, the

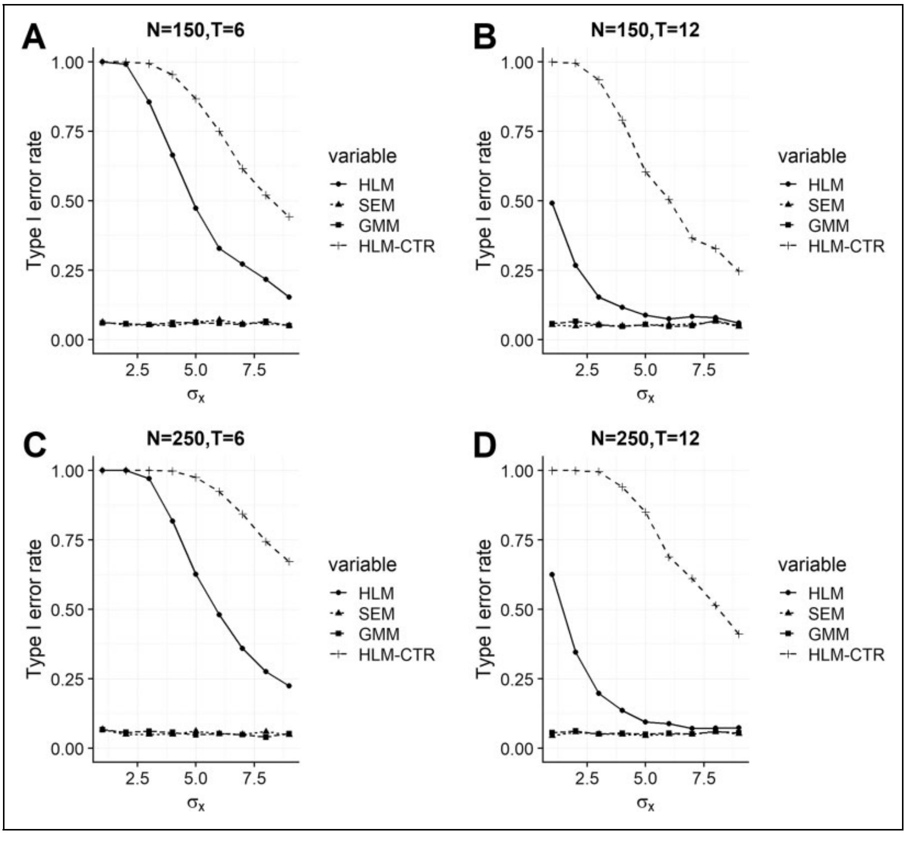

Figure 7. Type I error rate for $\hat{\gamma}$ as σ_x increases across patterns of N and T.

magnitude of bias decreases with the number of time points, T, while the number of units, N, does not have much impact on the bias of uncentered HLM approach.

Type I Error Rates. Type I error rates for testing the estimate of the lagged dependent variable ($\hat{\gamma}$) as the magnitude of the true coefficient of lagged dependent variable increases are shown in Figure 9. This evaluation differs from the typical evaluation of Type I error rates, where it is assumed under the null hypothesis that the population parameter of interest equals zero. When studying dynamic systems, there are many interesting null hypotheses where the autoregressive coefficient in the data-generating process does not equal zero. For instance, when the absolute value of γ is less than 1.0, then the response variable converges to a stationary equilibrium. When γ equals 1.0, then the response variable follows a random walk. If γ is greater than 1.0, then the response variable exhibits explosive growth. As a result, it is important to understand the Type I error rate for the test of γ across a wide range of potential non-zero null hypotheses. Therefore, in our analyses, the null hypothesis assumes the population parameter of interest equals the true value of γ in each experimental condition. DeShon (2012) provides additional details on the macro behavioral dynamics for various values of γ.

Figure 9 shows that when the true coefficient of γ is small, Type I error rates are generally small for all of the approaches except the within-unit centered HLM. However, as γ increases, the Type I error rates for all of the approaches increase in varying degrees, with the most drastic increase in the within unit–centered and uncentered HLM approaches. When the true

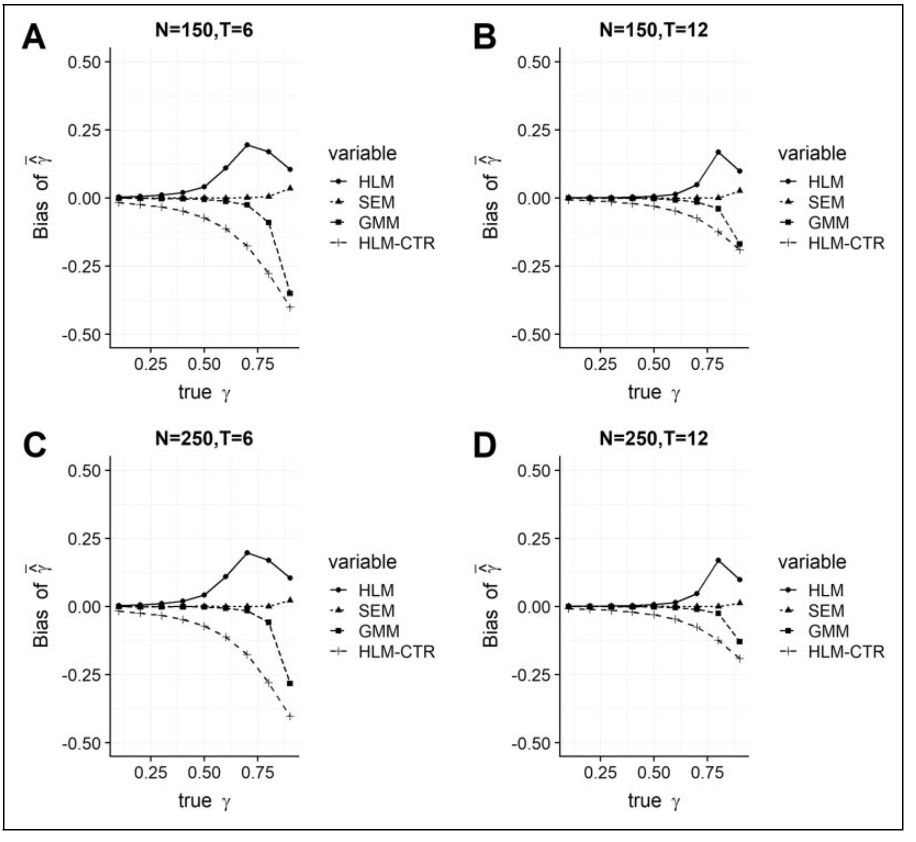

Figure 8. Bias in $\hat{\gamma}$ as γ increases across patterns of N and T.

coefficient of $\gamma > 0.8$, the Type I error rates of both the centered and uncentered HLM approaches are 100% across all conditions. That is, the significance test on the dynamic coefficient will always falsely reject the null hypotheses if the true γ is greater than 0.8. The SEM and GMM approaches also struggle, but to a lesser degree, to maintain nominal Type I error rates for larger values of γ, with the SEM Type I error rates generally having a slight advantage over the GMM error rates.

Additionally, the number of time points and the number of units have different effects on Type I error rates for the four approaches. For the centered and uncentered HLM approaches, although Type I error rates always increase from 0 to 1 as the true value of γ increases, high Type I error rates occur for larger values of true γ when there are more time points. The number of units, N, does not have much impact on the Type I error rates of both HLM approaches. For the GMM approach, Type I error rates slightly increase with the increase in the number of time points, T, while the Type I error rates in the SEM approach are generally not affected by the number of time points, T, or the number of units, N.

Additional Robustness Check

In Study 2, β and γ were constrained to sum to 1 to ensure a stationary data-generating process and constant equilibrium relationship between y and x. As true γ increases, the true β decreases, which reduces the amount of within-unit variation in y resulting from the

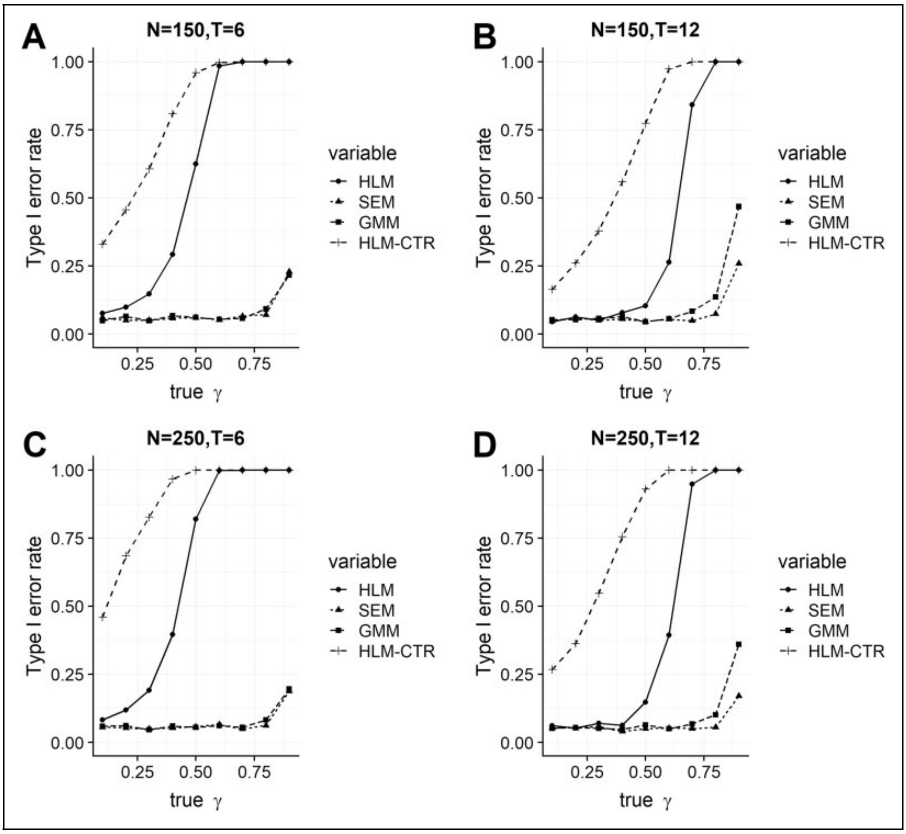

Figure 9. Type I error rate for $\hat{\gamma}$ as γ increases across patterns of *N* and *T*.

exogenous variable x and thus increases the correlation between y and the unobserved heterogeneity term. To ensure the aforementioned results are not completely driven by the constraint or the change in true β, we performed an additional simulation experiment with β fixed at 0.5 while γ ranged from 0.1 to 0.9. Bias and Type I error rates for testing the dynamic coefficient ($\hat{\gamma}$) when $N = 150$ and $T = 6$ are shown in Figure 10. In terms of bias, while all approaches generate smaller magnitudes of bias, the uncentered HLM approach is still upwardly biased, the within unit–centered HLM and GMM approaches are downwardly biased, and the SEM approach contains the smallest amount of bias. In terms of Type I error rates, the uncentered and within unit–centered HLM still generate substantial Type I error rates, especially when true γ is large. The Type I error rates for GMM and SEM also follow a similar pattern as before except that the error rates go down when true γ is large. Overall, the bias and Type I error rates in Figure 10 show very similar and consistent patterns as previously described, indicating that the results obtained are not an artifact of the constraints we put on the simulation parameters.

Discussion

Models of change and dynamics often include prior values of the response variable as predictors. Because longitudinal data consist of observations over time within units, it is necessary to account

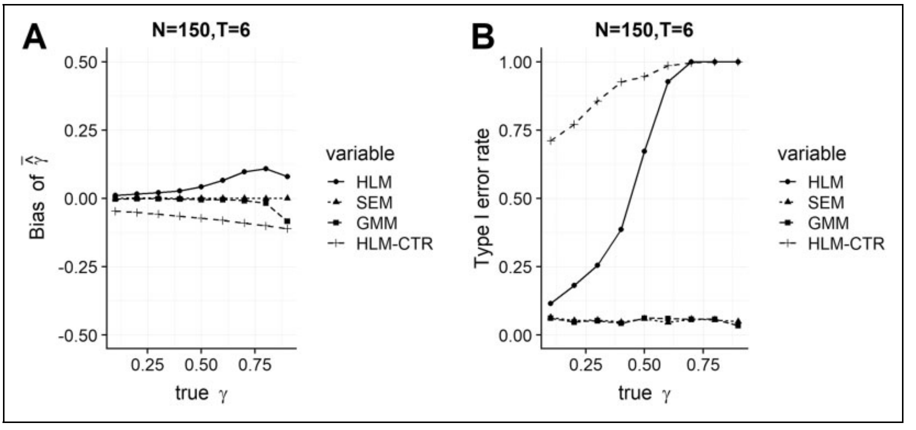

Figure 10. Bias and Type I error rate for $\hat{\gamma}$ as γ when β is fixed at 0.5.

for dependencies in the data resulting from nesting or clustering. Given the success of the HLM approach in the multilevel literature, it is the obvious choice for examining relationships in longitudinal data. HLM can be appropriate for some patterns in longitudinal data (e.g., growth), but when researchers model change or dynamics, key independence assumptions required by HLM are violated when lagged responses are included as predictors. The exact nature of the assumption violation depends on whether the prior values of the response variable are within-unit centered.

As reported in Table 3, if the prior response values are *not* within-unit centered, then these values must correlate with the between-unit effects (i.e., random intercepts). HLM, though, requires that the unit effects and the predictors in the model are independent. Since they are not, in this case, biased estimation and erroneous inferences result. This form of bias occurs when lagged response variables are included in HLM models with longitudinal data that contain between-unit heterogeneity (i.e., between unit differences)—we therefore refer to this bias as *multi-unit dynamic bias*. Our Monte Carlo simulations demonstrate that multi-unit bias causes the parameter estimate for the lagged DV (γ) to be upwardly biased, and under most conditions, the coefficients for the other predictor variables in the model are downwardly biased. Although our simulations show that the bias is mitigated with increases in T (possibly due to larger within-unit sample size to estimate the variance component/unobserved heterogeneity term), this form of bias is hardwired into HLM and does not completely go away even when we have large N or T.

If all the model predictors—including the lagged DV—are within-unit centered, then the correlation between the unit effect and the lagged DV is eliminated. However, within-unit centering the prior values of the response variable induces a correlation between those values and the errors, violating the common assumption that model predictors and the errors are uncorrelated. We refer to this form of bias as *dynamic heterogeneity bias*. In this case, the parameter estimate for the prior responses (γ) is downwardly biased as a function of $1/T$, irrespective of the sample size, and the coefficients for any other predictors in the model are also biased. Our simulation results show that the magnitude of bias and associated Type I errors are more severe (a) when within-unit variation (σ_x) is low, (b) when the true coefficient of lagged response variable (γ) is large, and (c) when number of time points (T) is relatively small. In practice, as *dynamic heterogeneity bias* is largely a function of $1/T$, the magnitude of bias for estimating lagged response variable and other predictors may be less of a concern once researchers have a relatively large T (e.g., T>25), as appeared in some recent organizational research (Bono, Glomb, Shen, Kim, & Koch, 2013; Lanaj et al., 2016). Taken together, the aforementioned issues demonstrate that HLM is rarely the appropriate approach when the inferential focus is change or dynamics.

The GMM instrumental variables approach developed by Arellano and Bond is widely used outside of organizational science to fit dynamic models. Our results show that in general, this approach works well and could be used as an alternate if desired. The approach, however, does have limitations. The GMM-IV approach is consistent (i.e., converges asymptotically to the population parameters) but tends to be biased in small samples, particularly when the autoregressive parameter (γ) is large. This can be seen in Figure 8 and in prior simulation work evaluating the GMM-IV approach (e.g., Blundell & Bond, 1998; Kiviet & Phillips, 2014). Another downside of the GMM-IV approach is that the estimates are not efficient (e.g., Ahn & Schmidt, 1995), meaning that it requires more information (e.g., larger sample sizes) to achieve a given level of precision than more efficient estimators such as maximum likelihood (Anderson & Hsiao, 1982). Finally, selecting the instruments to be used in the GMM-IV method is not clear-cut, and using either too few or too many instruments can increase the magnitude of small sample bias (Roodman, 2009).

Based on the results presented here, we recommend the Bollen and Brand (2010) approach using structural equation modeling when researchers are interested in change or dynamics. SEM is familiar to most organizational scientists, and the Bollen and Brand approach can be easily generalized to more complex models, including, for instance, reciprocal relations, and this approach performs extremely well across a wide range of realistic conditions. Importantly, the dynamic parameters in the Bollen and Brand SEM approach are purely within-unit—making within-unit centering of any variables in the model unnecessary.

Finally, some limitations with the current study introduce future avenues for research. First, the simulation setup is a large N, small T framework, which is commonly seen in organizational research. There is, however, an ever growing emphasis on event sampling in which researchers often have access to much larger T. Future research may explore how the methods proposed here perform under these large T scenarios. Second, although we only tested one type of SEM-based approach in this study, there are other recent developments that hold promise in modeling dynamics and dealing with the problems discussed previously. For example, simulation evidence has shown that the dynamic structural equation model is free from the Nickell bias and performs relatively well when modeling dynamics (Asparouhov, Hamaker, & Muthén, 2018). Finally, in this study, we focus on the bias in the fixed effect of the slope and assume that unobserved heterogeneity exists in the intercept but not the slope. That is, the slopes for the lagged response variable and the predictors are constant across units, which is not always true in organizational research. Future research may explore how the models proposed here perform when there is also variation in the slopes. Nevertheless, our study took a major step in clarifying different sources of bias when modeling change and dynamics and provided alternative approaches to exploring these patterns.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Richard P. DeShon https://orcid.org/0000-0002-3906-6766
Christopher R. Dishop https://orcid.org/0000-0003-4174-9790

Note

1. We use the term *hierarchical linear model* to refer to a class of models also known as random coefficient models rather than software used to estimate these models.

References

Ahn, S. C., & Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, *68*(1), 5-27. doi:10.1016/0304-4076(94)01641-C

Anderson, T., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, *18*(1), 47-82. doi:10.1016/0304-4076(82)90095-

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277-297. doi:10.2307/229796810.2307/2297968

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359-388.

Barnes, C. M., Schaubroeck, J., Huth, M., & Ghumman, S. (2011). Lack of sleep and unethical conduct. *Organizational Behavior and Human Decision Processes*, *115*(2), 169-180.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*(1), 115-143. doi:10.1016/S0304-4076(98)00009-8

Bollen, K. A., & Brand, J. E. (2010). A general panel model with random and fixed effects: A structural equations approach. *Social Forces*, *89*(1), 1-34.

Bono, J. E., Glomb, T. M., Shen, W., Kim, E., & Koch, A. J. (2013). Building positive resources: Effects of positive events and positive reflection on work stress and health. *Academy of Management Journal*, *56*(6), 1601-1627. doi:10.5465/amj.2011.0272

Bryk, A., & Raudenbush, S. W. (1992). *Hierarcal linear models for social and behavioral research: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Call, M. L., Nyberg, A. J., Ployhart, R. E., & Weekley, J. (2015). The dynamic nature of collective turnover and unit performance: The impact of time, quality, and replacements. *Academy of Management Journal*, *58*(4), 1208-1232.

Cao, X., Lemmon, M., Pan, X., Qian, M., & Tian, G. (2018). Political promotion, CEO incentives, and the relationship between pay and performance. *Management Science*. Retrieved from https://pubsonline.informs.org/doi/10.1287/mnsc.2017.2966

Chi, N.-W., Chang, H.-T., & Huang, H.-L. (2015). Can personality traits and daily positive mood buffer the harmful effects of daily negative mood on task performance and service sabotage? A self-control perspective. *Organizational Behavior and Human Decision Processes*, *131*, 1-15. doi:10.1016/j.obhdp.2015.07.005

Clark, T. S., & Linzer, D. A. (2015). Should i use fixed or random effects? *Political Science Research and Methods*, *3*(02), 399-408. doi:10.1017/psrm.2014.32

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin*, *74*, 68-80.

DeShon, R. P. (2012). Multivariate dynamics in organizational science. *The Oxford Handbook of Organizational Psychology*, *1*, 117-142.

D'Innocenzo, L., Luciano, M. M., Mathieu, J. E., Maynard, M. T., & Chen, G. (2016). Empowered to perform: A multilevel investigation of the influence of empowerment on performance in hospital units. *Academy of Management Journal*, *59*(4), 1290-1307.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121-138.

Gabriel, A. S., Diefendorff, J. M., Chandler, M. M., Moran, C. M., & Greguras, G. J. (2014). The dynamic relationships of work affect and job satisfaction with perceptions of fit. *Personnel Psychology*, *67*(2), 389-420.

Hurwicz, L. (1950). Least squares bias in time series. In T. C. Koopmans (Ed.), *Statistical inference in dynamic economic models*. New York, NY: Wiley.

Johnson, R. E., Lanaj, K., & Barnes, C. M. (2014). The good and bad of being fair: Effects of procedural and interpersonal justice behaviors on regulatory resources. *Journal of Applied Psychology*, *99*(4), 635-650.

Judge, T. A., Simon, L. S., Hurst, C., & Kelley, K. (2014). What I experienced yesterday is who I am today: Relationship of work motivations and behaviors to within-individual variation in the five-factor model of personality. *Journal of Applied Psychology*, *99*(2), 199-221.

Judson, R. A., & Owen, A. L. (1999). Estimating dynamic panel data models: A guide for macroeconomists. *Economics Letters*, *65*(1), 9-15. doi:10.1016/S0165-1765(99)00130-5

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics*, *68*(1), 53-78. doi:10.1016/0304-4076(94)01643-E1

Kiviet, J. F., & Phillips, G. D. (2014). Improved variance estimation of maximum likelihood estimators in stable first-order dynamic regression models. *Computational Statistics & Data Analysis*, *76*, 424-448. doi:10.1016/j.csda.2013.09.021

Lanaj, K., Johnson, R. E., & Lee, S. M. (2016). Benefits of transformational behaviors for leaders: A daily investigation of leader behaviors and need fulfillment. *Journal of Applied Psychology*, *101*(2), 237-251.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, *95*(2), 391-413. doi:10.1016/S0304-4076(99)00044-5

Li, J., Burch, T. C., & Lee, T. W. (2017). Intra-individual variability in job complexity over time: Examining the effect of job complexity trajectory on employee job strain. *Journal of Organizational Behavior*, *38*(5), 671-691. doi:10.1002/job.215810.1002/job.2158

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *68*, 304-305.

Marriott, F. H. C., & Pope, J. A. (1954). Bias in the estimation of autocorrelations. *Biometrika*, *41*, 390-402.

Mitchell, T. R., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review*, *26*(4), 530-547.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*(6), 1417-1426.

Pitariu, A. H., & Ployhart, R. E. (2010). Explaining change: Theorizing and testing dynamic mediated longitudinal relationships. *Journal of Management*, *36*(2), 405-429.

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, *36*(1), 94-120. doi:10.1177/0149206309352110

Rogosa, D. (1988). Myths about longitudinal research. In W. M. K. W. Schaie, R. T. Campbell, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171-209). New York, NY: Springer.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, *71*(1), 135-158. doi:10.1111/j.1468-0084.2008.00542.x

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Wainer, H. (1991). Adjusting for differential base-rates: Lord's paradox again. *Psychological Bulletin*, *109*, 147-151.

Wansbeek, T., & Bekker, P. (1996). On IV, GMM and ML in a dynamic panel data model. *Economics Letters*, *51*(2), 145-152. doi: 10.1016/0165-1765(95)00785-710

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, MA: MIT Press.

Zaheer, S., Albert, S., & Zaheer, A. (1999). Time scales and organizational theory. *Academy of Management Review*, *24*(4), 725-741.

Zhu, J., Wanberg, C. R., Harrison, D. A., & Diehn, E. W. (2016). Ups and downs of the expatriate experience? Understanding work adjustment trajectories and career outcomes. *Journal of Applied Psychology*, *101*(4): 549-568. doi10.1037/apl0000073

Ziliak, J. P. (1997). Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators. *Journal of Business and Economic Statistics*, *15*(4), 419-431.

Author Biographies

Ran Xu is a postdoctoral research associate in the Grado Department of Industrial and Systems Engineering at Virginia Tech. He will be an assistant professor in the Department of Allied Health Sciences at the University of Connecticut starting from Fall 2019. His research expertise includes advanced quantitative methods, computational social science, and social network analysis, with applications to education, science policy and health sciences. His research has appeared in journals such as *Social Networks, Journal of Policy Analysis and Management,* and *System Dynamics Review*.

Richard P. DeShon is a professor of Industrial/Organizational Psychology at Michigan State University. His research addresses individual and collective motivation, system dynamics, and psychological measurement.

Christopher R. Dishop is a PhD student in Organizational Psychology at Michigan State University. His research interests include motivation, cooperation, and dynamics.