



Published in final edited form as:

Annu Rev Psychol. 2011 ; 62: 583–619. doi:10.1146/annurev.psych.093008.100356.

The Disaggregation of Within-Person and Between-Person Effects in Longitudinal Models of Change

Patrick J. Curran and Daniel J. Bauer

Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599; curran@unc.edu

Abstract

Longitudinal models are becoming increasingly prevalent in the behavioral sciences, with key advantages including increased power, more comprehensive measurement, and establishment of temporal precedence. One particularly salient strength offered by longitudinal data is the ability to disaggregate between-person and within-person effects in the regression of an outcome on a time-varying covariate. However, the ability to disaggregate these effects has not been fully capitalized upon in many social science research applications. Two likely reasons for this omission are the general lack of discussion of disaggregating effects in the substantive literature and the need to overcome several remaining analytic challenges that limit existing quantitative methods used to isolate these effects in practice. This review explores both substantive and quantitative issues related to the disaggregation of effects over time, with a particular emphasis placed on the multilevel model. Existing analytic methods are reviewed, a general approach to the problem is proposed, and both the existing and proposed methods are demonstrated using several artificial data sets. Potential limitations and directions for future research are discussed, and recommendations for the disaggregation of effects in practice are offered.

Keywords

multilevel modeling; growth modeling; trajectory analysis; within-person effects

INTRODUCTION

Many central theories in psychology and allied fields either implicitly or explicitly focus on within-person processes. For example, when an individual engages in effective coping, this is thought to mitigate the effects of stress for this individual (e.g., Roth & Cohen 1986). Similarly, when a person experiences negative affect, this person is expected to be more likely to engage in alcohol or substance use (e.g., Kassel et al. 2010). Finally, when an individual exercises more, it is expected that his or her positive affect will subsequently increase (e.g., Penedo & Dahn 2005). These three examples all highlight that the underlying theory posits what will happen *within* a given individual (that is, with respect to intraindividual processes), but not across a *set* of individuals (that is, with respect to interindividual processes).

Copyright © 2010 by Annual Reviews. All rights reserved

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

Despite the fact that the majority of psychological theories posit within-person processes, the research conducted to empirically evaluate these theories often involves the collection and analysis of strictly between-person data. Such between-person data almost always take the form of cross-sectional (or single time point) assessments of behavior. However, as has long been known, such data are poorly suited for evaluating within-person processes (Molenaar 2004, Schaie 1965). For example, if at a single point in time one person reports being both depressed and alcohol dependent and another person reports being neither depressed nor alcohol dependent, this does not imply that either person will drink more alcohol when experiencing negative affect. Thus, theory explicitly posits an effect at one level of analysis, yet standard cross-sectional designs and associated statistical models test an effect at a different level of analysis (e.g., Curran & Willoughby 2003).

Fortunately, there is growing recognition in our field that greater emphasis must be placed on the study of within-person processes and that this can only be accomplished through the study of intraindividual differences in repeated measures data (Collins 2006; Molenaar 2004; Molenaar & Newell 2010; Nesselroade 1991a,b; Raudenbush 2001a,b). Long- and short-term longitudinal studies are therefore becoming increasingly prevalent, including both traditional designs (e.g., Goldstein 1981) as well as newer experience sampling and ecological momentary assessment approaches (e.g., Walls & Schafer 2006). Despite this encouraging trend, the importance of focusing on within-person processes is still not universally appreciated. Interestingly, it is common to see the articulated strengths of longitudinal data designs to include factors such as the establishment of temporal precedence, the reduction of alternative potential models, and increases in statistical power (e.g., Muthén & Curran 1997). However, it is much less common to see an emphasis placed on the fact that only longitudinal data allow for the proper separation of between-person and within-person effects and that this is critically needed for fully evaluating many theories in psychology.

We are thus faced with a curious juxtaposition of recent developments. On the one hand, it is comforting to see that a clear emphasis has been placed on the importance of collecting and analyzing longitudinal data; yet on the other hand, it does not appear that a similar emphasis has been placed on the testing of within- and between-person influences on behavior once such data are obtained. The net result is that, although empirical data are increasingly available that will allow for the direct disaggregation of within-person and between person effects, this important opportunity is not often fully capitalized upon, if capitalized upon at all.

There is certainly a variety of reasons why many researchers do not take full advantage of the data that are available to them, including the potentially high cost of conducting long-term studies and the possible introduction of selective attrition over time. However, one likely factor on which we focus here is the relative lack of attention that has been paid to these rather complex issues in both the substantive and quantitative disciplines of psychology. From a substantive perspective, it is sometimes difficult to fully articulate precisely in what ways a given influence on an outcome might vary in magnitude and form when looking within persons versus across persons. For example, one might be interested in studying the relation between anxiety and substance use (e.g., Kaplow et al. 2001). It can be quite challenging to unambiguously articulate the theoretically derived expected relations between variability in overall level of anxiety and substance use across individuals (the between-person effect) and a specific individual's variation in anxiety and variation in substance use (the within-person effect). This is even further exacerbated by the fact that these two levels of influence can operate simultaneously and even in opposite directions. We are quite sympathetic to this challenge, having wrestled with these same issues in our own substantive research.

From a quantitative perspective, undoubtedly much thoughtful and quality work has focused on these issues over the past several decades; indeed, this literature is too extensive to fully summarize here. However, there are two potential limitations of this existing work. First, many quantitative and statistically oriented resources are found in books and journals that are not typically read by substantively oriented psychologists, and (let's be honest here) they are not always written in a way that is widely accessible to nonmethodologists. There is thus a potential problem of ineffective dissemination. Second, and more importantly, we argue that much new work is needed to overcome several unresolved issues that commonly arise in applied research settings but have not yet been closely considered from a quantitative perspective. There is thus a clear limitation in the general applicability of current analytic methods relative to the types of data that are often collected in the behavioral sciences. Taken together, although repeated measures data are becoming increasingly common in the psychological sciences, much more emphasis is needed on methods for capitalizing on these data to better test our underlying theories and hypotheses.

The purpose of our review is to thoroughly explore both the conceptual and statistical issues related to the disaggregation of between-person and within-person influences in longitudinal data. We begin with a brief conceptual discussion of exactly why evaluating within-person processes is critical in many areas of the behavioral sciences. We describe the long-known issue of disaggregating within- and between-group processes, and we describe how these same issues apply to the individual. We then move to a more analytically oriented perspective and introduce the multilevel growth model. We define the model and review standard methods that are recommended for disaggregating between- and within-person effects in practice. We then propose a more general definition of these two types of effects to better understand when standard methods can and cannot be applied, and we describe new methods of disaggregation to augment existing techniques. We move to three empirical demonstrations based on simulated data, and we demonstrate the potential utility of our new methods of disaggregation. We conclude with a discussion of unresolved issues and recommendations for the use of these methods in practice.

THE DISAGGREGATION OF LEVELS OF EFFECTS

It is well known that when a set of measures is collected at a single point in time from multiple individuals, the resulting data provide information only about between-person relationships (e.g., Molenaar 2004, Raudenbush 2001b, Raudenbush & Bryk 2002, Singer & Willett 2003). The statistical models fitted to such data are necessarily limited to between-person inferences, and thus estimation and interpretation can proceed in a rather straightforward manner (albeit in a manner that often does not test our theories in the way we desire).

In contrast, when a set of measures is collected at multiple points in time from multiple individuals, the resulting data simultaneously contain information about both between-person and within-person differences (e.g., Raudenbush & Bryk 2002, p. 183). Such data provide the opportunity to identify relationships that hold within persons as well as relationships that hold across persons. Both types of relationships can have important implications for theory. However, the statistical models fitted to these data must be carefully specified to avoid confounding the two sources of variability. Further, the substantive interpretation of results can be more challenging given the need to simultaneously consider effects operating at two levels of analysis. To think further about these issues, it is helpful to consider a specific case.

An example from the medical literature nicely illustrates the need to disaggregate levels of effect. Empirical evidence has shown that an individual is more likely to experience a heart

attack while exercising (i.e., the within-person effect), but at the same time people who exercise more tend to have a lower risk of heart attack (i.e., the between-person effect) (e.g., Curfman 1993, Mittleman et al. 1993). Both the within-person and between-person findings are valid, and each has direct public health relevance. However, generalizing the between-person effect to the individual would be an error of inference (e.g., the more you exercise the more likely you are to suffer a heart attack). Further, examining only one level of this more complex two-level effect would necessarily limit the development of complete understanding of the true nature of these relations. The issues explicated in this example generalize directly to many (if not nearly all) areas of psychological research. As such, the psychological sciences can derive many benefits from the application of statistical models that generate separate and unambiguous estimates of within- and between-person effects. Yet such models are not as prevalent in the psychological sciences as they are in other related disciplines.

When considering how to disaggregate within- and between-person effects, we can begin by examining the much longer history of methodological developments for separating effects at different levels of analysis more generally. Interestingly, the problem of separating within- and between-person effects mirrors the problem of separating within- and between-group effects that has long been a focus of concern in sociology and education (e.g., Cronbach & Webb 1975, Duncan & Davis 1953, Firebaugh 1978, Mason et al. 1983, Raudenbush & Willms 1995, Robinson 1950). Because these fields are often concerned with macrolevel influences on individuals, such as teacher, school, or community effects, data are often collected in which multiple individuals are nested within each of many groups (Raudenbush & Bryk 2002, Raudenbush & Sampson 1999). Classic examples of nested/hierarchical data include children within classrooms, individuals within neighborhoods, spouses within marriages, and patients within therapists.

In these contexts, many substantive theories posit effects at both the individual and group levels. For example, positive behavior gains associated with a particular psychotherapeutic intervention may be influenced by characteristics of the individual patient (e.g., gender, ethnicity, baseline symptomatology), characteristics of the group within which the therapy was delivered (e.g., therapist experience, group size, group gender composition), or the interaction of characteristics of the patient with characteristics of the group. Thus, for many years, a distinction has been made in the study of hierarchically structured data between the examination of individual effects and contextual (or sometimes ecological) effects (e.g., Raudenbush & Sampson 1999).

Failing to recognize the important distinction between these effects can result in consequential errors of inference. In some cases, results obtained from individual data have been used to make inferences to the group level; more commonly, results obtained from group-level data are misattributed to individuals. This latter condition is known as the ecological fallacy and was first described more than half of a century ago by Robinson (1950).¹ Simply put, the ecological fallacy occurs when a researcher mistakenly believes that the observed relation between two variables at the aggregate level (that is, at the level of groups) also applies at the individual level (Firebaugh 1978, Robinson 1950, Schwartz 1994). Of course, the between-group and within-group relations may ultimately be the same, but the relation at one level is neither necessary nor sufficient to imply the same relation at another level.

¹Although Robinson (1950) is commonly credited with coining the term ecological fallacy, Schwartz (1994) notes that this term was not first used until several years later by Selvin (1958).

A classic example of the ecological fallacy is reflected in results published by Durkheim (1897) that suggested European countries with a higher proportion of Protestants were characterized by higher rates of suicide. One explanation offered to account for this observed relation was that people living under the harsh dictates of Protestantism were more likely to end their own lives. However, this is a classic case of the ecological fallacy. Specifically, there is no evidence that Protestant individuals are more likely to commit suicide than are non-Protestants within a given country. Further, there is equally no evidence to suggest that the proportion of Protestants plays any explanatory role at all; this may simply be a third-variable correlate that accounts for some other effect that was not included in the model.

Another more contemporary example comes from a study of psychostimulant prescription rates for black and white children diagnosed with attention-deficit/hyperactivity disorder (ADHD; Foster 2010). Although psychostimulants are a recommended treatment for ADHD, prescription rates at public agencies are lower for black than white children, reflecting broader racial disparities in health care. This difference, however, is more a consequence of between-agency differences than within-agency differences. For instance, a black child is more likely to be prescribed psychostimulants if he or she attends a clinic that services predominantly white children. Separating these two levels of effect is critical for better understanding the reasons behind racial disparities: Between-clinic differences likely reflect sources of institutional racism, such as residential segregation, whereas within-clinic differences may predominantly reflect the implicit prejudices of care providers.

A final example that is less relevant to the psychological sciences yet clearly highlights the issues at hand relates to the relation between body mass and life expectancy in mammals. Two facts have been well established (Millar & Zammuto 1983). First, on average, species that are characterized by larger body mass tend to have longer life expectancies than species with smaller body mass. So whales tend to live longer than cows who tend to live longer than ducks. However, on average, individual members within a species who are characterized by larger body mass tend to have shorter life expectancies relative to members of their own species. So fat ducks tend to have shorter life expectancies than skinny ducks. It would thus be an error to make an inference from the aggregate level (that larger species-specific body mass is associated with longer life expectancy) to the individual level (where the opposite effect actually holds). This is the heart of the ecological fallacy. Importantly, the ecological fallacy only applies when an aggregate relation is misattributed to the level of the individual. That is, the finding that species with larger body mass have longer life expectancies is unambiguously accurate at the level of the species. An error is made only when the group-level effect is applied to the individuals within the groups.²

In sum, more than half a century of both quantitative and substantive research has focused on the disaggregation of between- and within-group processes, and these methods have been used to great advantage for decades. Further, it has been long assumed that these same methods can also be used to distinguish within- and between-person effects given that the two data structures are quite similar (e.g., Enders & Tofighi 2007). In hierarchical data, individuals are nested within groups; in longitudinal data, repeated measures are nested within person. The extension of methods from one structure to the other is quite logical. However, as we demonstrate below, several key issues often arise with repeated measures data that, although less relevant in hierarchically structured data, can substantially

²The original work of Robinson (1950) only discussed the inappropriate inference of individual processes based on aggregate relations. In some social science disciplines it has been argued that there was an unnecessary “overcorrection” in moving away from aggregate studies to overcome these concerns and that certain fields need to move back to considering both individual and group-level effects (e.g., Pearce 2000).

complicate (if not wholly invalidate) the disaggregation of between- and within-person effects using existing methods.

Now we turn to a more detailed description of current analytic methods available for disaggregating levels of effect in longitudinal data. Although a variety of well-developed methods exist for analyzing such data structures, the multilevel model is extremely well suited for this endeavor, and hence it is our sole focus here.

THE MULTILEVEL GROWTH MODEL

We begin with a formal definition of the multilevel growth model. We briefly summarize this approach here, but see Bryk & Raudenbush (1987), Raudenbush (2001b), Raudenbush & Bryk (2002), and Singer & Willett (2003) for excellent in-depth overviews of these methods. Equations are necessary for formalizing these ideas, but we augment these with verbal descriptions and visual graphics whenever possible.

First, let us denote the repeated measure observed at time point t for individual i as y_{it} . The repeated measure might represent any psychologically relevant outcome such as substance use, self-esteem, depression, or academic achievement. In a linear growth model, the observed repeated measure is expressed as a simple linear function of time, given as

$$y_{it} = \beta_{0i} + \beta_{1i}x_{it} + r_{it}, \quad (1)$$

where β_{0i} and β_{1i} represent the intercept and linear slope for individual i , x_{it} is the observed value of time³ at assessment t for individual i , and r_{it} is the time- and individual-specific residual. This represents the within-person trajectory and is sometimes called the level-1 equation. Note that more complex within-person equations can be specified, for instance to allow for nonlinear patterns of change over time (e.g., a curved trajectory), but we retain the linear form here to simplify our exposition.

An important element of the growth model is that the values of the intercept and slope components vary randomly across persons. That is, some individuals might have larger versus smaller intercepts (or initial levels), and some individuals might change more rapidly versus less rapidly over time. This variability can be expressed as

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i} \end{aligned} \quad (2)$$

where γ_{00} and γ_{10} are the overall mean intercept and slope, and u_{0i} and u_{1i} are the individual-specific deviations from these means, respectively. This captures between-person (or interindividual) differences in within-person (or intraindividual) change and is sometimes called the level-2 equation.

The level-1 and level-2 equations are primarily of pedagogical value to allow for the within-person and between-person equations to be made explicit. However, the formal statistical model results from the substitution of Equation 2 into Equation 1 that in turn defines the reduced form expression:

³For ease of presentation we treat *time* and *age* as isomorphic. Many interesting challenges and opportunities arise when time of assessment and chronological age differ (e.g., Mehta & West 2000). However, treating these equivalently here in no way limits the generalizability of our findings.

$$y_{it} = (\gamma_{00} + \gamma_{10}x_{it}) + (u_{0i} + u_{1i}x_{it} + r_{it}), \quad (3)$$

The terms within the first set of parentheses are referred to as the fixed effects of the model, whereas the terms in the second set of parentheses are the random effects. The parameters that define the multilevel growth model described in Equations 1 and 2 are $E(\beta_{0i}) = \gamma_{00}$, $E(\beta_{1i}) = \gamma_{10}$, $\text{var}(u_{0i}) = \tau_{00}$, $\text{var}(u_{1i}) = \tau_{11}$, and $\text{var}(r_{it}) = \sigma_r^2$. The covariance between random effects is also commonly estimated as part of this model (e.g., $\text{cov}[u_{0i}, u_{1i}] = \tau_{10}$). Finally, although there are a number of alternative possible covariance structures for r_{it} , here we assume the residuals are independent and homoscedastic over time.

This model can be expanded to include one or more time-invariant covariates (TICs). Because TICs vary only across persons (e.g., gender, ethnicity, diagnostic status) and not within persons (i.e., take on different values for each person over time), their effects are strictly between-person. TICs thus enter into the level-2, or between-person, equations. For instance, denoting a single TIC as w_i , we can expand Equation 2 so that

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01}w_i + u_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11}w_i + u_{1i} \end{aligned} \quad (4)$$

where γ_{01} and γ_{11} represent the fixed effect regression of the random intercept and slope components on the TIC, respectively. These regression parameters reflect the expected change in the intercept and slope of the trajectory relative to a one-unit change in the TIC. It is clear that the predictor w_i is time invariant because the subscript is unique to individual i but is equal across all time points t .

Alternatively, one or more time-varying covariates (TVCs) can be incorporated into the level-1 equation that vary over both individual and time point. We denote the TVC as z_{it} , indicating that a unique value may be obtained at any time point t for any individual i . It is easy to see that TVCs simultaneously contain both within-person and between-person variability. For example, a simple expression for the TVC is given as

$$z_{it} = \bar{z}_i - r_{it}, \quad (5)$$

where \bar{z}_i is the person-specific mean of the TVC pooling over time, and r_{it} is the time-specific deviation of the TVC from the person-specific mean. It is thus clear that considering z_{it} in isolation embodies an aggregation of both within-person and between-person variability. As such, we must carefully consider the disaggregation of these two components.

For simplicity, let us consider how a TVC enters into a model that includes a random intercept but no random time slope (that is, we do not include x_{it} as a level-1 predictor of y_{it}). For example, we might want to use a diary study to examine how day-to-day fluctuations in anxiety (the TVC) predict daily levels of substance use (e.g., Hussong et al. 2001). Substance use might not be expected to change systematically with the passage of time when assessed on a daily basis, so only a random intercept is needed to capture individual variability in substance use over time.

The level-1 equation for this model is given as

$$y_{it} = \beta_{0i} + \beta_{1i}z_{it} + r_{it}, \quad (6)$$

where z_{ti} represents a measure on the covariate z at time t for individual i , and all else is defined as above. Although the influence of the TVC (i.e., β_{1i}) can itself be defined as random (Raudenbush & Bryk 2002, equation 6.21), for simplicity we assume this is a fixed effect. The corresponding level-2 equations are thus

$$\begin{aligned}\beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10}\end{aligned}\quad (7)$$

with reduced form

$$y_{it} = (\gamma_{00} + \gamma_{10}z_{it}) + (u_{0i} + r_{it}). \quad (8)$$

Conceptually, this is expressed in precisely the same way as an ordinary least squares regression would be, but with an additional residual term (i.e., u_{0i}) to account for the fact that there are unexplained differences among individuals in the average values of y_{it} . These unexplained differences arise from the collection of repeated observations taken on each individual.

As is well known in the quantitative literature (but less so in the substantive literature), the effect of the TVC on the outcome (i.e., γ_{10}) represents an aggregation of between-person and within-person influences of the TVC on the outcome (e.g., Raudenbush & Bryk 2002, equation 5.38). The reason is that z_{ti} varies both between individuals (in average level) and within individuals (across time). In some respects, these two types of differences mirror the classic distinction between traits and states (Nezlek 2007). Because z_{ti} is a combination of both sources of variability, when we estimate just one effect for z_{ti} , the result is an inextricable combination of potentially different effects operating at the two levels of analysis. To differentiate these effects, we must decompose z_{ti} into components that isolate between- and within-person differences, respectively. Fortunately, assuming certain conditions hold in the population, there are well-established methods for achieving this disaggregation of effects within the multilevel model. It is to this topic that we next turn.

TRADITIONAL METHODS FOR DISAGGREGATING BETWEEN- AND WITHIN-PERSON EFFECTS

It is well known that between- and within-person effects can be efficiently and unambiguously disaggregated within the multilevel model using the strategy of person-mean centering. Traditionally, the term centering is used to describe the rescaling of a random variable by deviating the observed values around the variable mean (e.g., Aiken & West 1991, pp. 28–48). For example, within the standard fixed-effects regression model, a predictor x_i is centered via $x'_i = x_i - \bar{x}$, where \bar{x} is the observed mean of x_i , and x'_i is the mean-deviated rescaling of x_i (see, e.g., Cohen et al. 2003, p. 261). By definition, the mean of a centered variable is equal to zero, and this offers both interpretational and sometimes computational advantages in a number of modeling applications.

However, centering becomes more complex when considering TVCs. This is because multiple repeated measures are nested within each individual, and there are thus two means to consider: the grand mean of the TVC pooling over all time points and all individuals, and each person-specific mean pooling over all time points within individual. There are two ways that we can center the TVC.

First, we can deviate the TVC around the grand mean pooling over all individuals. Here,

$$\dot{z}_{ti} = z_{ti} - \bar{z}_{..}, \quad (9)$$

where \dot{z}_{ti} represents the grand mean centered TVC, z_{ti} is the observed TVC, and $\bar{z}_{..}$ is the grand mean of z_{ti} pooling over all individuals and all time points. In other words, we simply compute the grand mean of the TVC and subtract this from each individual- and time-specific TVC score. Second, we can deviate the TVC around the person-specific mean of the TVC unique to each individual. Here,

$$\dot{z}_{ti} = z_{ti} - \bar{z}_i, \quad (10)$$

where \dot{z}_{ti} represents the person-mean centered TVC, z_{ti} is again the observed TVC, and \bar{z}_i is the person-specific mean for individual i . In other words, we subtract just the person-specific mean of the TVC from each of that same person's time-specific TVC scores. We can use z_{ti} , \dot{z}_{ti} , or \dot{z}_{ti} as the level-1 predictor in Equation 8, and each is associated with a potentially different inference with respect to the disaggregation of effects.

Methods exist that allow for the disaggregation of the between-person and within-person effects using z_{ti} , \dot{z}_{ti} , or \dot{z}_{ti} (Kreft et al. 1995, Raudenbush & Bryk 2002). However, direct estimates of these effects can be most easily obtained within the multilevel model by incorporating the person-mean centered TVC at level-1 (i.e., \dot{z}_{ti}) and the person-mean at level-2 (i.e., \bar{z}_i) (Raudenbush & Bryk 2002, equation 5.41). Specifically,

$$\begin{aligned} y_{ti} &= \beta_{0i} + \beta_{1i} \dot{z}_{ti} + r_{ti} \\ \beta_{0i} &= \gamma_{00} + \gamma_{01} \bar{z}_i + u_{0i}, \\ \beta_{1i} &= \gamma_{10} \end{aligned} \quad (11)$$

where all is defined as above. This requires three steps: We first compute the mean of the time-specific TVCs within each individual to obtain \bar{z}_i ; we then subtract that person-specific mean from each individual's time-specific TVC values to obtain \dot{z}_{ti} ; finally, we use both \bar{z}_i and \dot{z}_{ti} as predictors in our multilevel model.

The reduced form equation for this model is

$$y_{ti} = (\gamma_{00} + \gamma_{01} \bar{z}_i + \gamma_{10} \dot{z}_{ti}) + (u_{0i} + r_{ti}), \quad (12)$$

where γ_{00} is the intercept (or grand mean), γ_{01} is a direct estimate of the between-person effect, and γ_{10} is a direct estimate of the within-person effect. Following our earlier hypothetical example, γ_{01} would capture the relation between average levels of anxiety and average levels of substance use pooling over individuals. In contrast, γ_{10} would capture the mean relation between a given person's time-specific deviation in anxiety (relative to the overall level of anxiety) and the individual's time-specific substance use.

The approach we outline above is currently regarded as best practice for the disaggregation of between-person and within-person effects in multilevel growth models (e.g., Raudenbush & Bryk 2002, pp. 181-85; Singer & Willett 2003, pp. 173-77), and there is no question that this is a valid method for accomplishing these goals. As we describe in greater detail below, however, the validity of this approach heavily relies on a set of specific conditions that may or may not be met in practice. Further, we have found that these conditions are rarely, if ever, discussed in either the quantitative or applied literatures. To better define these specific

conditions, we next propose a more general framework for defining within-person and between-person effects. This framework both more formally establishes these expressions and allows us to explicate precisely under what conditions standard approaches are and are not valid.

A GENERAL DEFINITION OF WITHIN-PERSON AND BETWEEN-PERSON EFFECTS

The existing methods used to disaggregate within- and between-person effects implicitly assume that within- and between-person variability can be unambiguously and validly represented via \bar{z}_i and \dot{z}_{ti} (as we describe above). Indeed, the historical justification for using this approach has verged on tautology: You use \bar{z}_i and \dot{z}_{ti} to disaggregate between- and within-person effects because between- and within-person differences are disaggregated via \bar{z}_i and \dot{z}_{ti} . This method of disaggregation is indeed valid, albeit only under certain conditions. To better explicate these conditions, we propose a more general definition of between- and within-person components of the TVC. In an attempt to avoid the siren's song of tautology ourselves, we propose a new notation to reflect more broadly defined terms that do not rely on how the values are actually calculated. Once expressed in this way, we can then consider how these values are best estimated from empirical data.

First, we denote the between-person component of the TVC as zb_i and the within-person component as zw_{ti} . The z reflects that we are referencing the TVC z_{ti} ; the b and w denote between and within components of z , respectively; and the subscripts denote that the between component is unique to individual i and the within component is unique to time t for individual i . Our intent is that these more general expressions define the relevant components of the TVC in terms other than how these values are computed in sample data.

Once expressed in this way, the between- and within-person effects of the TVC on the outcome can be expressed via the model

$$y_{it} = (\gamma_{00} + \gamma_{01}zb_i + \gamma_{10}zw_{ti}) + (u_{0i} + r_{it}), \quad (13)$$

where γ_{01} represents the between-person effect and γ_{10} the within-person effect. Note this is a simple restatement of Equation 12, with the caveat that we no longer presume that \bar{z}_i and \dot{z}_{ti} are necessarily the best empirical representations of zb_i and zw_{ti} . As before, an important distinction to keep in mind here is that zb_i and zw_{ti} represent the between- and within-person components of the TVC itself, whereas γ_{01} and γ_{10} represent the between- and within-person components of the relationships between the TVC and the outcome. These different components are quite important to distinguish, and we return to this repeatedly throughout our review.

Now that we have a general notational scheme defining the disaggregation of TVC effects, we can more carefully consider the estimation of these effects under different population conditions. We consider three conditions here: when the TVC is unrelated to time, when the TVC is characterized by just a fixed effect of time, and when the TVC is characterized by both a fixed and random effect of time.

Disaggregation of Effects When the TVC is Unrelated to Time

A key aspect of our approach is to write an explicit model for the TVC itself. Given the historical presumption that \bar{z}_i and \dot{z}_{ti} are *prima facie* valid, there has not been a prior need to write a model for the TVC. However, such a model is necessary to better establish the

underlying conditions that are required to validly disaggregate the within-person and between-person levels of effect.

To do this, we begin by expressing variability in the TVC at the population level via a standard two-level model.⁴ The level-1 expression for the TVC is

$$z_{ti} = \beta_{0i} + r_{ti}, \quad (14)$$

where z_{ti} is the measure of the TVC at time t for individual i , β_{0i} is the person-specific mean of z_{ti} pooling over time, and r_{ti} is the time-specific deviation of the TVC from the mean of z_{ti} for individual i . Next, the level-2 expression is

$$\beta_{0i} = \gamma_{00} + u_{0i}, \quad (15)$$

where γ_{00} is the grand mean of the TVC pooling over both time and individual, and u_{0i} is the deviation of the person-specific mean from the grand mean. Finally, the reduced form is

$$z_{ti} = \gamma_{00} + u_{0i} + r_{ti}, \quad (16)$$

where all terms are defined as above.

Note that this is nothing more than a random intercept model written for the TVC instead of for the outcome as is usually done. The advantage of this expression is that we can clearly see that r_{ti} captures the within-person variability of the TVC around the person-specific mean (i.e., β_{0i}) and u_{0i} captures the between-person variability of the TVC around the grand mean (i.e., γ_{00}). Given this, we wish to define the within-person component of the TVC (i.e., $z_{w_{ti}}$) to solely reflect variability in r_{ti} and the between-person component of the TVC (i.e., z_{b_i}) to solely reflect variability in u_{0i} . We can now consider whether the traditional approach of setting $z_{b_i} = \bar{z}_i$ and $z_{w_{ti}} = \dot{z}_{ti}$ validly accomplishes this goal.

Let us first consider what \bar{z}_i represents in this case. Conceptually, we want to estimate the person-specific overall level of the TVC pooling over time. To do this, we can take the expected value of the reduced-form expression in Equation 16 for individual i . In other words, we want to compute the long run average of the TVC within each individual. This is given as

$$E_i(z_{ti}) = E_i(\gamma_{00} + u_{0i} + r_{ti}) = \gamma_{00} + u_{0i}, \quad (17)$$

where γ_{00} is the grand mean of the TVC and u_{0i} is the deviation of the person-specific mean from the grand mean. Importantly, because γ_{00} is constant across individuals, u_{0i} represents the individual-specific between-person component of the TVC. If we replace $E_i(z_{ti})$ in Equation 17 with the sample realization \bar{z}_i , we get

$$\bar{z}_i = \widehat{\gamma}_{00} + \widehat{u}_{0i}, \quad (18)$$

and with simple manipulation we get

⁴We chose to use the same notation in our model for z_{ti} as we did for y_{ti} to avoid unnecessary clutter in notation by specifying to which outcome each term belongs (e.g., $\gamma_{00}(y)$ vs $\gamma_{00}(z)$, etc.). As such, we assume the terms are implicitly differentiated with respect to the relevant outcome of interest.

$$\widehat{u}_{0i} = \bar{z}_i - \widehat{\gamma}_{00}, \quad (19)$$

which is our estimate of zb_i . Given that $\widehat{\gamma}_{00}$ is constant, the person-specific mean alone (without deviation by $\widehat{\gamma}_{00}$) provides a valid representation of the between-person component of the TVC unique to individual i when the model defined in Equation 16 holds in the population. Therefore, using \bar{z}_i as zb_i will produce a valid estimate of the between-person component of the TVC under these conditions.

Let us next consider the within-person component of the TVC. Conceptually, we want to isolate the within-person variability of the TVC around the person-specific level of the TVC pooling over time. Recall that above we noted that the level-1 residual term (i.e., r_{it}) captured with the within-person variability of the TVC around the person-specific mean. Given this, we can do a simple manipulation of Equation 16 to express the within-person residual as

$$r_{it} = z_{it} - (\gamma_{00} + u_{0i}), \quad (20)$$

highlighting that the within-person component of the TVC is indeed r_{it} . We saw in Equation 18 that the person-specific mean can be expressed as $\widehat{\gamma}_{00} + \widehat{u}_{0i}$, so we can in turn define the sample representation of \widehat{r}_{it} to be

$$\widehat{r}_{it} = z_{it} - (\widehat{\gamma}_{00} + \widehat{u}_{0i}) = z_{it} - \bar{z}_i. \quad (21)$$

Again assuming that Equation 16 holds in the population, the within-person component of the TVC can be computed by defining $zw_{it} = \dot{z}_{it}$, where \dot{z}_{it} is defined as above (e.g., $\dot{z}_{it} = z_{it} - \bar{z}_i$). Thus, we can obtain a valid estimate of the within-person effect using the traditional person-mean centering strategy.

A key component of these expressions is that we are defining the between- and within-person components of the TVC in terms of general expressions and then determining the appropriate sample realizations for these expressions. In this specific case, we find that computing these components as $zb_i = \bar{z}_i$ and $zw_{it} = \dot{z}_{it}$ meets the stated goals of the analysis. Expressing the model in this way helps us avoid the potential circularity that is sometimes present in prior discussions about methods for disaggregating TVC effects. More importantly, this allows us to generalize these expressions to conditions under which Equation 16 does not hold in the population.

More specifically, in the present case the population model for the TVC defined in Equation 16 is independent of the passage of time. In other words, although the TVC can take on a unique value at any given time point t , the conditional mean of the TVC is not systematically related to time; more succinctly, although there may be growth in the outcome (i.e., y_{it}), there is no growth in the TVC itself (i.e., z_{it}). However, in many longitudinal applications in the behavioral sciences, it could be quite likely (if not theoretically predicted) that the TVC is changing systematically with the passage of time. Such systematic growth might be less prevalent (if not wholly absent) in diary data or experience sampling designs in which observations are assessed daily, or even hourly. However, in designs in which assessments are made monthly or even yearly (e.g., Hussong et al. 2007, 2008), growth in the TVC might be fully expected and highly salient. Yet the standard methods used to disaggregate these effects implicitly assume the passage of time is

irrelevant with respect to the TVC. We must thus carefully consider what occurs when the TVC is indeed related to time.

Disaggregation of Effects When the TVC is Characterized by a Fixed-effect of Time

We begin by extending the model for the TVC presented in Equation 16 to include a main effect of time, but the magnitude of this effect is constant over individuals. Descriptively, this model implies that the conditional mean of the TVC is linearly changing with the passage of time, but that all individuals are changing at precisely the same rate. The level-1 model is thus

$$z_{ti} = \beta_{0i} + \beta_{1i}x_{ti} + r_{ti}, \quad (22)$$

where x_{ti} is the measure of time at time t for individual i , β_{1i} is the linear relation between time and the TVC for individual i , and all else is defined as above. The level-2 equations are

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} \end{aligned}, \quad (23)$$

where γ_{00} and γ_{10} represent the mean intercept and rate of change, respectively, and u_{0i} is the deviation of the intercept for individual i from the overall mean. Note that there is no corresponding u term for β_{1i} , indicating that the magnitude of the relation between time and the TVC is constant over all i . That is, individual trajectories on z_{ti} appear as parallel lines, with differences in level but not slope. Finally, the reduced form is

$$z_{ti} = \gamma_{00} + \gamma_{10}x_{ti} + u_{0i} + r_{ti}, \quad (24)$$

where all is defined as above.

As before, we wish to construct representations of the between- and within-person components of z_{ti} (i.e., zb_i and zw_{ti}) that will isolate the between-person variability in u_{0i} and the within-person variability in r_{ti} . Let us begin by re-expressing the between-person and within-person components of z_{ti} under this expanded model. The person-specific expected value of z_{ti} as defined in Equation 24 is now

$$E_i(z_{ti}) = (\gamma_{00} + u_{0i}) + \gamma_{10}E_i(x_{ti}). \quad (25)$$

Note that the between-person variability on z_{ti} is now both a reflection of u_{0i} (i.e., the first term) and the expected value of time (i.e., the second term). Ideally, we would prefer that our measure of between-person variability not depend on the timing of assessments and instead reflect differences in level, or u_{0i} , alone. We must thus take the influence of time into account in obtaining our sample estimates of zb_i .

Rearranging Equation 25 and inserting sample estimates, we obtain

$$\widehat{u}_{0i} = (\bar{z}_i - \widehat{\gamma}_{00}) - \widehat{\gamma}_{10}\bar{x}_i, \quad (26)$$

where \bar{x}_i is the mean of time for person i (e.g., $\bar{x}_i = \sum x_{ti}/T_i$ where T_i is the total number of time points for person i). Remember that, drawing on our general definitions above, we want to construct a variable zb_i to represent this component of the TVC, yet the person-mean of

the TVC contains additional variation due to potential individual differences in the mean value of time (i.e., $\hat{\gamma}_{10}\bar{x}_i$). Indeed, in this case, the only instance in which \bar{z}_i will provide an adequate measure for z_{bi} is when \bar{x}_i is constant across all individuals in the sample (i.e., the data are time structured); if \bar{x}_i varies across individuals, then setting $z_{bi} = \bar{z}_i$ will not isolate the between-person component of the TVC in the way that we desire.

Continuing on to the expression of the within-person component, after a bit of simple algebra we can represent the person- and time-specific residual defined in Equation 24 as

$$\hat{r}_{ii} = (z_{ii} - \bar{z}_i) - \hat{\gamma}_{10}(x_{ii} - \bar{x}_i), \quad (27)$$

where all remains defined as before. Note that the first term contains the person-mean centered TVC (because $\hat{z}_{ii} = z_{ii} - \bar{z}_i$) and that in isolation this is an imperfect representation of the within-person component of the TVC because it fails to consider the time trend reflected in the second term. Similar to what we found for the between-person component of the TVC, setting $z_{w_{ii}} = \hat{z}_{ii}$ will not isolate the within-person component of the TVC in the way that we desire. When the TVC is related to the passage of time, additional adjustments are needed to obtain ideal estimates of both z_{bi} and $z_{w_{ii}}$.

Disaggregation of Effects When the TVC is Characterized by a Fixed- and Random-Effect Time

We considered the condition in which the TVC was related to the passage of time, but the rate of change was constant in magnitude for all individuals. However, in many applications this time effect might vary randomly over individuals; indeed, in many conditions this might be expected (e.g., Hussong et al. 2007, 2008). Continuing with our hypothetical example, we might expect that not only does anxiety systematically change as a function of time, but the rates of change vary randomly over individual; some people may be changing at a faster rate, others at a slower rate, and others may not be changing at all. We can expand our equations to take this additional source of variability into account.

The level-1 model remains precisely as before:

$$z_{ii} = \beta_{0i} + \beta_{1i}x_{ii} + r_{ii}, \quad (28)$$

but we now expand the level-2 model to allow for person-specific deviations in both the intercept and slope components of the time trends:

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i} \end{aligned}, \quad (29)$$

where all is as defined above, but now u_{1i} represents the deviation of the person-specific slope from the overall mean slope. Finally, the reduced form is

$$z_{ii} = (\gamma_{00} + \gamma_{10}x_{ii}) + (u_{0i} + u_{1i}x_{ii} + r_{ii}), \quad (30)$$

where the first parenthetical term represents the fixed effects and the second the random effects.

In this setting, two random effects determine the between-person differences at any given point in time: between-person variability in the intercept and between-person variability in

the slope. Interestingly, given the random slope component (i.e., u_{1i}), the rank order of individuals can (and usually will) differ from one occasion to the next. This can be visualized by picturing a set of individual trajectories, each of which is defined by its own intercept and slope. Because some are changing at faster rates than others, the rank ordering of individuals on the TVC at a given point in time depends upon the specific point chosen. At one point in time there will be one rank ordering, and at another point in time there will be a different rank ordering.

As we discuss below, the time-dependent nature of the rank ordering makes it more difficult to conceptualize precisely what zb_i ought to represent. This is because the between-person component of the TVC captures between-person variability, yet this same variability changes at each time point in the presence of random growth (e.g., Biesanz et al. 2004). One reasonable way to define between-person differences in this context is as the difference in average levels of z_{it} that we would expect to observe at the average value of time. If we assume time (x_{it}) is scored so that zero is placed in the center of the time axis, then this is precisely what u_{0i} represents, so we can again isolate this term to determine how best to compute zb_i . However, our between-person component of the TVC is taken at the mean of time, and this value can (and likely would) change at any other point in time.

The expression for the person-specific expected value of the TVC is slightly more complex than before, but not terribly so:

$$\begin{aligned} E_i(z_{it}) &= (\gamma_{00} + u_{0i}) + \gamma_{10} E_i(x_{it}) + E_i(u_{1i} x_{it}) \\ &= (\gamma_{00} + u_{0i}) + (\gamma_{10} + u_{1i}) E_i(x_{it}). \end{aligned} \quad (31)$$

This expression highlights the influence of both the person-specific intercept (via u_{0i}) and person-specific slope (via u_{1i}) weighted by the person-mean of time. Thus, between-person differences can be represented in the sample via

$$\widehat{u}_{0i} = (\bar{z}_i - \widehat{\gamma}_{00}) - (\widehat{\gamma}_{10} + \widehat{u}_{1i}) \bar{x}_i, \quad (32)$$

which contains information about both the fixed and random effects of time. Clearly, \bar{z}_i does not isolate variability in \widehat{u}_{0i} , and hence a different measure for zb_i must be constructed. Similar to the previous case, time-structured data present an exception. In this case, if the data are time structured and time has been centered around the mean (as we have assumed), then this implies that $\bar{x}_i = 0$ for all i . Thus Equation 32 will simplify to the set of terms in the first parentheses, and \bar{z}_i will be an adequate sample representation for u_{0i} . However, this scenario rarely occurs in practice.

Moving on to the within-person component of the TVC, we can also isolate the individual- and time-specific residual such that

$$\widehat{r}_{it} = (z_{it} - \bar{z}_i) - (\widehat{\gamma}_{10} + \widehat{u}_{1i})(x_{it} - \bar{x}_i). \quad (33)$$

This also highlights the salient role of both the fixed and random effects defining the relation between the TVC and time. The first term again contains the person-mean centered TVC, or \dot{z}_{it} , and this continues to be an insufficient measure of zw_{it} because it fails to consider the individually varying time trend reflected in the second term.

A key issue to which we have already alluded relates to whether time is balanced or unbalanced. Because traditional methods for disaggregating between- and within-person

levels of influence assume no systematic relation between time and the TVC (i.e., Equation 16), there has been no need to consider the impact of different ways in which time might enter the model. However, when the TVC is related to time (i.e., Equations 24 and 30), we must more carefully evaluate in precisely what ways time can enter into the model. Of key importance here is whether the repeated measures data are collected using a design that is balanced or unbalanced with respect to time.

The Structure of Time: Balanced Versus Unbalanced

A design is time structured, or balanced with respect to time, if all individuals are the same age when assessed at the same time periods over the same total span of time (e.g., Bollen & Curran 2006, p. 75). This is a highly restrictive condition that is more prevalent in controlled lab-based designs but is relatively rare in most observational studies conducted in the behavioral sciences. For example, behavioral aggression in lab mice might be measured starting precisely at 28 days of age and reassessed every seven days for two months. There are no missing data, and all mice are the same age at each assessment. Although not common, there are situations when such designs also appear in studies of humans. One example is a birth cohort design in which a sample of individuals is collected from a single birth cohort (e.g., all children born in January of a given year) and is then followed annually over time. However, even in this situation we must make the unrealistic assumption of no missing data over time. Of importance to our discussion here, data that are balanced on time offer several simplifying conditions relevant to the separation of the TVC effects.

We showed above that when the TVC is related to time, both the time-specific value of time (i.e., x_{ti}) and the person-specific mean of time (i.e., \bar{x}_i) play a role in the disaggregation of effects (i.e., Equations 26 and 27, and Equations 32 and 33). An interesting characteristic of designs that are balanced on time is that all individuals have the same value of \bar{x}_i . It is easy to see why. The person-specific mean of time is defined as

$$\bar{x}_i = \left(\sum_{t=1}^T x_{ti} \right) / T, \quad (34)$$

where $t = 1, 2, \dots, T$ represents the observation number for the person. For balanced designs, the values for x_{ti} are identical across cases for any given value of t . For instance, at the first time point, or $t = 1$, everyone in the sample might be 16 years old, and additional observations from that point forward might be made on the entire sample at one-year intervals. As such, the person-mean of time is equal for all individuals; more formally, $\bar{x}_i = \bar{x}$ for all i . As we demonstrate empirically, this characteristic makes the disaggregation and interpretation of the between-person component of the TVC rather straightforward.

In contrast, a design is considered unbalanced with respect to time if all individuals are not assessed at all of the same points in time (e.g., Mehta & West 2000). Given this, in a design that is unbalanced with respect to time, individual ages will vary at any given assessment point. That is, x_{ti} is no longer constant over i for a given t . For example, one subset of observations may have been taken at ages 6, 7, 8, and 9, whereas another subset was taken at ages 8, 9, 10, and 11. This type of design is actually quite common in the behavioral sciences and often takes the form of a cohort-sequential (or accelerated longitudinal) design (e.g., Schaie 1965). Instead of all subjects being 6 years old at the first assessment, children might range between 6 and 10 years of age at first assessment. Thus multiple cohorts of children are combined within each assessment (e.g., cohort one subjects were age 6 at the first assessment, cohort two were age 7 at the first assessment, and so on).

Now consider a more general expression for the person-specific mean of time that allows for variability in time across observations:

$$\bar{x}_i = \left(\sum_{t=1}^{T_i} x_{ti} \right) / T_i. \quad (35)$$

Here $t = 1$ represents the first time of assessment, and T_i represents the total number of observations made on individual i . The values of x_{ti} need not be constant for a given t . To continue the example from above, in the case of a multiple cohort design, x_{ti} might take on values of 6, 7, 8, 9 for a person in cohort one and values of 8, 9, 10, 11 for a person in cohort two. Unlike balanced designs, in which the person-mean of time is constant over individual, the person-mean of time now varies over individuals. For example, an individual assessed at 6 through 9 has a midpoint of 7.5, but an individual assessed at 8 through 11 has a midpoint of 9.5. This has direct implications for how we disaggregate the between-person component of the TVC when the data are not balanced on time.

Summary

We have covered much ground thus far and here briefly summarize our key developments prior to examining how these impact the disaggregation of effects in practice. First, we proposed a general definition for the between-person and within-person components of the TVC and denoted their sample representations as zb_i and zw_{ti} , respectively. Second, we showed that under the assumption that the TVC is wholly unrelated to time, these components can be validly expressed via the person-mean (\bar{z}_i) and person-mean centered deviate (\dot{z}_{ti}). We also showed that when there are time trends in the TVC, \bar{z}_i and \dot{z}_{ti} are often poor choices for zb_i and zw_{ti} . An exception is the rather rare condition where data are time structured (i.e., observations are balanced on time, and there are no missing data). In this circumstance, zb_i can be validly defined as \bar{z}_i , even when there are time trends in z_{ti} . However, even in this case, \dot{z}_{ti} remains inadequate for zw_{ti} .

Thus, when the TVC model is unrelated to time, the standard methods currently recommended in practice provide valid estimates of within- and between-person effects. However, when the TVC is systematically related to time, the standard methods are no longer sufficient to accurately capture the between- and within-person components of the TVC, and additional analytic steps are needed to isolate these effects. Several empirical demonstrations below highlight how these issues are manifested in practice and illustrate alternative methods for computing zb_i and zw_{ti} .

EMPIRICAL DEMONSTRATIONS

Up to this point, we have primarily approached our thesis at the level of equations. To both augment our communication of these ideas and to empirically validate our analytic developments, we turn to three empirical demonstrations. We use artificially generated data so that we know precisely what is the population-generating model. This allows us to draw unambiguous conclusions about the extent to which a sample estimate is or is not recovering the known population parameter. We draw on characteristics of previously published applications of this type to define what we consider to be typical situations in which these methods might be applied in practice. However, all of our conclusions would hold equally across a wide range of alternative design characteristics (e.g., number of time points, spacing of time points, sample size, etc.).

We could consider six possible conditions: three types of growth in the TVC (no growth, growth with only a fixed effect, and growth with a fixed and random effect), each crossed with two types of structure of time (balanced or unbalanced). We focus here on the three that we believe are most common in practice: (a) no growth in the TVC with balanced time,⁵ and individually varying time trends in the TVC under structures of time that are either (b) balanced or (c) unbalanced.

Disaggregation of Effects With No Growth in the TVC and Time is Balanced

We begin by examining an artificial data set that was created to correspond to conditions under which the person-mean centering approach is expected to properly disaggregate within- and between-person effects. More specifically, we assume that Equations 13 and 16 hold in the population. For our initial data set we generated $n = 500$ simulated cases, each with $T = 9$ repeated measures. We scaled time so that the mean of time was zero (i.e., $t = -4, -3, -2, -1, 0, 1, 2, 3, 4$), although given the absence of growth in this condition, the scaling of time has no impact in the current model. Finally, because this design is balanced on time, all individuals are the same age, are assessed at the same points in time, and there are no missing data.

We can first consider the characteristics of the TVC itself prior to examining the simulated outcome variable. We generated the TVC to be independent of the passage of time; in other words, there is no systematic growth process that underlies z_{it} , consistent with Equation 16. This might be reflective of daily measures of anxiety in which anxiety varied both within and between individuals, but it did not systematically increase or decrease over time. This can be seen in the conditional distribution of the TVC as a function of time presented in Figure 1, in which the distribution of the TVC at each specific time point is nearly identical; that is, the mean of the TVC is independent of time.

The box plots in Figure 1 show the distributions of the TVC pooling over individuals within each time point. However, we can also examine the individual trajectories of the TVC over time. Figure 2 shows the model-implied trajectories of the TVC for 50 randomly selected observations. Two characteristics are particularly important. First, because there is no time trend in the population model, the estimated trajectories are perfectly flat with respect to time. That is, there is no systematic change in the TVC as a function of time. Second, there is substantial individual variability in the relative heights of the individual trajectories. That is, some observations reflect higher levels of the TVC, and others report lower levels. This between-person variability is captured in the random intercept term in Equation 15 from above. Extending our hypothetical example, this figure shows that although anxiety does not change systematically as a function of time, some people are reporting higher overall levels of anxiety, whereas others are not.

It is also helpful to consider the set of observations for just one individual plotted over all the time points; this highlights the within-person variability around each individual trajectory. For example, we could consider the nine repeated measures of anxiety taken on just one individual. The data for a single randomly chosen individual is presented in Figure 3, in which the observed TVC values are plotted against time. The points are the time-specific measures of the TVC, and the horizontal line demarcates the sample mean for the person, pooling over the set of TVCs. The horizontal line thus shows the overall level of anxiety for this individual, and the points show the time-specific values of anxiety relative to the overall level. We can see that the TVC does not appear to be related to time and that the

⁵Because time plays no role in the no-growth condition, whether the design is balanced or unbalanced is irrelevant in this situation. As such, although we focus on the balanced condition, all of our findings for the no-growth TVC model would directly generalize to the unbalanced condition as well.

time-specific measures of the TVC vary randomly around the person-specific mean. This is precisely what allows us to deviate each time-specific measure of the TVC from the person-mean to disaggregate the between-person and within-person effects.

Thus far we have considered only the overtime characteristics of the TVC itself. Next we turn to our simulated outcome, y_{it} , which was generated to be consistent with Equation 13; in words, this is a random intercept-only model for a continuously and normally distributed outcome variable with both a within-person and between-person effect of the single TVC z_{it} . In our hypothetical example, the outcome could represent daily alcohol use that varies both within and between individuals but does not systematically change over time. The overall intercept of the model for y_{it} was defined to be $\gamma_{00} = 5.0$, the within-person effect was $\gamma_{10} = -1.0$, and the between-person effect was $\gamma_{01} = 1.5$. Thus, higher time-specific deviations of the TVC from the overall person-mean are associated with lower values of the outcome, whereas higher overall person-means are associated with higher values of the outcome.

We chose these values to reflect the hypothetical relation that might be found between daily anxiety symptoms and daily alcohol use. More specifically, the positive between-person effect reflects that, on average, people who are more anxious tend to drink more alcohol; this might be attributable to a self-medication process, where alcohol is consumed to modulate anxiety symptoms (e.g., Kassel et al. 2010). In contrast, the negative within-person effect reflects that, on average, people tend to drink less alcohol on days when their anxiety is elevated relative to their typical stable level; this might be attributable to an individual avoiding alcohol-related social contexts on days when anxiety is particularly pronounced (e.g., Kaplow et al. 2001). Note that although theory is predictive of these relations, for our purposes here we consider these strictly hypothetical (although we would sure like to see this study done).

To begin, consider the simple bivariate scatter plot in Figure 4, where the TVC is plotted on the x -axis and the outcome on the y -axis. Although we see a generally positive trend, this is an inextricable aggregation of the between-person effect (which is positive) and the within-person effect (which is negative). Following our hypothetical example, we would conclude from the aggregate analysis that there is a positive relation between anxiety and alcohol use that is modest in size and holds across all individuals in the sample. However, we know the true relation to be patently different. To recover the more complex relation that truly exists, we must disaggregate the TVC into the between-person component (zb_i) and the within-person component (zw_{it}).

One way to get a better visual sense of these two effects is to plot the relationships observed at each level of analysis. Note that we are only using these plots to visually examine potential differences in levels of effect, and we will formally test these disaggregated effects through the parameterization of the multilevel model. To see the within-person effect, we can plot outcome y_{it} against the person-mean centered \hat{z}_{it} ; to see the between-person effect, we can plot the person-means \bar{y}_i against the person-means \bar{z}_i . Figure 5 presents the person-mean centered TVC plotted against the outcome, and Figure 6 presents the person-mean of the TVC plotted against the person-mean of the outcome.

These plots clearly reflect the strong negative within-person relation between the time-specific measure of the TVC and the outcome (Figure 5) and the strong positive between-person relation between the mean of the TVC and the mean of the outcome (Figure 6). This is of course precisely how we generated these data. We now use the techniques described above to obtain estimates of the between- and within-person effects via the multilevel model, in which $zb_i = \bar{z}_i$ and $zw_{it} = \hat{z}_{it}$ are included as separate predictors of y_{it} .

To do this, we fitted a multilevel model consistent with Equation 13 to formally test the between- and within-person influences of the TVC. Recall that 500 individuals were each assessed nine times, resulting in a total of 4500 person-time observations. We fitted a two-level model under full information maximum likelihood and obtained an estimate of the within-person effect of $\hat{\gamma}_{10} = -0.99$ ($se = 0.008$) and of the between-person effect of $\hat{\gamma}_{01} = 1.51$ ($se = 0.022$). Recall that the corresponding population values were $\gamma_{10} = -1.0$ and $\gamma_{01} = 1.5$, respectively; thus, as expected, we closely replicated these values in our artificial sample.⁶ Continuing with our hypothetical example, these results would reflect that, on average, people reporting higher overall levels of anxiety tended to drink more alcohol; but at the very same time, on average, people tended to drink less alcohol on days when they reported higher levels of anxiety. This nicely highlights that the first conclusion made with respect to *between* individual differences, and the second conclusion is made with respect to *within* individual differences.

As we fully expected based on prior analytic theory, the person-mean centering approach accurately recovered the known population-generating values. However, although comforting, this is at best a modest victory. That is, we generated a population model consistent with Equations 13 and 16, and then we fit a sample model that corresponded to these same generating equations. Had we found anything other than these results, you would do well to suspect that we made an error in our computer programming. However, we view this as an important endeavor in that it demonstrates that the existing methods work properly when the underlying assumptions are met. Further, it gooses us to think more carefully about the specific conditions under which person-mean centering is a valid method for disaggregating multiple levels of effect.

Disaggregation of Effects with Growth in the TVC and Time is Balanced

The second situation we consider is when there are both fixed and random effects of growth underlying the TVC and the design is balanced on time (i.e., Equation 30). Extending our hypothetical example, we remain interested in studying the relation between anxiety and alcohol use. However, we now want to consider the situation in which anxiety is not only increasing over time, but there are also individual differences in both starting point and rate of change. We thus defined a linear growth model to underlie the TVC itself based on the same sample size ($N = 500$) and same number of time points ($T = 9$) as before. The TVC in this second data set was defined to have an intercept equal to 25.0 and a linear slope equal to 1.0; these are arbitrary values, but they define a linear growth trajectory for the TVC. Further, we coded time so that the middle point was equal to zero, meaning that the intercept is defined as the mean of the outcome at the mean of time, and the TVC increased in value by one unit with each unit increase in time. Finally, we allowed for individual variability (that is, random effects) in both the starting point ($\tau_{00} = 4$) and rate of change over time ($\tau_{11} = 1$) and a level-1 residual equal to $\sigma^2 = 1$.

To better illustrate the implications of the inclusion of this time trend, Figure 7 presents the conditional distributions of the TVC as a function of time. It is clear that the time-specific means are (as we intended) increasing as a function of time. Further, note that the variance of the TVC varies as a function of time; this is also consistent with our population-generating model because there is a random slope component that differentially influences time-specific variability over time. In terms of our hypothetical example, both the mean and variance of anxiety are changing as a function of time; the mean is increasing linearly, and the variance is changing quadratically.

⁶Although there are also corresponding residual random effects, we do not focus on these here. As with the fixed effects, all random effects closely estimated the corresponding population values.

To see the influence of the random components on growth, in Figure 8 we present the individual model-implied trajectories of the TVC for 50 randomly drawn cases. This highlights not only the systematic increase in the TVC over time, but also the individual variability in starting point and rate of change. You can consider each of these lines as an individual's own trajectory of anxiety symptoms unfolding over the period of observation. On a related point, note that each trajectory spans the entire period of time, reflecting that these data are balanced with respect to time. Finally, relevant to later analysis, note that the relative rank ordering of values on the outcome changes over time. To see this, picture drawing a vertical line at each value of time; because the slopes are not parallel, the individual standing on the TVC varies at each vertical line drawn at a given value of time.

However, why would the systematic relation between the TVC and time potentially undermine the validity of the person-mean centering approach? Although we showed this analytically above (i.e., Equation 33), this threat to validity can be saliently visualized when examining the distribution of the TVCs over time for an individual case. In Figure 9, the TVC is plotted on the y -axis, time is plotted on the x -axis, and the horizontal line demarcates the person-specific mean of the set of TVCs. However, the positively sloped line is the regression line of best fit linking the TVC to time. This is consistent with the increasing value of the TVC associated with the passage of time; that is, the hypothetical individual is reporting progressively higher values of anxiety at each time point.

Importantly, note that the person-mean centering strategy deviates each TVC relative to the horizontal line because of the implicit assumption that the value of the TVC is independent of time. Yet it is clear from this plot that person-mean centering fails to differentiate within-person fluctuations around the time trend. Using existing standard methods, all of the values of the TVC falling below the person-mean receive a negative deviated score, and all of the values falling above the person-mean receive a positive deviated score. These values are incorrect for obtaining a sample estimate of the within-person variability of the TVC over time. Instead, we must deviate the time-specific values of the TVC not from the horizontal line but instead from the positively sloped regression line. Only this will properly isolate the within-person component of the TVC.

To demonstrate this, we first applied the standard methods for disaggregating the between- and within-person effects of the TVC on the outcome. Given that the TVC was generated to be related to time yet the standard methods assume no relation to time, we a priori expect these results to be biased. To evaluate this, we fitted precisely the same person-mean centered model to the second data set as we did to the first. Although in the first data set we nearly perfectly recovered the corresponding population parameters, this did not occur here.

The person-mean deviated TVC resulted in a highly biased estimate of the within-person effect. Specifically, the within-person effect was estimated to be $\hat{\gamma}_{01} = -0.07$ ($se = 0.006$), whereas the corresponding population value was $\gamma_{10} = -1.0$. Thus, applying the standard methods of person-mean centering to data in which the TVC varies as a function of time results in a within-person effect that drastically underestimates the known population value. In our hypothetical example, we would conclude that there was indeed a negative within-person effect, yet we would underestimate the magnitude of this effect by 93%. This is a striking amount of bias that occurs even under what are otherwise ideal conditions (e.g., large sample size, large numbers of repeated measures, no missing data).

In contrast to the highly biased within-person effect, we accurately recovered the population between-person effect; our obtained value was $\hat{\gamma}_{10} = 1.49$ ($se = 0.029$), whereas the corresponding population value was $\gamma_{10} = 1.5$. To better understand this accurate recovery, recall that we generated the TVC such that the mean of time was equal to zero (i.e., time was

centered around zero). As such, because this condition is balanced, $\bar{x}_i = \bar{x} = 0$ for all individuals. Thus the omitted second set of terms in Equation 32 (i.e., $[\gamma_{10} + \hat{u}_{1i}]\bar{x}_i$), drops out and the person-mean accurately recovers the between-person effect. Note, however, that this is strictly a function of the balanced design. If time were unbalanced (e.g., if there were missing data or a cohort-sequential design), then the person-mean would not accurately capture the between-person effect in this situation. Indeed, we demonstrate just this point in the next example.

Whereas in the balanced case the person-specific mean of time (\bar{x}_i) is constant over individual, the deviation of the individual value of time from the mean ($x_{ti} - \bar{x}_i$) is not. Thus the traditional method neglects the term $(\gamma_{10} + \hat{u}_{1i})(x_{ti} - \bar{x}_i)$ from Equation 33 in the calculation of the time-specific deviation of z_{ti} from the person-mean. This is why our sample estimate of the within-person effect was equal to -0.07 when the corresponding population value was equal to -1.0 . Fortunately, though, we can draw on our prior developments to obtain an unbiased estimate of this known population effect.

To do this, we need a person-specific estimate of $\gamma_{10} + u_{1i}$ to use in the calculation of $z_{w_{ti}}$. More specifically, instead of deviating the time-specific TVC measures with respect to the person-mean, we can deviate the TVCs with respect to the individual-specific regression line linking the TVC and time. This strategy can be more clearly understood by reconsidering Figure 9. Here we plotted the TVCs against time for a single individual, and we superimposed both a horizontal line representing the person-mean and the best-fitting regression line estimating the positive relation between time and the TVC. Whereas the traditional person-mean centering approach deviates the TVC with respect to the horizontal line, we can instead deviate the TVC with respect to the regression line. We refer to this strategy as detrending.

The general concept of detrending is far from novel, and it has been used in various forms in time-series analysis for decades (e.g., Chatfield 1996). However, to our knowledge there has been no prior discussion of applying these techniques in the multilevel model in order to disaggregate between- and within-person effects of a TVC on the outcome when the TVC itself is related to time. Our proposed approach for detrending is simple. We first regress the TVC on time separately for each individual using ordinary least squares (OLS). We then deviate each time-specific TVC not from the overall person-mean (as is done in the traditional approach) but instead from the model-implied value of the TVC specific to that particular unit of time. In other words, our deviated TVC measure is simply the residual (i.e., the observed minus expected value) from the regression of the TVC on time computed separately for each individual case.

We can present this more formally as a one-predictor regression equation estimated separately (case by case) for each individual in the sample. This is given as

$$z_{ti} = b_{0i} + b_{1i}x_{ti} + e_{ti}, \quad (36)$$

where z_{ti} is the time-specific measure of the TVC, x_{ti} is the measure of time, b_{0i} and b_{1i} are sample estimates of the intercept and the slope of the regression of the TVC on time, respectively, and e_{ti} is the time-specific residual.⁷ A trivial rearrangement of this equation shows that

⁷Here we use different notation to differentiate the OLS regression of the TVC on time (i.e., b_{0i} , b_{1i} , e_{ti}) from the multilevel growth model for the TVC (i.e., β_{0i} , β_{1i} , r_{ti}).

$$e_{ti} = z_{ti} - b_{0i} - b_{1i}x_{ti} = z_{ti} - \widehat{z}_{ti}, \quad (37)$$

where e_{ti} is the detrended rescaling of the TVC. In other words, the residual e_{ti} is computed by deviating the time-specific TVC from the model-implied value of the TVC that includes information about the specific value of time. Thus the TVC is deviated not relative to the horizontal line but instead relative to the regression line. We now define zw_{ti} as e_{ti} .

An interesting generalization can be seen here as well. We could fit the OLS regression of the TVC on time defined in Equation 36 to our initial artificial data set in which the TVC was unrelated to the passage of time. Given the structure of the data, there would be no $b_{1i}x_{ti}$ term in Equation 36, and this would simplify to

$$z_{ti} = b_{0i} + e_{ti}, \quad (38)$$

and the deviation of the TVC would be

$$e_{ti} = z_{ti} - b_{0i}, \quad (39)$$

which is precisely equal to the traditional person-mean centering approach we first described (because $b_{0i} = \bar{z}_i$ when there are no predictors in the regression equation). However, the more general conclusion is that the person-mean centering approach is equivalent to detrending but under the implicit assumption that there is no relation between the TVC and time, and thus b_{1i} is zero for all cases. Here we simply extend this approach to allow b_{1i} to take on some nonzero value from the data.

To examine the utility of this approach, we detrended the TVC in the second data set with respect to the regression line fitted to each case individually.⁸ Once detrended, we then used this rescaling of the TVC in precisely the same way as before; namely, we included the detrended TVC as the level-1 predictor (zw_{ti}), and we retained the OLS intercept from Equation 36 as the level-2 predictor (zb_i). Because in this balanced condition the OLS intercept is equal to the person-mean used in our initial model that we fitted to these data, we get the same estimate of the between-person effect as we did before: $\hat{\gamma} = 1.49$ ($se = 0.029$). However, whereas our prior estimate of the within-person effect was highly biased when using the person-mean centered TVC, we recover this with near-precision using the detrended TVC: $\hat{\gamma} = -0.99$ ($se = 0.018$). These results demonstrate that when the TVC is systematically related to the passage of time, it is critical that the TVC be deviated not with respect to the person-mean but instead with respect to the individual-specific regression linking the TVC and time.

In sum, this second artificial data set was generated so that there was a random growth process underlying the TVC. However, this was embedded in the unrealistic condition of complete and balanced data. Our third and final data example considers the same growth model for the TVC but embedded in a more realistic condition of unbalanced time.

Disaggregation of Effects with Growth in the TVC and Time is Unbalanced

An important characteristic of the first two artificial data sets is that each simulated subject was followed for precisely the same nine time periods. This is consistent with a birth-cohort

⁸This can easily be done in any commercial statistical package where a separate regression is estimated for each unique ID.

design in which an entire cohort of individuals is assessed at the same age at each assessment period and there are no missing data. Because we numerically coded time to range from -4 to 4 , the mean value (or midpoint) of time is equal to 0 for each of the 500 individuals. As such, every single person has the same mean of time, equal to zero. The person-mean of the TVC cannot then covary with the person-mean of time because all person-mean values of time are equal for all individuals.

However, as we described above, the time-balanced birth-cohort design is rare in many behavioral science research applications. Instead, multiple cohorts are often considered simultaneously, whether intentionally by design (e.g., one sample of 5-year-olds is recruited, one sample of 6-year-olds is recruited, etc.) or unintentionally by happenstance of the distribution of age within each assessment (e.g., inclusion criteria include children 5 to 9 years of age at first assessment). Further, given that missing data are endemic in longitudinal social science research, even a true birth cohort design will typically be unbalanced.

To simulate this much more realistic situation, we began with precisely the same empirical data as was used in our second example. However, we made one very simple yet critically important modification to this data set: we randomly divided the $N = 500$ individuals into six discrete groups, each representing one distinct cohort (there were 83 individuals in each of five cohorts and 85 in the sixth). Once we created the six groups, we then retained just the first through fourth assessments for the first cohort (i.e., time points $-4, -3, -2, -1$) and just the second through fifth assessments for the second cohort (i.e., time points $-3, -2, -1, 0$); we did this for each cohort, ending with the retention of the sixth through ninth assessments for the final cohort. There were thus still 500 individuals with the very same data as before, but here we only retained four assessments from any given individual, the specific four of which depended on the cohort to which the individual belonged.⁹ This design is unbalanced with respect to time.

Whereas in Figure 8 each individual trajectory spans all nine time points, here any given trajectory spans only four time points. Further, which four time points are spanned varies as a function of cohort membership. This can be seen in Figure 10, in which the trajectories of the TVC and time are shown for 50 random cases. Two implications arise from this unbalanced design.

First, recall that in the balanced case the mean of time (i.e., \bar{x}_i) was equal to zero across all 500 individuals. However, now the mean of time varies as a function of within which cohort the individuals reside. Specifically, the mean values of time for the six cohorts range from -2.5 to 2.5 by increments of 1 (e.g., $\bar{x}_i = -2.5$ for cohort 1; $\bar{x}_i = -1.5$ for cohort 2; and so on). Because the mean of time now varies over individual, we must account for this additional information in the disaggregation of our between- and within-person effects.

Second, even when the TVC is related to time, in the balanced condition there is just one unique value of the person-specific mean of the TVC pooling over the total period of time. That is, each person is characterized by a mean-value of the TVC pooling over the nine time points. However, when the TVC is related to time in the unbalanced condition, the person-specific mean value of the TVC varies as a function of precisely when in time the individual was assessed. For example, if the TVC is increasing over the nine time points, the person-specific mean of the TVC will also increase as the four-time-point assessment window increases (e.g., the mean of the TVC is directly related to the mean of time).

⁹Although we could have also introduced missing data within each cohort, this would not have influenced any of our subsequent conclusions, given that the data are already unbalanced with respect to time.

This can best be seen in the conditional distributions of the person-means of the TVC as a function of cohort membership; this is presented in Figure 11. To clarify, there were $N = 83$ individuals belonging to cohort 1 who were assessed at the first four time points (coded $-4, -3, -2, -1$); the first box plot in Figure 11 presents the distribution of the person-specific means of the TVC for these individuals, and this has an overall mean of 22.62. The second box plot presents the distribution of the person-specific means of the TVC for the next $N = 83$ individuals who belong to cohort 2 (and who were thus assessed between times -3 and 0), and this has an overall mean of 23.61; and so on. The horizontal line denotes the grand mean of the TVC, which is equal to 25. Notice that no cohort-specific mean is equal to the grand mean.

Returning to our hypothetical example, these data would reflect that earlier (and thus younger) cohorts are reporting less overall anxiety compared to the later cohorts. Interestingly, this is not some strange statistical artifact; this is an accurate reflection of the sample characteristics in that later cohorts do indeed report higher overall levels of anxiety than do earlier cohorts. However, the sole source of this difference is that the later cohorts are assessed at a later age than are the earlier cohorts, and anxiety is increasing with time. Thus person-mean values of anxiety are confounded with time. This is directly analogous to measuring height over time where one cohort was assessed between ages 5 and 10 and a second cohort between ages 9 and 14. Of course the second cohort reports higher values of average height—they are older, and children tend to increase in height with age. But this in no way implies that the second cohort would have been taller than the first had both cohorts been assessed at the same age. This is the crux of the challenge we face: We need to isolate the within-person and between-person differences in the TVC while adjusting for the different values of time at which the assessments were obtained.

Figure 11 clearly reflects that the cohort-specific mean of the person-means of the TVC increases monotonically as a function of the cohort to which individuals belong. Because cohort is directly related to time, the person-mean of the TVC is also unambiguously linked to the passage of time. It is very important to note that this is not a contrived or tortured example; indeed, this situation is almost universally encountered in any cohort-sequential design in which the TVC itself is related to the passage of time.

To examine the implications of this, we first used the standard person-mean centering approach to disaggregate the between-person and within-person influences of the TVC on the outcome. We thus fitted Equation 12 to the artificial data and (as expected) found significantly biased effects for both the within- and between-person influences. The within-person effect was $\hat{\gamma}_{10} = -0.24$ (compared to the population value of -1.0), and the between-person effect was $\hat{\gamma}_{01} = 0.71$ (compared to the population value of 1.50). Notice that whereas the person-mean successfully recovered the between effect in the balanced condition, this is now underestimated by more than 50% based on the very same data in the unbalanced condition. Thus under conditions that are likely common in many areas of psychological research, the standard methods for disaggregating effects are highly biased.

We next drew on our expressions for computing zb_i and zw_{it} in the presence of random growth to obtain the necessary disaggregated components of the TVC. In words, we simply regressed the TVC on time within each individual where time is grand-mean centered. We then retained the time-specific residuals as our estimate of zw_{it} (i.e., e_{it} from Equation 36), and we retained the sample estimate of the regression intercept as our estimate of zb_i (i.e., b_{0i} from Equation 36). Using these as predictors in the model for our outcome y_{it} , we obtained an estimate of the within-person effect of $\hat{\gamma}_{10} = -0.95$ ($se = 0.036$) and an estimate of the between-person effect of $\hat{\gamma}_{01} = 1.25$ ($se = 0.041$). Although the within-person effect was underestimated by 5% and the between-person effect by 17% relative to their

population counterparts, these estimated values are substantially more accurate than those obtained using traditional methods for disaggregating effects. This is because information about time (via x_{it} and \bar{x}_i) is explicitly considered in the computation of zb_i and zw_{it} , whereas this is omitted when using standard methods.

There are two related reasons why the between- and within-person effects were recovered with near-perfect precision in the balanced case but with only modest bias in the unbalanced case. First, all cases in the balanced condition had $T = 9$ repeated measures, and all cases in the unbalanced condition had $T = 4$ repeated measures. Thus the OLS estimates used as zb_i and zw_{it} are estimated with greater precision, given higher numbers of repeated measures. Second, and more importantly, recall that we are using the person-specific estimate of the intercept term (i.e., b_{0i}) from the regression of the TVC on time. As in any regression, the intercept reflects the mean of the TVC at the mean of time (i.e., since our coding of time means that $\bar{x} = 0$). In the balanced case, all individuals were observed across all time points, so b_{0i} was estimated within the range of observed data (that is, each individual was observed at $x_{it} = 0$). In contrast, in the unbalanced case, not every individual was observed when $x_{it} = 0$. For example, individuals in cohort 1 were observed at times $-4, -3, -2,$ and -1), yet the estimate of b_{0i} reflects the mean of the TVC when $x_{it} = 0$, which is outside of the range of observed data in this cohort. As such, the estimates of b_{0i} were projected beyond the window of observation for many individuals, thus further undermining the precision of estimates beyond simply having fewer repeated measures than in the balanced condition. Nevertheless, our obtained estimate for the between-person effect is still much improved by using b_{0i} compared to \bar{z}_i .

UNRESOLVED ISSUES AND FUTURE DIRECTIONS

Our goal in this review has been to explore the conditions under which traditional methods used to disaggregate between- and within-person effects are and are not valid and to propose new methods to augment existing techniques when needed. We believe that we have been able to meet these goals, although there remain a number of issues that must be considered both in terms of potential limitations to our proposed methods and as clear avenues for continued work and development. We briefly address several key remaining issues, although certainly more exist beyond these.

The Conditional Relation Between zb_i and \bar{x}_i in the Presence of Growth

Recall that for the most general case in which there were both fixed and random components of growth underlying the TVC, the between-person component of the TVC was given as

$$\hat{u}_{0i} = (\bar{z}_i - \hat{\gamma}_{00}) - (\hat{\gamma}_{10} + \hat{u}_{1i})\bar{x}_i, \quad (40)$$

where \hat{u}_{0i} was used as our estimate of zb_i . Note that the obtained value of \hat{u}_{0i} depends directly on \bar{x}_i ; that is, the individual-specific between-person component of z_{it} varies as a function of the person-mean of the TVC. Because we grand-mean centered time in our demonstrations, our between-person effect estimate was taken at the grand mean of time. But in the presence of random growth and unbalanced time, the value of \hat{u}_{0i} will change as a function of precisely where in time this is evaluated. For example, if the metric of time were rescaled such that $x_{it} = 0$ represented the initial assessment (instead of the middle assessment considered here), then \hat{u}_{0i} would change, as would $\hat{\gamma}_{01}$ (the between-person estimate of the TVC on the outcome); e.g., see Biesanz et al. (2004). Given this, we must broaden our conceptualization of the existence of a single between-person effect of the TVC on the outcome. Instead, we must consider this to be a conditional relation that varies as a function

of time. Both substantive and analytical issues extend from this, and each needs careful future considerations.

Alternative Functional Forms of Growth for the TVC

All of our work here has focused on a linear trend relating the TVC to time. Although our specific equations are thus limited to this linear trend, our more general concepts are not. For example, one conclusion we draw here is that the within-person component of the TVC should be obtained with respect to the trend relating the TVC to time and not with respect to the person-mean. This trend might be linear, quadratic, exponential, or any of a wide variety of functions. Our equations can be extended to a number of functions that are distinctly nonlinear with respect to time, and the methods to obtain sample estimates of the desired components of the TVC can be adjusted accordingly. However, further work is needed to understand the subtle nuances and potential complications that likely arise here.

Alternative Methods of Estimation of zb_i and zw_{ij}

We provided general definitions for zb_i and zw_{ij} and then focused solely on OLS estimation of these values via the standard person-specific regression model. We used OLS estimation here because this is consistent with the calculation of \bar{z}_i and \dot{z}_{ij} via the traditional methods of disaggregation. That is, we can compute \bar{z}_i and \dot{z}_{ij} from an OLS regression that has no predictors. However, OLS is just one of a variety of methods available to obtain sample estimates of these values. For example, one might consider using empirical Bayes estimates (e.g., Raudenbush & Bryk 2002, p. 46) calculated from a multilevel growth model fitted to the TVC itself; alternatively, factor score estimates (e.g., Bollen & Curran 2006, p. 52) could be obtained from a structural equation growth model fitted to the TVCs. Finally, it might be possible to calculate these values directly within the confines of a single model (e.g., Lüdtke et al. 2008). There are advantages and disadvantages to each of these alternatives to OLS, and future work is needed to more thoroughly study the conditions under which various methods might be optimal.

Discrete Distributions of the TVC

Throughout our review, we have made a simplifying assumption that both the outcome and the TVC are continuously and normally distributed. Interestingly, all of our developments extend directly to the generalized multilevel model in which the outcome measure is discrete (e.g., binary or ordinal); indeed, all of our work presented here stemmed from our attempts to overcome these problems when predicting binary drug use in our own data (Curran et al. 2010). However, complications are encountered when the TVC itself is discretely scaled. One reason is that, although we demonstrated using OLS estimation to obtain the desired components of the TVC, this method of estimation assumes continuously distributed outcomes. However, many binary TVCs may be more representative of a particular status at a particular time point (e.g., married versus single; Curran et al. 1998) and thus less likely to show systematic growth over time. More careful work is needed to understand how zb_i and zw_{ij} can best be obtained when the TVC is binary or ordinal.

Alternative Modeling Strategies

As we noted at the outset, we chose to place our sole emphasis on the multilevel model. There were a number of reasons for this, two of which were the generality of the multilevel modeling framework and the ubiquity of prior developments of disaggregating TVC effects within this approach. However, other modeling frameworks are available, a key example being the structural equation-based latent curve model (LCM). Whereas the multilevel model is motivated by the nesting of the repeated measures within an individual (e.g., Bryk & Raudenbush 1987), the LCM is motivated by the use of the repeated measures as

observed indicators of an underlying latent growth process (e.g., Meredith & Tisak 1990). As is well known, there is a great deal of overlap between the multilevel growth model and the LCM, although there are key points of divergence as well (Bauer 2003, Curran 2003, Raudenbush 2001a, Willett & Sayer 1994). Relevant to our discussion here, recent work has shown that the multilevel model and LCM handle the incorporation of TVCs in a radically different way despite being based on precisely the same empirical data (Curran et al. 2010***). Further, several different methods have been proposed to examine bidirectional and time-specific influences of one variable on another within the structural equation model (e.g., Bollen & Curran 2003, Cole et al. 2006, McArdle et al. 2002). Future work will do well to consider how the issues we have explored here are manifested within both modeling frameworks.

Striving for a Truly Multivariate Model of Change

Finally, the entire premise of our paper is that there exists some time trend in the TVC that must be isolated and removed from the observed data prior to estimating the multilevel model of interest. This literally takes the form of a manual manipulation of our observed data: We obtain our observed values of the TVC; we fit a regression model to the TVC with time as a predictor, and we retain the estimated intercept and residuals; and we use the intercept and residuals as new predictors in the multilevel model. However, as with any statistical model, this two-step approach is neither parsimonious nor statistically efficient (nor very pretty, to be completely candid). For example, although we use b_{0i} drawn from the OLS regressions as our estimate of zb_i , we do not consider imprecision in estimation of b_{0i} , and we thus treat these values as fixed and known (just as we do when using \bar{z}_i). We are thus discarding potentially meaningful information about within-person variability when calculating \bar{z}_i or b_{0i} .

What we ultimately desire here is a truly multivariate model that simultaneously relates the outcome to time, the TVC to time, and the outcome to the TVC. Although a multi-variate multilevel model is well developed and very powerful (e.g., MacCallum et al. 1997), this allows only for the relating of the TVC and the outcome strictly at the level of the trajectories. This approach does not allow for the addition of time-specific structural relations between the TVC and the outcome, which are necessary to obtain unambiguous insights into the within-person relation between the two constructs.

Although models such as these have been proposed in other analytic frameworks (e.g., Bollen & Curran 2003, Curran & Bollen 2001, McArdle et al. 2002), none of these have closely considered the disaggregation of between- and within-person effects. For example, although Curran & Bollen (2001) describe time-specific relations and trajectory-specific relations, no mention is made as to how these map onto the concept of within-person and between-person effects. Indeed, crossing the work of Curran & Bollen (2001) with Curran et al. (2010) raises several key questions as to precisely how within- and between-person effects might meaningfully map onto time- and trajectory-specific effects within the LCM (if they even can be mapped at all). Much more careful work is needed in the ongoing pursuit of a truly multivariate model that successfully disentangles within-person and between-person effects in an unambiguous and meaningful way.

RECOMMENDATIONS FOR USE IN PRACTICE

We conclude by offering several specific recommendations for separating and testing between- and within-person effects of a TVC on an outcome in practice. However, we cannot stress strongly enough that we view these recommendations as preliminary at best, and we do not intend for these to be taken as the new best practice strategies. Although we believe our recommendations are analytically informed, empirically supported, and

pragmatically useful, we would fully expect that future developments in any of the areas we described above would modify our proposed strategies.

First, we recommend that a random effects growth model first be fit to the TVC itself. Many quality resources exist that offer guidance in fitting and interpreting growth models within both the multilevel model (e.g., Raudenbush & Bryk 2002, Singer & Willett 2003) and structural equation model (Bollen & Curran 2006, Duncan et al. 2006, McArdle 2009). Second, if it is determined that no meaningful growth is evident in the TVC, then the standard methods of obtaining zb_i and zw_{ti} via \bar{z}_i and \bar{z}_{ti} may be used. Third, if evidence of growth in the TVC is found, then zb_i and zw_{ti} should be obtained using case-based regressions of the TVC on grand-mean centered time where b_{0i} and e_{ti} are estimated and retained. Finally, because in the absence of growth in the TVC $\bar{z}_i = \hat{b}_{0i}$ and $\bar{z}_{ti} = \hat{e}_{ti}$, as a sensitivity analysis the multilevel model for the outcome could be run using each approach as the obtained estimate of zb_i and zw_{ti} . Convergence in results lends greater confidence to the lack of a potentially biasing relation between time and the TVC; divergence in results prompts more careful scrutiny as to the possible presence of growth in the TVC that may not be evident in the TVC growth model results. Only through the careful and thoughtful consideration of the role that time plays with respect to both the TVC and the outcome can valid and reliable estimates of between-person and within-person effects be obtained.

Acknowledgments

The authors thank Andrea Hussong for her helpful guidance and valuable insights. This work was partially supported by NIH grant DA013148 awarded to the first author. Sample data and computer code can be obtained from either author.

LITERATURE CITED

- Aiken, L.S.; West, S.G. *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage; 1991.
- Bauer DJ. Estimating multilevel linear models as structural equation models. *J. Educ. Behav. Stat.* 2003; 28:135–167.
- Biesanz JC, Deeb-Sossa N, Aubrecht AM, Bollen KA, Curran PJ. The role of coding time in estimating and interpreting growth curve models. *Psychol. Methods.* 2004; 9:30–52. [PubMed: 15053718]
- Bollen KA, Curran PJ. Autoregressive latent trajectory (ALT) models: a synthesis of two traditions. *Sociol. Methods Res.* 2003; 32:336–383.
- Bollen, K.A.; Curran, P.J. *Latent Curve Models: A Structural Equation Approach*. Wiley Series on Probability and Mathematical Statistics. Hoboken, NJ: Wiley-Intersci; 2006.
- Bryk AS, Raudenbush SW. Application of hierarchical linear models to assessing change. *Psychol. Bull.* 1987; 101:147–158.
- Chatfield, C. *The Analysis of Time Series*. 5th ed.. Boca Raton, FL: Chapman & Hall; 1996.
- Cohen, J.; Cohen, P.; West, S.; Aiken, L. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd ed.. Hillsdale, NJ: Erlbaum; 2003.
- Cole DA, Nolen-Hoeksema S, Girgus J, Paul G. Stress exposure and stress generation in child and adolescent depression: a latent trait-state-error approach to longitudinal analyses. *J. Abnorm. Psychol.* 2006; 115:40–51. [PubMed: 16492094]
- Collins LM. Analysis of longitudinal data: the integration of theoretical model, temporal design, and statistical model. *Annu. Rev. Psychol.* 2006; 57:505–528. [PubMed: 16318605]
- Cronbach LJ, Webb N. Between-class and within-class effects in a reported aptitude \times treatment interaction: reanalysis of a study by GL Anderson. *J. Educ. Psychol.* 1975; 67:717–724.
- Curfman GD. Is exercise beneficial—or hazardous—to your heart? *N. Engl. J Med.* 1993; 329:1730–1731. [PubMed: 8232463]

- Curran PJ. Have multilevel models been structural equation models all along? *Multivariate Behav. Res.* 2003; 38:529–569.
- Curran, PJ.; Bollen, KA. The best of both worlds: combining autoregressive and latent curve models. In: Collins, LM.; Sayer, AG., editors. *New Methods for the Analysis of Change*. Washington, DC: Am. Psychol. Assoc.; 2001. p. 105-136.
- Curran PJ, Hussong AM, Bauer DJ, Chassin L, Sher K, Zucker R. The within-person and between-person relations between internalizing and externalizing symptomatology and drug use. Manuscript submitted. 2010
- Curran PJ, Lee T, MacCallum R. Disaggregating between-person and within-person effects in multilevel and structural equation models. Manuscript submitted. 2010
- Curran PJ, Muthén BO, Harford TC. The influence of changes in marital status on developmental trajectories of alcohol use in young adults. *J. Stud. Alcohol.* 1998; 59:647–658. [PubMed: 9811086]
- Curran PJ, Willoughby MJ. Reconciling theoretical and statistical models of developmental processes. *Dev. Psychopathol.* 2003; 15:581–612. [PubMed: 14582933]
- Duncan OD, Davis B. An alternative to ecological correlation. *Am. Sociol. Rev.* 1953; 18:665–666.
- Duncan, TE.; Duncan, SC.; Strycker, LA. *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications*. 2nd ed.. Mahwah, NJ: Erlbaum; 2006.
- Durkheim, E. *Le Suicide*. Paris: F. Alcan. Transl. JA Spalding, 1951. Toronto: Free Press, Collier-MacMillan; 1897.
- Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol. Methods.* 2007; 12:121–138. [PubMed: 17563168]
- Firebaugh G. A rule for inferring individual-level relationships from aggregate data. *Am. Sociol. Rev.* 1978; 43:557–572.
- Foster M. Race or place: racial disparities in children’s use of psychotropic medications. Manuscript submitted. 2010
- Goldstein, H. *The Design and Analysis of Longitudinal Studies: Their Role in the Measurement of Change*. New York: Academic; 1981.
- Hussong AM, Cai L, Curran PJ, Flora DB, Chassin L, Zucker RA. Disaggregating the distal, proximal, and time-varying effects of parent alcoholism on children’s internalizing symptoms. *J. Abnormal. Child Psychol.* 2008; 36:335–346.
- Hussong AM, Hicks RE, Levy SA, Curran PJ. Specifying the relations between affect and heavy alcohol use among young adults. *J. Abnorm. Psychol.* 2001; 110:449–461. [PubMed: 11502088]
- Hussong AM, Wirth RJ, Edwards MC, Curran PJ, Zucker RA, Chassin LA. Externalizing symptoms among children of alcoholic parents: entry points for an antisocial pathway to alcoholism. *J. Abnorm. Psychol.* 2007; 116:529–542. [PubMed: 17696709]
- Kaplow JB, Curran PJ, Costello EJ. The prospective relation between dimensions of anxiety and the initiation of adolescent alcohol use. *J. Clin. Child Psychol.* 2001; 30:316–326. [PubMed: 11501249]
- Kaplow JB, Curran PJ, Dodge K. Conduct Probl. Prev. Res. Group. Child, parent, and peer predictors of early-onset substance use: a multi-site longitudinal study. *J. Abnorm. Child Psychol.* 2002; 30:199–216. [PubMed: 12041707]
- Kassel, DJ.; Hussong, AM.; Wardle, MC.; Veilleux, JC.; Heinz, A., et al. Affective influences in drug use etiology. In: Scheier, LM., editor. *Handbook of Drug Use Etiology: Theory, Methods and Empirical Findings*. Washington, DC: Am. Psychol. Assoc.; 2010. p. 183-206.
- Kreft IGG, de Leeuw J, Aiken LS. The effect of different forms of centering in hierarchical linear models. *Multivariate Behav. Res.* 1995; 30:1–21.
- Lüdke O, Marsh HW, Robitzsch A, Trautwein U, Asparouhov T, Muthén B. The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. *Psychol. Methods.* 2008; 13:203–229. [PubMed: 18778152]
- MacCallum RC, Kim C, Malarkey W, Kiecolt-Glaser J. Studying multivariate change using multilevel models and latent curve models. *Multivariate Behav. Res.* 1997; 32:215–253.

- Mason WM, Wong GY, Entwisle B. Contextual analysis through the multilevel linear model. *Sociol. Method.* 1983; 14:72–103.
- McArdle JJ. Latent variable modeling of differences in changes with longitudinal data. *Annu. Rev. Psychol.* 2009; 60:577–605. [PubMed: 18817479]
- McArdle JJ, Ferrer-Caja E, Hamagami F, Woodcock RW. Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life-span. *Dev. Psychol.* 2002; 38:115–142. [PubMed: 11806695]
- Mehta PD, West SG. Putting the individual back into individual growth curves. *Psychol. Methods.* 2000; 5:23–43. [PubMed: 10937321]
- Meredith W, Tisak J. Latent curve analysis. *Psychometrika.* 1990; 55:107–122.
- Millar JS, Zammuto RM. Life histories of mammals: an analysis of life tables. *Ecology.* 1983; 64:631–635.
- Mittleman MA, Maclure M, Tofler GH, Sherwood JB, Goldberg RJ, Muller JE. Triggering of acute myocardial infarction by heavy physical exertion—protection against triggering by regular exertion. *N. Engl. J. Med.* 1993; 329:1677–1683. [PubMed: 8232456]
- Molenaar PCM. A manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever. *Meas.: Interdiscip. Res. Perspect.* 2004; 2:201–218.
- Molenaar, PCM.; Newell, KM. *Individual Pathways of Change: Statistical Models for Analyzing Learning and Development.* Washington, DC: Am. Psychol. Assoc.; 2010.
- Muthén BO, Curran P. General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation. *Psychol. Methods.* 1997; 2:371–402.
- Nesselroade, JR. Interindividual differences in intraindividual change. In: Collins, LM.; Horn, JL., editors. *Best Methods for the Analysis of Change: Recent Advances, Unanswered Questions, Future Directions.* Washington, DC: Am. Psychol. Assoc.; 1991a. p. 92-105.
- Nesselroade, JR. The warp and the woof of the developmental fabric. In: Downs, R.; Liben, L.; Palermo, DS., editors. *Visions of Aesthetics, the Environment, and Development: The Legacy of Joachim F. Wohlwill.* Hillsdale, NJ: Erlbaum; 1991b. p. 213-240.
- Nezlek JB. A multilevel framework for understanding relationships among traits, states, situations and behaviors. *Eur. J. Personal.* 2007; 21:789–810.
- Pearce N. The ecological fallacy strikes back. *J. Epidemiol. Community Health.* 2000; 54:326–327. [PubMed: 10814650]
- Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. *Curr. Opin. Psychiatry.* 2005; 18:189–193. [PubMed: 16639173]
- Raudenbush, SW. Toward a coherent framework for comparing trajectories of individual change. In: Collins, L.; Sayer, A., editors. *New Methods for the Analysis of Change.* Washington, DC: Am. Psychol. Assoc.; 2001a. p. 35-64.
- Raudenbush SW. Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annu. Rev. Psychol.* 2001b; 52:501–525. [PubMed: 11148315]
- Raudenbush, SW.; Bryk, AS. *Hierarchical Linear Models.* 2nd ed.. Thousand Oaks, CA: Sage; 2002.
- Raudenbush SW, Sampson RJ. Ecometrics: toward a science of assessing ecological settings, with application to the systematic observation of neighborhoods. *Sociol. Method.* 1999; 29:1–41.
- Raudenbush SW, Willms JD. The estimation of school effects. *J. Educ. Behav. Stat.* 1995; 20:307–335.
- Robinson WS. Ecological correlations and the behavior of individuals. *Am. Sociol. Rev.* 1950; 15:351–357.
- Roth S, Cohen LJ. Approach, avoidance, and coping with stress. *Am. Psychol.* 1986; 41:813–819. [PubMed: 3740641]
- Schaie KW. A general model for the study of developmental problems. *Psychol. Bull.* 1965; 64:92–107. [PubMed: 14320080]
- Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *Am. J. Public Health.* 1994; 84:819–824. [PubMed: 8179055]
- Selvin H. Durkheim's suicide and problems of empirical research. *Am. J. Sociol.* 1958; 63:607–619.

- Singer, JD.; Willett, JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. New York: Oxford Univ. Press; 2003.
- Walls, TA.; Schafer, JL. Models for Intensive Longitudinal Data. London: Oxford Univ. Press; 2006.
- Willett JB, Sayer AG. Using covariance structure analysis to detect correlates and predictors of individual change over time. Psychol. Bull. 1994; 116:363–381.

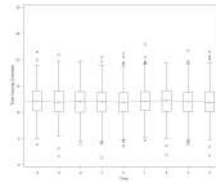


Figure 1.
The time-specific distributions of the TVC (z_{it}) for the first artificial data set.

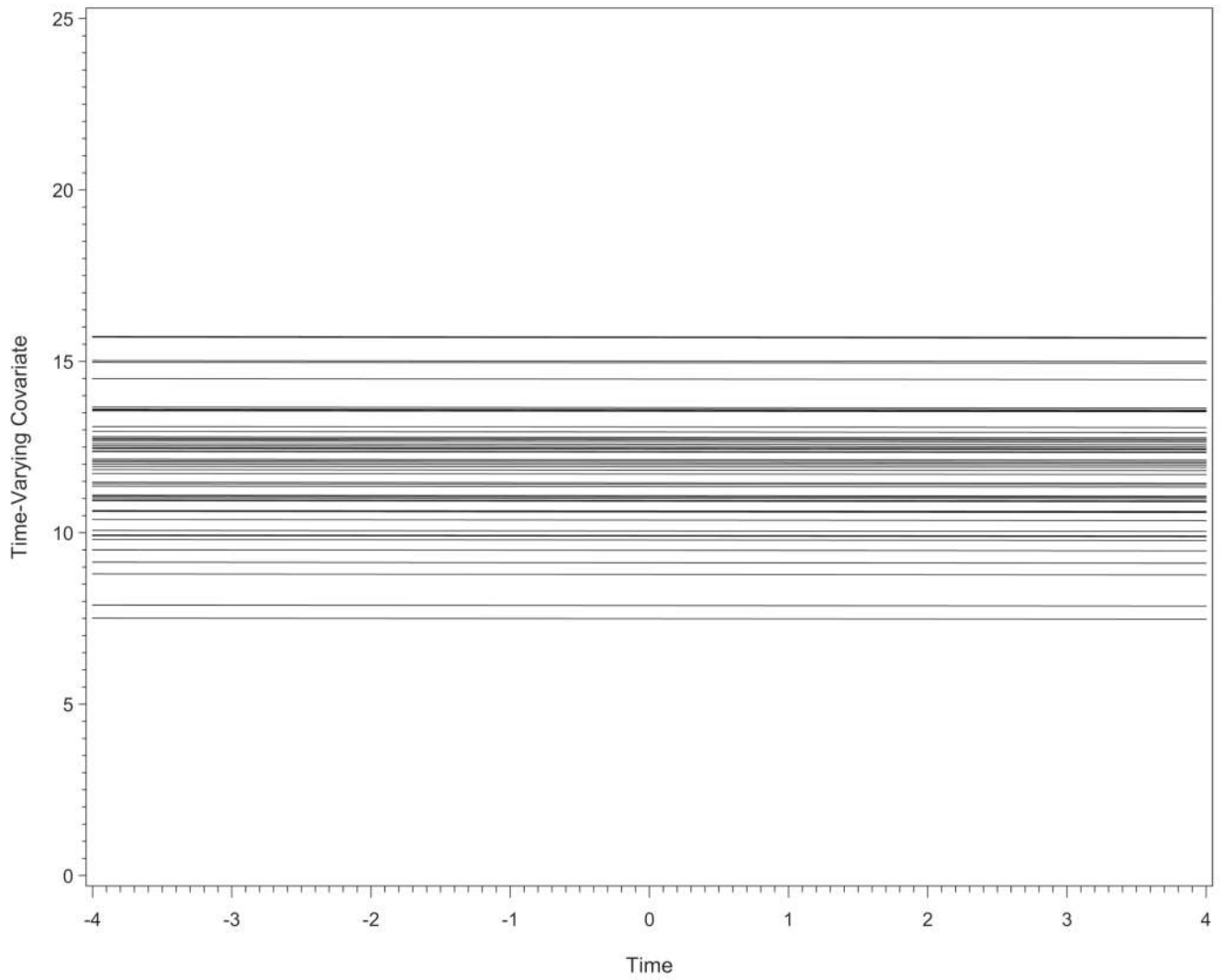


Figure 2. Model-implied growth trajectories for the TVC (z_{it}) over time for 50 randomly drawn observations from the first artificial data set.

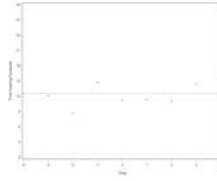


Figure 3. The time-specific values of the TVC over time for a randomly drawn case from the first artificial data set.

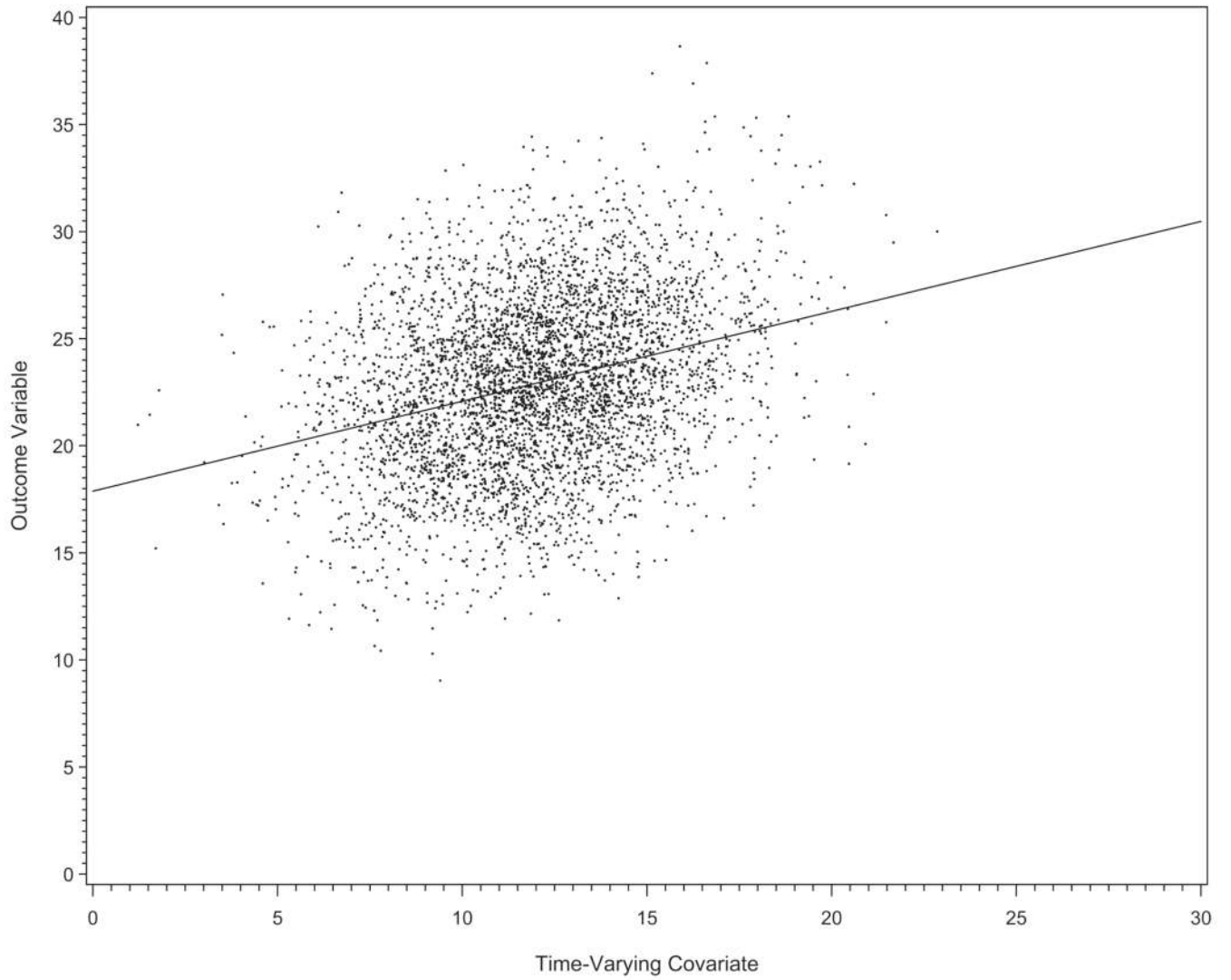


Figure 4. The bivariate distribution between the outcome (i.e., y_{it}) and the time-varying covariate (i.e., z_{it}).

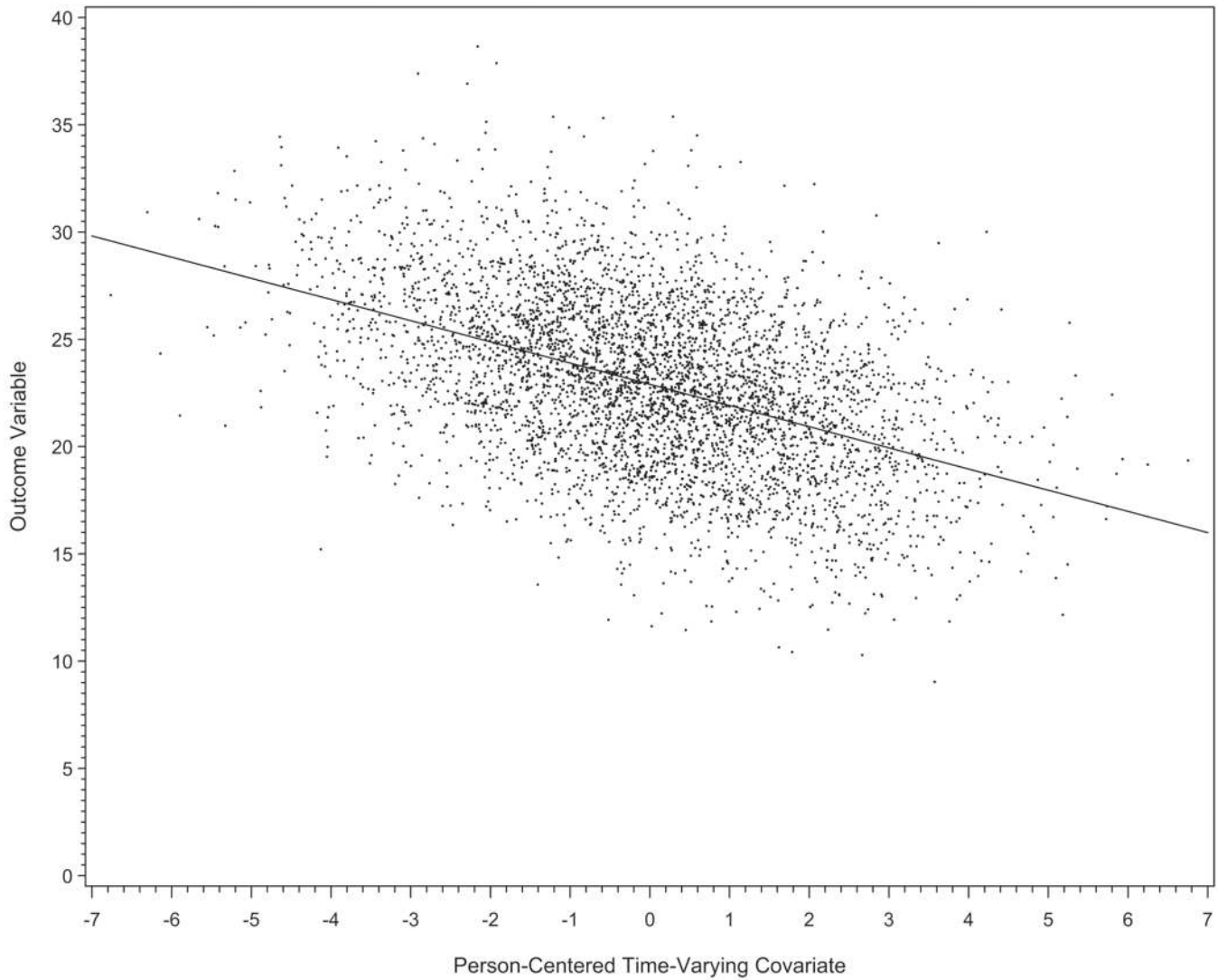


Figure 5. The bivariate distribution between the outcome (y_{it}) and the person-mean centered time-varying covariate (z_{it}).

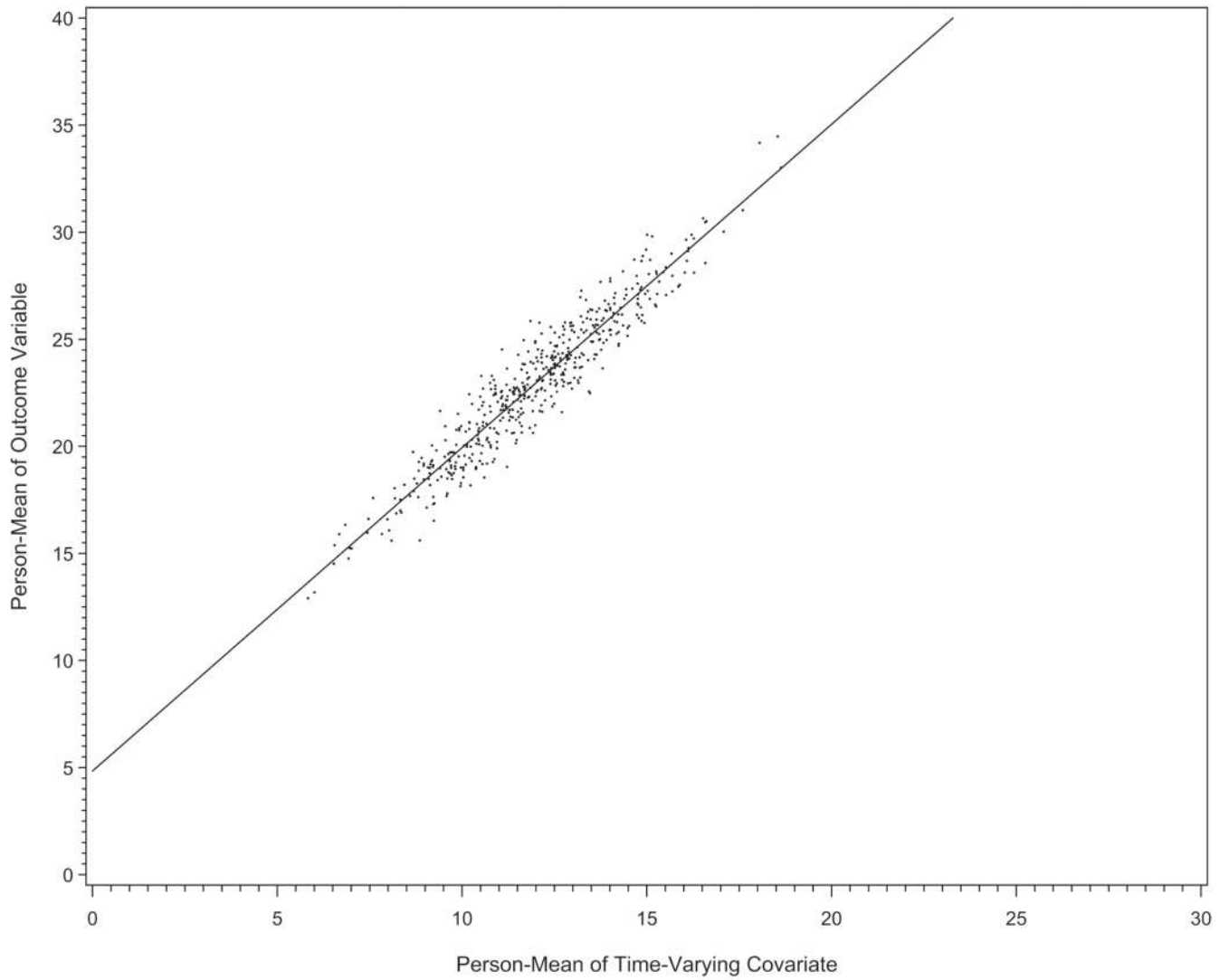


Figure 6. The bivariate distribution between the person-mean of the outcome (\bar{y}_i) and the person-mean of the time-varying covariate (\bar{z}_i).

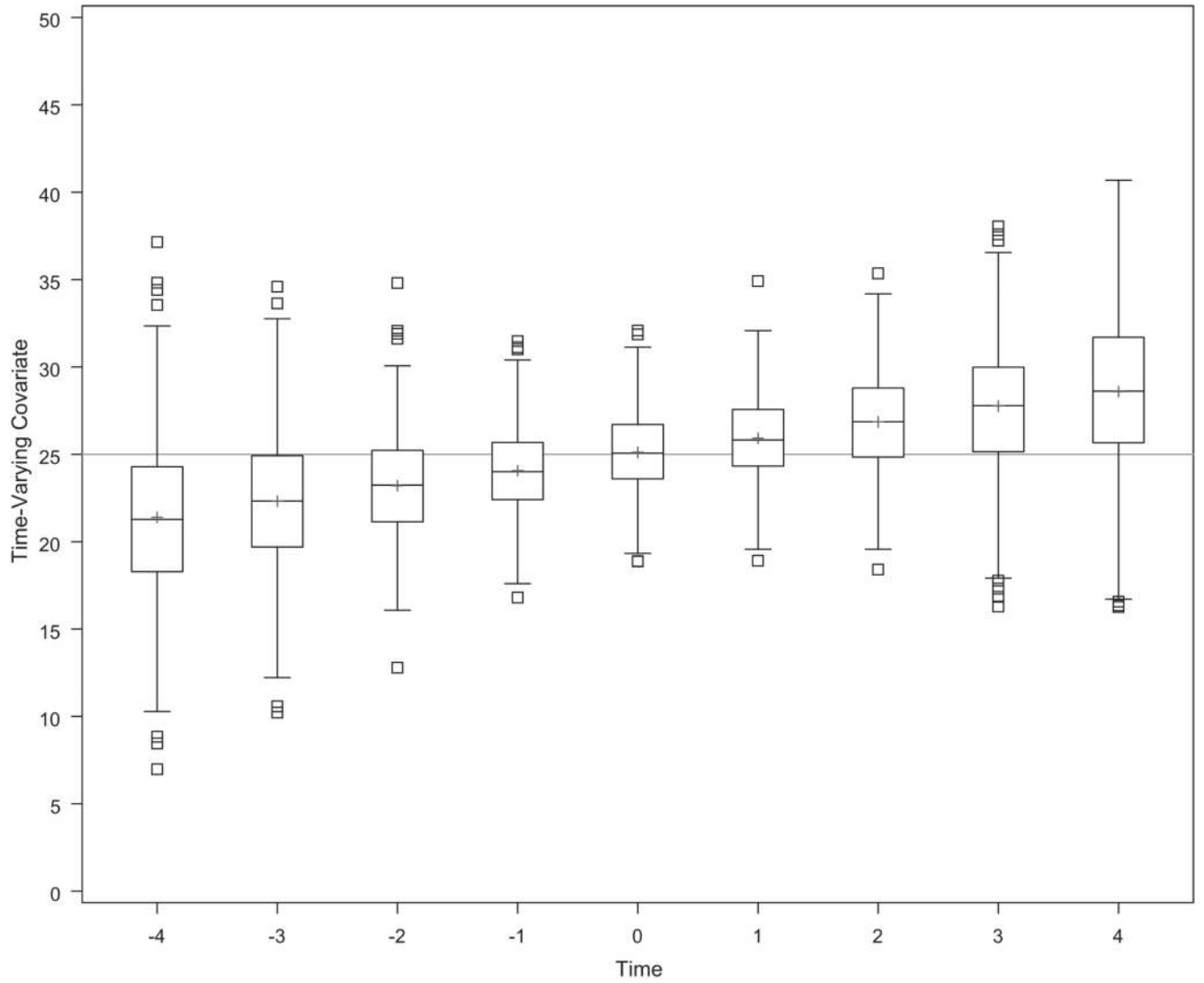


Figure 7. The time-specific distributions of the TVC (z_{it}) for the second artificial data set.

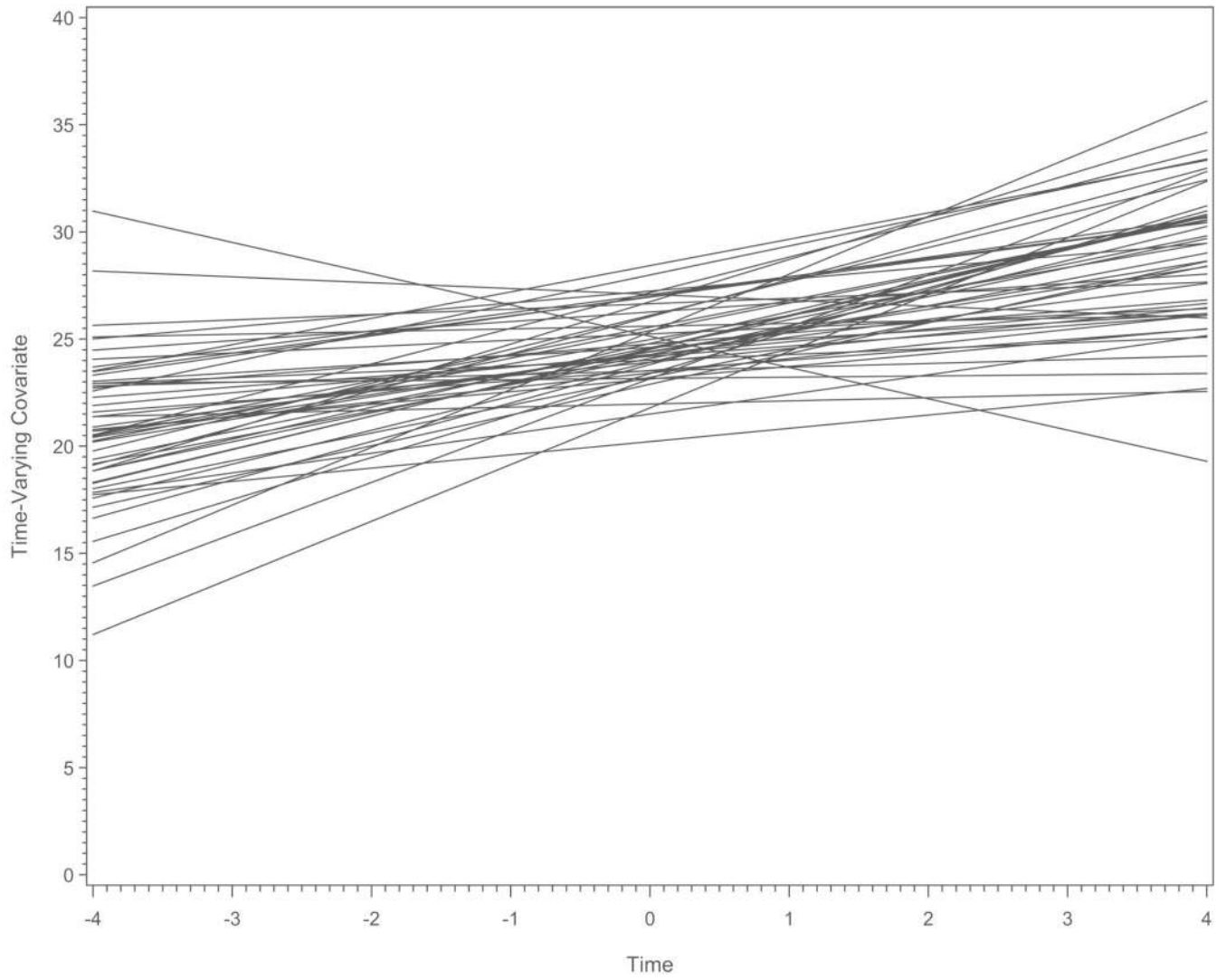


Figure 8. Model-implied growth trajectories for the TVC (i.e., z_{it}) over time for 50 randomly drawn observations from the second artificial data set.

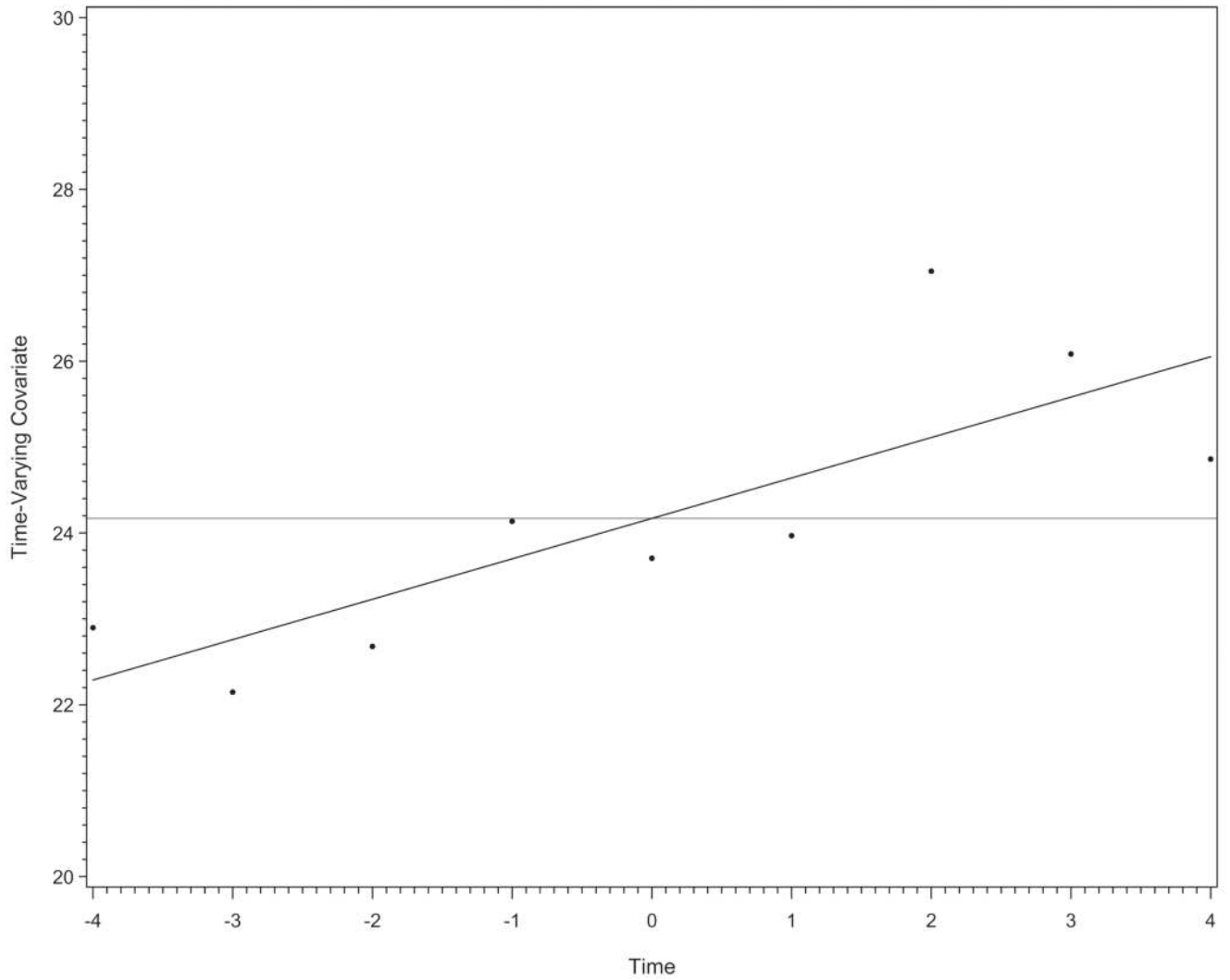


Figure 9. The time-specific values of the TVC over time for a randomly drawn case from the second artificial data set.

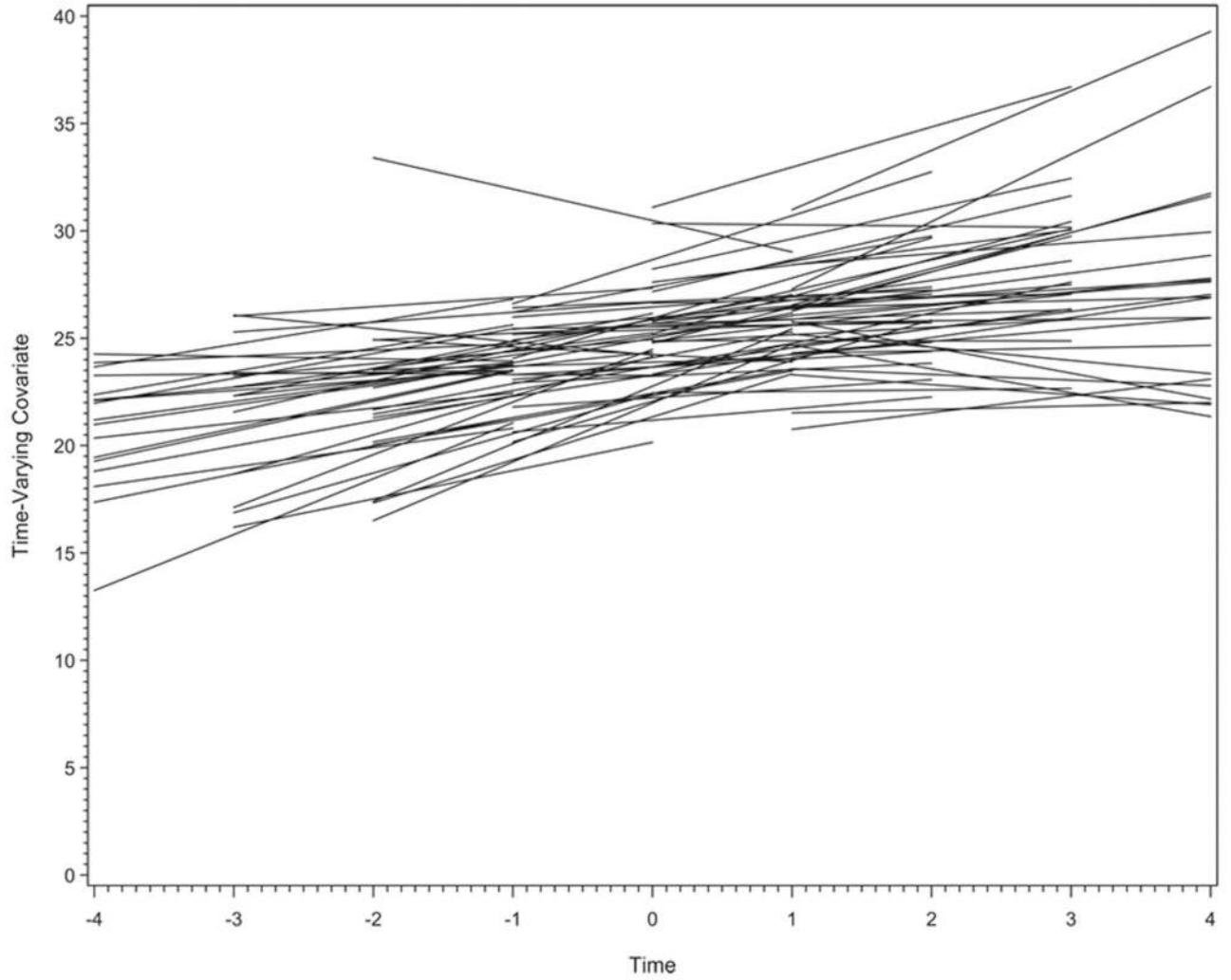


Figure 10. Model-implied growth trajectories for the TVC (i.e., z_{it}) over time for 50 randomly drawn observations from the third artificial data set.

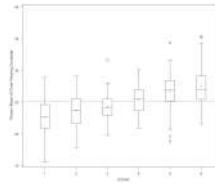


Figure 11. The cohort-specific distributions of the person-means of the time-varying covariate (i.e., \bar{z}_i) pooling over time and within cohort for the third artificial data set.