Journal of Consulting and Clinical Psychology

Dynamic Models of Individual Change in Psychotherapy Process Research

Fredrik Falkenström, Steven Finkel, Rolf Sandell, Julian A. Rubel, and Rolf Holmqvist

Online First Publication, April 10, 2017. http://dx.doi.org/10.1037/ccp0000203

Journal of Consulting and Clinical Psychology

0022-006X/17/$12.00 http://dx.doi.org/10.1037/ccp0000203

Dynamic Models of Individual Change in Psychotherapy Process Research

Fredrik Falkenström
Karolinska Institutet and Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden

Steven Finkel
University of Pittsburgh

Rolf Sandell
Lund University

Julian A. Rubel
University of Trier

Rolf Holmqvist
Linköping University

Objective: There is a need for rigorous methods to study the mechanisms that lead to individual-level change (i.e., process-outcome research). We argue that panel data (i.e., longitudinal study of a number of individuals) methods have 3 major advantages for psychotherapy researchers: (1) enabling microanalytic study of psychotherapeutic processes in a clinically intuitive way, (2) modeling lagged associations over time to ensure direction of causality, and (3) isolating within-patient changes over time from between-patient differences, thereby protecting against confounding influences because of the effects of unobserved stable attributes of individuals. However, dynamic panel data methods present a complex set of analytical challenges. We focus on 2 particular issues: (1) how long-term trajectories in the variables of interest over the study period should be handled, and (2) how the use of a lagged dependent variable as a predictor in regression-based dynamic panel models induces endogeneity (i.e., violation of independence between predictor and model error term) that must be taken into account in order to appropriately isolate within- and between-person effects. ***Method:*** An example from a study of working alliance in psychotherapy in primary care in Sweden is used to illustrate some of these analytic decisions and their impact on parameter estimates. ***Results:*** Estimates were strongly influenced by the way linear trajectories were handled; that is, whether variables were "detrended" or not. ***Conclusions:*** The issue of when detrending should be done is discussed, and recommendations for research are provided.

What is the public health significance of this article?
This article provides recommendations on how to study psychotherapy processes using dynamic panel data models to strengthen causal inferences. Accurate estimates of what drives individual development in psychotherapy are needed to generate recommendations on what therapists should focus on in therapy. Using the alliance-outcome association as an example, we show that estimated effect sizes may vary greatly depending on which modeling approach is used, with the decision on whether to remove time-trends from the outcome variable making the largest difference.

Keywords: panel data, structural equation modeling, cross-lagged panel model, mechanisms of change, process-outcome research

Supplemental materials: http://dx.doi.org/10.1037/ccp0000203.supp

Although psychotherapy research has provided compelling evidence for the average efficacy and effectiveness of psychological treatments, the factors that lead to—or prevent—individual change are still not well understood (Lambert, 2013). For instance, if patients receiving cognitive therapy (CT) for depression rate their symptoms as less severe posttreatment than patients ran-

Fredrik Falkenström, Centre for Psychotherapy Research and Education, Centre for Psychiatry Research, Department of Clinical Neuroscience (CNS), Karolinska Institutet, and Stockholm Health Care Services, Stockholm County Council, Sweden; Steven Finkel, Department of Political Science, University of Pittsburgh; Rolf Sandell, Department of Psychology, Lund University; Julian A. Rubel, Department of Psychology, University of Trier; Rolf Holmqvist, Department of Behavioral Sciences and Learning, Linköping University.

Correspondence concerning this article should be addressed to Fredrik Falkenström, Centre for Psychotherapy Research and Education, Centre for Psychiatry Research, Department of Clinical Neuroscience (CNS), Karolinska Institutet & Stockholm Health Care Services, Stockholm County Council, Lijeholmstorget 7, Hiss B, SE-117 63 Stockholm, Sweden. E-mail: fredrik.falkenstrom@ki.se

domly allocated to a waiting-list for the same time-period, we may conclude that CT is more effective than no treatment, but we still do not know what component(s) of CT that are responsible for this effect. It is possible to go on with experimental component designs, in which groups are randomized to versions of a treatment that differ only with respect to one specific component. However, experiments require a lot of time, money, and effort to achieve sufficient statistical power to find small to moderate group differences. Furthermore, there are many research questions in psychotherapy research that are difficult to test using experimental designs. Sometimes this is because of ethical problems; for example, in withholding active treatment to patients or deliberately providing inferior forms of hypothesized ingredients (e.g., a poor therapeutic relationship). Sometimes, too, it is impossible to credibly manipulate a certain variable, such as the competence or the conviction with which therapists deliver their interventions.

For these reasons, psychotherapy researchers also use nonexperimental designs for studying change mechanisms. The methodology used in the vast majority of process-outcome studies summarized in the sixth edition of Bergin and Garfield's *Handbook of Psychotherapy and Behavior Change* (Crits-Christoph, Connolly Gibbons, & Mukherjee, 2013) consisted of sampling one or a few sessions from each therapy (assuming that these are representative of the entire therapy process) and correlating the session process measure (or average of sessions if more than one session is sampled) with changes in symptoms over the course of the study. In a few studies (e.g., DeRubeis & Feeley, 1990), the session measures were correlated with changes in symptoms subsequent to the session in which process was measured.

It is well known that the major problem with these types of correlational designs is that correlation cannot be assumed to imply causation. Even if the process variable (e.g., the working alliance) is measured early in therapy and outcome is measured at the end of treatment, we cannot know whether better outcome is a result of a higher level of the process variable or whether both process and outcome are caused by some unmeasured patient characteristic or by prior improvement of symptoms. For example, more intelligent or self-aware patients may develop stronger worker alliances with therapists, while also being more likely to independently develop coping strategies for their depression over time. Any variable that is associated with both process and outcome variables could be a confounder that leads to a spurious association between process and outcome. However, using more intensive longitudinal designs, it is possible to eliminate some of these validity threats to correlational designs. These designs are typically based on either time-series analysis, which involves measuring a single case repeatedly, or panel data analysis, which involves measuring a number of cases repeatedly. During the last few years, researchers have begun to use longitudinal methods for sequential analysis relating processes to outcome (e.g., Falkenström, Granström, & Holmqvist, 2013; Hoffart, Øktedalen, Langkaas, & Wampold, 2013; Zilcha-Mano & Errázuriz, 2015).

In this article, we will outline the advantages and complexities associated with psychotherapy process-outcome research using panel data. We will focus in particular on so-called "dynamic" panel models in which time-lagged associations, including the effect of the dependent variable on itself over time, are modeled explicitly, arguing that this class of models is most suitable to psychotherapy process data and is also strongest for drawing causal conclusions. We are most concerned with the increasingly common situation with session-wise measurements of outcome and one or more psychotherapy processes in relatively large samples undergoing brief psychotherapy (up to a maximum of around 15–20 sessions). The analysis of panels with more than 20 timepoints, and with the possibility of small *N*, has been discussed elsewhere (e.g., Ramseyer, Kupper, Caspar, Znoj, & Tschacher, 2014) and will not be covered here.

Advantages of Panel Data

Panel data offer three major advantages for psychotherapy process-outcome research: (1) allowing the microanalysis of processes as they occur during sessions, which is intuitively appealing to practitioners, (2) modeling lagged associations over time to ensure direction of causality, and (3) isolating within-patient changes over time from stable between-patient differences to protect against confounding influences because of the effects of unobserved stable attributes of individuals.

Microanalytic Studies of Processes and Mechanisms of Change

An advantage of process-outcome research is that it becomes possible to do more fine-grained studies of change processes that are intuitively appealing to clinicians. Change is often conceptualized as relatively long-term change; that is, change from pre- to posttherapy or as a trajectory across the whole treatment. However, change can also be conceptualized in the short-term; as change from one session to the next or even as change from one moment to the next within a session. Research on long-term change is important for political decisions, such as what therapy forms should be provided within state-financed health care or what therapy to offer to which patient (Delgadillo, Moreea, & Lutz, 2016). However, if the researcher is interested in understanding how to actually conduct therapy, this level of analysis is often too coarse to provide useful information. For example, the question of whether CT for depression is associated with better outcome than nonspecific supportive therapy is not very informative for clinicians practicing CT wanting to get research-informed advice on how to improve their practice. For this reason, clinicians are often more interested in processes as they occur during sessions and how these are related to change in the short-term, which in turn builds up to produce overall outcome. We believe that the closer research gets to the actual behaviors and experiences of therapists and patients during sessions, and the effect therapist behaviors have on patient experience, the easier it is for practitioners to understand the relevance of research. For example, Hoffart (2016) showed with this microanalytic approach that anxiety in panic disorder is maintained by patients' catastrophic cognitions regarding bodily sensations (e.g., "My heart is pumping too fast, I am going to have a heart attack") and not by patients' beliefs of being unable to effectively engage in coping behaviors (self-efficacy). Given their nature as within-patient associations, these results have direct implications for the treatment of patients with panic disorder and emphasize the importance of targeting catastrophic cognitions compared with self-efficacy.

Time-Lagged Associations Ensure Direction of Causality

A distinct advantage of longitudinal research in general is the possibility of testing for causal direction. A general requirement for causal interpretation is that cause should precede effect, and this is not possible to ensure in cross-sectional designs. Models may be estimated with time-lagged associations between variables, thus ensuring that X at one point in time (cause) precedes outcome Y (effect) at some future point. Moreover, in so-called cross-lagged models, in which time-lagged effects of two (or more) variables are tested simultaneously, it is possible to test the (time-lagged) effect of one variable while holding the (time-lagged) effect of the other variable constant, thereby strengthening conclusions about which variable "drives" development of the other (Finkel, 1995).

Isolating the Within-Person Variance Over Time

When doing nonexperimental process-outcome research, our analyses may suffer from omitted variable bias that can seriously distort findings and lead to incorrect causal inferences, conclusions and recommendations. Omitted variable bias is a term referring to the problem of unmeasured "third" variables that are causally related to the outcome variable, and also correlated with the predictor. The predictor is thus confounded with this third variable, and the estimated relationship between predictor and outcome is spurious. Psychotherapy process researchers have tried to cope with this by measuring probable confounders and adjusting analyses for their effects, but the risk that there is some important covariate that is unknown to the researcher is always present and limits the confidence of causal conclusions. However, using repeated measures/panel data designs, it is possible to rule out one class of omitted variables, namely ones that are stable over time (e.g., personality traits, intelligence). This is because of the possibility of isolating the within-patient overtime variance in the variables studied from between-patient differences, making the within-patient process independent of any stable confounder at the patient level. Importantly, among between-patient differences, any higher-level factors are included such as therapist differences (Baldwin & Imel, 2013) or organizational effects such as organizational climate or culture (Falkenström, Grant, & Holmqvist, 2016). These higher-level factors may moderate within-patient effects, but they cannot confound (i.e., explain) them, which is one of the most important advantages of the panel design.

For instance, if a researcher is studying positive emotions in therapy session-by-session, the "within" part will consist of the relative positivity in a particular session compared with other sessions of the particular patient, while the "between" part will consist of the average positivity of one patient compared with the other patients in the dataset. The within-patient effect is thus concerned with changes in positive emotions from session-to-session, while the between-patient effect is about how much positive emotions patients experience on average during therapy, relative to one another. If the researcher is interested in studying how change in positive emotions during therapy impacts outcome, it is imperative to disentangle the variance in session-to-session changes "within" individuals from the variation between persons, which may be caused by stable "between" person-level characteristics such as differential temperament or personality traits, or higher-level variables such as differential therapist effectiveness or organizational climate. This way, stable confounders are ruled out, even if not measured and included in the analysis.

Fixed or random effects? The oldest approach to disaggregate within- and between-person variance is the Least Squares Dummy Variable (LSDV) regression model, in which the between-patient effect is represented by *N* dummy variables (one for each patient in the sample, or N-1 if there is an overall intercept in the model)

$$Y_{i,t} = \beta_{0i} + \beta_1 X_{i,t} + e_{i,t}. \quad (1)$$

where $Y_{i,t}$ is the outcome variable for individual i at time t, β_{0i} is the person-specific effect which in this case is represented by N dummy variables, $X_{i,t}$ is the time-varying predictor (or a vector of predictors), and $e_{i,t}$ is the error term. The model may be estimated by ordinary least squares regression (OLS). Because this model may be computationally burdensome if there are many individuals in the dataset, it is usually estimated by centering all variables on each participant's own mean; that is, subtracting the respective individual's mean value across all time-points from each session-specific score of that person:

$$Y_{i,t} - \bar{Y}_i = (\beta_{0i} - \bar{\beta}_{0i}) + \beta_1(X_{i,t} - \bar{X}_i) + (e_{i,t} - \bar{e}_{i,t}). \quad (2)$$

Because β_{0i} is a constant, $\bar{\beta}_{0i} = \beta_{0i}$ and thus this term is eliminated. This method, referred to as the "Fixed Effects" model in the econometrics literature, yields the same estimates as the LSDV method and will effectively isolate the within-person variance from the between-person component because β_{0i} is eliminated via the centering process. The downside of this method is that the researcher cannot incorporate any stable between-person predictors in the analysis, because there is no between-person variation left.[1]

In psychotherapy research, random effects, multilevel, or mixed models (Raudenbush & Bryk, 2002; Snijders & Bosker, 2011) have become state-of-the-art for the last decade or so. This is because psychotherapy data is almost invariably hierarchical, with repeated measures being nested within therapies/patients, and patients often nested within therapists. This violates the independence assumptions of most conventional statistical models (e.g., linear regression analysis), with attendant risk for increased Type-I error rate. Multilevel models accommodate such data structures by estimating one or more additional variance components for each level of nesting, thereby explicitly modeling the nested structure. However, it seems to be less well known among psychotherapy researchers that there are other ways than adding random effects for protecting against the increased Type-I error rate associated with ignoring nesting. Because all between-patient variation is removed in the fixed effects model, the estimated coefficients of this model are not affected by higher-order nesting (e.g., within therapists).[2] However, if the researcher is interested in *modeling* variables at higher levels (e.g., therapist effects) rather than just controlling for them, random effects are needed. There is also the

[1] The fixed effects model can, however, incorporate between-person predictors as moderators of within-person effects.

[2] Still, higher-order nesting may affect *SE*s of model coefficients, but these can be adjusted by using cluster-robust *SE*s (Huber, 1967).

possibility that *relationships among variables* at the session-to-session level vary at higher levels, so that treatment mechanisms are different for different patients. To study these issues, random effects models provide greater flexibility in model specification than the LSDV or the more general person-mean centered "Fixed Effects" model.

In the random effects model, differences between units (e.g., patients in a repeated-measures framework) are modeled by a single variance component, while in the Fixed Effects model each unit has its own fixed parameter (similar to analysis of variance [ANOVA]). Thus, β_{0i} in Equation 1 is a between-patient variance component (a "random intercept"), rather than *N* dummy variables as in the fixed effects model. Comparing these models, it can easily be seen that the random effects model is more parsimonious (and statistically "efficient"), in that only one extra parameter is estimated while in the fixed effects model the number of parameters is equal to the number of units. However, the increased parsimony of the random effects model comes with the price of stronger assumptions, namely that the differences among units has a known distribution of some form (usually normal), and that there is zero correlation between the random effect and lower-level predictors. The latter assumption is the most potentially problematic, in that the unmodified random effects model assumes away the possibility that between-patient differences in the dependent variable (which comprise the random effect) are related to between-patient differences in the time-varying predictor. For instance, if the researcher is predicting symptoms of depression by prior negative thoughts, it is highly likely that average between-patient differences in depressive symptoms will be correlated with between-patient differences in negative thinking.

The hybrid random effects model. A way of dealing with this problem is referred to as the "hybrid random effects" model (Firebaugh, Warner, & Massoglia, 2013). Here, predictors are person-mean centered as in the fixed effects model to remove between-person variance, while the between-person component of the dependent variable is handled by a random intercept. In this model, it is possible to add the person-mean of the time-varying covariate (i.e., $\bar{X}_i$) as an additional predictor in order to estimate the between-person effect on outcome:

$$\begin{aligned} &\text{Level 1 model:} && Y_{i,t} = \beta_{0i} + \beta_1(X_{i,t} - \bar{X}_i) + e_{i,t} \\ &\text{Level 2 model:} && \beta_{0i} = \gamma_{00} + \gamma_{01}\bar{X}_i + u_{0i} \\ & && \beta_1 = \gamma_{10}. \\ &\text{Mixed model:} && Y_{i,t} = \gamma_{00} + \gamma_{01}\bar{X}_i + \gamma_{10}(X_{i,t} - \bar{X}_i) + u_{0i} + e_{i,t}. \end{aligned} \tag{3}$$

The hybrid alternative is a widely utilized model in recent studies of within-patient processes in psychotherapy (e.g., Curran & Bauer, 2011; Hoffart, 2016; Rubel, Rosenbaum, & Lutz, 2017). It is also widely used in studies disaggregating effects at other levels, for example, between therapists and within therapist (patient) levels (Baldwin, Wampold, & Imel, 2007). The key advantage, as can be seen from the "mixed model" formulation of Equation 3, is the simultaneous estimation of within-patient effects that match what would be obtained with fixed effects methods, while estimating between-patient effects in the same model. In addition, these models have the capacity to test whether within-patient effects vary between higher-level units (e.g., patients, therapists, clinics; Baird & Maxwell, 2016). An example is provided in a study of the relationship between quality of working alliance and next-session symptom change, in which it was shown that this relationship was stronger for some patients than for others and that this variation was predicted by differential levels of personality problems among patients (Falkenström et al., 2013):

$$\begin{aligned} &\text{Level 1 model:} && Y_{i,t} = \beta_{0i} + \beta_{1i}(X_{i,t} - \bar{X}_i) + e_{i,t} \\ &\text{Level 2 model:} && \beta_{0i} = \gamma_{00} + \gamma_{01}\bar{X}_i + u_{0i} \\ & && \beta_{1i} = \gamma_{10} + u_{1i}. \end{aligned} \tag{4}$$

The only difference between Equations 3 and 4 is the addition in equation 4 of u_{1i}, which estimates the variation in the within-patient effect β_1 around its grand mean γ_{10}. This is an important extension of the random effects model, because it is reasonable to think that the effect of psychotherapy processes is not uniform across patients. It is also possible to add Level 2 predictor(s) of the variance in β_1 (e.g., $\beta_{1i} = \gamma_{10} + \gamma_{11}Z_i + u_{1i}$). In addition, this formulation can easily accommodate higher-level effects, for example, therapist effects (Baldwin & Imel, 2013), such that the patient-level intercept (and/or slopes) at Level 2 itself can be predicted by a "Level 3" dummy variable for each therapist or a random therapist effect, and/or by a series of "Level 3" variables on which therapists themselves may differ (competence, gender, etc.).

Complications With Panel Data

The preceding section has clarified some distinct advantages of dynamic panel data models. The following section outlines some of the complexities with such models. We focus in particular on two issues: (1) how to handle long-term trajectories over the study period, and (2) how to estimate dynamic models while simultaneously disaggregating within- from between-patient effects.

Stationarity Considerations in Panel Data Models

Because many panel models have derived from the time-series literature, some of the same statistical assumptions should apply for the analysis of the relatively short-term repeated measurements that comprise panel data. One such assumption often mentioned (but seldom explained) is the stationarity assumption(s) (Wooldridge, 2012). According to the "covariance stationarity" assumption, the mean and variance of a variable should be stable over time, and the correlation between two measurements at different time-points should only depend on the distance between them. The different aspects of (non-) stationarity have different implications. Nonstationarity of *variances* (i.e., the diagonal elements of the covariance matrix) is a well-known phenomenon in the statistical literature called heteroscedasticity (in this case longitudinal heteroscedasticity, as opposed to cross-sectional heteroscedasticity which concerns different variances among persons). If the model used does not take heteroscedasticity into account, the estimated residuals will also be heteroscedastic. This may result in incorrect standard errors and consequently biased statistical significance tests. However, with many statistical methods (e.g., structural equation modeling [SEM] or mixed models incorporated in SAS, Stata, SPSS, etc.) it is possible to estimate different residual variances at different occasions, thus modeling heteroscedasticity explicitly.

If the *covariances* between a time-varying predictor and the outcome (i.e., the off-diagonal elements of the covariance matrix) are not stable over time, but the model estimated assumes stability (i.e., only one covariance is estimated between the two variables for the whole time-period) the estimated coefficient may be misleading. For example, *structure* may be important early in psychotherapy, more or less independently of therapy orientation, while later in treatment it may be less important or even detrimental (e.g., by hindering patient autonomy). If a model assuming stationarity of covariances is applied, the resulting coefficient will reflect a compromise between the covariances among measures in the early sessions and the (lack of) covariances among measures in the latter sessions. The researcher may erroneously conclude that there is a small effect of structure on symptom change throughout therapy, while the true effect is large in the early sessions and small or negative later in treatment. In this example, violation of the assumption of stationarity of covariances has problematic consequences. In other applications, however, such change in the effect of the time-varying predictor over time may be nonexistent or statistically negligible, in which case the model would satisfy the classical stationarity assumptions. Importantly, differences in the size of the time-varying predictors can be accommodated (and tested) relatively easily in the SEMs we present below.

However, the most difficult issue, especially in the context of treatment studies, is with *changes in means* over time. It is straightforward to show that two completely unrelated variables become strongly correlated if a common time-trend is added to them, resulting in spurious correlation (Granger & Newbold, 1974). For example, when studying the effect of changes in positive emotions on symptoms during psychotherapy, the researcher may suspect that between-person differences in intelligence, or a personality disposition such as openness, cause not just average between-patient differences in positive emotions and symptoms but also *continuous changes* in positive emotions and symptoms over time. In this case, intelligence/openness will act as a confounder of the relationship between positivity and symptoms. One method of protecting against the phenomenon of spurious correlation in time-series analysis is to remove time-trends in the data (what is referred to as "detrending"). When doing so, it becomes possible to control for case-specific patterns over time. Psychology researchers often use some form of detrending, either manually by using (patient-specific) residuals from the separate regression of each variable on a grand-mean centered continuous time variable (Curran & Bauer, 2011), by entering a time variable as a random covariate in a traditional multilevel random intercept model (Wang & Maxwell, 2015), or by adding a latent (random) time-trend to a longitudinal SEM (Bollen & Curran, 2004). These methods all have the effect of removing individual time-trends, so that what is left to analyze are (assumed to be) stationary deviations from the trend.

Detrending or Not Detrending?

Detrending is sometimes described as "controlling for the effect of time," an expression that seems to imply that time has a causal effect of its own on the studied phenomena. For example, in an uncontrolled observational study of some psychotherapeutic treatment, time-trends may represent either so-called "spontaneous recovery" or a genuine treatment effect, or both. Although spontaneous recovery may be thought of as merely "caused" by the passage of time, it is more likely that it represents other factors that require time to have effect, for example, internal healing processes of the body/mind or external changes (friends intervening, finding a romantic partner, etc.). To the extent that the latter causes of the time-trend in depressive symptoms are also correlated with time-trends in a predictor such as, for example, the working alliance between patient and therapist, these causes will act as confounders of whatever alliance effect may be estimated in the analysis.

Detrending will effectively protect against unobserved confounders that are correlated with a linear (or curvilinear) time variable, as in the previous example with intelligence/openness being confounded with positive emotions predicting symptoms. However, as noted in the example with spontaneous recovery and working alliance, if the time-trend is a mix of spontaneous recovery and treatment effect, *removing time-trends will remove both, including the very effect that the researcher is interested in explaining* (Wang & Maxwell, 2015). Effect sizes from such an analysis are then likely to be suppressed. In the context of psychotherapy process-outcome research, what may be most important to protect against is spontaneous recovery. In a randomized trial, spontaneous recovery is best controlled for by including an untreated control group. If an untreated control group is used, it is possible (and advisable) to detrend against the average trajectory of that control group. However, for ethical reasons it is often problematic to include untreated control groups when effective treatments are known to exist. Furthermore, even if it were possible to observe an untreated control group over time, many if not most process variables cannot be observed in such control groups (e.g., there will be no working alliance without a therapeutic dyad). For this reason, researchers need to think carefully about whether and how detrending should be done in a given research context.

Figure 1 shows the problem of choosing whether to detrend or not in the form of a path diagram. If the correlation between the time-trends of X and Y is caused by an unobserved confounder (as in Figure 1), detrending should be done. However, if the phenomenon under study is such that the effect of X on Y is causing the time-trends in both variables, detrending will result in statistical overcontrol. For example, if a better quality of patient-therapist working alliance causes reduction in depressive symptoms by the following session and reduction of depressive symptoms causes improvement of working alliance (e.g., Falkenström, Ekeblad, & Holmqvist, 2016), then in successful therapy cases this bidirec-

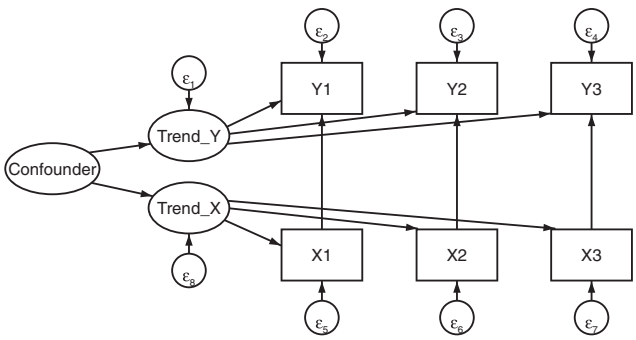

Figure 1. Path diagram showing linear time-trend as a proxy for unmeasured confounder(s).

tional influence will result in gradual decrease of symptoms over time together with gradual improvement of the working alliance, and vice versa in unsuccessful ones. This means that there will be correlated trajectories that are caused by the studied phenomenon rather than by a confounder, and detrending will result in an inability to estimate some or all of the effects in which the researcher is interested. The problem is knowing if there are confounders associated with the time-trends, which is more of a theoretical than a statistical problem. We consider these issues in more depth with an empirical example in Data Analytic Example.

Complications in Estimating Dynamic Panel Models While Controlling for Unobserved Person-Level Confounders

The simplest panel data model is the static model (see Figure 2a). In this model, the time-varying predictor X is assumed to 1) affect the dependent variable Y instantaneously, and 2) there is no indirect effect of X_t on Y_{t+1} via Y_t, which means that the effect of X_t on Y_t will disappear immediately at time t + 1. We argue that in psychotherapy this model is usually unrealistic, in that the effects in which we are most interested do not have instantaneous impact and then dissipate immediately. Rather, there are usually both lagged effects that cumulate over time, and effects that linger for some time period after their initial introduction. For example, coping skills may need to be practiced for some time before a change in symptoms can be observed, but after that has happened the effect is not likely to disappear immediately until coping skills are improved further. Figure 2b shows a time-lagged model, in which the effect of X on Y is lagged. However, this effect is still assumed to vanish immediately at the next measurement. If it is likely that the effect of X on Y will linger on in future time periods, the effect of Y on future values of itself (i.e., Y_{t-1}, Y_{t-2}, etc. on Y_t) may be included in the model in order to take this into account. Such a dynamic model is depicted graphically in Figure 2c.

State-dependence versus unobserved heterogeneity. When estimating dynamic panel models, it becomes possible to distinguish *state-dependence*; that is, the value of the outcome at a particular time-point depends on its own value at the previous time-point, from *unobserved heterogeneity*, which is the dependence of all measurements on an unobserved factor (or factors) affecting all observations. The difference here is in the causal assumptions implied by the two models. As an example, if the outcome were depression scores measured repeatedly over time, state-dependence would imply that depression at a particular time-point has a causal effect on depression at the next time-point. In accordance with this assumption is the behavioral model of depression (Jacobson, Martell, & Dimidjian, 2001): The depressed person becomes passive and avoidant, which means reduced positive reinforcement (i.e., fewer stimulating activities), which in turn increases the likelihood of being depressed at the next measurement. In this case, there would thus be a causal effect of depression at one time-point on depression at the next.[3] On the other hand, unobserved heterogeneity would be exemplified by a biomedical explanation for depression, in which case depression at all measurement occasions is caused by the same underlying disease process. Statistically, state-dependence is modeled via an autoregressive ("lagged dependent variable") model (Figure 3a), while unobserved heterogeneity is modeled using a random intercept or fixed effects/LSDV model (Figure 3b). In the latter, the (random) intercept represents the underlying stable process that affects all measurements equally.

Endogeneity problems in the dynamic panel model. Autoregressive and fixed effects/random intercept models can also be combined in the same model and tested simultaneously (Figure 3c) in what econometricians have formally labeled the "dynamic panel model" (Nickell, 1981). However, there is a well-known problem in the estimation of the dynamic panel model: Because the intercept is a direct cause of Y at every point in time, there is an intrinsic correlation between lagged Y (Y_{t-1}) and the composite residual (i.e., the random intercept and the person-time error term) of its equation. This results in what is known as "endogeneity," violating one of the basic assumptions of regression analysis (Baltagi, 2013), and leading to potentially massive bias in the estimation of model parameters. It may be surmised that person-mean centering lagged Y could alleviate this problem, in that the "between" component of lagged Y, which correlates with the person-level intercept, would thus be removed from consideration. However, in this case $Y_{i,t-1} - \bar{Y}_i$ is still intrinsically related to its equation's (person-mean centered) error term because some elements of the average time-specific error term are also determinants of the average lagged Y.[4] For this reason, dynamic panel models—even those which isolate within-person processes via person-mean centering or dummy variable methods—suffer *by construction* from endogeneity problems related to correlations between predictors and the error term of outcome equations. The result of endogeneity tends to be that the effect of the lagged dependent variable is estimated as too large, and the effect of other regressors as too small (Allison, 2015). Because of endogeneity, the general recommendation is to only use a lagged dependent variable as predictor if there are strong theoretical reasons to assume a causal influence of the outcome variable on future values of itself (Rabe-Hesketh & Skrondal, 2012). Whenever a dynamic panel model is adopted, however, the endogeneity problem necessitates alternative estimation procedures to the fixed/random effects (or hybrid) models that have been described thus far. There are two known solutions; one is the econometric solution called instrumental variable regression, and the other is to estimate the between-person component and its relationship with lagged Y using latent variable modeling in a SEM framework. Because instrumental variable regression is highly complex, is not commonly used in psychotherapy research, and recent research (Moral-Benito, Allison, & Williams, 2016) has shown that SEM is superior in almost all respects, we will not discuss this method further. For a description and psychotherapy

[3] Recent psychological work also conceptualizes the *size* of the state dependence/autoregressive parameter as an indicator of "inertia" or "regulatory weakness," with higher values meaning that individuals are slow to return to their equilibrium psychological state after an external shock or treatment (Jongerling, Laurenceau, & Hamaker, 2015; Kuppens, Allen, & Sheeber, 2010).

[4] This can be seen by taking the dynamic panel: $Y_{i,t} = \beta_{0i} + \beta_1 X_{i,t} + \beta_2 Y_{i,t-1} + e_{i,t}$ and subtracting the "between" equation from it to yield: $Y_{i,t} - \bar{Y}_i = (\beta_{0i} - \bar{\beta}_{0i}) + \beta_1(X_{i,t} - \bar{X}_i) + \beta_2(Y_{i,t-1} - \bar{Y}_i) + (e_{i,t} - \bar{e}_i)$. Although endogeneity because of the possible relationship between β_{0i} and $X_{i,t}$ has now been eliminated from consideration (because $\beta_{0i} = \bar{\beta}_{0i}$ and hence drops out of the model), the "demeaned" lagged Y term ($Y_{i,t-1} - \bar{Y}_i$) is still related to the demeaned error term ($e_{i,t} - \bar{e}_{i,t}$), given that some component of $\bar{e}_i$ is related to $\bar{Y}_i$.

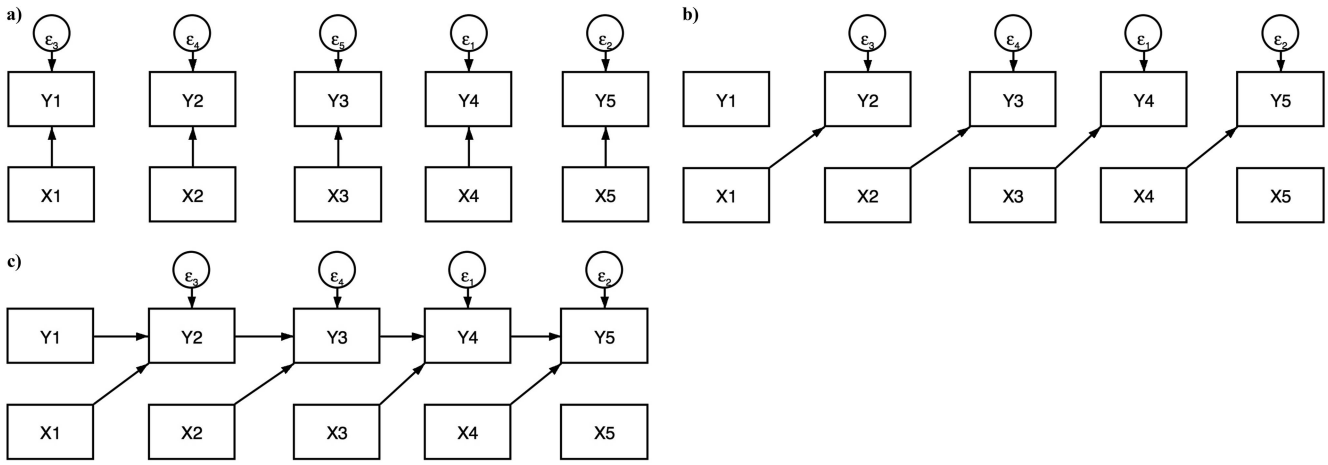

Figure 2. Types of panel data models: (a) static model, (b) time-lagged model, and (c) dynamic model.

research application of this method, see Falkenström, Ekeblad, et al. (2016).

SEM. Allison and his colleagues (Allison, 2016; Moral-Benito, 2013; Williams, Allison, & Moral-Benito, 2015) have noted that the dynamic panel model may be estimated in a SEM framework, where the autoregressive effect of one variable on later versions of itself can be modeled, along with unobserved heterogeneity and free of endogeneity bias, without having to resort to the complex instrumental variable approach. This is accomplished by including the random intercept as a latent variable that "causes" Y at each point in time, thus explicitly accounting for the potential correlation between lagged Y (i.e., Y_{t-1}) and the error term within the SEM framework (Figure 4). In addition, SEM has several other attractive features, such as the possibility of estimating cross-lagged relationships within the same model (making possible inference about which variable "leads" development), and the

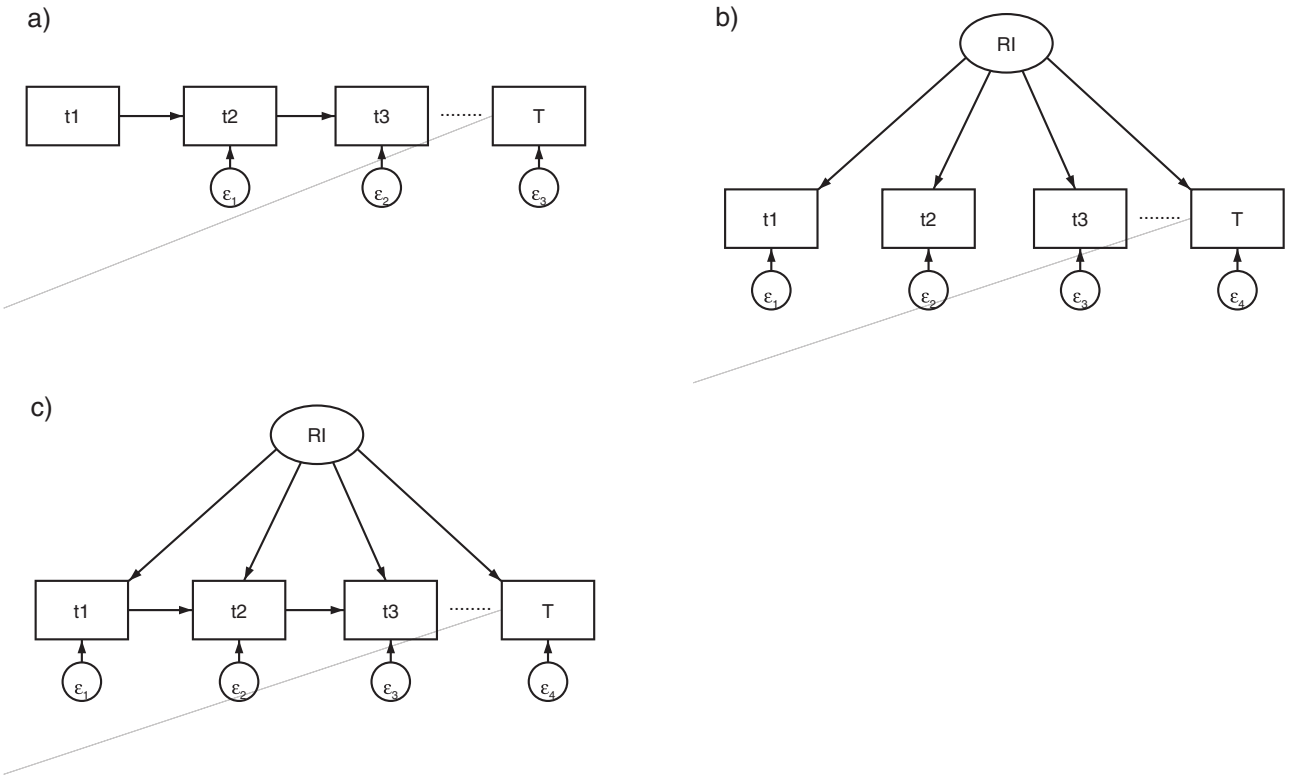

Figure 3. Path diagrams of (a) autoregressive process modeling state dependence, (b) random intercept model of unobserved heterogeneity, and (c) combined state-dependence and unobserved heterogeneity model.

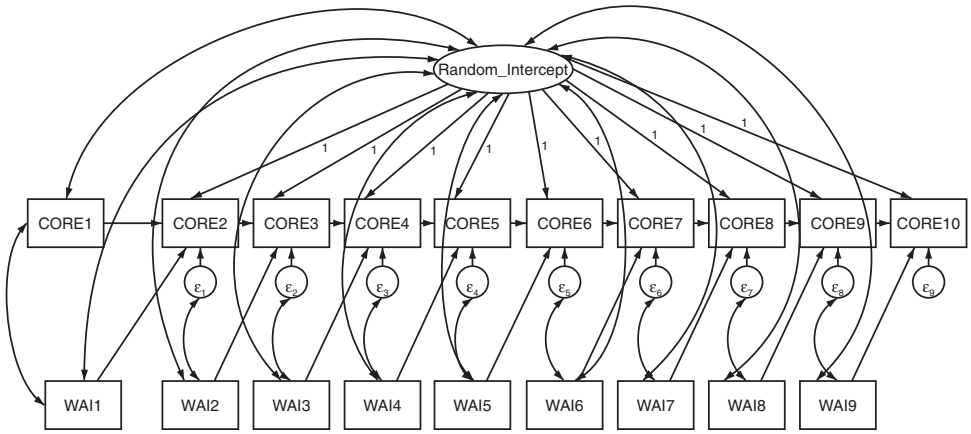

Figure 4. Paul Allison's Maximum Likelihood Structural Equation Model (standard version, not predetermined).

possibility of including between-person as well as within-person predictors. An example of an SEM for the analysis of within-patient change in psychotherapy is presented in the data analytic example below.

Data Analytic Example

The considerations discussed above will be illustrated using data collected on patients (total $N = 1{,}095$) undergoing counseling and psychotherapy in Swedish primary care. This dataset was used to study the dynamic relationship between working alliance and symptoms/well-being (Falkenström et al., 2013). Symptoms/well-being was measured before each session using the Clinical Outcome in Routine Evaluation—Outcome Measure (CORE-OM; Evans et al., 2002), and quality of working alliance was rated by the patients after each session using the Working Alliance Inventory—Short-Form Revised (WAI-SR; Hatcher & Gillaspy, 2006). The patients were a mixed diagnosis primary care sample treated with a mix of therapeutic interventions, mostly of cognitive–behavioral or psychodynamic orientation. More details about the sample, patients, and measures can be found in Falkenström et al. (2013).

The total number of repeated measurements varied between 1 and 37, with a strongly unbalanced design (i.e., the number of sessions attended varied widely, with most patients attending only a few session). For ease of presentation, we limit the analysis to only include Sessions 1–10 (which of course means that results only generalize to these sessions). The detrending method used in the original analyses requires patients to have at least three observations on each measure, which reduced the sample to 636 therapies.[5] In the reanalyses, the sample size will vary somewhat depending on which estimator is used.[6]

Before analyzing the data, it is important to think through the hypothesized causal structure of the measurements. This means that the researcher needs to consider not just the timing of measurements, but also the time-period to which the measures are referring. With observational measures the time-period usually refers to the exact time of measurement, but questionnaires often ask participants to think of a specific time when rating. Because the CORE-OM was filled out before the session, and the instrument asks for symptoms experienced during the last week, and the WAI-SR was filled out immediately after the session, it is unlikely that the alliance as measured by the WAI-SR filled out after a particular session would be causing symptoms reported before the same session. Thus, the causal flow is hypothesized to run from WAI-SR at session t to CORE-OM at session $t + 1$. Only one lag of WAI-SR (and CORE-OM in the dynamic models) are used as predictors in order to keep the model as simple as possible.

The results of several analyses are shown in Table 1.[7] We begin with the basic hybrid random effects model, first without detrending and then adding time as a (random) covariate (following Wang & Maxwell, 2015) to show the effect of detrending. We next introduce the lagged dependent variable ("dynamic panel") model using SEM estimation. A latent (random) trajectory is then added to demonstrate detrending in a SEM framework. Hybrid random effects models were estimated using Stata 13.1 (StataCorp, 2013) while SEM was estimated using Mplus 7.3 (Muthén & Muthén, 1998–2012).

Hybrid Random Effects Model

As described previously, the hybrid random effects model estimates a random intercept for the dependent variable (Y), while the predictor (X) is person-mean centered to isolate the within-person effect (and ensuring zero correlation between predictor and random intercept). With therapies of only a single session disappearing because of lagging, 815 therapies with a total of 3301 observations were available for analysis using this method. Estimating

[5] All models assume large samples, although what this means specifically varies. Simple multilevel models have been shown to work with samples as small as $N = 30$–50 (Maas & Hox, 2005). A recent simulation study (Moral-Benito et al., 2016) shows that ML-SEM worked well with sample sizes down to $N = 100$, although the specific sample size requirement depends highly on the number of time-points, model complexity, and prevalence of missingness (Wolf, Harrington, Clark, & Miller, 2013).

[6] All models assume that data is Missing-At-Random, which is a relatively weak assumption (Enders, 2010). Some sensitivity tests of this assumption were described in the original publication; see Falkenström et al. (2013) for more details.

[7] The Stata and Mplus code for all models is available online in the supplemental materials.

Table 1
Results Showing Estimates for the Effect of the Working Alliance on Next-Session Symptoms/ Well-Being Using Different Models

Model	Coefficient	*SE*	*p*	95% CI Lower	95% CI Upper
Time-lagged models					
Hybrid random effects model	−1.32	.12	<.001	−1.56	−1.08
Hybrid random effects model, detrended	−.47	.12	<.001	−.71	−.23
Dynamic models					
ML-SEM, WAI-SR$_{t-1}$ exogenous	−1.30	.05	<.001	−1.40	−1.21
ML-SEM, WAI-SR$_{t-1}$ exogenous, detrended	−.39	.14	<.001	−.65	−.12
ML-SEM, WAI-SR$_{t-1}$ predetermined	−1.42	.06	<.001	−1.53	−1.30

Note. ML-SEM = Maximum Likelihood–Structural Equations Model; WAI-SR = Working Alliance Inventory - Short form Revised.

equation 3 yielded an effect of WAI-SR$_{t-1}$ on CORE-OM$_t$ (γ_{10}) of −1.32 ($SE = 0.12$, $p < .001$).

Adding a lagged dependent variable to the hybrid random effects model is, as discussed, not recommended because of endogeneity. When this endogeneity was ignored and the model estimated in any case, the estimate for γ_{10} was reduced to −0.89 ($SE = 0.12$, $p < .001$). The effect of the lagged dependent variable was 0.21 ($SE = 0.02$, $p < .001$).

To detrend following Wang and Maxwell (2015), a linear time variable with fixed and random effects are added to Equation 3[8]:

$$\text{Level 1 model: } CORE-OM_{i,t} = \beta_{0i} + \beta_1(WAI_{i,t} - \overline{WAI_i}) + \beta_2(Time) + e_{i,t}$$

$$\text{Level 2 model: } \beta_{0i} = \gamma_{00} + \gamma_{01}\overline{WAI_i} + u_{0i}$$
$$\beta_1 = \gamma_{10}$$
$$\beta_2 = \gamma_{20} + u_{2i}$$

$$\text{Mixed model: } CORE-OM_{i,t} = \gamma_{00} + \gamma_{01}\overline{WAI_i} + \gamma_{10}(WAI_{i,t} - \overline{WAI_i}) + \gamma_{20}(Time) + u_{0i} + u_{2i} + e_{i,t}.$$

In this model, γ_{20} estimates the fixed effect of Time (i.e., the average linear trajectory among all patients across all sessions) and u_{2i} estimates the random effect of Time (i.e., the variation among patients around the average linear trajectory). The estimate for the effect of WAI-SR$_{t-1}$ on CORE-OM$_t$ (γ_{10}) in this model was −0.47 ($SE = 0.12$, $p < .001$); that is, about 64% smaller than the nondetrended estimate. The effect of Time was statistically significant and fairly strong, both the fixed ($\gamma_{20} = -0.60$, $SE = 0.05$, $p < .001$) and the random ($u_{2i} = 0.47$, $SE = 0.07$) effects. These results illustrate the potentially dramatic effects that may be seen when variables are detrended. As such, it is imperative to assess the detrended model in terms of its theoretical plausibility, which, as will be discussed further below, may not be compelling in this particular case.

Maximum Likelihood–Structural Equations Model (ML-SEM)

The ML-SEM model is the most flexible model of the ones reviewed, with several advantages such as the possibility of modeling fully bivariate/multivariate relationships among variables and eliminating endogeneity bias when using lagged dependent variables as predictors. The simplest model is one that mimics a standard regression model; that is, with only one variable modeled as dependent variable while the other one is a predictor. As shown in Figure 4, the between-person effect is modeled as a latent variable, or factor, which subsumes the variance that is common to all observations for each patient (i.e., between-patient variability)—leaving only time-specific (i.e., within-patient) variance at each given point in time. Fixing the factor loadings at one means that all observations contribute equally to the factor. Because the first observation of CORE-OM does not have any path from a prior variable pointing at it, its factor loading is likely to differ from the other observations and thus it is treated as exogenous although with a freely estimated correlation with the between-person effect. It is important that to remove the between-patient effect from WAI-SR, all observations must be allowed to correlate with the CORE-OM between-patient factor. The effect of WAI-SR$_{t-1}$ on CORE-OM$_t$ is modeled as a regression path, while the contemporaneous correlation between WAI-SR$_t$ and CORE-OM$_t$ is modeled as a residual correlation (because it needs to be taken into account while not being of primary interest in this particular analysis).

The effect of WAI-SR$_{t-1}$ on CORE-OM$_t$, controlling for the effect of CORE-OM$_{t-1}$ was estimated to be −1.30 ($SE = 0.05$, $p < .001$, 95% confidence interval [CI] −1.40, −1.21); that is, very similar to the nondetrended hybrid random effects model, despite being adjusted for the effect of CORE-OM$_{t-1}$ which in this model was estimated as 0.33 ($SE = 0.02$, $p < .001$, 95% CI 0.30, 0.36).

Detrending can be handled in dynamic SEM models by adding a growth curve component to the model, such that time has an effect on the outcome that randomly varies across individual as an addition to the rest of the model.[9] The addition of random slopes for linear time is the equivalent to using a time variable as a

[8] It is of course possible to detrend against other kinds of trajectory shapes than a linear one (e.g., quadratic, cubic, etc.). For ease of presentation, we only present linear detrending here (which, incidentally, was the one having the strongest effect on results in the data used).

[9] There are other SEM models for dynamic panel data in which detrending is the default (e.g., the "autoregressive latent trajectory" model of Bollen and Curran [2004], or the "structured residual" model of Curran, Lee, Howard, Lane, and MacCallum [2012]).

random covariate in the regression models. The result for WAI-SR$_{t-1}$ on CORE-OM$_t$ using this model was −0.39 (SE = 0.14, p = .005, 95% CI −0.65, −0.12), again showing a substantially smaller effect (70%) when using detrending.

Modeling Reverse Causation Using SEM

As discussed previously, a distinct advantage of dynamic panel models is the possibility of determining causal direction. In one sense this is accomplished in the SEM models described so far, because the predictor is lagged and controlled for the effect of prior values of the dependent variable. However, if we believe that there is simultaneous causal influence from X to Y and also from Y to X (sometimes referred to as dynamic feedback loops), this will create further endogeneity problems, given the intrinsic relationship between the explanatory variables and the error terms in their respective equations. Fortunately, it is possible to model these feedback effects by estimating correlations between the error term of Y and future values of X.

In the alliance-outcome literature the idea of reverse causation from symptoms/well-being to alliance is theoretically plausible and has empirical support (e.g., Falkenström et al., 2013). In the primary care psychotherapy data, the contemporaneous relationship between WAI-SR and CORE-OM is interpreted as reverse causation (CORE-OM$_t$ → WAI-SR$_t$) because the CORE-OM is filled out before the session and refers to the week prior to the session while the WAI-SR is filled out after the session.

The potential effect from CORE-OM$_t$ to WAI-SR$_t$ is already taken into account in the ML-SEM model used so far by the correlation between the error term of CORE-OM$_t$ and WAI-SR$_t$. Still, reverse causation in the form of earlier lags of CORE-OM affecting WAI-SR may also be present, and will then need to be modeled by estimating correlations between error terms of CORE-OM and all future observations of WAI-SR. This yielded a slightly larger effect for WAI-SR$_{t-1}$ (−1.42, SE = 0.06, $p < .001$, 95% CI −1.53, −1.30) than the model in which reverse causation was not modeled.[10]

Relaxing Assumptions of Stationarity of Variances and Covariances in SEM

In SEM, it is possible to relax and test assumptions of stationarity of variances and covariances. Stationarity of variances can be tested by comparing a model with residual variances constrained to equality with a model for which residual variances are allowed to differ. These models are nested and can thus be compared using a χ^2 difference test with degrees of freedom equal to the difference in number of parameters estimated. This test was not statistically significant, $\Delta\chi^2(8) = 10.94$, $p = .20$, which means that the assumption of stationarity of (residual) variances held in this sample. Therefore, the more parsimonious model assuming stationarity is preferred to the one estimating separate error variances for each time-point. Stationarity of covariances is tested by comparing a model that estimates a single coefficient for the effect of WAI-SR$_{t-1}$ on CORE-OM$_t$ with one that estimates different coefficients for different time-points. The result was that the model with different coefficients at different time-points was significantly better than the model with coefficients constrained to equality over time, $\Delta\chi^2(8) = 33.73$, $p < .001$). Inspecting estimates showed that the alliance effect on subsequent symptoms was increasing almost linearly, from −1.22 between Sessions 1 and 2, to −1.52 between Sessions 9 and 10. Because the stationarity assumption in this case did not hold, we should interpret the model with separate coefficients for the different time-points rather than the one in which these are constrained to equality. This means that all the coefficients in Table 1 should be interpreted as showing an average effect of the alliance in one session on symptoms in the following session, while the true effect is slightly smaller in the initial sessions and increasing up to Session 9–10 (because we are only using data for Sessions 1–10, we cannot extrapolate beyond this point). In this case, relaxing the assumption of stationarity of covariances adds nuance but does not fundamentally change the substantive findings, while in other contexts the differences in causal effects over time may be more critical.

Discussion

Psychotherapy research is moving beyond simple questions such as "does psychotherapy work?" and process-outcome research seems to be gaining in popularity among psychotherapy researchers. Panel data models are clinically intuitive because of the focus on associations relatively close in time. They are also stronger for causal interpretation because of the possibilities of ruling out some classes of unobserved confounders, and of modeling cross-lagged associations to test which variable is "driving" development of the other. At the same time, fundamental complexities associated with separating between- from within-patient effects in the context of a dynamic model (i.e., one in which autoregression of the dependent variable is modeled explicitly) do not seem to be well known among psychotherapy researchers. In addition, there are issues of when, whether, and how to adjust for time-trends in the data.

Our data analytic example gave some interesting examples of the kinds of differences that various data-analytic choices can result in. When inspecting the results of Table 1, it is apparent that detrending resulted in the WAI-SR effect dropping markedly, with 64% reduction in the hybrid random effects model and 70% reduction in SEM. As discussed, detrending is potentially problematic in the context of a treatment study. On the one hand, detrending strengthens causal interpretation in the sense that the observed effect cannot be confounded with irrelevant variables having a linear effect on the outcome over time. On the other hand, detrending will suppress estimates when the phenomena of interest are likely to cause linear trends over time, as in the example of the alliance-outcome relationship in which bidirectional influence between symptoms and alliance will generate correlated trends. A possible solution to this dilemma is to report nondetrended estimates, then use detrending as a sensitivity test to see whether the effect found remains in the presence of detrending. If it does, stronger causal statements can be made than if the effect disappears after detrending.

[10] Detrending the predetermined model resulted in essentially the same estimates as the detrended model without the residual covariances between CORE-OM and future observations of WAI-SR.

Conclusions and Recommendations for Future Research

We recommend using SEM for dynamic panel models to overcome the issue of endogeneity when entering a lagged dependent variable as a predictor. As our data analytic example showed, parameter estimates will differ substantially if this issue is ignored. However, there are situations that may preclude the use of SEM, such as when the number of measurements is large compared with the number of cases, and/or when data is strongly unbalanced, in which case SEM estimation may not converge. Such situations may not be uncommon in psychotherapy process research, so an important task for future research is to assess how much bias the hybrid random effects model with an additional lagged dependent variable term included, is likely to produce, and what conditions influence the amount and direction of bias.

Of the data analytic choices discussed, the largest difference was between detrending and not detrending. A reasonable question then is which of the models comes closest to "the truth"? Unfortunately, although this a reasonable question, it is not as easy to answer as one would wish. Obviously, the choice of model has important consequences, but testing the methods/models on simulated data cannot solve the problem for the simple reason that the data has to be generated assuming one of the models—which, when tested, will unavoidably yield the "best" parameter estimates. Therefore, the choice must be made primarily on rational/theoretical grounds, that is, careful thought based on theory and research situation being given to what kinds, if any, counterfactual trends in the absence of treatment would need to be controlled. In a psychotherapy process-outcome study in which outcome is symptoms/well-being, differential treatment effects will yield trends over time, and removing these by detrending Y will result in "throwing out the baby with the bathwater". Predictors in such a study may either be expected to show no linear trend (e.g., psychotherapeutic techniques), or may show a trend that to a large extent is caused by the treatment (e.g., working alliance). In these situations, as is arguably the case in the empirical example presented in this article, it would not make sense to detrend either X or Y.

Below follows a list of recommendations for researchers interested in applying dynamic panel models in their work:

(1) Is there theoretical reason to believe that the dependent variable causes itself over time? If not, do not include a lagged dependent variable as predictor. Statistical dependency of outcomes over time is handled well by standard multilevel modeling; specifically, by the hybrid random effects model, which estimates both within-patient and between patient effects, and which does not require explicit modeling of autoregression.

(2) If there is a theoretical need for including a lagged dependent variable as predictor, SEM is required to avoid endogeneity bias.

(3) If it is theorized that there is reverse causation from the dependent variable to future realizations of the predictor, then variations of the SEM model need to be used that take the additional endogeneity between lagged X and current or prior error terms of Y into account.

(4) If there is no lagged dependent variable included among the predictors, reverse causation is unlikely, and it is unlikely that relationships between variables change in any major way over time, the hybrid random effects model is the best analytic choice because of its relative simplicity and flexibility.

(5) If there is a risk, theoretically, that trajectories over time represent unmeasured confounder(s), then detrending is needed. If the answer is no, and time trends in X and Y are both the product of some experimental or natural treatment that is of primary theoretical interest, then detrending should be avoided. Detrending could still be used as a sensitivity test to explore whether the effect of interest is robust to removal of trends. An effect that is robust to detrending can be more confidently (though not absolutely) interpreted as causal.

(6) When detrending is used, the "time as covariate" model (Wang & Maxwell, 2015) is recommended for fixed/random effects models. The equivalent in SEM is to add a growth curve model "on top" of the cross-lagged part of the SEM.

(7) Is it theoretically plausible that relationships between variables change over time in any major way? In that case, SEM should be used, and a model estimating separate covariances should be compared with a model constraining covariances over time. Similarly, models estimating separate error variances over time can be compared with models constraining error variances to be equal. In our experience, the issue of error variances changing over time seldom has any major impact on results, but it should nevertheless be tested.

Dynamic panel data methods have the potential of improving psychotherapy process-outcome research considerably, because of the combination of clinical relevance and causal stringency. We hope that the issues discussed in this paper lead researchers to adopt these methods in future work, while at the same time attending to the theoretical choices involved and the analytic complexities that these models entail.

References

Allison, P. (2015). *Don't put lagged dependent variables in mixed models*. Retrieved from http://m.statisticalhorizons.com/?url=http%3A%2F%2Fstatisticalhorizons.com%2Flagged-dependent-variables&width=412

Allison, P. (2016). *Maximum Likelihood for dynamic panel models with cross-lagged effects*. Manuscript submitted for publication.

Baird, R., & Maxwell, S. E. (2016). Performance of time-varying predictors in multilevel models under an assumption of fixed or random effects. *Psychological Methods, 21,* 175–188. http://dx.doi.org/10.1037/met0000070

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 258–297). New York, NY: Wiley.

Baldwin, S. A., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology, 75,* 842–852. http://dx.doi.org/10.1037/0022-006X.75.6.842

Baltagi, B. H. (2013). *Econometric analysis of panel data*. Chichester, United Kingdom: Wiley.

Bollen, K., & Curran, P. J. (2004). Autoregressive Latent Trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods & Research, 32,* 336–383. http://dx.doi.org/10.1177/0049124103260222

Crits-Christoph, P., Connolly Gibbons, M. B., & Mukherjee, D. (2013). Psychotherapy process-outcome research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 298–340). New York, NY: Wiley.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62,* 583–619. http://dx.doi.org/10.1146/annurev.psych.093008.100356

Curran, P. J., Lee, T., Howard, A. L., Lane, S., & MacCallum, R. (2012). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. *Advances in Longitudinal Methods in the Social and Behavioral Sciences,* 217–253.

Delgadillo, J., Moreea, O., & Lutz, W. (2016). Different people respond differently to therapy: A demonstration using patient profiling and risk stratification. *Behaviour Research and Therapy, 79,* 15–22. http://dx.doi.org/10.1016/j.brat.2016.02.003

DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research, 14,* 469–482. http://dx.doi.org/10.1007/BF01172968

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *The British Journal of Psychiatry, 180,* 51–60. http://dx.doi.org/10.1192/bjp.180.1.51

Falkenström, F., Ekeblad, A., & Holmqvist, R. (2016). Improvement of the working alliance in one treatment session predicts improvement of depressive symptoms by the next session. *Journal of Consulting and Clinical Psychology, 84,* 738–751. http://dx.doi.org/10.1037/ccp0000119

Falkenström, F., Granström, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology, 60,* 317–328. http://dx.doi.org/10.1037/a0032258

Falkenström, F., Grant, J., & Holmqvist, R. (2016). Review of organizational effects on the outcome of mental health treatments. *Psychotherapy Research, 26,* 1–15. http://dx.doi.org/10.1080/10503307.2016.1158883

Finkel, S. (1995). *Causal analysis with panel data*. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412983594

Firebaugh, G., Warner, C., & Massoglia, M. (2013). Fixed effects, random effects, and hybrid models for causal analysis. In L. S. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 113–132). Dordrecht, the Netherlands: Springer Netherlands. http://dx.doi.org/10.1007/978-94-007-6094-3_7

Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics, 2,* 111–120. http://dx.doi.org/10.1016/0304-4076(74)90034-7

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16,* 12–25. http://dx.doi.org/10.1080/10503300500352500

Hoffart, A. (2016). Cognitive models for panic disorder with agoraphobia: A study of disaggregated within-person effects. *Journal of Consulting and Clinical Psychology, 84,* 839–844. http://dx.doi.org/10.1037/ccp0000114

Hoffart, A., Øktedalen, T., Langkaas, T. F., & Wampold, B. E. (2013). Alliance and outcome in varying imagery procedures for PTSD: A study of within-person processes. *Journal of Counseling Psychology, 60,* 471–482. http://dx.doi.org/10.1037/a0033604

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 221–233). Berkeley, CA: University of California Press.

Jacobson, N. S., Martell, C. R., & Dimidjian, S. (2001). Behavioral activation treatment for depression: Returning to contextual roots. *Clinical Psychology: Science and Practice, 8,* 255–270. http://dx.doi.org/10.1093/clipsy.8.3.255

Jongerling, J., Laurenceau, J.-P., & Hamaker, E. L. (2015). A Multilevel AR(1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research, 50,* 334–349. http://dx.doi.org/10.1080/00273171.2014.1003772

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science, 21,* 984–991. http://dx.doi.org/10.1177/0956797610372634

Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 169–218). New York, NY: Wiley.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1,* 86–92. http://dx.doi.org/10.1027/1614-2241.1.3.86

Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics, 31,* 451–472. http://dx.doi.org/10.1080/07350015.2013.818003

Moral-Benito, E., Allison, P., & Williams, R. (2016). Dynamic panel Data Modeling using maximum likelihood: An alternative to Arellano-Bond. Retrieved from https://www3.nd.edu/~rwilliam/dynamic/Benito_Allison_Williams.pdf

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide*. Los Angeles, CA: Author.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica, 49,* 1417–1426. http://dx.doi.org/10.2307/1911408

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata. Vol. I. Continuous responses* (3rd ed.). College Station, TX: Stata Press Publication.

Ramseyer, F., Kupper, Z., Caspar, F., Znoj, H., & Tschacher, W. (2014). Time-series panel analysis (TSPA): Multivariate modeling of temporal associations in psychotherapy process. *Journal of Consulting and Clinical Psychology, 82,* 828–838. http://dx.doi.org/10.1037/a0037168

Raudenbush, S. W., & Bryk, A. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage Publications.

Rubel, J. A., Rosenbaum, D., & Lutz, W. (2017). Patients' in-session experiences and symptom change: Session-to-session effects on a within- and between-patient level. *Behaviour Research and Therapy, 90,* 58–66. http://dx.doi.org/10.1016/j.brat.2016.12.007

Snijders, T. A. B., & Bosker, R. J. (2011). Multilevel analysis: An introduction to basic and advanced multilevel modeling (2nd ed.). London, England: Sage Publications.

StataCorp. (2013). *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.

Wang, L. P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods, 20,* 63–83. http://dx.doi.org/10.1037/met0000030

Williams, R., Allison, P., & Moral-Benito, E. (2015, July). *Linear dynamic panel-data estimation using Maximum Likelihood and Structural Equa-*

tion Modeling. Paper presented at the Stata Users Conference, Columbus, OH.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for Structural Equation Models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 76,* 913–934. http://dx.doi.org/10.1177/0013164413495237

Wooldridge, J. M. (2012). *Introductory econometrics: A modern approach* (5th ed.). Mason, OH: Cengage Learning.

Zilcha-Mano, S., & Errázuriz, P. (2015). One size does not fit all: Examining heterogeneity and identifying moderators of the alliance-outcome association. *Journal of Counseling Psychology, 62,* 579–591. http://dx.doi.org/10.1037/cou0000103

Received October 18, 2016
Revision received January 24, 2017
Accepted February 1, 2017 ■