Equivalent Approaches to Dealing with Unobserved Heterogeneity in Cross-Lagged Panel Models?

Investigating the Benefits and Drawbacks of the Latent Curve Model with Structured Residuals and Random Intercept Cross-Lagged Panel Model

Henrik Kenneth Andersen

22 May, 2021

Abstract

Panel models in structural equation modeling that combine static and dynamic components for investigating reciprocal relations while controlling for time-invariant unobserved heterogeneity are becoming increasingly popular. Recently, the Latent Curve Model with Structured Residuals and the Random-Intercept Cross-Lagged Panel Model were suggested as 'residual-level' versions of the more traditional Autoregressive Latent Trajectory and Dynamic Panel Models, respectively. Their main benefit is that they allow for a more straightforward interpretation of the trajectory factors. It is not widely known, however, that the residual-level models place potentially strong assumptions on the initial conditions, i.e., the process that was occurring before the observation period began. If the process under investigation is not both stationary and at equilibrium then the residual-level models are not appropriate. They then do not control for all time-invariant unobserved heterogeneity and can result in biased cross-lagged and autoregressive estimates. I demonstrate this using the problem behavior of cigarette smoking amongst adolescents: because the mean and variance of this process changes as a young person's smoking behavior develops, early stages of this process should not be examined using the residual-level models. This issue potentially exists for a wide variety of psychological and sociological subjects, essentially whenever the process under investigation is changing over the course of the observation period. This paper discusses strategies to help researchers decide which model to use when, and compares some of their relative advantages and drawbacks. An amendment to the residual-level models is suggested in which the latent individual effects are allowed to covary with the initial residuals. This makes the residual-level models robust to violations of the assumptions surrounding the initial conditions, while retaining their other beneficial aspects.

Introduction

Dynamic linear panel models in the structural equation modeling (SEM) framework that examine (cross-) lagged relations while taking time-invariant unobserved heterogeneity into account have become increasingly popular over the last 15 years or so. Models like the Autoregressive Latent Trajectory (ALT, K. Bollen and Curran 2004; Curran and Bollen 2001) and the Dynamic Panel Model (DPM, P. Allison, Williams, and Moral-Benito 2017; Moral-Benito, Allison, and Williams 2019; Williams, Allison, and Moral-Benito 2018), as well as the General Cross-Lagged Model (GCLM, Zyphur et al. 2020a, 2020b) combine aspects of static and dynamic panel models to allow researchers to control for time-invariant unobserved factors while examining within-unit processes; either of a single construct over time, or the reciprocal relations between two or more constructs (Voelkle and Oud 2015).[1]

---

[1] I focus on the ALT and DPM because of they are arguably the closest observation-level analogues to the residual-level models introduced below. The GCLM also investigates cross-lagged relationships while controlling for time-invariant unobserved heterogeneity but it goes a step further, including a moving average (MA) component (regressing the current observation on

The Random-Intercept Cross-Lagged Panel Model (RI-CLPM, Hamaker, Kuiper, and Grasman 2015) and the Latent Curve Model with Structured Residuals (LCM-SR, Curran et al. 2014) were recently suggested as novel modifications of the conventional ALT and DPM.[2] These models specify the autoregressive (AR) and, if present, cross-lagged (CL) effects at the residual- or disturbance-level, so to speak, i.e., what is left over of the time-varying observed variables after they have been regressed on the latent individual effects. Their advocates highlight the fact that the 'residual-level' models provide the benefit of cleanly separating stable between-person differences from within-person fluctuations (Usami, Murayama, and Hamaker 2019), thereby making the mean structure easier to interpret (Curran et al. 2014).

However, the residual-level models are re-expressions of the ALT and DPM, and there are a number of different versions of these 'observation-level' models depending on the assumptions placed on the initial conditions, i.e., the unobserved process prior to the first observation. Ou et al. (2016) explain that there are at least three versions of the ALT or DPM,[3] depending on the assumptions regarding the initial conditions. The so-called 'predetermined' model is the most flexible and general option because it places no restrictions on the initial observations, treating them as exogenous. The typical 'constrained' model assumes the AR process is stationary and that it has been going on long enough to have reached equilibrium (Ou et al. 2016; K. Bollen and Curran 2004; Hamaker 2005). The benefit of the constrained models is that if the assumptions about the initial conditions hold, they are more parsimonious. Still another version assumes the cumulative effects of the earlier realizations are negligible, or that the observation period captures a distinct process separate from what came before it (Ou et al. 2016).

In this paper, I show that while the RI-CLPM and LCM-SR can offer benefits compared to their more conventional counterparts, they are re-expressions specifically of the constrained versions of the observation-level models. This is shown algebraically and demonstrated with examples using real data. This means that the RI-CLPM and the LCM-SR assume that the process being modeled is stationary, and has been going on for enough time to have reached its equilibrium state. If this is not the case, then the RI-CLPM and LCM-SR can yield misleading results.

The main goals of this paper are to:

- emphasize that while it is widely known that observation-level models can be re-expressed as residual-level models (see Jongerling and Hamaker 2011; Hamaker 2005; Hsiao 2014, 86f.; Ou et al. 2016), they result in 'constrained' versions of the observation-level counterparts,
- raise awareness about the assumptions inherent to the constrained, and thus also residual-level models,
- demonstrate the consequences of a violation of the assumptions, and
- provide some suggestions for dealing with these issues.

The paper is structured as follows. First, I briefly discuss the synthesis of static (e.g., random and fixed effects, latent growth curves) and dynamic (e.g., autoregressive) panel models. Then, I show algebraically that the residual-level models (LCM-SR and RI-CLPM) can be rewritten as constrained versions of the observation-level counterparts (ALT and DPM). I discuss the implicit assumptions of the constrained and residual-level models and demonstrate violations of these assumptions using a brief simulated example. I then look at the topic of problem behaviors in adolescents as an empirical example of a process that is problematic to examine using the constrained and residual-level models. I discuss some potential strategies for dealing with a violation of assumptions, focusing on an approach that involves allowing the latent individual effects to covary with the initial residuals (i.e., the deviations from those individual effects). This accounts for situations in which the process under investigation is not at equilibrium and thus makes the residual-level models more robust to violated assumptions. I conclude with a brief outlook and a discussion of some potential next steps.

---

the previous residuals or 'impulses' as they are referred to). The DPM could be thought of a special case of the GCLM in which the MA structure is dropped and the individual effects are assumed time invariant, (though relaxing the assumption of time-invariant individual effects is always an option, see K. Bollen and Brand (2010)).

[2]The main difference between the DPM and ALT is that the DPM accounts for time-invariant unobserved heterogeneity in level, but not in growth. This is the main difference between the RI-CLPM and LCM-SR, as well.

[3]Hsiao (2014) goes into even more granular detail (p. 87ff.).

Background

The main benefit of longitudinal or panel data is that they allow researchers to control for all time-invariant unobserved heterogeneity, i.e., omitted variables that are constant over time, like genetic factors, personality traits, intelligence, motivation (K. Bollen and Brand 2010). These time-invariant omitted variables are sometimes referred to as the 'individual effects' (Hsiao 2014, the term I prefer), 'unobserved effects' (Wooldridge 2002), 'unit effects' (Zyphur et al. 2020a, 2020b; Skrondal and Rabe-Hesketh 2004), 'individual trajectories' (Bianconcini and Bollen 2018; K. Bollen and Curran 2004), etc. Models that control for such time-invariant unobserved heterogeneity are referred to as *fixed effects* (FE) models. While there is typically some confusion surrounding the term, the modern sign of a fixed effects model is that it assumes the exogenous model covariates are related to the stable (unobserved) time-invariant component(s) (Wooldridge 2002). This is often a much more realistic assumption than the alternative, embodied by random effects (RE) models, which is to assume there is no correlation between the individual effects and the model covariates (K. Bollen and Brand 2010; Andersen 2021). For example, if we were looking at the relationship between social isolation and subjective well-being (Seifert and Andersen 2020), then an FE regression model would allow and control for the possibility that a person's social isolation at a given point in time is correlated with underlying personality traits like extraversion, or neuroticism, even if we do not have observed measures of these traits.

The implementation of basic fixed effects regression in structural equation modeling (SEM) has been discussed, for example, in K. Bollen and Brand (2010); Teachman et al. (2001); Andersen (2021); P. D. Allison (2009). It works by specifying a latent variable to represent the individual effects. This latent variable is treated as a random variable with a constant effect on the time-varying observed dependent variable (this can be relaxed, see K. Bollen and Brand 2010; Zyphur et al. 2020a), that is allowed to covary freely with the model covariates (P. D. Allison 2009; K. Bollen and Brand 2010). In the majority of social science applications based on observational data, the model covariates are also treated as random variables. By including the latent individual effects along with the time-varying covariates in the equation for the time-varying dependent variable, the correlation between the covariates and the individual effects can be partialled out, giving the estimated coefficients their interpretation as 'within-unit' effects (i.e., cleansed of the between-unit component, Rüttenauer and Ludwig 2020; K. Bollen and Brand 2010, 27; Zyphur et al. 2020b, 7). Note that constraining the correlations between the latent individual effects and the model covariates to zero is equivalent to an RE model (K. Bollen and Brand 2010).

Another type of model popular in the SEM framework looks at and accounts for stable differences between units not just in level, but also in slope. These are referred to as Latent Growth Curve (LGC), or Latent Curve Models (LCM, Meredith and Tisak 1990; K. A. Bollen 2006; K. Bollen and Zimmer 2010). These models typically have two latent variables, representing individual intercepts and slopes, thereby allowing each unit to have its own trajectory. Often, linear trajectories are assumed, but higher order polynomials can also be modeled (Kim et al. 2018). Fixed effects variants of these models that control for time-invariant unobserved heterogeneity can be specified by allowing the time-varying covariates to correlate with the latent trajectory components.

So, the FE, RE and LGC models are all examples of panel models that account for time-invariant unobserved heterogeneity, a likely and common source of bias (Rüttenauer and Ludwig 2020; Andersen 2021). They are sometimes referred to as 'static' panel models because they account for stable differences (Zyphur et al. 2020b). Another type of panel regression model popular in the social sciences, and within the SEM framework in particular, looks at the lagged effects of a covariate on some other dependent variable, while holding the level of the dependent variable at a previous point in time constant. Models that include the lagged dependent variable in the equation for the current dependent variable, like autoregressive (AR) models, are sometimes referred to as 'dynamic' panel models (Zyphur et al. 2020b). Cross-Lagged Panel Models (CLPM) expand on the idea of AR models by including a second dynamic process and examining the bidirectional or reciprocal relations between the two. They work by regressing each of the variables, call them x and y, on their own lagged values (the AR components), as well as on the lagged values of the opposite variable (the cross-lagged (CL) components). By looking at *lagged* relations, they have the advantage of fulfilling one of the main stipulations of Granger's concept of causality: that the supposed

cause come before the outcome (Granger 1969; Zyphur et al. 2020b; Hamaker, Kuiper, and Grasman 2015). Typically, CLPMs are used to investigate the question of whether x causes y or the other way around. It is common practice to compare the standardized coefficients $x_{t-1} \rightarrow y_t$ and $y_{t-1} \rightarrow x_t$, declare the stronger of the two the 'causal winner' (Hamaker, Kuiper, and Grasman 2015) and take this as evidence that, say, x 'causes' y, and not the other way around.

Around 20 years ago, panel models in SEM were suggested that combine both static and dynamic processes; "the best of both worlds" (Curran and Bollen 2001; K. Bollen and Curran 2004; K. Bollen and Zimmer 2010). The ALT model examines autoregressive and (cross-) lagged relations while controlling for time-invariant unobserved heterogeneity. Again, if we were interested in the effect of social isolation on subjective health, then we could use a bivariate ALT to look at the lagged effect of one's degree of social isolation on their subjective well-being, holding their previous level of subjective health constant, while also allowing for the possibility that social isolation is related to a variety of stable omitted variables. Other models that can be seen as special cases of the ALT have been suggested in recent years, such as the Dynamic Panel Model (DPM, P. Allison, Williams, and Moral-Benito 2017; Williams, Allison, and Moral-Benito 2018; Moral-Benito, Allison, and Williams 2019) and certain versions of the General Cross-Lagged Model (GCLM, Zyphur et al. 2020a, 2020b). These, like a traditional FE model, control for individual differences in level, but not trajectory.

More recently, a new approach to combining static and dynamic models was suggested by Curran et al. (2014) with the LCM-SR and Hamaker, Kuiper, and Grasman (2015) with the RI-CLPM. Like the ALT, DPM and GCLM, these models also look at the lagged relationships between two (or more) variables, controlling for each of the variables' own previous levels while accounting for stable between-unit differences (in level for the RI-CLPM and in both level and trajectory for the LCM-SR). These models provide the benefit of "cleanly separating" (Curran et al. 2014, 885) the static from the dynamic processes; thereby allowing for a more intuitive interpretation of the between-unit individual effects components. They do so by specifying the AR and CL effects at the *residual-level*, i.e., what is left over of the observed time-varying dependent variables after they have been regressed on the latent individual effects. Because of this, I refer to these type of models as 'residual-level' models, in contrast with the 'observation-level' models discussed earlier. The residuals can be thought of as within-unit centered scores (Usami, Murayama, and Hamaker 2019, 4), which brings to mind the method of 'demeaning' or 'detrending' observed variables to eliminate between-unit variance, which is the typical approach to least-squares based FE estimation Rüttenauer and Ludwig (2020).

The RI-CLPM is explicitly presented as an extension and improvement of the CLPM, which includes AR and CL effects, but does not account for time-invariant unobserved heterogeneity. In the abstract, Hamaker, Kuiper, and Grasman (2015) write: "if stability of constructs is to some extent of a trait-like, time-invariant nature, the autoregressive relationships of the CLPM fail to adequately account for this" (2015, 102) and later in the introduction that "not only should we *account* for stability, but we need to *account for the right kind* of stability" (2015, 102, emphasis in original). And, as noted, the article by Curran et al. (2014) discusses that the LCM-SR offers benefits compared to the ALT in terms of the interpretability of the individual effects. This raises the following questions: Are the residual- and observation-level models equivalent? Why does Hamaker, Kuiper, and Grasman (2015) present the RI-CLPM as a solution to the problem of the CLPM ignoring time-invariant unobserved heterogeneity when the ALT has been around for decades? Is there any reason to prefer the residual-level models to the traditional observation-level counterparts if the within-unit effects are of main substantive interest? The following section will begin by reiterating the relationship between the observation- and residual-level models (this has been done before elsewhere, for example, Hsiao (2014); Jongerling and Hamaker (2011); Hamaker (2005); Curran and Bollen (2001); K. Bollen and Curran (2004); Ou et al. (2016)) before moving on to highlight the implications of such.

Analytical comparison

The ALT and DPM model the autoregressive effects and the effects of the other covariates at the observed variable level.[4] The DPM assumes there are stable differences between individuals while the ALT takes individual differences in both intercept and slope (varying trajectories) into account. We can write a simple version of the ALT as

$$y_{it} = \eta_{1_i} + t\eta_{2_i} + \rho y_{it-1} + \nu_{it}, \ t = 1, \ldots, T \tag{1}$$

where y_{it} is the observed dependent variable for person $i = 1, \ldots, N$ at time $t = 0, \ldots, T$. Notice the initial observation is $t = 0$, but Equation (1) starts at $t = 1$ because there is no previous observation to regress y_{i0} on. I will come back to this momentarily. η_{1_i} and η_{2_i} are the latent individual intercepts and slopes, respectively (although their interpretation as such is not straightforward, as will be discussed below). I will refer to these individually as the individual intercepts and slopes, or together as the individual effects (a term used in Hsiao 2014). y_{it-1} is the lagged dependent variable and ρ is the autoregressive coefficient, which, for the sake of simplicity, will be assumed time-invariant throughout, i.e., $\rho_t = \rho$. ν_{it} is the idiosyncratic error with $\mathbb{E}(\nu_{it}) = 0$ (K. Bollen and Curran 2004). We could easily extend this model to include other time-varying (either contemporaneous or lagged) covariates as well as time-invariant covariates (K. Bollen and Brand 2010).

In Equation (1) and throughout, I omit time-specific or 'occasion effects' (Zyphur et al. 2020b), sometimes written α_t, τ_t, μ_t, etc. These represent shifts in the mean of the observed variable at a given point in time (Zyphur et al. 2020b; P. Allison, Williams, and Moral-Benito 2017). Leaving them out of the equations puts the focus on the covariance structure, and is equivalent to analyzing centered data (Hamaker, Kuiper, and Grasman 2015). These effects can be recovered in typical SEM software by simply turning on the mean structure (e.g., in `lavaan` by adding the optional argument `meanstructure = TRUE` in the fitting function). By default, the means of the latent individual effects are fixed to zero to identify an intercept per observed timepoint. This is desirable because it places no constraints on the means (Hamaker, Kuiper, and Grasman 2015), but it does go against the usual convention in LCMs and ALT models of constraining the time-specific intercepts to zero and attempting to model the mean structure entirely with the latent intercept and slope factors, which, as Zyphur et al. (2020b) note, is a problematic assumption. This is likely why the DPM explicitly allows for occasion effects (see Equations (1) and (2) in P. Allison, Williams, and Moral-Benito 2017).

An analogous DPM would be

$$y_{it} = \eta_i + \rho y_{it-1} + \nu_{it}, \ t = 1, \ldots, T. \tag{2}$$

Notice Equations (1) and (2) apply for $t = 1, \ldots, T$, so we have to decide what to do with y_{i0} since we do not have an observation for $y_{i,-1}$. The most straightforward approach is to simply treat y_{i0} as *predetermined*. That is, we enter y_{i0} into the model as an exogenous variable, with an unconditional mean and individual deviation from it (K. Bollen and Curran 2004). We make no assumptions about the process before the observations began and allow y_{i0} to covary freely with the individual effects. A four-wave version of a predetermined observation-level model with individual effects (Equation (1) describes the model for $t = 1, \ldots, T$) is shown in Figure 1a.

The drawback to treating the initial observation as predetermined is that it may use up 'too many' degrees of freedom, depending on the specific model one is trying to run and number of waves of data one has available.

---

[4]These models can be extended to include multiple-indicator measurement models, see, for example, Bianconcini and Bollen (2018); Mulder and Hamaker (2020).

For example, a three-wave ALT is not over-identified, even if we constrain the autoregressive effect and error variances to be equal over time (K. Bollen and Curran 2004).[5]

A solution put forth by K. Bollen and Curran (2004); Curran and Bollen (2001) and further discussed in Jongerling and Hamaker (2011); Hamaker (2005) is to make some assumptions about the process *before the observation period began* and work out fixed values for the associations between the initial observation and the individual intercept and slope. This is referred to as a *constrained* model. Namely, if we assume the autoregressive effect is constant over time, $\rho_t = \rho$, and less than one in absolute value, $|\rho| < 1$, then we can begin substituting the equations for the previous *unobserved* realizations into the first one:

$$
\begin{aligned}
y_{i0} &= \eta_{1_i} + 0\eta_{2_i} + \rho y_{i,-1} + \nu_{i0} \\
&= \eta_{1_i} + \rho(\eta_{1_i} - 1\eta_{2_i} + \rho y_{i,-2} + \nu_{i,-1}) + \nu_{i0} \\
&= \eta - 1_i + \rho(\eta_{1_i} - 1\eta_{2_i} + \rho(\eta_{1_i} - 2\eta_{2_i} + \rho y_{i,-3} + \nu_{i,-2}) + \nu_{i,-1}) + \nu_{i0} \\
&= \dots
\end{aligned}
\tag{3}
$$

and so on, back to $-\infty$,[6] which gives

$$
y_{i0} = \sum_{j=0}^{\infty} \rho^j \eta_{1_i} + \sum_{j=0}^{\infty} (t-j)\rho^j \eta_{2_i} + \sum_{j=0}^{\infty} \rho^j \nu_{i(t-j)}.
\tag{4}
$$

Given that $\rho_t = \rho$ and $|\rho| < 1$, this converges to

$$
y_{i0} = (1-\rho)^{-1}\eta_{1_i} - \rho(1-\rho)^{-2}\eta_{2_i} + \sum_{j=0}^{\infty} \rho^j \nu_{i(t-j)}.
\tag{5}
$$

The final term can be recognized as a moving average process of infinite order (MA(∞)), which can be written as a first-order autoregressive (AR(1)) process in the form of $\varepsilon_{it} = \rho\varepsilon_{it-1} + \nu_{it}$ (Hamaker and Dolan 2009). This means we can write Equation (3) as

$$
y_{i0} = (1-\rho)^{-1}\eta_{1_i} - \rho(1-\rho)^{-2}\eta_{2_i} + \varepsilon_{i0}.
\tag{6}
$$

So, we can constrain the factor loadings of $\eta_{1_i} \to y_{i0}$ and $\eta_{2_i} \to y_{i0}$ to these fixed values, saving two degrees of freedom. For the constrained DPM, we drop the latent slope and are left with just $y_{i0} = (1-\rho)^{-1}\eta_i + \varepsilon_{i0}$. A four-wave version of the constrained observation-level model with individual effects is shown in Figure 1b.

Now, let us keep this result in mind while turning to the residual-level LCM-SR and RI-CLPM. A LCM-SR that is analogous to the ALT described in Equation (1) is
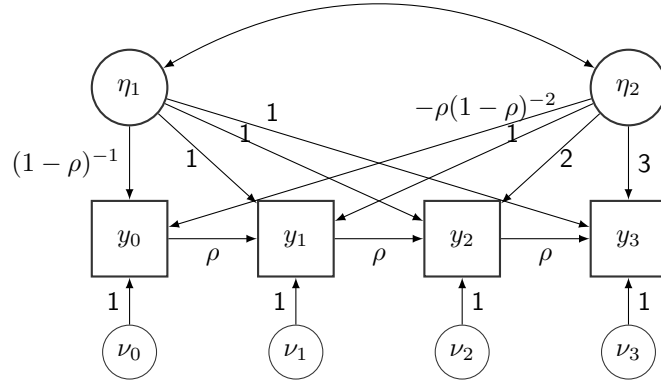
$$
\begin{aligned}
y_{it} &= \alpha_{1_i} + t\alpha_{2_i} + \varepsilon_{it}, \ t = 0, \dots, T \\
\varepsilon_{it} &= \rho\varepsilon_{it-1} + \nu_{it}, \ t = 1, \dots, T.
\end{aligned}
\tag{7}
$$

---

[5]Depending on whether we treat y_{i0} as fixed or let the variance and covariances be estimated as part of the model, we will have either zero or -2 degrees of freedom.

[6]As Hamaker (2005) mentions in endnote 2, we do not actually have to go back to $-\infty$ for this approach to make sense, because if $|\rho| < 1$, then ρ^j approaches zero as j gets larger and we can say at some point, probably not too far in the past (especially when $|\rho|$ is not too close to one), the y_{it-j} realization has essentially no effect on y_{it}.
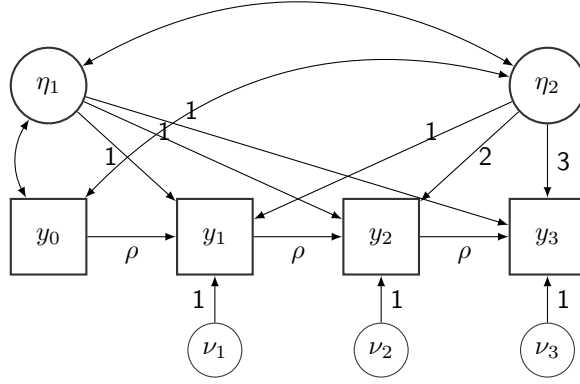
(a) Predetermined observation-level

(b) Constrained observation-level
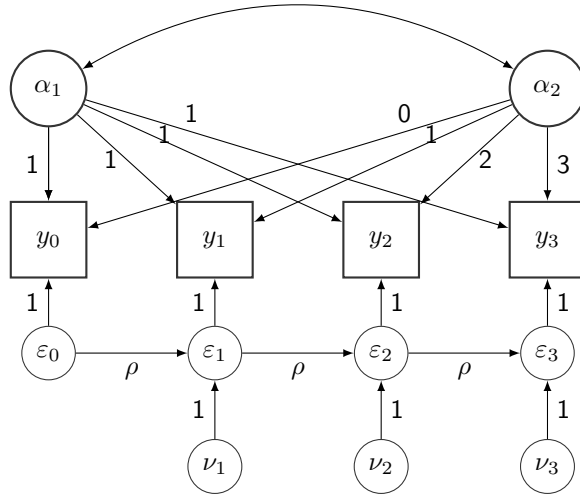
(c) Residual-level

Figure 1: Autoregressive models with individual effects

The RI-CLPM is essentially the same, but with no latent slope,

$$y_{it} = \alpha_i + \varepsilon_{it}, \ t = 0, \ldots, T$$
$$\varepsilon_{it} = \rho\varepsilon_{it-1} + \nu_{it}, \ t = 1, \ldots, T$$

(8)

where $\mathbb{E}(\nu_{it}) = \mathbb{E}(\varepsilon_{it}) = 0$ (Hamaker 2005). From these residual-level models, we can see the benefit is that the individual effects are separated from the autoregression, which makes them easier to interpret (Usami, Murayama, and Hamaker 2019). For example, the expected value of y_{it} in the LCM-SR would be

$$\mathbb{E}(y_{it}) = \mathbb{E}(\alpha_{1_i}) + t\,\mathbb{E}(\alpha_{2_i}) + \mathbb{E}(\varepsilon_{it})$$
$$= \mathbb{E}(\alpha_{1_i}) + t\,\mathbb{E}(\alpha_{2_i})$$

(9)

where $\mathbb{E}(\alpha_{1_i})$ can be thought of as the individual's intercept and $\mathbb{E}(\alpha_{2_i})$ can be thought of as their slope (Ou et al. 2016; Hamaker 2005). For the ALT and DPM, the same is not true, as

$$\mathbb{E}(y_{it}) = \mathbb{E}(\eta_{1_i}) + t\,\mathbb{E}(\eta_{2_i}) + \rho\,\mathbb{E}(y_{it-1}) + \mathbb{E}(\nu_{it})$$
$$= \mathbb{E}(\eta_{1_i}) + t\,\mathbb{E}(\eta_{2_i}) + \rho\,\mathbb{E}(\eta_{1_i} + (t-1)\eta_{2_i} + \rho y_{it-2} + \nu_{it-1}) + \mathbb{E}(\nu_{it})$$
$$= \ldots$$

(10)

means the expected value of the dependent variable is more complicated than $\mathbb{E}(\eta_{1_i})$ and $t\,\mathbb{E}(\eta_{2_i})$ and these cannot be interpreted as the intercept and slope factors. So, while the residual-level models allow for an easier interpretation of the latent individual effects, attempting to model the mean structure solely by the latent individual effects, as is common practice in LCMs and ALT models, is problematic because it ignores occasion effects (Zyphur et al. 2020b).

Note that we can take the residual-level models and insert the equation for the residuals into the equation for the observations. For the LCM-SR in what we can call *reduced-form*, we have

$$y_{it} = \alpha_{1_i} + t\alpha_{2_i} + \rho\varepsilon_{it-1} + \nu_{it}$$
$$= \alpha_{1_i} + t\alpha_{2_i} + \rho(y_{it-1} - \alpha_{1_i} - (t-1)\alpha_{2_i}) + \nu_{it}$$
$$= \alpha_{1_i} + t\alpha_{2_i} + \rho(y_{it-1} - \alpha_{1_i} - t\alpha_{2_i} + \alpha_{2_i}) + \nu_{it}$$
$$= (1-\rho)\alpha_{1_i} + \rho\alpha_{2_i} + t(1-\rho)\alpha_{2_i} + \rho y_{it-1} + \nu_{it} \ t = 1, \ldots, T.$$

(11)

From here, we define $\eta_{1_i} = (1-\rho)\alpha_{1_i} + \rho\alpha_{2_i}$ and $\eta_{2_i} = (1-\rho)\alpha_{2_i}$ (Hsiao 2014; Ou et al. 2016) to yield

$$y_{it} = \eta_{1_i} + t\eta_{2_i} + \rho y_{it-1} + \nu_{it}, \ t = 1, \ldots, T$$

(12)

which is clearly Equation (1) for the ALT for $t = 1, \ldots, T$. For the initial timepoint, we have

$$y_{i0} = \alpha_{1_i} + \varepsilon_{i0}$$

(13)

and, by rearranging the definition above, we have $\alpha_{1_i} = (1-\rho)^{-1}\eta_{1_i} - \rho(1-\rho)^{-2}\eta_{2_i}$, which we can substitute in for

8

$$y_{i0} = (1 - \rho)^{-1}\eta_{1_i} - \rho(1 - \rho)^{-2}\eta_{2_i} + \nu_{i0} \tag{14}$$

since without ε_{it-1}, we just have $\varepsilon_{i0} = \nu_{i0}$. This is exactly the equation for the first observation in the *constrained* ALT from Equation (6). The same goes for the RI-CLPM, which is equivalent to a constrained type of DPM. To see how, take the RI-CLPM in reduced-form

$$
\begin{aligned}
y_{it} &= \alpha_i + \rho\varepsilon_{it-1} + \nu_{it}, \ t = 1, \dots, T \\
&= \alpha_i + \rho(y_{it-1} - \alpha_i) + \nu_{it} \\
&= (1 - \rho)\alpha_i + \rho y_{it-1} + \nu_{it}
\end{aligned}
\tag{15}
$$

and define $\eta_i = (1 - \rho)\alpha_i$ for

$$y_{it} = \eta_i + \rho y_{it-1} + \nu_{it}, \ t = 1, \dots, T \tag{16}$$

which is the DPM for $t = 1, \dots, T$. For y_{i0}, we have

$$
\begin{aligned}
y_{i0} &= \alpha_i + \varepsilon_{i0} \\
&= (1 - \rho)^{-1}\eta_i + \nu_{i0}
\end{aligned}
\tag{17}
$$

which is the equation for y_{i0} in the constrained version of the DPM. The equivalence holds for cross-lagged or 'two-sided' models as well. To see how, take the RI-CLPM,

$$
\begin{aligned}
y_{it} &= \alpha_i^{(y)} + \varepsilon_{it} \\
x_{it} &= \alpha_i^{(x)} + \delta_{it}, \ t = 0, \dots, T, \\
\varepsilon_{it} &= \rho\varepsilon_{it-1} + \beta\delta_{it-1} + \nu_{it} \\
\delta_{it} &= \varphi\delta_{it-1} + \gamma\varepsilon_{it-1} + \upsilon_{it}, \ t = 1, \dots, T
\end{aligned}
\tag{18}
$$

where the superscripts (y) and (x) are meant to differentiate the individual intercepts of the two variables. The reduced-form RI-CLPM is achieved by inserting the equations for the residuals into the equations for the observed variables:

$$
\begin{aligned}
y_{it} &= \rho y_{it-1} + \beta x_{it-1} + (1 - \rho)\alpha_i^{(y)} - \beta\alpha_i^{(x)} + \nu_{it} \\
x_{it} &= \varphi x_{it-1} + \theta y_{it-1} + (1 - \varphi)\alpha_i^{(x)} - \gamma\alpha_i^{(y)} + \upsilon_{it}, \ t = 1, \dots, T.
\end{aligned}
\tag{19}
$$

We can turn the RI-CLPM into a constrained observation-level model (we could call it a constrained two-sided DPM) by a reparameterization such that, in each equation, we collect the linear combinations of $\alpha_i^{(y)}$ and $\alpha_i^{(x)}$ into one latent variable

$$y_{i0} = \varpi^{(y)}\eta_i^{(y)} + \omega^{(x)}\eta_i^{(x)} + \nu_{i0}$$
$$x_{i0} = \varpi^{(x)}\eta_i^{(x)} + \omega^{(y)}\eta_i^{(y)} + \upsilon_{i0},$$
$$y_{it} = \rho y_{it-1} + \beta x_{it-1} + \eta_i^{(y)} + \nu_{it} \tag{20}$$
$$x_{it} = \varphi x_{it-1} + \theta y_{it-1} + \eta_i^{(x)} + \upsilon_{it}, \ t = 1, \dots, T$$

where $\eta^{(y)} = (1 - \rho)\alpha_i^{(y)} - \beta\alpha_i^{(x)}$ and $\eta_i^{(x)} = (1 - \varphi)\alpha_i^{(x)} - \theta\alpha_i^{(y)}$ and

$$\varpi^{(y)} = \frac{1 - \varphi}{(1 - \rho)(1 - \varphi) - \beta\theta}, \ \omega^{(y)} = \frac{\beta}{(1 - \rho)(1 - \varphi) - \beta\theta},$$
$$\varpi^{(x)} = \frac{1 - \rho}{(1 - \rho)(1 - \varphi) - \beta\theta}, \ \omega^{(x)} = \frac{\pi}{(1 - \rho)(1 - \varphi) - \beta\theta}. \tag{21}$$

This shows that the full two-sided RI-CLPM can be re-expressed as a corresponding DPM in which the initial variables, y_{i0} and x_{i0} are each linear functions of both $\eta_i^{(y)}$ and $\eta_i^{(x)}$, the latent individual effects. The coefficients in those linear functions are themselves nonlinear functions of the cross-lagged and autoregressive coefficients.

It is possible to estimate the constrained model directly in an SEM package by imposing the constraints outlined in the equations above. And we could do the same to work out similar (but more complicated) re-expressions of the LCM-SR as a constrained two-sided ALT. But this would be ultimately unnecessary. The point of all this algebra is to make clear that the residual-level models are just the observation-level models with constraints on the relationships between the individual effects variable(s) and the initial observations.

Implications

The previous section showed that the residual-level models are equivalent to constrained observation-level models, and that these differ from the predetermined observation-level models *solely* in their treatment of the initial observations.

This has two important implications. First, it means the residual models offer no benefit in terms of the estimated within-unit coefficients. Afterall, when we write an observation-level model as a residual-level model, like in Equations (3) and (4), the coefficient(s) did not change, we just constrained the initial factor loading(s) and redefined the individual effects. As we will see below, if the assumptions concerning the initial conditions hold, then the residual and constrained models are essentially identical to the predetermined model in terms of estimated AR and CL effects; the predetermined model just uses more degrees of freedom in the process. Secondly, those assumption placed on the initial conditions are potentially very restrictive. In this section, we will discuss these assumptions in detail and look at situation in which they are likely not met.

For the sake of simplicity, let us focus on the DPM and RI-CLPM outlined in Equations (2) and (8). Most of what can be shown for the models with individual intercepts can be shown as well for those with individual intercepts and slopes, but it becomes more complicated and less intuitive. The empirical examples that follow will show that the results hold for the ALT and LCM-SR, as well.

From any regression model, we know that the model covariates must be independent of the error. If they are not, then estimated coefficients can be biased and inconsistent (Wooldridge 2002). Let us examine the initial equation in the constrained model (which is, as we have established, equivalent to the residual-level model) since it is the only thing differentiating the constrained from the predetermined model. For y_{i0}, we have $y_{i0} = (1 - \rho)^{-1}\eta_i + \nu_{i0}$. Here, the covariate is η_i, even though it is unobserved, and we have to establish

$\text{Cov}(\eta_i, \nu_{i0}) = 0$ for consistency (the stronger assumption, $\mathbb{E}(\nu_{i0}|\eta_i) = 0$ is needed for unbiasedness, but for most practical purposes we will be interested primarily in the covariance).

So, we can write

$$y_{i0} = (1 - \rho)^{-1}\eta_i + \nu_{i0}$$
$$\nu_{i0} = y_{i0} - (1 - \rho)^{-1}\eta_i \tag{22}$$

and

$$
\begin{aligned}
\text{Cov}(\nu_{i0}, \eta_i) &= \mathbb{E}[(y_{i0} - (1-\rho)^{-1}\eta_i)\eta_i] - \mathbb{E}(y_{i0} - (1-\rho)^{-1}\eta_i)\,\mathbb{E}(\eta_i) \\
&= \mathbb{E}(y_{i0}\eta_i - (1-\rho)^{-1}\eta_i^2) - \mathbb{E}(y_{i0})\,\mathbb{E}(\eta_i) + (1-\rho)^{-1}\,\mathbb{E}(\eta_i)^2 \\
&= \mathbb{E}(y_{i0}\eta_i) - \mathbb{E}(y_{i0})\,\mathbb{E}(\eta_i) - (1-\rho)^{-1}[\mathbb{E}(\eta_i^2) - \mathbb{E}(\eta_i)^2] \\
&= \text{Cov}(y_{i0}\eta_i) - (1-\rho)^{-1}\,\text{Var}(\eta_i)
\end{aligned} \tag{23}
$$

so that for this equation to equal zero, the covariance between the initial observation and the individual effects *must* equal $(1-\rho)^{-1}\,\text{Var}(\eta_i)$, which is the assumption of the constrained and residual-level models. We arrive at this covariance by assuming $\rho_t = \rho$ and $|\rho| < 1$ and continually substituting in the equation for the previous variable into the current one. But it is crucial to note that this is an *assumption*. Usami, Murayama, and Hamaker (2019) address this when they write: "Note that in the RI-CLPM, the initial within-person centered scores [...] are not correlated with the stable trait factors [...] by definition. The reason for this is that the model is based on decomposing the observed variance into between-person/trait-like variance and within-person/state-like variance. There is no reason to assume that an individual's temporal deviation at the first measurement from the person's trait score is dependent on that trait score (especially not when the measurements started at a relatively arbitrary moment in time)" (2019, 640).[7]

The problem is that the covariance between the initial observation and the individual effects is *unobserved*. So how can we know whether it is appropriate to fix the covariance to $(1-\rho)^{-1}\,\text{Var}(\eta_i)$? Fortunately, the assumptions surrounding the initial conditions impact the model-implied moments of the residual and constrained models, so that looking at the empirically observed moments can help diagnose whether these have been violated.

Namely, residual and constrained models assume the process y_{it} is *covariance stationary* and at *equilibrium*. Only then will $\text{Cov}(y_{i0}, \eta_i) = (1-\rho)^{-1}\,\text{Var}(\eta_i)$. Covariance stationary means that the expected value of the dependent variable, $\mathbb{E}(y_{it})$, is constant over time, the variance, $\text{Var}(y_{it})$, is constant over time (assuming the variance of the idiosyncratic error is constant over time, as well) and autocovariance, $\text{Cov}(y_{it}, y_{it+h})$, depends on the lag, h, but not on the specific point in time, t (Pickup 2015; Hamaker and Dolan 2009). Equilibrium means that the process has been going on long enough to have reached these conditions (Greenberg and Kessler 1982).[8]

To see why this is the case, let us look at the model-implied moments of the constrained model. We will focus on the constrained model rather than the residual-level one to emphasize the fact that stationarity and equilibrium boil down to the constraint placed on the initial factor loading and that freeing this factor loading gives us back the more flexible predetermined model.[9] In the Appendix, the model-implied moments of both the constrained and residual-level models will be shown in more detail.

---

[7]The notation used in Usami, Murayama, and Hamaker (2019) was removed to avoid confusion.

[8]As discussed in Hamaker and Dolan (2009), models with individual slopes are sometimes referred to as having a 'deterministic trend.' Because of this trend, the means of the observed variable are always increasing or decreasing. Such processes are referred to as 'trend stationary' if, once the trend has been removed (i.e., by differencing or detrending, Brüderl and Ludwig 2015; Hamaker and Dolan 2009), the means do not change over time, though the variance may still be time-varying.

[9]In fact, the model-implied moments are much easier to derive using the residual-level formulation.

Compared to the residual-level formulation, it is not as obvious in the constrained model why fixing the initial factor loading to $(1-\rho)^{-1}$ and the subsequent observations to 1.0 should imply constant means, variances and autocovariances. For the sake of exposition, let us assume there are no occasion effects, i.e., the mean structure can be modeled entirely with just the latent intercept. As noted earlier, this is often an unrealistic assumption, and it is probably wiser to avoid placing restrictions on the mean structure by either running the model on centered data, or allowing an intercept per point in time to be estimated freely (Zyphur et al. 2020b; Hamaker, Kuiper, and Grasman 2015). Assuming there are no occasion effects, taking the expectation of y_{it} gives us

$$
\begin{aligned}
\mathbb{E}(y_{it}) &= \mathbb{E}(\eta_i + \rho y_{it-1} + \nu_{it}) \\
&= \mathbb{E}(\eta_i) + \rho \, \mathbb{E}(\eta_i + \rho y_{it-2} + \nu_{it-1}) \\
&= \mathbb{E}(\eta_i) + \rho \, \mathbb{E}(\eta_i) + \rho^2 \, \mathbb{E}(\eta_i + \rho y_{it-3} + \nu_{it-2}) \\
&= \ldots \\
&= \sum_{j=0}^{\infty} \rho^j \, \mathbb{E}(\eta_i) \\
&= (1-\rho)^{-1} \, \mathbb{E}(\eta_i)
\end{aligned}
\tag{24}
$$

since $\mathbb{E}(\nu_{it}) = 0$, $\forall \, t$. Again, as long as $\rho_t = \rho$ and $|\rho| < 1$, then this should hold for any t.

By proceeding in a similar fashion (see the Appendix for more detail), we get the other model-implied moments:

$$
\begin{aligned}
\mathrm{Var}(y_{it}) &= (1-\rho)^{-2} \, \mathrm{Var}(\eta_i) - (\rho^2-1)^{-1} \, \mathrm{Var}(\nu_{it}) \\
\mathrm{Cov}(y_{it}, y_{it-1}) &= (1-\rho)^{-2} \, \mathrm{Var}(\eta_i) - \rho(\rho^2-1)^{-1} \, \mathrm{Var}(\nu_{it})
\end{aligned}
\tag{25}
$$

where we can find the model-implied covariance between observations separated by more than one point in time by replacing the ρ in front of the brackets in the second term with ρ^j, where j is the distance in time between the two.

As noted above, the covariance between the observed variable and the individual intercept is also assumed constant at

$$
\mathrm{Cov}(y_{it}, \eta_i) = (1-\rho)^{-1} \, \mathrm{Var}(\eta_i).
\tag{26}
$$

This can be shown by continual substitution of the previous equations into the current one, or even more easily by looking at Figure 2, where by path tracing, we can see $\mathrm{Cov}(y_3, \eta)$, for example, is $\mathrm{Var}(\eta) + \rho \, \mathrm{Var}(\eta) + \rho^2 \, \mathrm{Var}(\eta) + \rho^3 (1-\rho)^{-1} \, \mathrm{Var}(\eta) = (1-\rho)^{-1} \, \mathrm{Var}(\eta)$.[10]

So, we can look at the empirical means, variances and covariances of the observed variable over time and diagnose whether we should fix the initial covariance, $\mathrm{Cov}(y_{i0}, \eta_i)$ to $(1-\rho)^{-1} \, \mathrm{Var}(\eta_i)$, or whether it would be preferable to switch to a predetermined model and allow this to be estimated freely. The following section will show using a brief simulation how inspecting the empirical means, variances and covariances can help determine whether a residual or constrained model is appropriate, or not.

---

[10]Consider an example where $\rho = 0.5$ and $\mathrm{Var}(\eta) = 2$, then $2 + 0.5 \cdot 2 + 0.5^2 \cdot 2 + 0.5^3 (1-0.5)^{-1} \cdot 2 = (1-0.5)^{-1} \cdot 2 = 4$.
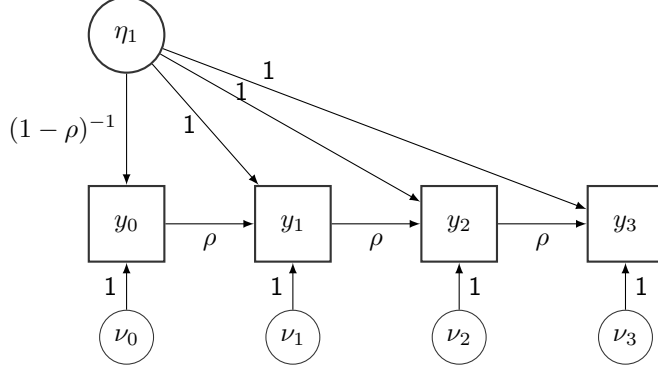
Figure 2: Constant covariance between y_{it} and η_i in the constrained model

A brief simulated example

To demonstrate what has been shown up until now, we can look at a brief example using simulated data based on the equation $y_{it} = \eta_i + \rho y_{it-1} + \nu_{it}$, with $n = 10{,}000$, $t = 15$, $\rho = 0.5$, $y_0 \sim N(0, 1)$, $\eta \sim N(5, 1)$ and $\nu_t \sim N(0, 0.25)$. These values were chosen mostly arbitrarily, though the fact that $\mathbb{E}(\nu_t) = 0$ means there are no occasion effects. Code for this simulated example, as well as the empirical examples to follow can be found at https://github.com/henrik-andersen/equivalent-approaches-clpm/tree/main/r-files.

From the values used in the simulation, and using what was worked out above, the stationary state should be

$$\mathbb{E}(y_{it}) = (1 - \rho)^{-1} \mathbb{E}(\eta_i) = (1 - 0.5)^{-1} \cdot 5 = 10, \tag{27}$$

$$\text{Var}(y_{it}) = (1 - \rho)^{-2} \text{Var}(\eta_i) - \frac{\text{Var}(\nu_{it})}{\rho^2 - 1} = (1 - 0.5)^{-2} \cdot 1 - \frac{0.25}{0.5^2 - 1} = 4.33, \tag{28}$$

$$\text{Cov}(y_{it}, y_{it-1}) = (1 - \rho)^{-2} \text{Var}(\eta_i) - \rho \frac{\text{Var}(\nu_{it})}{\rho^2 - 1} = (1 - 0.5)^{-2} \cdot 1 - 0.5 \frac{0.25}{0.5^2 - 1} = 4.17, \tag{29}$$

$$\text{Cov}(y_{it}, \eta_i) = (1 - \rho)^{-1} \text{Var}(\eta_i) = (1 - 0.5)^{-1} \cdot 1 = 2. \tag{30}$$

Figure 3 shows five randomly drawn cases from the simulated data. As can be seen, the process begins with $y_{i0} \sim N(0, 1)$ and, because the process is stationary, quickly moves towards its equilibrium state. This sets in after around the 5th or 6th timepoint. After that, the process is centered around $\mathbb{E}(y_{it}) = 10$, with a constant variance. Each individual has their own expected value, based on their own value of η_i. For example, case number 1039, (circle) has a corresponding individual effect of $\eta_{i=1039} = 3.989$. Thus, their equilibrium state is centered over time around $(1 - \rho)^{-1} \eta_{1039} = (1 - 0.5)^{-1} 3.989 = 7.978$.

Once the process has reached equilibrum, the model-implied moments discussed above are appropriate. Importantly, once the process is at equilibrium, $\text{Cov}(y_{it}, \eta_i) = (1 - \rho)^{-1} \text{Var}(\eta_i)$. However, this equilibrium state is not reached immediately, but rather requires a 'spin-up' phase. I.e., in the initial phase, up until about $t = 6$ or so, the covariance, $\text{Cov}(y_{it}, \eta_i)$, does not equal $(1 - \rho)^{-1} \text{Var}(\eta_i)$. For example, at $t = 2$, the covariance is only around $1.53 \neq (1 - 0.5)^{-1} \cdot 1 = 2$. This means $\text{Cov}(\nu_{it}, \eta_i) \neq 0$ and thus the coefficient of interest, ρ, should we attempt to model the process using the constrained or residual-level model, will be biased. Again, all of this boils down to the treatment of the initial observation, and the constraint of fixing the initial factor loading to $(1 - \rho)^1$ so that $\mathbb{E}(y_{it})$, $\text{Var}(y_{it})$, $\text{Cov}(y_{it}, y_{it+h})$ and $\text{Cov}(y_{it}, \eta_i)$ are all constant.

Usami, Murayama, and Hamaker (2019) even comment on this, saying the "treatment of the initial observation is an important theoretical and practical problem for some cross-lagged models like the ALT model that includes common factors but do not separate stable between-person differences from within-person fluctuations over time" (Usami, Murayama, and Hamaker 2019, 641). This seems to suggest that the RI-CLPM is
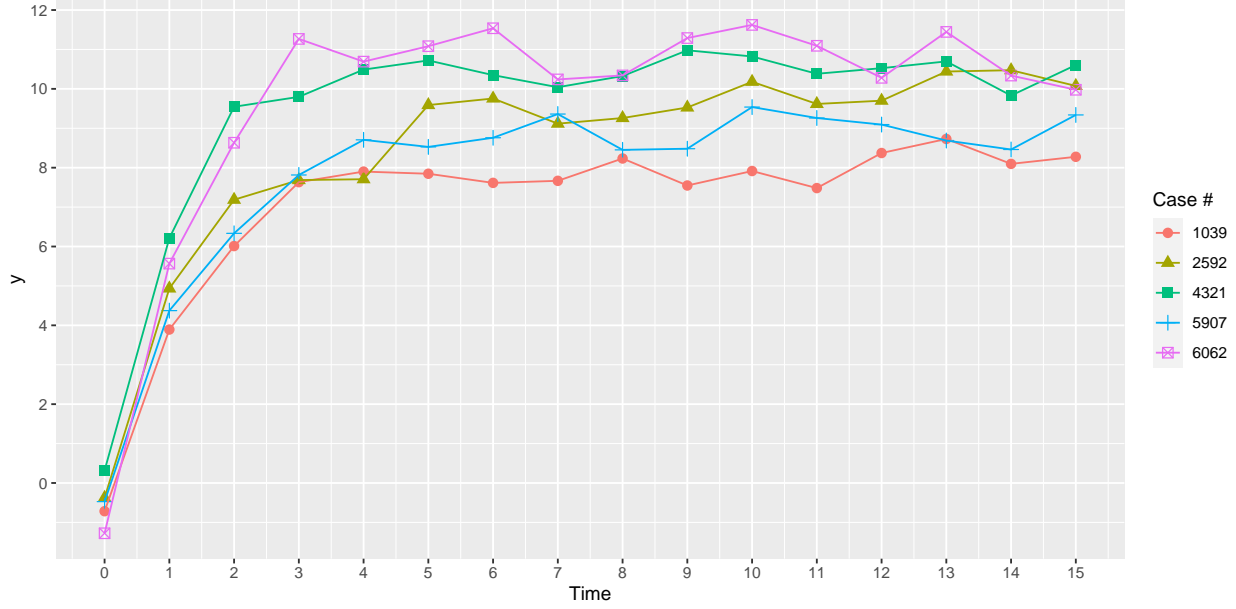
13

Figure 3: Simulated example, subsample of five randomly chosen cases

spared from having to deal with this issue, but as we have seen in the above example, it is absolutely not; it just ignores it. For the residual-level models to be appropriate, we must be sure that the process we are modeling is indeed stationary and at equilibrium. If it is not, the residual-level models will return biased results for the AR effects as well as potentially those of the other model covariates.

Even though the covariance between the observed variable and the individual intercept is arguably the most important for the unbiasedness of the estimated coefficients, it is unobserved, so it is impossible to assess the assumption surrounding this covariance in applied settings. But as we have seen, the way we treat the initial observation affects the model-implied moments for which we have observed couterparts, as well. This means that the means, variances and autocovariances of the observed variable can help us diagnose situations in which it is not appropriate to assume stationarity and equilibrium.

In applied cases with large sample sizes, it may not be feasible to look at the trajectories of each of the observed units, and plotting a subset of them, like in Figure 3, may not always reveal unambiguous information. So, it may be more helpful to simply plot the means, variances and autocovariances of the entire sample. Figure 4 does this, and summarizes what was shown in Figure 3 for the entire simulated sample. The dots represent the observed means and variances, while the horizontal lines represent the model-implied means and variances of the residual- and constrained models, which are constants.

From this as well, we can see that the equilibrium values outlined above set in around the 7th or so timepoint. But at $t = 3$, for example, the observed mean is only 8.74, while the constrained and residual model-implied value is $\mathbb{E}(y_{it}) = 10$, the variance is 3.42 (model-implied: $\mathrm{Var}(y_{it}) = 4.33$), the covariance between y_{i3} and the previous timepoint, y_{i2}, is just 2.84 (model-implied: $\mathrm{Cov}(y_{it}, y_{it-1}) = 4.17$) and the covariance between the observed variable and the individual effects (normally not directly observable) is 1.76 (model-implied: $\mathrm{Cov}(y_{it}, \eta_i) = 2$). The further along in time we go, the closer these values come to the model-implied counterparts, but in the initial phase, these values are still quite far off, leading to misfit and ultimately also bias, since we are trying to establish the within-unit effects by removing the covariance between the model covariates (e.g., the lagged dependent variable) and the individual effects. For example, if we were focused on the autoregressive effect, ρ, then

14

(a) Observed and model-implied means

(b) Observed and model-implied variances

(c) Observed and model-implied autocovariances

(d) Observed and model-implied covariances with individual intercepts (only observed in simulations)
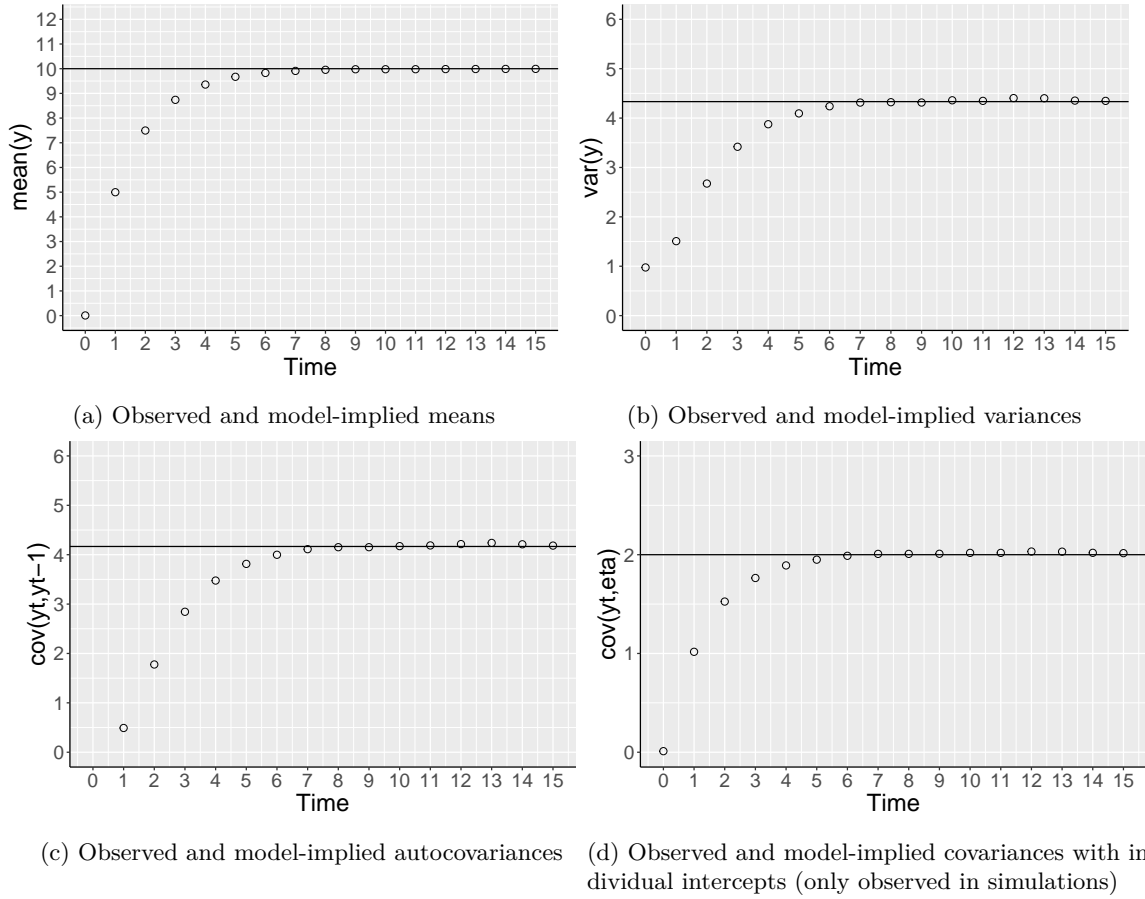
Figure 4: Difference between observed and residual-level model-implied means, variances and covariances (simulated example, dots are observed, lines are model-implied)

$$
\begin{aligned}
\mathrm{Cov}(y_{it}, y_{it-1}) &= \mathbb{E}(y_{it} y_{it-1}) - \mathbb{E}(y_{it})\,\mathbb{E}(y_{it-1}) \\
&= \mathbb{E}[(\rho y_{it-1} + \eta_i + \nu_{it}) y_{it-1}] - \mathbb{E}(\rho y_{it-1} + \eta_i + \nu_{it})\,\mathbb{E}(y_{it-1}) \\
&= \mathbb{E}(\rho y_{it-1}^2 + \eta_i y_{it-1} + \nu_{it} y_{it-1}) - \rho\,\mathbb{E}(y_{it-1})^2 - \mathbb{E}(\eta_i)\,\mathbb{E}(y_{it-1}) - \mathbb{E}(\nu_{it})\,\mathbb{E}(y_{it-1}) \\
&= \rho(\mathbb{E}(y_{it-1}^2) - \mathbb{E}(y_{it-1})^2) + \mathbb{E}(\eta_i y_{it-1}) - \mathbb{E}(\eta_i)\,\mathbb{E}(y_{it-1}) \\
&= \rho\,\mathrm{Var}(y_{it-1}) + \mathrm{Cov}(\eta_i, y_{it-1}) \\
\hat{\rho} &= \frac{\mathrm{Cov}(y_{it}, y_{it-1}) - \mathrm{Cov}(\eta_i, y_{it-1})}{\mathrm{Var}(y_{it-1})}
\end{aligned}
\tag{31}
$$

(Hayduk et al. 2003), where $\hat{\rho}$ would be the estimated autoregressive coefficient. Obviously, if we fix the value of $\mathrm{Cov}(\eta_i, y_{it-1})$ to $(1-\rho)^{-1}\,\mathrm{Var}(\eta_i)$ and our assumption about stationarity and/or equilibrium does not hold, then we will be subtracting off too much or too little (in the simulation above, it is too much since $\mathrm{Cov}(\eta_i, y_{it-1}) < (1-\rho)^{-1}\,\mathrm{Var}(\eta_i)$ in the first few realizations) from the autocovariance and the estimated coefficient for ρ will be biased. If we consider the fixed effects regression in SEM, which works by allowing the model covariates to freely covary with the latent individual effects, then we can go so far as to say that the residual-level models *are not fixed effects models* in the sense that they control for all time-invariant unobserved heterogeneity *unless the data is stationary and at equilibrium*. We can extrapolate from here what an happen to the effects of other covariates. Essentially, whenever the covariance between the lagged dependent variable and the individual effects is wrong, the the effects of the other covariates will likely be wrong. And in two-sided residual-level models, all this applies equally to the other time-varying variable. If it is not at equilibrium, then we can expect the effects to be biased in the other direction, as well.

Example: Use of cigarettes and alcohol in adolescents and young adults

In this section, I use the example of cigarette smoking amongst adolescents and young adults (as well as the reciprocal relations with drinking and the effect of parental monitoring) to demonstrate that:

1. the residual-level models are identical to their constrained observation-level counterparts, and
2. that the implicit assumptions placed on the initial conditions can have serious consequences for the estimated coefficients.

While smoking amongst adolescents is not necessarily a strictly psychological issue, it and other related behaviours are often looked at in combination with psychological and sociological constructs. Keijsers (2016), for example, used the RI-CLPM to examine the reciprocal relations between parental monitoring and adolescent problem behaviours, which can encompass substance abuse and smoking, risky sexual behaviour, shoplifting, etc. Other studies, like that of Barnett et al. (2013); Nkansah-Amankra and Minelli (2016); Bierhoff et al. (2019); Tarter et al. (2006) look at the effects of future time perspective, depression, anxiety, ADHD, etc. on substance use and problem behaviours in adolescents. Lloyd-Richardson et al. (2002) looked at the effect of psychological and social influences at various life-stages on substance abuse and Nelson, Ryzin, and Dishion (2014) compared the trajectories of alcohol, marijuana nd smoking behaviour of 12- to 24-year-olds, based on demographic characteristics like sex and race. Another sociological example includes Schuler et al. (2019), who looked at perceived substance abuse amongst peers and family members and its influence on adolescents' own substance use and abuse.

I use 14 (but focus on six) waves of the National Longitudinal Survey of Youth 1997 (NLYS97) study, spanning the years 1997 to 2011 (Bureau of Labor Statistics 2017). These are the years for which there is yearly data on both smoking and drinking (after 2011, the variables are measured every two years). The NLYS study (especially the previous version started in 1979) has been used in some prominent papers on panel SEM, see, for example, K. Bollen and Brand (2010); P. D. Allison (2009), and it is freely available to

download. The NLYS97 study encompasses a probability sample of 8984 youths between the ages of 12 and 17 in 1997 in the United States.

The variable I have renamed `cig` is a measure of the typical number of cigarettes smoked per smoking occasion in the last 30 days, measured on an interval scale from zero (none) to 15, where the final category encompasses 15 cigarettes or more per occasion. This variable demonstrates nicely the issue described above.

It is usually plausible to think that each individual has a relatively stable 'usual' pattern of smoking. Some tend to smoke more than others, some smoke a lot and some do not smoke at all, and these differences are influenced by a range of factors, likely including family norms, childhood experiences, predisposition, etc. Individuals may even follow trajectories, where they smoke more or less over time. Some of the influences of these stable components may be measurable directly in a survey, but other like, say, potential genetic factors, would be difficult or impossible to feasibly observe.

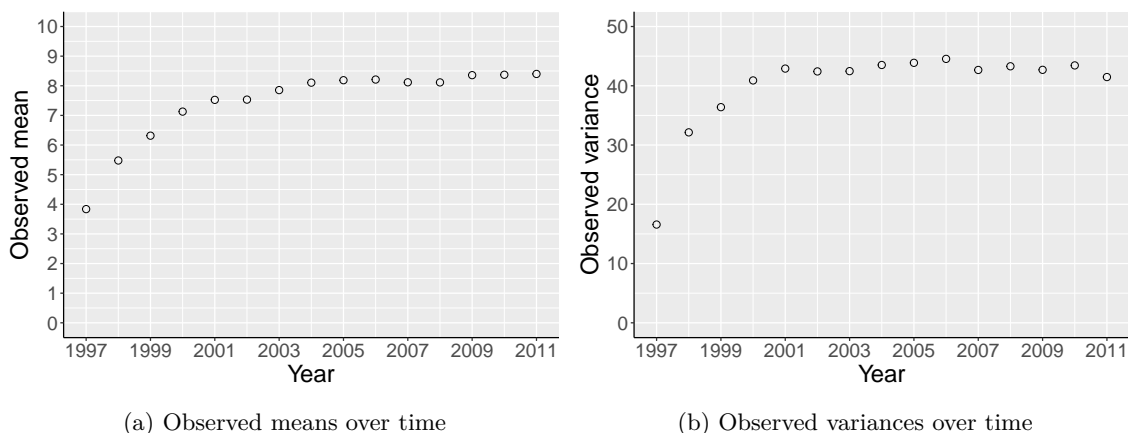(a) Observed means over time

(b) Observed variances over time

Figure 5: Number of cigarettes smoked per occasion in the last 30 days

However, young children, except in rare cases, do not smoke cigarettes. An individual's pattern of smoking typically only emerges in adolescence and, for a given youth, this pattern likely changes over time. The development of smoking behaviour therefore likely follows a similar trajectory to the one shown in Figure 3. At ages 12-14, the age at which the majority of the NLYS97 sample was surveyed for the first time, we would not expect much variance. Most youths in this age group do not smoke much, if at all. The overall average number of cigarettes smoked per occasion will be low at the first timepoint, as well. As these youths become older, some of their smoking behaviour will change; some start going to parties where they might smoke many cigarettes on a given occasion. Therefore, the variance as well as the average number of cigarettes smoked per timepoint will both increase over time. We would not necessarily expect these to increase indefinitely, but rather settle into a 'normal' pattern of behaviour. At this point, the process will have reached equilibrium, where it could stay for some time, though likely not forever.

Figure 5a shows the actual observed mean of `cig` from 1997 to 2011, while Figure 5b shows the overall variance over this timespan. As can be seen, the observed variances do, in fact, increase rapidly over the years 1997 to about 2001. After that, the variance stabilizes and remains fairly constant. The means follow a similar trajectory: the average number of cigarettes smoked per occasion is around 4 in 1997, when respondents were between the ages of 12-14. The means increase rapidly to eight or nine cigarettes per occasion over the years from about 1997 to 2002. The average remains fairly stable for the rest of the observed timespan, with perhaps a slight positive trend.

This means that over the course of these 14 years, the observed variable `cig` does not reach equilibrium until around 2002-2005. We do not see the individual trajectories in Figure 5 (with a sample size of nearly 9000 it would be unfeasible), but plotting the observed means and variances gives us an impression of the overall trend. So, what happens when we run an RI-CLPM or LCM-SR on the data from 1997 to 2002?

To demonstrate what has been discussed up until now, I run four 'scenarios': 1) a simple autoregressive one-sided model with individual intercepts, 2) an extended autoregressive one-sided model with individual

intercepts and a time-varying and time-invariant covariate, 3) an extended one-sided autoregressive model with individual intercepts and slopes and 4) a full cross-lagged model with individual intercepts. For each of these four types of models, I run a residual-level model, a constrained model and a predetermined model. I will show that in each case, the residual and constrained model are equivalent. The predetermined model differs from the other two in the way it treats the initial observations, and so it will differ, sometimes substantially, from the other two. Within each of the four scenarios, the residual and constrained models are nested within the predetermined one, so a chi-square difference test will be carried out to help put the differences into perspective. Throughout, I use full-information maximum likelihood (FIML) to account for missing values and panel attrition.

Scenario 1: Simple autoregressive one-sided models with individual intercepts

Let us start with the simplest one-sided autoregressive model with individual intercepts and no covariates. For the residual-level model, we have:

$$
\begin{aligned}
cig_{it} &= \alpha_i + \varepsilon_{it}, \ t = 1997, \ldots, 2002, \\
\varepsilon_{it} &= \rho \varepsilon_{it-1} + \nu_{it}, \ t = 1998, \ldots, 2002
\end{aligned}
\tag{32}
$$

where $\mathrm{Cov}(\varepsilon_{i97}, \alpha_i) = 0$ by assumption. For the constrained observation-level model, we have:

$$
\begin{aligned}
cig_{i97} &= (1 - \rho)^{-1} \eta_i^{(c)} + \nu_{i97}^{(c)}, \\
cig_{it} &= \rho cig_{it-1} + \eta_i^{(c)} + \nu_{it}^{(c)}, \ t = 1998, \ldots, 2002
\end{aligned}
\tag{33}
$$

where we can use the superscript (c) to denote the constrained model. For the predetermined observation-level model, we have:

$$
cig_{it} = \rho^{(p)} cig_{it-1} + \eta_i^{(p)} + \nu_{it}^{(p)}, \ t = 1998, \ldots, 2002
\tag{34}
$$

where (p) denotes the predetermined model and $\mathrm{Cov}(cig_{i97}, \eta_i^{(p)}) \neq 0$, i.e., the individual effects are allowed to covary with the initial observation.

Table 1: Model Results: Simple AR(1) models with individual intercepts

		Residual-level		Constrained observation-level		Predetermined observation-level	
		Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
		Regression Slopes					
$e_t \mid cig_t$	$\leftarrow$						
	$e_{t-1} \mid cig_{t-1}$	0.572***	0.013	0.572***	0.013	0.352***	0.017
		(Residual) Variances					
	$\cdot\ e_{97} \mid cig_{97}$	8.513***	0.614	8.513***	0.616	15.938***	0.557
	$\cdot\ e_{98} \mid cig_{98}$	21.685***	0.787	21.685***	0.788	17.722***	0.721
	$\cdot\ e_{99} \mid cig_{99}$	20.399***	0.692	20.399***	0.692	16.492***	0.657
	$\cdot\ e_{00} \mid cig_{00}$	21.994***	0.710	21.994***	0.710	18.238***	0.690
	$\cdot\ e_{01} \mid cig_{01}$	23.151***	0.721	23.151***	0.721	20.207***	0.714
	$\cdot\ e_{02} \mid cig_{02}$	21.512***	0.662	21.512***	0.662	18.604***	0.641
	$alpha_1 \mid eta_1$	8.123***	0.593	1.486***	0.188	8.504***	0.607
		(Residual) Coariances					
$alpha_1 \mid eta_1$	$\longleftrightarrow$						
	$\cdot\ e_{97} \mid cig_{97}$	0.00+	–	–	–	4.906***	0.391
		Fit Indices					
	χ^2(df)	222.240(13)		222.240(13)		37.246(12)	
	CFI	0.957		0.957		0.995	
	RMSEA	0.058		0.058		0.021	

Estimator: FIML, $n = 4759$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. +: fixed parameter. A dot ($\cdot$) indicates a residual (co)variance. `e` is the structured residual for `cig`. `alpha1` and `eta1` are latent intercepts for residual- and observation-level models, respectively.

The `lavaan` code, as well as the fitted models (i.e., detailed summaries of the model results, saved as `R` objects) can be found at https://github.com/henrik-andersen/equivalent-approaches-clpm/tree/main/r-files. Table 1 shows the results of this first example. As we can see, the residual-level model (the fitted model is called `m1a.fit` below) is exactly equivalent to the constrained observation-level model (`m1b.fit`) with 13 degrees of freedom, $\chi^2 = 222.240$, $p < 0.001$, and an estimated coefficient of $\rho = 0.572$, $p < 0.001$. This model suggests a relatively strong autoregressive effect (standardized coefficients ranged between 0.338 and 0.581) that is highly significant.

The variance of the individual effects in the residual-level model differs from the variance in the constrained observation-level model, as it should. We can recover one from the other easily, though. For example, we defined above $\eta^{(c)} = (1 - \rho)\alpha$, so $\mathrm{Var}(\eta^{(c)}) = (1 - \rho)^2 \mathrm{Var}(\alpha) = (1 - 0.572)^2 \cdot 8.123 = 1.488$ which is the estimated variance of the individual effects in the constrained model (with a slight deviation due to rounding).

The predetermined observation-level model (`m1c.fit`), on the other hand, differs from the other two in exactly one degree of freedom: the covariance between the individual effects and the initial observation is estimated freely, using one degree of freedom in the process. Crucially, the estimated autoregressive effect in the predetermined model is much smaller at $\rho^{(p)} = 0.352$, $p < 0.001$ and the model fit is noticeably better with a chi-square statistic of just 37.246. A χ^2 difference test shows the difference between the predetermined, on the one hand, and residual and constrained models, on the other hand, is significant (below is the output from the `anova()` function in `R` used on the fitted `lavaan` models):

```
## Chi-Squared Difference Test
##
##         Df   AIC   BIC   Chisq Chisq diff Df diff Pr(>Chisq)
## m1c.fit 12 84891 84988  37.246
## m1a.fit 13 85074 85164 222.240     184.99       1  < 2.2e-16 ***
## m1b.fit 13 85074 85164 222.240       0.00       0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From this, we can conclude that fixing the covariance between the initial observation and the individual intercept to $(1 - \rho)^{-1} \mathrm{Var}(\eta_i^{(c)})$ (or, equivalently $\mathrm{Var}(\alpha_i)$) leads to a model that fits significantly worse than a model that simply estimates this covariance freely.

Scenario 2: Extended autoregressive one-sided models with individual intercepts

Now let us move on to an extended one-sided model with individual intercepts and two covariates besides the lagged dependent variable: `drink` for the typical amount of alcoholic drinks consumed per session in the last 30 days (measured similarly to `cig` on an interval scale ranging from 0: no alcoholic drinks to 15: 15 alcoholic drinks or more per session) and `mon` for the degree of parental monitoring by the respondent's mother as reported by the respondent in 1997 (ranging from 0 to 16, with higher values indicating more parental monitoring). Perhaps we are interested testing the hypothesis that drinking is a gateway to smoking, and we are also interested in the role of parental monitoring as it affects smoking behaviour in adolescents. The variable `drink` is a time-varying covariate, while `mon` varies only across respondents but not over time. While it is plausible that the degree of parental monitoring can change over time, it was not measured throughout the entire panel, and it will be treated as time-invariant here.

For the residual-level, we have:

$$
\begin{aligned}
cig_{it} &= \alpha_i + \varepsilon_{it}, \ t = 1997, \ldots, 2002, \\
\varepsilon_{it} &= \rho\varepsilon_{it-1} + \beta drink_{it-1} + \gamma mon_i + \nu_{it}, \ t = 1998, \ldots, 2002
\end{aligned}
\tag{35}
$$

again, where $\text{Cov}(\varepsilon_{i97}, \alpha_i) = 0$. For the constrained observation-level model, we have:

$$
\begin{aligned}
cig_{i97} &= (1 - \rho)^{-1}\eta_i^{(c)} + \nu_{i97}^{(c)}, \\
cig_{it} &= \rho cig_{it-1} + \beta drink_{it-1} + \gamma mon_i + \eta_i^{(c)} + \nu_{it}^{(c)}, \ t = 1998, \ldots, 2002
\end{aligned}
\tag{36}
$$

and for the predetermined model, we have:

$$
cig_{it} = \rho^{(p)} cig_{it-1} + \beta^{(p)} drink_{it-1} + \gamma^{(p)} mon_i + \eta_i^{(p)} + \nu_{it}^{(p)}, \ t = 1998, \ldots, 2002
\tag{37}
$$

where $\text{Cov}(cig_{i97}, \eta_i^{(p)}) \neq 0$.

Table 2: Model Results: Extended AR(1) models with individual intercepts

		Residual-level		Constrained observation-level		Predetermined observation-level	
		Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
				Regression Slopes			
e_t/cig_t	$\leftarrow$						
	e_{t-1}/cig_{t-1}	0.557***	0.015	0.557***	0.015	0.345***	0.017
	$drink_{t-1}$	0.025	0.020	0.024	0.021	0.043*	0.021
	mon	-0.121***	0.019	-0.121***	0.019	-0.150***	0.024
				(Residual) Variances			
	$\cdot\, e_{97}/\cdot\, cig_{97}$	8.818***	0.621	8.818***	0.623	16.046***	0.564
	$\cdot\, e_{98}/\cdot\, cig_{98}$	21.018***	0.766	21.018***	0.769	17.263***	0.702
	$\cdot\, e_{99}/\cdot\, cig_{99}$	20.235***	0.685	20.235***	0.686	16.454***	0.650
	$\cdot\, e_{00}/\cdot\, cig_{00}$	21.624***	0.698	21.624***	0.698	18.044***	0.677
	$\cdot\, e_{01}/\cdot\, cig_{01}$	22.947***	0.713	22.947***	0.713	20.093***	0.706
	$\cdot\, e_{02}/\cdot\, cig_{02}$	21.493***	0.661	21.493***	0.662	18.654***	0.638
	$alpha_1/eta_1$	8.070***	0.587	1.582***	0.205	8.319***	0.591
				(Residual) Coariances			
$alpha_1/eta_1$	$\longleftrightarrow$						
	$\cdot\, e_{97}/\cdot\, cig_{97}$	0.00$^+$	–	–	–	4.823***	0.389
				Fit Indices			
	χ^2(df)	334.438(36)		334.438(36)		147.963(35)	
	CFI	0.957		0.957		0.984	
	RMSEA	0.032		0.032		0.020	

Estimator: FIML, $n = 8305$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. $^+$: fixed parameter. A dot ($\cdot$) indicates a residual (co)variance. `e` is the structured residual for `cig`. `alpha1` and `eta1` are latent intercepts for residual- and observed-level models, respectively.

From Table 2, we again see that the extended residual-level model (`m2a.fit`) and constrained observation-level model (`m2b.fit`) are equivalent in terms of model fit and degrees of freedom. The estimated coefficients are the same as well (except for a discrepancy in the third position after the comma for the coefficient for $drink_{it-1}$ and its standard error).[11] As before, the estimated autoregressive coefficient in the residual-level and constrained model, $cig_{it-1} \rightarrow cig_{it}$, is much stronger than that of the predetermined model. The residual and constrained models also suggest that the lagged effect of alcohol consumption on cigarette smoking is non-significant. The predetermined model (`m2c.fit`), on the other hand, suggests this effect is significant at the 5% level. If we were to compare the three models, we would probably conclude that the lagged effect of alcohol is not worth interpreting substantively, since it is small in any case and only significant in the predetermined model. But most researchers do not compare the residual with the predetermined approach. A researcher who had ran the residual or constrained model would conclude the effect of alcohol is not significantly different than zero. A researcher who had ran the predetermined model could say there is a small but significant effect of alcohol consumption on cigarette smoking. As for the time-invariant parental monitoring variable, the effect is negative in all models and significant, but slightly stronger in the predetermined model.

Again, the models differ in only one degree of freedom; the covariance between the initial observation and the individual intercept. This one difference has a large impact, however. The results of the second chi-square difference test show that the predetermined model displays a significantly better fit.

```
## Chi-Squared Difference Test
##
##          Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## m2c.fit 35 190650 191037 147.96
## m2a.fit 36 190835 191214 334.44     186.47       1  < 2.2e-16 ***
## m2b.fit 36 190835 191214 334.44       0.00       0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Remember the previous section concluded by arguing the residual-level models are not fixed effects models, in the sense that they do not control for all time-invariant unobserved heterogeneity, unless the assumptions concerning the initial conditions hold. This can be shown easily by attempting to allow the time-invariant covariate `mon` to covary with the latent individual intercepts. Afterall, it is widely known that in fixed effects models, time-invariant covariates are perfectly collinear with the individual effects (K. Bollen and Brand 2010). Trying to allow the time-invariant covariates and the latent individual effects to covary will cause an error because it is then not possible to get separate estimates for their respective effects.

If we try to estimate the residual and constrained models (called `m2ax` and `m2bx`) again using maximum likelihood, but where `mon` is allowed to covary with the individual intercepts, the models run without issue. It follows therefore, that these cannot be considered fixed effects models. However, the predetermined model (`m2cx`) — a fixed effects model — rightfully returns an error:

```
In lav_model_vcov(lavmodel = lavmodel, lavsamplestats = lavsamplestats,  :
  lavaan WARNING:
    The variance-covariance matrix of the estimated parameters (vcov)
    does not appear to be positive definite! The smallest eigenvalue
    (= 1.956191e-14) is close to zero. This may be a symptom that the
    model is not identified.
```

---

[11]I believe this small discrepancy has something to due with using FIML to estimate the models, as I have not noticed such discrepancies before using maximum likelihood.

Scenario 3: Extended autoregressive one-sided models with individual intercepts and slopes

Next, we can add a latent slope to the extended models in the previous scenario. For the residual-level model, something akin to a one-sided LCM-SR, we have

$$
\begin{aligned}
cig_{it} &= \alpha_{1_i} + \pi\alpha_{2_i} + \varepsilon_{it}, \ t = 1997, \ldots, 2002, \\
\varepsilon_{it} &= \rho\varepsilon_{it-1} + \beta drink_{it-1} + \gamma mon_i + \nu_{it}, \ t = 1998, \ldots, 2002
\end{aligned}
\tag{38}
$$

and $\pi = t - 1997$, so that at the first observation in 1997, the effect of the latent slope is $1997 - 1997 = 0$. Importantly, $\text{Cov}(\alpha_{1_i}, \varepsilon_{i97}) = \text{Cov}(\alpha_{2_i}, \varepsilon_{i97}) = 0$. For the equivalent constrained model, we have

$$
\begin{aligned}
cig_{i97} &= (1-\rho)^{-1}\eta_{1_i}^{(c)} - \rho(1-\rho)^{-2}\eta_{2_i}^{(c)} + \nu_{i97}^{(c)}, \\
cig_{it} &= \rho cig_{it-1} + \beta drink_{it-1} + \gamma mon_i + \eta_{1_i}^{(c)} + \pi\eta_{2_i}^{(c)} + \nu_{it}^{(c)}, \ t = 1998, \ldots, 2002
\end{aligned}
\tag{39}
$$

where we derived the factor loadings for the initial observation cig_{i97} above. For the predetermined model, we have

$$
cig_{it} = \rho^{(p)} cig_{it-1} + \beta^{(p)} drink_{it-1} + \gamma^{(p)} mon_i + \eta_{1_i}^{(p)} + \pi\eta_{2_i}^{(p)} + \nu_{it}^{(p)}, \ t = 1998, \ldots, 2002
\tag{40}
$$

and $\text{Cov}(\eta_{1_i}^{(p)}, cig_{i97}) \neq 0$ and $\text{Cov}(\eta_{2_i}^{(p)}, cig_{i97}) \neq 0$.

Table 3: Model Results: Extended AR(1) models with individual intercepts and slopes

		Residual-level		Constrained observation-level		Predetermined observation-level	
		Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
					Regression Slopes		
e_t/cig_t	$\leftarrow$						
	e_{t-1}/cig_{t-1}	0.377***	0.025	0.377***	0.026	0.256***	0.030
	$drink_{t-1}$	−0.038	0.027	−0.038	0.027	−0.038	0.025
	mon	−0.162***	0.025	−0.162***	0.025	−0.173***	0.028
					(Residual) Variances		
	$\cdot\, e_{97}/\cdot\, cig_{97}$	10.264***	1.132	10.265***	1.137	15.968***	0.560
	$\cdot\, e_{98}/\cdot\, cig_{98}$	19.810***	0.790	19.810***	0.792	15.328***	1.058
	$\cdot\, e_{99}/\cdot\, cig_{99}$	18.469***	0.671	18.469***	0.672	15.722***	0.730
	$\cdot\, e_{00}/\cdot\, cig_{00}$	18.951***	0.714	18.951***	0.717	17.072***	0.683
	$\cdot\, e_{01}/\cdot\, cig_{01}$	19.840***	0.769	19.840***	0.772	18.443***	0.776
	$\cdot\, e_{02}/\cdot\, cig_{02}$	16.737***	0.831	16.737***	0.837	15.686***	1.023
	$alpha_1/eta_1$	6.908***	1.085	3.066***	0.476	12.897***	1.993
	$alpha_2/eta_2$	0.431***	0.101	0.167**	0.049	0.408**	0.149
					(Residual) Coariances		
$alpha_1/eta_1$	$\longleftrightarrow$						
	$alpha_2$ / eta_2	0.692**	0.262	0.370***	0.079	−0.812*	0.402
	$\cdot\, e_{97}$ / $\cdot\, cig_{97}$	0.00^+	–	–	–	7.263***	0.750
$alpha_2/eta_2$	$\longleftrightarrow$						
	$\cdot\, e_{97}$ / $\cdot\, cig_{97}$	0.00^+	–	–	–	−0.477**	0.180
					Fit Indices		
	χ^2(df)	164.321(29)		164.321(29)		67.240(27)	
	CFI	0.980		0.980		0.994	
	RMSEA	0.024		0.024		0.013	

Estimator: FIML, $n = 8305$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. $^+$: fixed parameter. A dot ($\cdot$) indicates a residual (co)variance. `e` is the structured residual for `cig`. `alpha1`, `alpha2` and `eta1`, `eta2` are latent intercepts and slopes for residual- and observed-level models, respectively.

Table 3 again shows the equivalence of the residual and constrained models in terms of degrees of freedom and model fit, as well as the estimated coefficients. The predetermined model differs in two degrees of freedom: $\mathrm{Cov}(cig_{i97}, \eta_{1_i}^{(p)})$ and $\mathrm{Cov}(cig_{i97}, \eta_{2_i}^{(p)})$ are freely estimated in the predetermined model and fixed in the residual and constrained models. Interestingly, the models with individual intercepts and slopes differ less substantively: the models agree at least in terms of sign and significance, with a positive and significant autoregressive effect, a negative and non-significant lagged effect of alcohol use, and a negative and significant effect of parental monitoring. The coefficients still differ, however, and the autoregressive effect is still larger in the residual and constrained models compared to the predetermined one. The effect of parental monitoring is slightly stronger in the predetermined model, as well.

In any case, the predetermined model is still arguably preferable to the others, according to the chi-square difference test below. While the residual and predetermined models are more parsimonious, they fit significantly worse than the predetermined model.

```
## Chi-Squared Difference Test
##
##         Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## m3c.fit 27 190585 191028  67.24
## m3a.fit 29 190679 191107 164.32     97.081       2  < 2.2e-16 ***
## m3b.fit 29 190679 191107 164.32      0.000       0
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

Scenario 4: Cross-lagged panel models with individual intercepts

Let us now look at the reciprocal relations between cigarette and alcohol use amongst adolescents in a full cross-lagged model. Because the variance of the latent slope for cigarette smoking in models `m3a`, `m3b` and `m3c` was fairly low, let us drop the latent slopes and compare a RI-CLPM with a constrained and predetermined observation-level CLPM with individual effects, which we could call a two-sided DPM. For the RI-CLPM, we have

$$
\begin{aligned}
cig_{it} &= \alpha_i^{(y)} + \varepsilon_{it} \\
drink_{it} &= \alpha_i^{(x)} + \delta_{it}, \ t = 1997, \dots, 2002, \\
\varepsilon_{it} &= \rho \varepsilon_{it-1} + \beta \delta_{it-1} + \nu_{it} \\
\delta_{it} &= \varphi \delta_{it-1} + \theta \varepsilon_{it-1} + \upsilon_{it}, \ t = 1998, \dots, 2002
\end{aligned}
\tag{41}
$$

and $\mathrm{Cov}(\alpha_i^{(y)}, \varepsilon_{i97}) = \mathrm{Cov}(\alpha_i^{(y)}, \delta_{i97}) = \mathrm{Cov}(\alpha_i^{(x)}, \varepsilon_{i97}) = \mathrm{Cov}(\alpha_i^{(x)}, \delta_{i97}) = 0$, i.e., the initial deviations from the individual effects are assumed to be unrelated to the individual effects themselves (P. Allison, Williams, and Moral-Benito 2017). The constrained model is

$$
\begin{aligned}
cig_{i97} &= \varpi^{(y)} \alpha_i^{(c,y)} + \omega^{(x)} \eta_i^{(c,x)} + \nu_{i97}^{(c)} \\
drink_{i97} &= \varpi^{(x)} \alpha_i^{(c,x)} + \omega^{(y)} \eta_i^{(c,y)} + \upsilon_{i97}^{(c)}, \\
cig_{it} &= \rho cig_{it-1} + \beta drink_{it-1} + \eta_i^{(c,y)} + \nu_{it} \\
drink_{it} &= \varphi drink_{it-1} + \theta cig_{it-1} + \eta_i^{(c,x)} + \upsilon_{it}, \ t = 1998, \dots, 2002
\end{aligned}
\tag{42}
$$

where $\eta_i^{(c,y)} = (1 - \rho)\alpha_i^{(y)} - \beta \alpha_i^{(x)}$ and $\eta_i^{(c,x)} = (1 - \varphi)\alpha_i^{(x)} - \theta \alpha_i^{(y)}$ and the ϖs and ωs are defined as above in Equation (21). Finally, for the predetermined model, we have simply

$$cig_{it} = \rho^{(p)} cig_{it-1} + \beta^{(p)} drink_{it-1} + \eta_i^{(p,y)} + \nu_{it} \tag{43}$$

$$drink_{it} = \varphi^{(p)} drink_{it-1} + \theta^{(p)} cig_{it-1} + \eta_i^{(p,x)} + v_{it}, \ t = 1998, \dots, 2002 \tag{44}$$

where the superscripts (p, y) and (p, x) denote the individual effects from the predetermined cross-lagged model for cigarette smoking and alcohol consumption, respectively.

Table 4: Model Results: Full cross-lagged models with individual intercepts

		Residual-level		Constrained observation-level		Predetermined observation-level	
		Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
				Regression Slopes			
$e_t \mid cig_t$	$\leftarrow$						
	$e_{t-1} \mid cig_{t-1}$	0.567***	0.014	0.567***	0.014	0.342***	0.017
	$d_{t-1} \mid drink_{t-1}$	0.079***	0.023	0.079**	0.023	0.087***	0.022
$d_t \mid drink_t$	$\leftarrow$						
	$d_{t-1} \mid drink_{t-1}$	0.221***	0.014	0.221***	0.014	0.198***	0.013
	$e_{t-1} \mid cig_{t-1}$	0.024**	0.009	0.024*	0.009	0.007	0.010
				(Residual) Variances			
	alphay/etay	8.157***	0.589	3.066***	0.476	8.592***	0.601
	alphax/etax	2.624***	0.142	0.167**	0.049	2.063***	0.145
				(Residual) Coariances			
alphay | etay	$\longleftrightarrow$						
	alphax | etax	1.205***	0.228	0.164	0.131	0.827**	0.240
	$\cdot\, e_{97} \mid cig_{97}$	0.00+				4.918***	0.388
	$\cdot\, d_{97} \mid drink_{97}$	0.00+				0.917**	0.305
alphax | etax	$\longleftrightarrow$						
	$\cdot\, e_{97} \mid cig_{97}$	0.00+				0.793***	0.227
	$\cdot\, d_{97} \mid drink_{97}$	0.00+				1.403***	0.158
				Fit Indices			
	χ^2(df)	366.116(53)		366.116(53)		122.839(49)	
	CFI	0.960		0.960		0.990	
	RMSEA	0.029		0.029		0.014	

Estimator: FIML, $n = 8305$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. +: fixed parameter. A dot ($\cdot$) indicates a residual (co)variance. `e` and `d` are the structured residuals for `cig` and `drink`, respectively. `alphay`, `etay` and `alphax`, `etax` are latent intercepts for `cig` and `drink` in the residual- and observed-level models, respectively. Error variances not shown due to space constraints.

Table 4 shows the results of th full cross-lagged models with individual intercepts and no time-invariant covariates. They examine the 'within-person' relations between cigarette smoking and alcohol use in adolescents and young adults. The residual-level model is a true RI-CLPM, the constrained and predetermined models are akin to a 'two-sided' DPM, though P. Allison, Williams, and Moral-Benito (2017) only ever really discuss the predetermined variant.

Here, again, we see the autoregressive coefficient for smoking being estimated much higher in the residual and constrained models than in the predetermined one. The lagged effect of drinking on smoking is comparable between the models, but the effect is slightly higher in the predetermined one. The autoregressive effect of drinking is also comparable between the models, while the lagged effect of smoking on drinking is only significant in the residual and constrained formulations.

The fit is fairly drastically better in the predetermined model, and the chi-square difference is again significant, as can be seen in the output below. The models differ in four degrees of freedom. Those are the freely estimated covariances in the predetermined model between the initial observations cig_{i97} and $drink_{i97}$ and the two latent individual intercept variables, $\eta_i^{(p,y)}$ and $\eta_i^{(p,y)}$.

```
## Chi-Squared Difference Test
##
##          Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## m4c.fit 49 184656 184938 122.84
## m4a.fit 53 184891 185146 366.12     243.28       4  < 2.2e-16 ***
## m4b.fit 53 184891 185146 366.12       0.00       0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

These examples all underline the importance of the assumptions surrounding the initial conditions. While it is likely not wise to try to compare results between the tables (we do not know which is the 'best' modeling approach, e.g., with or without a latent slope, in a one- or two-sided formulation), but the residual and constrained models are nested within the predetermined one within a given table, so those comparisons are valid. In each case, the predetermined model fit significantly better than the residual and constrained ones. This should come as no surprise given Figure 5: the data does not support the assumptions inherent to those models.

To emphasize one of the points made above, that the residual-level models do not offer an advantage in terms of the 'within-person' effects, let us compare the residual, constrained and predetermined models in a final example, using the years between 2005 and 2010. From Figure 5, we see that at that point, the data essentially embodies the assumptions of the residual and constrained models.

Table 5: Model Results: Full cross-lagged models with individual intercepts, years 2005-2010

		Residual-level		Constrained observation-level		Predetermined observation-level	
		Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
		Regression Slopes					
$e_t \mid cig_t$	$\leftarrow$						
	$e_{t-1} \mid cig_{t-1}$	0.166***	0.015	0.166***	0.016	0.163***	0.015
	$d_{t-1} \mid drink_{t-1}$	0.030	0.025	0.030	0.025	0.039	0.025
$d_t \mid drink_t$	$\leftarrow$						
	$d_{t-1} \mid drink_{t-1}$	0.122***	0.010	0.122***	0.010	0.125***	0.010
	$e_{t-1} \mid cig_{t-1}$	0.008	0.006	0.008	0.006	0.009	0.006
		(Residual) Variances					
	alphay/etay	29.816***	0.775	20.657***	1.007	21.083***	1.013
	alphax/etax	2.661***	0.066	2.025***	0.078	1.990***	0.082
		(Residual) Coariances					
alphay $\mid$ etay	$\longleftrightarrow$						
	alphax $\mid$ etax	1.820***	0.180	1.056***	0.245	0.989***	0.255
	$\cdot\ e_{05} \mid cig_{05}$	0.000+				24.429***	0.847
	$\cdot\ d_{05} \mid drink_{05}$	0.000+				1.227**	0.252
alphax $\mid$ etax	$\longleftrightarrow$						
	$\cdot\ e_{05} \mid cig_{05}$	0.000+				1.562***	0.279
	$\cdot\ d_{05} \mid drink_{05}$	0.000+				2.368***	0.080
		Fit Indices					
	$\chi^2(df)$	155.624(53)		155.624(53)		145.882(49)	
	CFI	0.994		0.994		0.995	
	RMSEA	0.016		0.016		0.016	

Estimator: FIML, $n = 7418$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. +: fixed parameter. A dot ($\cdot$) indicates a residual (co)variance. e and d are the structured residuals for cig and drink, respectively. alphay, etay and alphax, etax are latent intercepts for cig and drink in the residual- and observed-level models, respectively. Error variances not shown due to space constraints.

Table 5 shows that once the process has essentially reached equilibrium, the residual, constrained and pre-determined models are very comparable in terms of model fit and estimated coefficients. The predetermined model still has a slight advantage in terms of fit; the chi-square difference test is significant, but only at the 5% level. This should not be surprising since a model that puts less constraints will inevitably tend to fit better, even if this is due to largely to sampling error.

```
## Chi-Squared Difference Test
##
##          Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## m6c.fit 49 211490 211774 145.88
## m6a.fit 53 211492 211748 155.62     9.7413       4    0.04502 *
## m6b.fit 53 211492 211748 155.62     0.0002       0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Practical strategies

In this section I suggest some practical modeling strategies. First, there are likely substantive aspects to consider. Is it plausible that the process under investigation is not stationary or at equilibrium? For many topics, these assumptions may be unproblematic. But a wide range of topics in, say, developmental psychology, are likely consistent with the trajectories shown in Figure 5. In these cases, the residual-level models should not be applied without careful consideration. For example, Voelkle and Oud (2015) re-analyzed studies by Bradway and Thomas (1962); McArdle et al. (2000) on the development of short-term memory over time and found a strong nonlinear trend in youths towards their long-term equilibrium memory scores (see Figure 5 in Voelkle and Oud 2015). This is particularly relevant because a not insignificant number of papers citing the Hamaker, Kuiper, and Grasman (2015) article are from the field of developmental psychology, or at least deal with topic of adolescent development. In fact, as of writing the first few pages of citing research on Google Scholar are filled with articles on the development of children and adolescents. However, the issue may not be restricted to youth and adolescent development; research on aging or life transitions, for example, could potentially also be affected.

Another difficultly could arise when attempting to examine the effects of interventions. Consider an example by Voelkle et al. (2018), in which they examine (among other things) social anxiety as a dynamic process. Introducing an intervention to lower anxiety around the time of the initial measurement, such as administering serotonin, could lead to an initial measurement of social anxiety that was well below an individual's long-term 'usual' levels. Over the course of the study, anxiety would likely move back towards its equilibrium state as the effects of the intervention wore off. For this and other related topics, it is likely wise to spend a good amount of time considering the initial conditions.

It is likely possible to identify potential issues by looking at the observed means and variances over time. If these are not stable over time, it would seem unwise to nevertheless place constraints on the model as if they were. The autocorrelations can be examined in a simple covariance matrix of the observed data. The covariances between the dependent variable(s) should only depend on the temporal distance between the two, but not on the specific timepoint.

At an even more basic level, a researcher should know in advance what the focus of the investigation is. I mention this not to be condescending, but only to point out that the between-person components may not be of central interest. If the within-person cross-lagged effects are the focus, then the residual-level models do not offer much in the way of real benefits compared to the observation-level models, which are much more straightforward to code in any case. If the between-unit components are of interest (e.g., examining differing levels or trajectories across subgroups), then that is a strong argument for sticking with the residual-level approach.

And notice that we can achieve a type of 'predetermined residual-level model' by relaxing the assumptions on the initial conditions. Remember from Equation (23) that $\text{Cov}(\nu_{i0}, \eta_i)$ was only zero if $\text{Cov}(y_{i0}, \eta_i) = (1 - \rho)^{-1} \text{Var}(\eta_i)$. If the process under investigation is stationary but not at equilibrium, then the individual effects will be correlated with the initial residuals, ν_{i0}. In the residual-level model, this is equivalent to saying $\text{Cov}(\alpha_i, \varepsilon_{i0}) \neq 0$,[12] so we have to account for this covariance. We can do so by including it (along with $\text{Cov}(\alpha_{2_i}, \varepsilon_{i0})$, if necessary, e.g., in the one-sided LCM-SR). This captures the extent to which ε_{i0} is not just a random temporal deviation from the stable characteristics, itself unrelated to those characteristics. Doing so gives us *exactly* a predetermined observation-level model in terms of model fit, degrees of freedom and estimated coefficients (though the standardized effects may differ). It also does not change the model's structure, still we just have $\mathbb{E}(y_{it}) = \mathbb{E}(\alpha_i)$, so this type of model may be better for those interested in both the within- and between-unit components.

To achieve a kind of 'predetermined RI-CLPM,' analogous to the normal RI-CLPM shown in Table 4, we change the following two lines of `lavaan` code from

---

[12]Since $y_{i0} = \alpha_i + \varepsilon_{i0} = (1 - \rho)^{-1}\eta_i + (y_{i0} - (1 - \rho)^{-1}\eta_i) = (1 - \rho)^{-1}\eta_i + \nu_{i0}$.

```
alphay ~~ 0*e97 + 0*d97
alphax ~~ 0*e97 + 0*d97
```

to

```
alphay ~~ e97 + d97
alphax ~~ e97 + d97
```

in order to allow the initial deviations to correlate with the individual intercepts. Again, this gives us exactly the analogous predetermined model.

Interestingly, the default behavior of the `sem` wrapper in `lavaan` as well as Mplus and Stata is to allow exogenous latent variables to covary (Muthén and Muthén 1998--2017; Rosseel et al. 2020; StataCorp 2019). And, in order to estimate the residual-level models in structural-form (the most common way), new latent variables must be generated that are identical to the residuals (e.g., $\varepsilon_{it} = y_{it} - \alpha_{1_i} - t\alpha_{2_i}$), while setting the variances of the actual residuals to zero.[13] So, estimating the residual-level models in structural-form means that the individual effects and the structured residuals are recognized as exogenous latent variables and will thus be set to covary by default in the most common SEM software. In order to estimate the residual-level model in the intended way (the way it is described in Usami, Murayama, and Hamaker (2019), p. 640, where the initial deviations are not meant to correlate with the individual effects), the user must actually override this behavior and constrain the covariance(s) to zero.

Another possibility to make the residual-level models more robust was mentioned by Zyphur et al. (2020a), p. 660. There, they note that fixing the factor loadings of the individual intercepts to one at all points in time means we are treating the observed variable as "mean stationary" and that to relax this assumption, we can simply allow the factor loadings to be estimated freely. The method they suggest is to freely estimate all factor loadings except for the final one at $t = T$ (in order to supply the latent variable with a scale).[14] This essentially gives us a model that goes above and beyond the flexibility of the predetermined models discussed so far, allowing the individual effects to vary at each point in time. The drawback to this is that it goes against the point of the residual-level models by making the mean structure less intuitive, since $\mathbb{E}(y_{i1}) = \lambda_1 \mathbb{E}(\alpha_i)$, $\mathbb{E}(y_{i2}) = \lambda_2 \mathbb{E}(\alpha_i)$, etc., where λ_t are the factor loadings. And if one is turning to a constrained ALT or DPM to save degrees of freedom, then trying to freely estimate each factor loading would be counterproductive.

Ultimately, when a researcher has enough degrees of freedom, there is no real need to constrain the covariances of the initial observations and individual effects. If the process under investigation conforms to the residual-level assumptions, then these unrestricted correlations will be estimated accordingly. And due to sampling error, the fit of the predetermined model will almost inevitably be at least marginally better. When the process violates the assumptions discussed above, the discrepancy in the models can be quite large, both in terms of fit and substantive results, as has been shown above. Of course, there is nothing stopping a researcher from running both types of models, a residual-level and a predetermined observation-level, and comparing the two.

Conclusion

It has been shown previously with varying degrees of explicitness that the residual-level panel models are algebraically equivalent to their observation-level counterparts. However, I believe it is not yet widely recognized that they are re-expressions specifically of the constrained observation-level models. The goal of this article was to draw attention to this fact and point out the assumptions that go along it.

---

[13]See, for example, the supplementary matrials from Berry and Willoughby (2017), or the github repository associated with Mulder and Hamaker (2020)

[14]K. Bollen and Brand (2010) also discusses freely estimated factor loadings of the individual effects as a way to relax assumptions in static panel models in SEM.

So, while the residual-level models provide benefits in terms of allowing for better interpretability of the growth trajectory parameters, they come at the cost some perhaps restrictive assumptions. If the process being observed is nonstationary, stationary but not yet at equilibrium (perhaps due to an intervention), or has been temporarily knocked out of equilibrium, the residual-level models could return biased AR and CL estimates which are typically of central substantive interest. In the case of cigarette smoking amongst adolescents, it was shown that the estimated autoregressive coefficients tended to be larger in the residual and constrained model than in the predetermined one. In the models with only latent individual intercepts, the lagged effect of alcohol consumption on smoking was positive and significant in each of the predetermined models, while for the residual and constrained models, it was only significant in the cross-lagged specification.

In most cases, there is nothing stopping one from running both the residual- and analogous observation-level model, or alternatively, estimating the residual-level model with and without the covariances between the initial deviations and the individual effects fixed to zero. It has been shown here that the residual-level models are nested in their observation-level counterparts, so a likelihood ratio test can easily be performed to assess whether the observed deviation from the model-implied constraints is likely random or systematic. If the fit of the predetermined model is significantly better, then it could be an indication that the assumptions of stationarity and equilibrium are not met.

Over a range of substantive applications, the residual-level models are likely well-suited. As Usami, Murayama, and Hamaker (2019) say, there is often "no reason to assume that an individual's temporal deviation at the first measurement from the person's trait score is dependent on that trait score" (Usami, Murayama, and Hamaker 2019, 640). In these cases, the residual-level models, but also the constrained observation-level models are more parsimonious and are thus likely preferable. However, as Ou et al. (2016) note, for many applications the issue of the initial conditions is often regarded as a "trivial methodological detail" (Ou et al. 2016, 179). Based on what was shown in this article, the assumptions regarding the initial conditions should be of central concern when it comes to those dynamic panel models discussed here.

I end by noting some limitations of this analytical paper and discussing some potential next steps. First, the models discussed here included only observed indicators. One of the main strengths of SEM is the ability to model latent variables to account for measurement error in observed variables. Adding measurement error into the mix means we have to account for another component, i.e., deviations from the individual trajectory due to measurement error rather than potentially substantive sources. Secondly, while the equivalence of the residual- and constrained observed-level models was established, a full assessment of the various models' performance under different scenarios was not conducted. Recent papers by Zyphur et al. (2020a); Zyphur et al. (2020b) attempt to formulate a generalized cross-lagged panel model and show that cross-lagged panel models with individual effects can be extended even further, for example by including not just an autoregressive but also a moving average (MA) structure. The equivalence and/or benefits and drawbacks of residual- and observation-level models in this context could be investigated as well.

As Box famously put it, "all models are wrong" (1976), and care should be taken when specifying any statistical model. The goal of this article was not to discredit the residual-level models but rather draw attention to a detail I believe has not been given enough attention in the past. Failure to take assumptions regarding the initial conditions into account could lead to incorrect conclusions regarding the substantively interesting within-person effects. At the same time, while the predetermined models offer flexibility and are robust to a number of scenarios, there are drawbacks associated with them as well (Voelkle 2008). It is my hope that this article could provide researchers with some further practical guidance for their empirical projects.

References

Allison, Paul D. 2009. *Fixed Effects Regression Models.* Thousand Oaks: Sage Publications. https://doi.org/https://dx.doi.org/10.4135/9781412993869.

Allison, Paul, Richard Williams, and Enrique Moral-Benito. 2017. "Maximum Likelihood for Cross-Lagged Panel Models with Fixed Effects." *Socius* 3: 1–17. https://doi.org/10.1177/2378023117710578.

Andersen, Henrik. 2021. "A Closer Look at Fixed-Effects Regression in Structural Equation Modeling." *Structural Equation Modeling: A Multidisciplinary Journal.*

Barnett, Elizabeth, Donna Spruijt-Metz, Jennifer B. Unger, Louise Ann Rohrbach, Ping Sun, and Steve Sussman. 2013. "Bidirectional Associations Between Time Perspective and Substance Use Among Continuation High-School Students." *Substance Use and Misuse* 48: 574–80.

Berry, Daniel, and Michael Willoughby. 2017. "On the Practical Interpretability of Cross-Lagged Panel Models: Rethinking a Developmental Workhorse." *Child Development* 88 (4): 1186–206.

Bianconcini, Silvia, and Kenneth A. Bollen. 2018. "The Latent Variable-Autoregressive Latent Trajectory Model: A General Framework for Longitudinal Data Analysis." *Structural Equation Modeling: A Multidisciplinary Journal* 25 (5): 791–808. https://doi.org/https://doi.org/10.1080/10705511.2018.1426467.

Bierhoff, Jennifer, Regine Haardörfer, Michael Windle, and Carla J. Berg. 2019. "Psychological Risk Factors for Alcohol, Cannibis and Various Tobacco Use Among Young Adults: A Longitudinal Analysis." *Substance Use and Misuse*, 1–11.

Bollen, Kenneth A. 2006. *Latent Curve Models: A Structural Equation Approach.* Hoboken, NJ: Wiley.

Bollen, Kenneth, and Jennie Brand. 2010. "A General Panel Model with Random and Fixed Effects: A Structural Equations Approach." *Social Forces* 89 (1): 1–34.

Bollen, Kenneth, and Patrick Curran. 2004. "Autoregressive Latent Trajectory (ALT) Models: A Synthesis of Two Traditions." *Sociological Methods and Research* 32 (3): 336–83. https://doi.org/10.1177/0049124103260222.

Bollen, Kenneth, and Catherine Zimmer. 2010. "An Overview of the Autoregressive Latent Trajectory (ALT) Model." In *Longitudinal Research with Latent Variables*, edited by K. van Montfort, J. Oud, and A. Satorra, 153–76. Berlin, Heidelberg: Springer. https://doi.org/https://doi.org/10.1007/978-3-642-11760-2_5.

Box, George E. P. 1976. "Science and Statistics." *Journal of the American Statistical Association* 71 (356): 791–99.

Bradway, K. P., and C. W. Thomas. 1962. "Intelligence at Adulthood: A Twenty-Five Year Follow-up." *Journal of Educational Psychology* 53: 1–14. https://doi.org/10.1037/h0045764.

Brüderl, Josef, and Volker Ludwig. 2015. "Fixed-Effects Panel Regression." In *The Sage Handbook of Regression Analysis and Causal Inference*, edited by Henning Best and Christof Wolf, 327–57. London, Thousand Oaks: Sage Publications.

Bureau of Labor Statistics. 2017. "National Longitudinal Survey of Youth 1997 Cohort, 1997-2017." The Ohio State University: Center for Human Resource Research.

Curran, Patrick, and Kenneth Bollen. 2001. "The Best of Both Worlds: Combining Autoregressive and Latent Curve Models." In *New Methods for the Analysis of Change*, edited by L. Collins and A. Sayer, 107–35. Washington, DC: American Psychological Press. https://doi.org/%20doi:10.1037/10409-004.

Curran, Patrick, Andrea Howard, Sierra Bainter, Stephanie Lane, and James McGinley. 2014. "The Separation of Between-Person and Within-Person Components of Individual Change over Time: A Latent Growth Curve Model with Structured Residuals." *Journal of Consulting and Clinical Psychology* 82 (5): 879–94. https://doi.org/10.1037/a0035297.

Granger, Clive. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37 (3): 424–38.

Greenberg, David, and Roland Kessler. 1982. "Equilibrium and Identification in Linear Panel Models." *Sociological Methods and Research* 10 (4): 435–51.

Hamaker, Ellen. 2005. "Conditions for the Equivalence of the Autoregressive Latent Trajectory Model and a Latent Growth Curve Model with Autoregressive Disturbances." *Sociological Methods and Research* 33 (3): 404–16. https://doi.org/10.1177/0049124104270220.

Hamaker, Ellen, and Conor Dolan. 2009. "Idiographic Data Analysis: Quantitative Methods — from Simple to Advanced." In *Dynamic Process Methodology in the Social and Developmental Sciences*, edited by J. Valsiner, P. Molenaar, M. Lyra, and N. Chaudhary, 191–216. New York: Springer-Verlag.

Hamaker, Ellen, Rebecca Kuiper, and Raoul Grasman. 2015. "A Critique of the Cross-Lagged Panel Model." *Psychological Methods* 20 (1): 102–16. https://doi.org/10.1037/a0038889.

Hayduk, Leslie, Greta Cummings, Rainer Stratkotter, Melanie Nimmo, Kostyantyn Grygoryev, Donna Dosman, Michael Gillespie, Hannah Pazderka-Robinson, and Kwame Boadu. 2003. "Pearl's D-Separation: One More Step into Causal Thinking." *Structural Equation Modeling* 10 (2): 289–311.

Hsiao, Cheng. 2014. *Analysis of Panel Data. Third Edition.* New York: Cambridge University Press.

Jongerling, Joran, and Ellen Hamaker. 2011. "On the Trajectories of the Predetermined ALT Model: What Are We Really Modeling?" *Structural Equation Modeling: A Multidisciplinary Journal* 18 (3): 370–82.

Keijsers, Loes. 2016. "Parental Monitoring and Adolescent Problem Behaviors: How Much Do We Really Know?" *International Journal of Behavioral Development* 40 (3): 271–81.

Kim, Minjung, Hsien-Yuan Hsu, Oi-man Kwok, and Sunmi Seo. 2018. "The Optimal Starting Model to Search for the Accurate Growth Trajectory in Latent Growth Models." *Frontiers in Psychology* 9. https://doi.org/10.3389/fpsyg.2018.00349.

Lloyd-Richardson, E., G. Papandonatos, A. Kazura, C. Stanton, and R. Niaura. 2002. "Differentiating Stages of Smoking Intensity Among Adolescents: Stage-Specific Psychological and Social Influences." *Journal of Consulting and Clinical Psychology* 70: 998–1009.

McArdle, J. J., F. Hamagami, W. Meredith, and K. P. Bradway. 2000. "Modeling the Dynamic Hypotheses of Gf-Gc Theory Using Longitudinal Life-Span Data." *Learning and Individual Differences* 12: 53–79. https://doi.org/10.1016/s1041-6080(00)00036-4.

Meredith, William, and John Tisak. 1990. "Latent Curve Analysis." *Psychometrika* 55 (1): 107–22.

Moral-Benito, Enrique, Paul Allison, and Richard Williams. 2019. "Dynamic Panel Data Modeling Using Maximum Likelihood: An Alternative to Arellano-Bond." *Applied Economics* 51 (20): 2221–32. https://doi.org/https://doi.org/10.1080/00036846.2018.1540854.

Mulder, Jeroen D., and Ellen L. Hamaker. 2020. "Three Extensions of the Random Intercept Cross-Lagged Panel Model." *Structural Equation Modeling: A Multidisciplinary Journal.*

Muthén, Linda, and Bengt Muthén. 1998--2017. *Mplus User's Guide. Eighth Edition.* Los Angeles, California: Muthén & Muthén.

Nelson, Sarah E., Mark J. Van Ryzin, and Thomas J. Dishion. 2014. "Alcohol, Marijuana, and Tobacco Use Trajectories from Age 12 to 24 Years: Demographics and Young Adult Substance Use Problems." *Development and Psychopathology* 27: 253–77.

Nkansah-Amankra, Stephen, and Mark Minelli. 2016. ""Gateway Hypothesis" and Early Drug Use: Additional Findings from Tracking a Population-Based Sample of Adolescents to Adulthood." *Preventive Medicine Reports* 4: 134–41.

Ou, Lu, Sy-Miin Chow, Linying Ji, and Peter C. M. Molenaar. 2016. "(Re)evaluating the Implications of the Autoregressive Latent Trajectory Model Through Liklihood Ratio Tests of Its Initial Conditions."

Multivariate Behavioral Research 52 (2): 178–99. https://doi.org/https://doi.org/10.1080/00273171.2016.1259980.

Pickup, Mark. 2015. *Introduction to Time Series Analysis*. Los Angeles, London: Sage.

Rosseel, Yves, Terrence D. Jorgensen, Daniel Oberski, Jarrett Byrnes, Leonard Vanbrabant, Victoria Savalei, Ed Merkle, et al. 2020. *Lavaan: Latent Variable Analysis*. https://CRAN.R-project.org/package=lavaan.

Rüttenauer, Tobias, and Volker Ludwig. 2020. "Fixed Effects Individual Slopes: Accounting and Testing for Heterogeneous Effects in Panel Data or Other Multilevel Models." *Sociological Methods and Research* OnlineFirst: 1–42. https://doi.org/https://doi.org/10.1177/0049124120926211.

Schuler, Megan S., Joan S. Tucker, Eric R. Pedersen, and Elizabeth J. D'Amico. 2019. "Relative Influence of Perceived Peer and Family Substance Use on Adolescent Alcohol, Cigarette, and Marijuana Use Across Middle and High School." *Addictive Behaviours* 88: 99–105.

Seifert, Nico, and Henrik K. Andersen. 2020. "The Impact of Social Isolation on Subjective Well-Being." https://www.researchgate.net/publication/348548953_The_impact_of_social_isolation_on_subjective_well-being.

Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, London, New York: Chapman & Hall.

StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC.

Tarter, Ralph, Michael Vanyukov, Levent Kirisci, Maureen Reynolds, and Duncan Clark. 2006. "Predictors of Marijuana Use in Adolescents Before and After Licit Drug Use: Examination of the Gateway Hypothesis." *American Journal of Psychiatry* 163 (12): 2134–40.

Teachman, Jay, Greg Duncan, Jean Yeung, and Dan Levy. 2001. "Covariance Structure Models for Fixed and Random Effects." *Sociological Methods and Research* 30 (2): 271–88.

Usami, Satoshi, Kou Murayama, and Ellen L. Hamaker. 2019. "A Unified Framework of Longitudinal Models to Examine Reciprocal Relations." *Psychological Methods* 24 (5): 637–57. https://doi.org/http://dx.doi.org/10.1037/met0000210.

Voelkle, Manuel C. 2008. "Reconsidering the Use of Autoregressive Latent Trajectory (ALT) Models." *Multivariate Behavioral Research* 43 (4): 564–91. https://doi.org/https://doi.org/10.1080/00273170802490665.

Voelkle, Manuel C., Christian Gische, Charles C. Driver, and Ulman Lindenberger. 2018. "The Role of Time in the Quest for Understanding Psychological Mechanisms." *Multivariate Behavioral Research* 53 (6). https://doi.org/https://doi.org/10.1080/00273171.2018.1496813.

Voelkle, Manuel C., and Johan H. L. Oud. 2015. "Relating Latent Change Score and Continuous Time Models." *Structural Equation Modeling: A Multidisciplinary Journal* 22 (3): 366–81.

Williams, Richard, Paul Allison, and Enrique Moral-Benito. 2018. "Linear Dynamic Panel-Data Estimation Using Maximum Likelihood and Structural Equation Modeling." *The Stata Journal* 18 (2): 293–326. https://doi.org/https://doi.org/10.1177/1536867X1801800201.

Wooldridge, Jeffery. 2002. *Econometric Analysis of Cross Sectional and Panel Data*. Cambridge, Massachusetts: The MIT Press.

Zyphur, Michael J., Paul D. Allison, Louis Tay, Manuel C. Voelkle, Kristopher J. Preacher, Zhen Zhang, Ellen L. Hamaker, et al. 2020a. "From Data to Causes i: Building a General Cross-Lagged Panel Model (GCLM)." *Organizational Research Methods* 23 (4): 651–87. https://doi.org/10.1177/1094428119847278.

———, et al. 2020b. "From Data to Causes II: Comparing Approaches to Panel Data Analysis." *Organizational Research Methods* 23 (4): 688–716. https://doi.org/10.1177/1094428119847280.

Appendix

Here I show how to arrive at the model-implied moments of (simple versions of) the constrained and residual-level models in detail. First, let us consider the scenario in which we assume there are no 'occasion effects' (Zyphur et al. 2020b), though this scenario is often implausible in real applications. If there are no occasion effects, then we could try to model the mean structure of the observed variable entirely with the individual effects. We do so by turning the mean structure on, but also constraining the intercepts of the observed endogenous variables to zero.

For the constrained model, the model-implied expected value, $\mathbb{E}(y_{it})$, was shown in the main text. It was $\mathbb{E}(y_{it}) = (1 - \rho)^{-1} \mathbb{E}(\eta_i)$. In terms of the residual-level model, the equivalent expression is

$$
\begin{aligned}
\mathbb{E}(y_{it}) &= \mathbb{E}(\alpha_i + \varepsilon_{it}) \\
&= \mathbb{E}(\alpha_i)
\end{aligned}
\tag{1}
$$

if $\mathbb{E}(\varepsilon_{it}) = 0$, which is equal to $(1 - \rho)^{-1} \mathbb{E}(\eta_i)$.

Again, this equation lacks an occasion effect, which we could call μ_t. The occasion effect absorbs the expectation of the error, $\mathbb{E}(\nu_{it})$, which we could call a 'global shock,' i.e., something unmodeled that affects all units at a given point in time (Zyphur et al. 2020b). It also absorbs the expected value of α_i or η_i, since a model trying to estimate intercepts per timepoint as well as a latent mean are under-identified. So, for the residual-level model, we could define μ_t as

$$
\begin{aligned}
y_{it} &= \alpha_i + \varepsilon_{it} \\
&= \alpha_i + \mathbb{E}(\alpha_i) - \mathbb{E}(\alpha_i) + \varepsilon_{it} + \mathbb{E}(\varepsilon_{it}) - \mathbb{E}(\varepsilon_{it}) \\
&= \mu_t + \alpha_i^* + \varepsilon_{it}^*
\end{aligned}
\tag{2}
$$

where $\mu_t = \mathbb{E}(\alpha_i) + \mathbb{E}(\varepsilon_{it})$, $\alpha_i^* = \alpha_i - \mathbb{E}(\alpha_i)$ and $\varepsilon_{it}^* = \varepsilon_{it} - \mathbb{E}(\varepsilon_{it})$ so that $\mathbb{E}(\alpha_i^*) = \mathbb{E}(\varepsilon_{it}^*) = 0$ (Wooldridge 2002, 257).

With occasion effects, the mean structure changes. For a predetermined model, we have (dropping the notation with the stars from above)

$$
\begin{aligned}
y_{i0} &= \mu_0 + \nu_{i0} \\
\mathbb{E}(y_{i0}) &= \mu_0 \\
\mu_0 &= \mathbb{E}(y_{i0}), \\
y_{it} &= \mu_t + \rho y_{it-1} + \eta_i + \nu_{it} \\
\mathbb{E}(y_{it}) &= \mu_t + \rho\, \mathbb{E}(y_{it-1}) \\
\mu_t &= \mathbb{E}(y_{it}) - \rho\, \mathbb{E}(y_{it-1})
\end{aligned}
\tag{3}
$$

since we are allowing μ_t to absorb $\mathbb{E}(\eta_i)$ and $\mathbb{E}(\nu_{it})$. For the constrained model, we have

$$
\begin{aligned}
y_{i0} &= \mu_0 + (1 - \rho)^{-1}\eta_i + \nu_{i0} \\
\mathbb{E}(y_{i0}) &= \mu_0 \\
\mu_0 &= \mathbb{E}(y_{i0}), \\
y_{it} &= \mu_t + \rho y_{it-1} + \eta_i + \nu_{it} \\
\mathbb{E}(y_{it}) &= \mu_t + \rho y_{it-1} \\
\mu_t &= \mathbb{E}(y_{it}) - \rho\,\mathbb{E}(y_{it-1})
\end{aligned}
\tag{4}
$$

and for the residual-level model,

$$
\begin{aligned}
y_{it} &= \mu_t + \alpha_i + \rho\varepsilon_{it-1} + \nu_{it} \\
\mathbb{E}(y_{it}) &= \mu_t \\
\mu_t &= \mathbb{E}(y_{it})
\end{aligned}
\tag{5}
$$

which demonstrate the benefit of the residual-level model, which is to make the intercepts easier to interpret; they are just the means per timepoint.

Now, for the sake of simplicity, let us assume we are grand mean centering the observed variable at each timepoint so that we can ignore the occasion effects without having to try to model the mean structure entirely by the latent individual effects. This is equivalent to allowing each observed variable to have a freely estimated intercept per timepoint (Hamaker, Kuiper, and Grasman 2015, 103) but it simplifies the equations. Going back to the constrained model, the model-implied variance is

$$
\begin{aligned}
\mathrm{Var}(y_{it}) =&\, \mathbb{E}(y_{it}^2) - \mathbb{E}(y_{it})^2 \\
=&\, \mathbb{E}[(\eta_i + \rho y_{it-1} + \nu_{it})^2] - \mathbb{E}(\eta_i + \rho y_{it-1} + \nu_{it})^2 \\
=&\, \mathbb{E}[(\eta_i + \rho(\eta_i + \rho y_{it-2} + \nu_{it-1}) + \nu_{it})^2] - \mathbb{E}(\eta_i + \rho(\eta_i + \rho y_{it-2} + \nu_{it-1}) + \nu_{it})^2 \\
=&\, \ldots \\
=&\, \mathbb{E}[((\rho^t(1-\rho)^{-1} + \rho^{t-1} + \ldots + \rho^1 + 1)\eta_i + \rho^t\nu_{i0} + \rho^{t-1}\nu_{i1} + \ldots + \rho^1\nu_{it-1} + \nu_{it})^2] \\
&\, - \mathbb{E}[(\rho^t(1-\rho)^{-1} + \rho^{t-1} + \ldots + \rho + 1)\eta_i + \rho^t\nu_{i0} + \rho^{t-1}\nu_{i1} + \ldots + \rho\nu_{it-1} + \nu_{it}]^2 \\
=&\, (1-\rho)^{-2}\,\mathbb{E}(\eta_i^2) - (1-\rho)^{-2}\,\mathbb{E}(\eta_i)^2 + \rho^{2\cdot t}\,\mathbb{E}(\nu_{i0}^2) + \rho^{2\cdot(t-1)}\,\mathbb{E}(\nu_{i1}^2) + \ldots + \rho^2\,\mathbb{E}(\nu_{it-1}) + \mathbb{E}(\nu_{it}) \\
=&\, (1-\rho)^{-2}\,\mathrm{Var}(\eta_i) - (\rho^2 - 1)^{-1}\,\mathrm{Var}(\nu_{it}),
\end{aligned}
\tag{6}
$$

which results if we assume $\mathbb{E}(\nu_{it}) = 0$, $\forall\, t$ and $\mathbb{E}(\nu_{it}\nu_{is}) = 0$, $t \neq s$. The last term, $-(\rho^2 - 1)^{-1}$, is what the sequence $\rho^{2\cdot t}\,\mathbb{E}(\nu_{i0}^2) + \rho^{2(t-1)}\,\mathbb{E}(\nu_{i1}^2) + \ldots + \rho^2\,\mathbb{E}(\nu_{it-1}^2) + \mathbb{E}(\nu_{it}^2) = \sum_{j=0}^{t}\rho^{j\cdot 2}\,\mathbb{E}(\nu_{it-j}^2)$ approaches if we treat $\mathbb{E}(\nu_{it}^2) = \mathrm{Var}(\nu_{it})$ as constant over time. The variance of the observed variable is actually much easier to derive if we start from the residual-level formulation:

$$
\begin{aligned}
\mathrm{Var}(y_{it}) =&\, \mathbb{E}(y_{it}^2) - \mathbb{E}(y_{it})^2 \\
=&\, \mathbb{E}[(\alpha_i + \varepsilon_{it})^2] - \mathbb{E}(\alpha_i + \varepsilon_{it})^2 \\
=&\, \mathbb{E}(\alpha_i^2 + 2\alpha_i\varepsilon_{it} + \varepsilon_{it}^2) - \mathbb{E}(\alpha_i)^2 \\
=&\, \mathbb{E}(\alpha_i^2) - \mathbb{E}(\alpha_i)^2 + \mathbb{E}(\varepsilon_{it}^2) \\
=&\, \mathrm{Var}(\alpha_i) + \mathrm{Var}(\varepsilon_{it}),
\end{aligned}
\tag{7}
$$

as long as $\mathbb{E}(\alpha_i\varepsilon_{it}) = 0$ and $\mathbb{E}(\varepsilon_{it}) = 0$. Then, recall $\alpha_i = (1-\rho)^{-1}\eta_i$, so

$$
\begin{aligned}
\mathrm{Var}(\alpha_i) &= \mathbb{E}(\alpha_i^2) - \mathbb{E}(\alpha_i)^2 \\
&= \mathbb{E}[((1-\rho)^{-1}\eta_i)^2] - \mathbb{E}[(1-\rho)^{-1}\eta_i]^2 \\
&= (1-\rho)^{-2}\,\mathbb{E}(\eta_i^2) - (1-\rho)^{-2}\,\mathbb{E}(\eta_i)^2 \\
&= (1-\rho)^{-2}\,\mathrm{Var}(\eta_i)
\end{aligned}
\tag{8}
$$

and

$$
\begin{aligned}
\mathrm{Var}(\varepsilon_{it}) &= \mathbb{E}[(\rho\varepsilon_{it-1} + \nu_{it})^2] \\
&= \mathbb{E}(\rho^2\varepsilon_{it-1}^2 + 2\rho\varepsilon_{it-1}\nu_{it} + \nu_{it}^2) \\
&= \rho^2\,\mathbb{E}[(\rho\varepsilon_{it-2} + \nu_{it-1})^2] + \mathbb{E}(\nu_{it}^2) \\
&= \rho^2\,\mathbb{E}(\rho^2\varepsilon_{it-2}^2 + 2\rho\varepsilon_{it-2}\nu_{it-1} + \nu_{it-1}^2) + \mathbb{E}(\nu_{it}^2) \\
&= \rho^4\,\mathbb{E}(\nu_{it-2}^2) + \rho^2\,\mathbb{E}(\nu_{it-1}^2) + \mathbb{E}(\nu_{it}^2) \\
&= \ldots
\end{aligned}
\tag{9}
$$

as long as $\mathbb{E}(\varepsilon_{it-1}\nu_{it}) = 0$, $\forall\, t$ and where we can notice that if $|\rho| < 1$ and we go back far enough in time, then the first term will approach zero, and the remaining terms can be rewritten as

$$
\begin{aligned}
\mathrm{Var}(\varepsilon_{it}) &= \sum_{j=0}^{\infty} \rho^{j\cdot 2}\,\mathrm{Var}(\nu_{it}) \\
&= -(\rho^2 - 1)^{-1}\,\mathrm{Var}(\nu_{it})
\end{aligned}
\tag{10}
$$

again assuming $\mathrm{Var}(\nu_{it})$ is constant over time, which is the same expression we got by starting with the constrained model. The covariance between, say, y_{it} and y_{it-1}, is easiest to show by starting with the residual-level model:

$$
\begin{aligned}
\mathrm{Cov}(y_{it}, y_{it-1}) &= \mathbb{E}(y_{it}y_{it-1}) - \mathbb{E}(y_{it})\,\mathbb{E}(y_{it-1}) \\
&= \mathbb{E}[(\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{it-1})] - \mathbb{E}(y_{it})\,\mathbb{E}(y_{it-1}) \\
&= \mathbb{E}(\alpha_i^2 + \alpha_i\varepsilon_{it} + \varepsilon_{it-1}\alpha_i + \varepsilon_{it}\varepsilon_{it-1}) - \mathbb{E}(y_{it})\,\mathbb{E}(y_{it-1}) \\
&= \mathbb{E}(\alpha_i^2) + \mathbb{E}(\varepsilon_{it}\varepsilon_{it-1}) - \mathbb{E}(y_{it})\,\mathbb{E}(y_{it-1}).
\end{aligned}
\tag{11}
$$

From the main text, we know the expectation of y is constant over time, so $\mathbb{E}(y_{it}) = \mathbb{E}(y_{it-1})$ and $\mathbb{E}(y_{it})\,\mathbb{E}(y_{it-1}) = \mathbb{E}(y_{it})^2$ and we can expand $\mathbb{E}(\varepsilon_{it}\varepsilon_{it-1})$ for

$$
\begin{aligned}
\mathrm{Cov}(y_{it}, y_{it-1}) &= \mathbb{E}(\alpha_i^2) - \mathbb{E}(y_{it})^2 + \mathbb{E}[(\rho\varepsilon_{it-1} + \nu_{it})\varepsilon_{it-1}] \\
&= \mathrm{Var}(\alpha_i) + \rho\,\mathbb{E}(\varepsilon_{it-1}^2) \\
&= \mathrm{Var}(\alpha_i) + \rho\,\mathrm{Var}(\varepsilon_{it-1})
\end{aligned}
\tag{12}
$$

where the assumption of constant error variance means $\mathbb{E}(\varepsilon_{it-1}^2) = \mathbb{E}(\varepsilon_{it}^2) = \mathrm{Var}(\varepsilon_{it})$, i.e., it does not matter what specific timepoint we are looking at. In terms of the original individual effects and error term, we have

$$\begin{aligned}
\text{Cov}(y_{it}, y_{it-1}) &= (1-\rho)^{-2}\,\text{Var}(\eta_i) + \rho\,\text{Var}(\varepsilon_{it}) \\
&= (1-\rho)^{-2}\,\text{Var}(\eta_i) - \rho(\rho^2 - 1)^{-1}\,\text{Var}(\nu_{it}).
\end{aligned} \tag{13}$$

The covariance between the dependent variable separated by more than just one timepoint can be derived in a straightforward fashion by replacing the ρ in front of the brackets in the second term with ρ^j, where j is the distance in time between the two.