One does not simply correct for serial dependence

Sigert Ariens[*,1] Janne K. Adolf[*] Eva Ceulemans[*]

∗ KU Leuven, Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, Leuven, 3000, Belgium

1 sigert.ariens@kuleuven.be

Abstract

Serial dependence is present in most time series data sets collected in psychological research. This paper investigates the implications of various approaches for handling such serial dependence, when one is interested in the linear effect of a time-varying covariate on the time-varying criterion. Specifically, the serial dependence is either neglected, corrected for by specifying autocorrelated residuals, or modeled by including a lagged version of the criterion as an additional predictor. Using both empirical and simulated data, we showcase that the obtained results depend considerably on which approach is selected. We discuss how these differences can be explained by understanding the restrictions imposed under the various approaches. Based on the insight that all three approaches are restricted versions of an autoregressive distributed lag model, we demonstrate that accessible statistical tools, such as information criteria and likelihood-ratio tests can be used to justify a chosen approach empirically.

Keywords

Psychological dynamics, serial dependence, time series analysis, intensive longitudinal data.

Acknowledgements

This work was supported by a research fellowship from the German Research Foundation (DFG) awarded to Janne Adolf, by research grants from the Fund for Scientific Research-Flanders (FWO, Project No. G.074319N), the Research Council of KU Leuven (C14/19/054 and iBOF/21/090), awarded to Eva Ceulemans.

Declarations of interest

The authors declare the absence of any financial, intellectual, or other conflicts of interest which may have biased any aspect of this manuscript.

Open practice statement

The materials used in this manuscript are available online through the open science framework. These supplementary materials include the scripts used to generate and analyze simulation data, the experimental data, and scripts used to analyze the experimental data.

Ethics

This manuscript uses data collected during an experimental study. This study was approved by the local ethics committee (Social and Societal Ethics Committee at the KU Leuven; case number G-2020-2772-R2(MIN)).

1 Introduction

In recent years, psychological science has started to devote increasing attention towards studying intra-individual variability in psychological phenomena over time (Molenaar, 2004). Affective science in particular has seen a steep rise of studies investigating within-person linear relationships between momentary affect and time-varying covariates of interest, such as influential events (e.g. Vanhasbroeck et al., 2022; Takano et al., 2014; Silk et al., 2003; Lafit et al., 2021). This entails gathering intensive longitudinal data (ILD) which consist of many repeated measurements (Ariens et al., 2020) of both the criterion and the covariate.

To study the within-person relationship of interest, researchers often specify linear regression (LR) models, regressing the criterion y_t, e.g. negative affect, on the covariate x_t. Nevertheless, the ubiquitous phenomenon that observations closer in time are typically more similar than those further apart can imply a form of dependence between the observations which is not accommodated for by the classical LR model. Indeed, a textbook issue for regression based analysis of repeated measures designs is the possibility that the model residuals are correlated over time, violating the independence assumption and detrimenting inference (Wooldridge, 2012, p.412-414).

Reviewing empirical studies, one can broadly delineate three approaches in how researchers have addressed the issue of serial dependence. Firstly, some researchers *neglect* the serial dependence by fitting a LR model, and proceeding with inference about the relationships in the data (e.g. Silk et al., 2003; Gard et al., 2014). A second approach consists of adjusting the model specification to *correct* for the autocorrelation in the residuals. The most popular specification, which we will focus on in this manuscript, is the specification of an autoregressive (AR) equation for how the residuals behave over time (e.g. Takano et al., 2014; Myin-Germeys et al., 2001; Asparouhov and Muthén, 2021; Ravindran et al., 2020; Mak and Schneider, 2020; Dixon-Gordon and Laws, 2021; Johnson et al., 2020; Kung et al., 2021; Whelen and Strunk, 2021). The idea is that with the temporal dependence controlled for, inference on the regression effect of X on Y can safely proceed. A third option is to include additional effects in the model, by adding *lagged* versions of the observed criterion and sometimes also the covariate as predictor variables (e.g. Stevenson et al., 2022; Hamaker et al., 2018; McNeish and Hamaker, 2020). Adding the lagged criterion makes it explicit that one expects current criterion scores to depend on previous ones. We term these three approaches *neglecting*, *correcting*, and *lagged observed variables* respectively.

The question of when which approach should be used for a given data analytic problem naturally rises. Answering this question is important, since conclusions about the effects of the covariate on the criterion can differ depending on which approach is selected (e.g. Asparouhov and Muthén, 2020). While the neglecting approach can be understood as a model without serial dependence and thus as a special case of both correction and lagged observed variable approaches, the differences between the *correction* approach and the *lagged observed variables* appear less well understood in the behavioral sciences literature.

In this paper we therefore provide an overview of the differences between these approaches. This is accomplished by considering the approaches and the models they imply as restricted versions of a more general *autoregressive distributed lag* (ADL) model (e.g. Hendry et al., 1984), and studying the restrictions each approach imposes. Importantly, recapitulating econometric literature (e.g. Hoover, 1988), we demonstrate that the correcting approach imposes non-linear (so-called *common-factor* (CF)) restrictions on the parameters of the ADL model. We highlight that these CF restrictions imply a particular form of dynamics which may or may not be valid for the relationships in the data at hand. We go on by illustrating that simply picking an approach and imposing restrictions without testing their adequacy is a poor strategy, since imposing invalid restrictions can bias the estimated regression parameters as well as the estimated SEs, seriously distorting inference. We demonstrate such misspecification biases in a simulation study and pay attention to model selection procedures, which can be used by applied scientists to justify a chosen approach empirically. Specifically, we discuss information criteria (AIC and BIC) and formal likelihood-ratio tests of the restrictions imposed under the approaches. The relevance of these topics for the behavioral literature is emphasized by applying the various approaches to data from a behavioral experiment. We place the reader in the position of an applied scientist modeling the relationship between a contextual cue and affect over time. We find substantial differences in the estimated effects depending on which serial dependence approach one employs. Moreover, the size of these differences in parameter estimates directly corresponds to the extent that the restrictions implied by the various approaches appear violated.

To make the materials as accessible as possible, we focus on $N = 1$ versions of the employed models. We nevertheless show that the restrictions and considerations for misspecification biases extend straightforwardly to multilevel models in Appendix 3. In this appendix, we also provide simulation evidence by reanalysing data generated by Asparouhov and Muthén (2020).

The structure of the manuscript is as follows: In section 2, we tie the various approaches to different models, providing the necessary background for the remainder of the paper. In section 3, we discuss how the various approaches can be considered as special cases of the ADL model, paying particular attention to the CF restriction. Section 4 discusses estimation, model comparison, and misspecification biases. In section 5 we present the $N = 1$ simulation study. Section 6 contains our data analytic example. The discussion focuses on a number of important implications of these findings for the analysis of ILD in the behavioral sciences.

2 Regression models for the analysis of ILD

Regarding notation, we use β to denote a regression effect. We specify which predictor it relates to in subscript. Whenever this predictor is actually a lagged variable, we include L in the subscript. In the superscript we specify which model the regression effect belongs to.

2.1 Linear regressions

We start from the LR model, which posits a linear relationship between a covariate X and a criterion Y. When the LR is applied to time series x_t and y_t where $t \in \mathbb{N}$ indexes discrete time, the formulation is straightforward:

$$\begin{aligned} y_t &= \beta_0^{LR} + \beta_x^{LR} x_t + v_t \\ v_t &\overset{iid}{\sim} N(0, \sigma_v^2) \end{aligned} \tag{1}$$

The model thus assumes that the criterion score at time t is a linear function of the covariate at time t and the error, v_t. The parameters β_0^{LR} and β_x^{LR} are intercept and slope parameters respectively. The errors v_t are assumed to be independently and identically distributed (*iid*) according to a normal distribution with zero mean and variance σ_v^2. When applying such models to time series, one may find after fitting the model that the observed errors or regression *residuals*, $\hat{v}_t$, are correlated with themselves at previous time points, which violates the independence assumption. When one ignores this assumption violation and proceeds with inference, one follows the *neglecting* approach.

2.2 Autoregressive residuals

Many researchers opt to 'correct' for serial dependence in the residuals of a LR model by specifying an AR model (Hamilton, 1994 p.53-56) for how they behave over time. Such a model can be formulated as:

$$\begin{aligned} y_t &= \beta_0^{CF} + \beta_x^{CF} x_t + u_t \\ u_t &= \beta_{Lu}^{CF} u_{t-1} + v_t \\ v_t &\overset{iid}{\sim} N(0, \sigma_v^2), \end{aligned} \tag{2}$$

and has been termed the autocorrelation correction LR (ACLR, McGuirk and Spanos, 2009) model. The reason for the superscript CF for the ACLR model parameters will become evident in section 3.1. Again,

β_0^{CF} is an intercept and β_x^{CF} a slope parameter. The error process u_t is assumed to follow an AR equation, with AR effect β_{Lu}^{CF}. Throughout the manuscript we confine ourselves to AR processes with stable and finite means, variances, and auto(co)variances, and as such the AR effects in this manuscript are assumed to be smaller than 1 in absolute value (for more details about this restriction, see Hamilton, 1994, p. 45-47). Furthermore, we confine ourselves to first-order or AR(1) processes, the temporal dependence of which is captured by a single AR parameter, linking current values of the variable to preceding values (at time t-1).

2.3 Lagged criterion variables

The third strategy for accommodating serial dependence is to specify a lagged effect of the criterion variable (LCV). Such a model can be written as:

$$\begin{aligned} y_t &= \beta_0^{LCV} + \beta_x^{LCV} x_t + \beta_{Ly}^{LCV} y_{t-1} + v_t \\ v_t &\overset{iid}{\sim} N(0, \sigma_v^2) \end{aligned} \tag{3}$$

The model formulation is similar to the LR model, but now y_{t-1} serves as a predictor for y_t in addition to the covariate x_t. The effect of y_{t-1}, β_{Ly}^{LCV}, is an AR effect conditional on x_t.

2.4 Distributed lags: Unifying the approaches

Less common in the behavioral sciences is to allow for lagged covariate effects on top of lagged criterion effects. Nevertheless these models, which are termed ADL models, will play a crucial role in what follows. The simplest ADL model may be formulated as:

$$\begin{aligned} y_t &= \beta_0^{ADL} + \beta_x^{ADL} x_t + \beta_{Ly}^{ADL} y_{t-1} + \beta_{Lx}^{ADL} x_{t-1} + v_t \\ v_t &\overset{iid}{\sim} N(0, \sigma_v^2) \end{aligned} \tag{4}$$

Each of the previous three approaches can be considered a special case of this ADL model (see also Hendry et al., 1984), which brings us to the next section.

3 Restrictions on ADL parameters

It is evident that the full or free ADL model contains the LCV model as a special case. Specifically, the LCV model imposes a *restriction* on the effect of x_{t-1} , namely that this effect equals zero, $\beta_{Lx}^{ADL} = 0$. The LR model is also a special case of the ADL model in that it imposes two restrictions, namely that $\beta_{Lx}^{ADL} = 0$ and $\beta_{Ly}^{ADL} = 0$. As will be shown in the next section, the nature of the restriction is a bit more involved for the ACLR model. To facilitate its exposition, and since the presence of an intercept term has no implications for the results presented in the manuscript, we assume that the data is centered prior to analysis such that the various intercept terms become 0.

3.1 The common factor restriction

The simplest way to notice that the regression parameters are constrained in a nonlinear way when a LR with AR errors is specified (see Hoover, 1988), is by noting that equation 2 specifies two equalities for u_t: $u_t = \beta_{Lu}^{CF} u_{t-1} + v_t$ and $u_t = y_t - \beta_x^{CF} x_t$. If the equations are assumed to describe the evolution of y_t and u_t for all time points, it is obvious that $u_{t-1} = y_{t-1} - \beta_x^{CF} x_{t-1}$ also holds. Substituting the first equality for u_t, and subsequently the second equality for u_{t-1} we find that:

$$
\begin{aligned}
y_t &= \beta_x^{CF} x_t + u_t && \text{|Equation 2} \\
y_t &= \beta_x^{CF} x_t + \beta_{Lu}^{CF} u_{t-1} + v_t && \text{|Substitute } u_t \\
y_t &= \beta_x^{CF} x_t + \beta_{Lu}^{CF} (y_{t-1} - \beta_x^{CF} x_{t-1}) + v_t. && \text{|Substitute } u_{t-1}
\end{aligned}
$$

Writing these terms out, we obtain that:

$$y_t = \beta_{Lu}^{CF} y_{t-1} + \beta_x^{CF} x_t - \beta_{Lu}^{CF} \beta_x^{CF} x_{t-1} + v_t. \tag{5}$$

The ACLR model can thus be written as an ADL model:

$$y_t = \beta_{Ly}^{ADL} y_{t-1} + \beta_x^{ADL} x_t + \beta_{Lx}^{ADL} x_{t-1} + v_t. \tag{6}$$

by imposing the nonlinear CF restriction $\beta_{Lx}^{ADL} = -\beta_x^{ADL} \beta_{Ly}^{ADL}$ (Hendry and Mizon, 1978), with $\beta_{Lu}^{CF} = \beta_{Ly}^{ADL}$.

To understand why this restriction is referred to as a 'common factor' restriction - and has nothing to do with common factors as in common latent variables -, we introduce the *lag operator L*. The operator simply returns the variable at the previous time point, and thus acts on a variable x_t such that $Lx_t = x_{t-1}$ (see Hamilton, 1994, chapter 2). This allows us to represent an AR(1) process $y_t = \beta_{Ly} y_{t-1} + v_t$ simply as[1]

[1] The term $(1 - \beta_{Ly} L)$ is a *lag polynomial*, and if $|\beta_{Ly}| < 1$, it is invertible (Hamilton, 1994, p.28-29). It is well known that inverting an AR(1) process yields an equivalent infinite order moving average (MA) process: $y_t = v_t(1 - \beta_{Ly} L)^{-1}$. Here, $(1 - \beta_{Ly} L)^{-1} = (1 + (\beta_{Ly} L) + (\beta_{Ly} L)^2 + ...)$, and as such inverting the polynomial corresponds to solving by recursive substitution.

$(1-\beta_{Ly}L)y_t = v_t$. In terms of lag operators, the ADL under the CF restriction can be written as:

$$\begin{aligned} y_t &= \beta_{Ly}y_{t-1} + \beta_x x_t - \beta_x\beta_{Ly}x_{t-1} + v_t \\ y_t - \beta_{Ly}y_{t-1} &= \beta_x x_t - \beta_x\beta_{Ly}x_{t-1} + v_t \\ (1-\beta_{Ly}L)y_t &= \beta_x(1-\beta_{Ly}L)x_t + v_t \end{aligned} \tag{7}$$

The common factor corresponds to the presence of the factor $(1-\beta_{Ly}L)$ for both y_t and x_t (Hoover, 1988).

3.2 Implications of the CF restriction

The role of the nonlinear CF restriction $\beta_{Lx}^{ADL} = -\beta_{Ly}^{ADL}\beta_x^{ADL}$ might seem rather abstract, but is actually very concrete: It ensures that x_t has only a transient effect on y_t. That is, the constrained direct effect of x_{t-1} on y_t cancels out its indirect effect via y_{t-1}, which equals $\beta_{Ly}^{ADL}\beta_x^{ADL}$. This cancelling out implies that the influence of x_t on y_t does not accumulate over time:

$$\begin{aligned} (1-\beta_{Ly}L)y_t &= \beta_x(1-\beta_{Ly}L)x_t + v_t \\ (1-\beta_{Ly}L)^{-1}(1-\beta_{Ly}L)y_t &= \beta_x(1-\beta_{Ly}L)^{-1}(1-\beta_{Ly}L)x_t + (1-\beta_{Ly}L)^{-1}v_t \\ y_t &= \beta_x x_t + (1-\beta_{Ly}L)^{-1}v_t \\ y_t &= \beta_x x_t + v_t + \beta_{Ly}v_{t-1} + \beta_{Ly}^2 v_{t-2} + ... \\ y_t &= \beta_x x_t + \sum_{i=0}^{\infty}\beta_{Ly}^i v_{t-i} \end{aligned} \tag{8}$$

Hence, if a CF restriction holds, previous covariate values x_{t-1}, x_{t-2}, ..., x_{t-i} do not enter into the equation. In other words, x_t has no lasting effects on y_t. Only the variation in y_t due to the error process v_t - and hence the variation independent of x_t - is carried over in time. In contrast, both the LCV and ADL models imply such carry over effects of covariate influences, to the extent that β_{Lx}^{ADL} is different from $-\beta_{Ly}^{ADL}\beta_x^{ADL}$ (see Appendix 1).

Whether the CF restriction is appropriate thus depends on whether we may presume that a particular covariate has a purely transient effect which does not last in the above sense. From a substantive perspective, this appears overly restrictive for many commonly used covariates in affective science. Imagine $y_1, y_2, ..., y_t$ reflect measurements of negative affect, and $x_1, x_2, ..., x_t$ reflect emotionally relevant events. The CF restriction implies that the effects of the events on negative affect do *not* last, while the effects of all unmeasured factors in v_t do.

A CF specification nevertheless makes sense in some situations, which we can investigate by formulating particular models for how the covariate behaves over time. For instance the CF restriction holds when the covariate is time itself ($x_t = t$, see Hendry and Juselius, 2000, equation 7). Such a model is used to model a linear time trend in the data around which the process displays AR(1) dynamics. One can think of this situation as first detrending the data to then yield the AR(1) dynamics contained in the error terms of this regression (Michaelides and Spanos, 2020; Wooldridge, 2012, p.368).

A second plausible model for the behavior of a covariate is an AR(1) process. Indeed the covariate may also display serial dependence and and one might think of joint VAR(1) process having generated the covariate and criterion data. In this VAR(1) process, the contemporaneous effect of x_t on y_t is accommodated by a contemporaneous covariance between the error terms (see Appendix 2). In this case, we can delineate two situations in which the CF restriction holds. The first situation is that y_t and x_t have identical AR effects and no crossregressive effects, implying that y_t and x_t have equal lag one autocorrelations (Appendix 2a; see also Spanos, 1987; Mizon, 1995). A second situation pertains to unequal AR effects, but then the CF restriction implies the presence of an additional compensatory crossregressive effect from x_{t-1} to y_t. In this situation, the lag one autocorrelations of y_t and x_t will also differ (Appendix 2b; see also McGuirk and Spanos, 2009).

The above examples show that CF constrained model structures can imply highly restrictive relationships for how criterion and covariate behave over time, which may not be valid for a particular data analytic problem (as stressed in the econometric literature; Hendry and Mizon, 1978; Hoover, 1988; Mizon, 1995; Spanos, 1987; McGuirk and Spanos, 2009). The utility of the CF restriction for psychological science can thus best be gauged by providing *empirical evidence* for its absence or presence.

4 Estimation, model comparison, and misspecification biases

4.1 Estimation

The parameters of the considered models can be estimated in a variety of ways[2]. In this paper we employ maximum likelihood (ML) estimation as implemented in the Mplus (Muthén and Muthén, 2017, ver. 8.5) software, placing constraints on the regression parameters where appropriate. For instance, when estimating the parameters of an ACLR model, we specify a CF restricted ADL with the non-linear constraint on β_{Lx}^{ADL}. We opt for ML estimation since we shall use the log-likelihood values for model comparison and selection

[2]Ordinary least squares is the classical approach for estimating the parameters of the ADL, LR, and LCV models. Generalized least squares is the classical method employed for the ACLR (Aitken, 1936, Cochrane and Orcutt, 1949)

(see next subsection). It is important to note that the implications of the different approaches that we aim to point out are related purely to the model structures rather than the method by which their parameters are estimated. It is straightforward to replicate the various results presented in the manuscript using other estimation methods. For example, in our reanalysis of the multilevel data generated by Asparouhov and Muthén, 2020 in appendix 2, we employ Bayesian estimation as done in their original paper.

4.2 Model selection

Model selection procedures enable to empirically decide among the various model structures discussed. In this paper we discuss two approaches, one based on information criteria, and the other based on formal statistical testing.

Information criteria do not require the considered models to be nested, and as such can be used to directly compare the various approaches. The two most popular criteria are the Bayesian information criterion (BIC, Schwarz, 1978) and the Akaike information criterion (AIC, Akaike, 1974). For the ADL model for instance, these information criteria are calculated as $BIC^{ADL} = -2\,\ell^{ADL} + p^{ADL}\,ln(N)$ and $AIC^{ADL} = -2\,\ell^{ADL} + p^{ADL}\,2$ respectively. After computing the AIC (resp. BIC) values of all considered models, the model with the lowest value is selected. AIC and BIC thus differ in the imposed penalty term, in that AIC only considers the number of model parameters p, whereas BIC also takes the number of time points N into account. For the sample sizes we consider in this paper, the penalty value will be lower for AIC, implying that AIC will favor more general models than BIC.

Likelihood-ratio tests can also be considered for the models we discuss, since each model is nested in the free ADL model. The free ADL model can thus be used as a full model, and a chosen competitor as the alternative model. Concretely, a likelihood-ratio test of the CF restriction (e.g. Burridge, 1981) tests the null hypothesis $\mathcal{H}_0 : \beta^{ADL}_{Lx} + \beta^{ADL}_{x}\beta^{ADL}_{Ly} = 0$ against the alternative $\mathcal{H}_1 : \beta^{ADL}_{Lx} + \beta^{ADL}_{x}\beta^{ADL}_{Ly} \neq 0$. The test statistic λ is obtained by calculating the difference between the log-likelihood values associated with the ML solutions of the CF constrained and the free ADL model, ℓ^{CF} and ℓ^{ADL}, and multiplying this difference by -2, hence $\lambda = -2(\ell^{CF} - \ell^{ADL})$. The test statistic is compared with a central χ^2 distribution with a single degree of freedom, stemming from the additional free parameter β_{Lx} in the free ADL relative to the CF constrained ADL. Although other tests can be considered (e.g. Sargan, 1980), MCMC studies on the performance of likelihood-ratio tests of the CF restriction have reported to perform adequately (e.g. Garijo and Lacambra, 2011) even when the number of time points are relatively limited (Mur and Angulo (2006) considered $T = 25$, for instance). Tests of the other restrictions can be formulated in analogous ways, comparing the log-likelihood of a model under the desired restrictions with the one of the free ADL model using a central χ^2 distribution with degrees of freedom equalling the difference in the number of freely estimated parameters.

4.3 Misspecification biases

As highlighted previously, imposing CF and other restrictions without putting them to a test can have detrimental effects for estimation if restrictions are invalid. Specifically, regression effect and SE estimates can be biased. We refer to these as misspecification biases. Biased regression effect estimates can arise due to omitting a relevant predictor (omitted variable bias, Wooldridge, 2012, p. 88-92), or due to 'forcing' a relationship to hold which does not (e.g., the CF restriction on the ADL parameters). Biased SE estimates can arise if temporal dependence is not properly modeled leading to correlated errors. This is particularly relevant in the context of a LR fit to serially dependent data (the neglecting approach). It is important to note that, unlike small sample biases that are known to affect AR modeling (see Maeshiro, 2000; Krone et al., 2017), the discussed misspecification biases do not diminish as sample size increases. They instead persist as we illustrate in the following section.

To already provide two more concrete examples in the context of the models we discuss, omitted variable bias is for instance to be expected when data generated by a free ADL model is fit with the LCV model. The data generating model implies a nonzero β_{Lx}, yet the LCV estimation model incorrectly restricts $\beta_{Lx} = 0$. Since the omitted x_{t-1} also correlates with y_{t-1}, $\hat{\beta}^{LCV}_{Ly}$ will display bias. An example of forcing an invalid restriction to hold would occur when the the ACLR model is fit to the data generated by the free ADL model (Spanos, 1987). The ACLR constrains β_{Lx} in the ADL in a specific way, forcing $\beta^{ADL}_{Lx} = -\beta^{ADL}_{Ly}\beta^{ADL}_{x}$ as we have seen. If however $\beta^{ADL}_{Lx} \neq -\beta^{ADL}_{Ly}\beta^{ADL}_{x}$, the estimation model is misspecified. As such, the estimates of the ACLR model parameters will deviate from the true values to the extent the restriction is invalid.

5 Simulation study

5.1 Problem

The purpose of our simulation study is to compare the performance of four approaches for dealing with serial dependence in ILD: neglecting (LR), correcting (ACLR), and two approaches that include lagged observed variables (ADL and LCV). Modeling performance is cast in terms of biases in the regression effect estimates and SE estimates, and estimation precision. The theory above allows us to make specific predictions about the emergence of misspecification biases (see section on expectations). As for model selection, we assess AIC and BIC performance in selecting the true data generating model, and the performance of likelihood-ratio tests, particularly when testing for the adequacy of a CF restriction and LCV restriction.

5.2 Design

The present simulation study is based on a recently reported simulation study by Asparouhov and Muthén (2020) in which they pitted a multilevel LCV (DSEM) against a multilevel ACLR (RDSEM). We build on their design, refining and extending it in a number of ways. Importantly, although we focus on $N = 1$ models, the results we report here generalize to the multilevel case (see reanalysis of the original datasets in appendix 3).

5.2.1 Data generation

Concerning data generation, we include the free ADL as a simulation model, on top of the LCV and ACLR. We further distinguish 2 conditions for the ADL simulation models, one where the CF restrictions are 'far' from holding (ADL), and one in which the CF restrictions are only 'slightly' violated (ADL_s). Our data generating models can thus be seen as ADL models differing in their implications for β_{Lx} (see Table 1).

Next to the data simulation model, we vary two more data characteristics, further extending the design of Asparouhov and Muthén (2020) . Firstly, we include 3 settings for T, $\{50, 150, 1000\}$. Additionally, we vary the serial in dependence in the covariate, γ. We generate the covariate according to an AR(1) model with AR effect $\gamma = .7$ or $\gamma = 0$, implying lag-one autocorrelations of the same size. We thus have 24 (4 x 3 x 2) data cells in our simulation design. Per design cell, we generated 500 data replications using Mplus (Muthén and Muthén, 2017, version 8.5). Data were generated with an intercept of 0, and were centered prior to analysis. For all data simulation conditions, we set $\beta_{Ly} = .7$.

5.2.2 Estimation and Analysis

Per replicated dataset we fit four estimation models: LR, ACLR, LCV, and ADL. In total we have 96 (24 x 4) estimation conditions, the data from each cell being estimated by the 4 estimation models. Model fitting was also conducted in Mplus, and the Mplus results were processed in R (R Core Team, 2021, version 4.1.1). In each case, the intercept was fixed to 0 during estimation. Per design cell and estimation model, we extracted the mean of each parameter estimate and SE estimates over replications from the Mplus output. To assess biases of the regression effect estimators, the average of the parameter estimates was compared with the parameter value underlying the data generating model. For biases in the SE estimators, the average of the SE estimates was compared with the empirical SD (i.e. the observed precision) of the parameter estimates.

Regarding model selection and comparison, we extracted BICs and log-likelihoods for each estimated

Models	β_x	β_{Ly}	Restriction on β_{Lx}	β_{Lx}
ADL	1	0.7	None	0.5
ADL_s	1	0.7	None	-0.5
ACLR	1	0.7	$-\beta_x\beta_{Ly}$	-0.7
LCV	1	0.7	0	0

Table 1: Data generating models of the simulation study with utilized parameter values, and the restrictions implied for β_{Lx}.

model. As an index of performance, we calculated the proportion of cases that each estimation model attained the lowest BIC when fit to data generated by a particular simulation model. For the LR tests, our index of performance is obtained by calculating the proportion of times $\mathcal{H}_0$ was rejected, allowing us to assess the power and type 1 error rates, depending on whether the data were generated under one of the constrained models or an ADL respectively. All utilized scripts can be found in the supplementary materials.

5.3 Expectations

5.3.1 Misspecification biases

For our specific design, we can summarize our expectations on (mainly) misspecification biases per estimation model, with two exceptions. The first obviously concerns the cases where the data simulation model is fitted and for which we expect no misspecification biases in neither regression effect nor SE estimates. This case is not re-iterated in the following list. The second exception, which is detailed in the following, concerns the situation where the LR is fitted to data from the ACLR, which is different from the situation where data come from the full ADL or the LCV model.

- ADL (fitted to LCV or ACLR data): Since the various models are nested in the ADL and it thus represents an over-specified model, we do not expect to find any biases when the ADL is fit to data generated by the competing models.

- LCV (fitted to ADL or ACLR data): In these cases x_{t-1} is omitted from the model while it has unique effects on y_t *conditional on* the predictors in the estimation model, x_t and y_{t-1}. It is also correlated with y_{t-1}. We therefore expect omitted variable bias for the effect estimate of y_{t-1}, $\hat{\beta}_{Ly}^{LCV}$. If x_t is autocorrelated (i.e., x_t and x_{t-1} and consequently x_t and y_{t-1} become correlated, see Ariens et al. (2022)), we additionally expect omitted variable bias for the effect estimate of x_t, $\hat{\beta}_{x}^{LCV}$ (see Section 4.1.3). We do not have specific bias expectations for SE estimates.

- ACLR (fitted to ADL or LCV data): The corresponding CF restricted ADL imposes the restriction $\beta_{Lx}^{ADL} = -\beta_{Ly}^{ADL}\beta_{x}^{ADL}$. We thus expect misspecification biases in both $\hat{\beta}_{x}^{CF}$ and $\hat{\beta}_{Ly}^{CF}$. If data are generated under the ADL_s, we expect biases to be less pronounced, because the restriction is less strongly violated. We do not have specific expectations for the SE estimates.

- LR: If fitted to ADL or LCV data, x_{t-1} and/or y_{t-1} are omitted while both or at least y_{t-1} have unique effects on y_t conditional on the predictor in the LR. However, only if x_t is autocorrelated will it be correlated with x_{t-1} and y_{t-1}, and will there be omitted variable bias for $\hat{\beta}_{x}^{LR}$.

- LR: If fitted to CF restricted ADL data, x_{t-1} and y_{t-1} are again omitted, but only the part of y_{t-1} independent of x_{t-1} has a unique effect on y_t conditional on the predictor in the LR. This part of y_{t-1} is thus per definition not correlated with x_t, also if x_t is autocorrelated, so we expect no biases in the regression effect estimates. In fact, it is actually simpler to think explicitly of the ACLR as underlying the data: In this case we omit only u_{t-1}, which is under no circumstances correlated with x_t, hence, there cannot be omitted variable bias. For the SE estimate of $\hat{\beta}_x^{LR}$ we expect - generally under the LR model based on various textbook warnings - bias, at least if the covariate is autocorrelated (Woolridge, p. 413 f.).

Importantly, we expect all misspecification biases to persist in large samples.

5.3.2 Model selection

Regarding model selection, both for BIC and likelihood-ratio tests, we expect good model selection performance in large samples, with a slight attenuation in small samples (e.g. Mur and Angulo, 2006). Specifically, we expect nominal type 1 error rates ($\alpha = .05$) and low type 2 error rates for the likelihood-ratio analysis, and we expect AIC and BIC to accurately select the correct model over competitor models. We have no particular expectations with respect to γ. We are interested in how the performance of a likelihood-ratio test of a CF restriction depends on the extent of misspecification, i.e. the comparison between ADL and ADL_s data simulation conditions.

5.4 Results

For reasons of presentability, we report for misspecification biases the conditions $T = 50$ (small samples) and $T = 1000$ (large samples). The supplementary materials contain the results for all conditions.

5.4.1 Misspecification biases and estimation precision

Estimation condition		ADL		LCV		ACLR		LR
γ	T	β_{Ly}	β_x	β_{Ly}	β_x	β_{Ly}	β_x	β_x
ADL								
0.0	50	-0.03 (0.07) [-0.01]	0.00 (0.14) [0.00]	0.07 (0.07) [0.01]	-0.01 (0.16) [0.00]	0.05 (0.09) [0.01]	-0.57 (0.16) [0.03]	-0.06 (0.29) [0.01]
0.7	50	-0.01 (0.04) [-0.01]	0.00 (0.14) [0.00]	0.06 (0.04) [0.01]	0.23 (0.13) [0.01]	0.20 (0.06) [0.01]	-0.33 (0.23) [0.05]	1.29 (0.35) [-0.19]
0.0	1000	0.00 (0.01) [0.00]	0.00 (0.03) [0.00]	0.09 (0.01) [0.00]	0.00 (0.03) [0.00]	0.09 (0.02) [0.00]	-0.58 (0.03) [0.01]	-0.01 (0.07) [0.00]
0.7	1000	0.00 (0.01) [0.00]	0.00 (0.03) [0.00]	0.07 (0.01) [0.00]	0.23 (0.03) [0.00]	0.23 (0.01) [0.00]	-0.34 (0.05) [0.01]	1.63 (0.08) [-0.05]
ADLs								
0.0	50	-0.06 (0.10) [-0.01]	0.00 (0.14) [0.00]	-0.22 (0.09) [-0.03]	0.01 (0.15) [-0.01]	-0.05 (0.11) [-0.01]	-0.09 (0.12) [0.00]	0.00 (0.19) [-0.01]
0.7	50	-0.07 (0.10) [-0.01]	0.00 (0.14) [0.00]	-0.22 (0.08) [-0.03]	-0.16 (0.13) [-0.03]	-0.02 (0.10) [-0.01]	-0.02 (0.15) [0.00]	0.21 (0.15) [-0.10]
0.0	1000	0.00 (0.02) [0.00]	0.00 (0.03) [0.00]	-0.17 (0.02) [-0.01]	0.00 (0.03) [0.00]	0.00 (0.02) [0.00]	-0.09 (0.03) [0.00]	0.00 (0.04) [0.00]
0.7	1000	0.00 (0.02) [0.00]	0.00 (0.03) [0.00]	-0.17 (0.02) [0.00]	-0.20 (0.03) [0.00]	0.04 (0.02) [0.00]	-0.02 (0.03) [0.00]	0.27 (0.03) [-0.02]
LCV								
0.0	50	-0.04 (0.09) [-0.01]	0.00 (0.14) [0.00]	-0.03 (0.08) [-0.01]	0.00 (0.14) [0.00]	0.00 (0.10) [0.00]	-0.33 (0.13) [0.01]	-0.03 (0.23) [0.00]
0.7	50	-0.03 (0.07) [-0.01]	0.00 (0.14) [0.00]	-0.02 (0.05) [-0.01]	0.01 (0.12) [0.00]	0.15 (0.08) [0.00]	-0.17 (0.18) [0.02]	0.75 (0.23) [-0.13]
0.0	1000	0.00 (0.02) [0.00]	0.00 (0.03) [0.00]	0.00 (0.02) [0.00]	0.00 (0.03) [0.00]	0.05 (0.02) [0.00]	-0.34 (0.03) [0.00]	0.00 (0.05) [0.00]
0.7	1000	0.00 (0.01) [0.00]	0.00 (0.03) [0.00]	0.00 (0.01) [0.00]	0.00 (0.03) [0.00]	0.19 (0.01) [0.00]	-0.18 (0.04) [0.01]	0.95 (0.05) [-0.04]
ACLR								
0.0	50	-0.06 (0.11) [-0.01]	0.00 (0.14) [0.00]	-0.29 (0.09) [-0.04]	0.01 (0.16) [-0.01]	-0.06 (0.11) [-0.01]	0.00 (0.12) [0.00]	0.01 (0.19) [-0.01]
0.7	50	-0.08 (0.11) [-0.01]	0.00 (0.14) [0.00]	-0.30 (0.10) [-0.04]	-0.25 (0.14) [-0.05]	-0.07 (0.11) [-0.01]	0.00 (0.14) [0.00]	0.00 (0.14) [-0.09]
0.0	1000	0.00 (0.02) [0.00]	0.00 (0.03) [0.00]	-0.24 (0.02) [-0.01]	0.00 (0.04) [0.00]	0.00 (0.02) [0.00]	0.00 (0.03) [0.00]	0.00 (0.04) [0.00]
0.7	1000	0.00 (0.02) [0.00]	0.00 (0.03) [0.00]	-0.24 (0.02) [-0.01]	-0.32 (0.03) [-0.01]	0.00 (0.02) [0.00]	0.00 (0.03) [0.00]	0.00 (0.03) [-0.02]

Table 2: Results for misspecification biases. Columns denote estimation models, rows denote simulation conditions. Per estimation model, we report bias of parameter estimates, (SD of parameter estimates), and [bias of SE estimates] for the parameter given in the corresponding column.

Table 2 displays the results regarding misspecification biases and estimation precision, which we briefly summarize. The results in general confirm our expectations regarding misspecification biases. Firstly, the ADL does not display evidence of misspecification bias. Small sample biases in $\hat{\beta}_{Ly}^{ADL}$ appear to be present, as expected, and can also be found for correctly specified LCV and ACLR models. Estimating the ACLR model parameters results in biased regression effect estimates when data was generated by the LCV or the ADL conditions. We observe biases both for $\hat{\beta}_{Ly}^{CF}$ and $\hat{\beta}_{x}^{CF}$. Especially the latter effect estimate is biased. The covariate being autocorrelated changes the bias pattern. In the ADL_s condition, biases remain present, but are attenuated, confirming our expectation that the extent of bias depends on the extent of misspecification. These biases persists in large T situations. The LCV displays biases in $\hat{\beta}_{Ly}^{LCV}$ when data is generated by the ADL or ACLR models. Biases in $\hat{\beta}_{x}^{LCV}$ are however found only if the covariate is autocorrelated. Again these biases persist in large samples. The LR displays biases in $\hat{\beta}_{x}^{LR}$ when data is generated by the ADL or LCV models, but only when the covariate is autocorrelated. In these cases we see that the SE estimates also display bias, they are in fact deflated, see Wooldridge (2012), p. 414. These biases persist in large samples. When data is generated under the ACLR, $\hat{\beta}_{x}^{LR}$ appears unbiased regardless of autocorrelation in the covariate.

Looking at estimation precision, we note that the observed SD of $\hat{\beta}_{x}^{LR}$ always appears larger compared to those of the well-specified competing models. This indicates that, although omitting temporal relationships present in the data does not always cause misspecification bias, it can improve estimation precision[3]. These precision differences are substantial in small samples.

5.4.2 Model selection

Figure 1 panels A and B show the proportion of cases each estimation model was selected based on AIC and BIC respectively, with panels showing the different data generating models. In panel C, the results of the likelihood-ratio tests are presented. We first remark on likelihood-ratio test performance in terms of type 1 and 2 error rates, after which we discuss the information criteria results.

Likelihood-ratio tests of the CF restriction

- For data generated by the ACLR model we find that likelihood-ratio tests of the CF restriction have nominal type 1 error rates. For $\gamma = 0$, the type 1 error rates are .07, .06, and .05 for T = 50, 150, and 1000 respectively. $\gamma = .7$ does not alter these results, where we find .06, .06, and .06 respectively.

- For data generated by the ADL, we find that likelihood-ratio tests of the CF restriction do not result in type 2 errors, the tests correctly refuted the null hypothesis for all datasets.

[3] In general however, the inclusion of *irrelevant* predictors is known to decrease estimation precision (Wooldridge, 2012, p. 88)

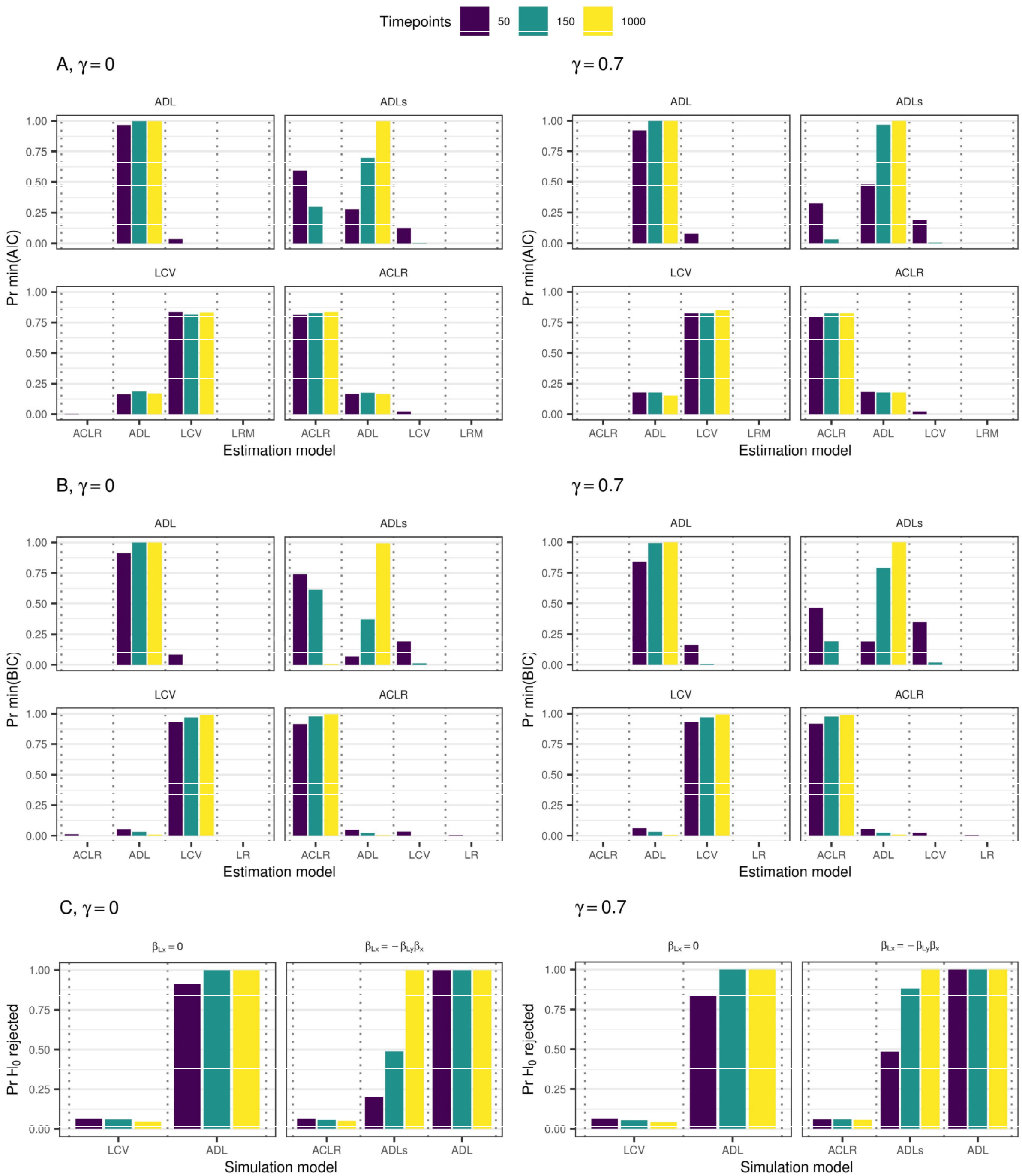

Figure 1: Results of the model comparison analysis. In panels Aand B, we report the results for the AIC and BIC analysis respectively, and in panel C the results for the likelihood-ratio tests are shown. In panels A and B, the proportion of cases where AIC or BIC was lowest (y-axis) over estimation models (x-axis) is displayed. The plots are faceted by simulation condition, further separating the different simulation models and the case where the covariate was not autocorrelated (left, $\gamma = 0$) with cases where the covariate was autocorrelated (right, $\gamma = .7$). Colors denote settings for T. In panel B, the proportion of cases $\mathcal{H}_0$ was rejected is displayed (y-axis). Here, the plots are faceted by which $\mathcal{H}_0$ was tested, with on the x-axis the simulation model employed. Dotted lines are included within plots to facilitate discrimination.

- For data generated by the ADL_s, we find that likelihood-ratio tests of the CF result in increased type 2 error rates in small samples. For $\gamma = 0$, the type 2 error rates are .8, .5, and 0 for T = 50, 150, and 1000 respectively. For $\gamma = .7$, the type 2 error rates are .51, .12, and 0 respectively.

Likelihood-ratio tests of the LCV restriction

- For data generated by the LCV, we find that likelihood-ratio tests of the LCV restriction have nominal type 1 error rates. For $\gamma = 0$, the type 1 error rates are .07, .06, and .05 for T = 50, 150, and 1000 respectively. $\gamma = .7$ does not alter these results, where we find .06, .05, and .04 respectively.
- For data generated by the ADL, we find that for $\gamma = 0$, the type 2 error rates are .09, 0, and 0 for T = 50, 150, and 1000 respectively. For $\gamma = 0.7$, we find .16, 0, and 0.

Information criteria results

Figure 1, panels A and B, show that AIC and BIC respectively yield fairly similar model selection results as using likelihood-ratio tests. Yet there are some discernible differences. For smaller sample sizes (i.e. $T = 50$ or $T = 150$) and when data was generated under either the LCV or ACLR models, AIC tends to incorrectly favor the more general ADL compared to the likelihood-ratio tests, and especially compared to BIC. If the data were generated under an ADL, we see that AIC tends to outperform BIC, at least for $T = 50$. For data generated under the ADL_s condition, we observe worse AIC and BICperformance in that they both often erroneously select the ACLR or LCV models, as was also the case for the likelihood-ratio tests. In comparative terms, AIC performed best in this condition, followed by the likelihood-ratio tests and BIC. In large T situations, BIC and the likelihood-ratio tests appear to perform particularly well, while AIC still often incorrectly selects the ADL model. It is known that AIC tends to overestimate model orders in AR settings, even in large T situations (Talata, 2005)

To conclude this section, the results suggest that these accessible model selection tools can perform well in selecting the simulation model. However, model selection accuracy appears strongly dependent on both the effect sizes and the number of available time points: 'slight' violations of the CF restriction resulted in worse model comparison performance, particularly in small samples.

6 A data-analytic example

In this section we analyze data from a simple experimental paradigm, a variant of which is reported in Vanhasbroeck et al. (2022). The purpose is to make the role of the restrictions more tangible, and emphasize the relevance of these issues for applied work. In the experiment, a gambling paradigm, participants (N =

89) were asked to repeatedly chose between 4 choice options (doors), behind which were located monetary gains or losses. Unbeknownst to the participants, the series of gains or losses was predetermined by the experimenters. Figure 2 panel A displays an example of an experimental trial, showing that the cumulative gains or losses were available as cue at the top of the screen. After each trial they were asked to report their affective state on the evaluative space grid (Larsen et al., 2009, see Figure 2 panel A). The next trial commenced 1 second after the affective evaluation, the participant's reported position on the grid disappearing with each new trial. As such, positive and negative emotions were measured over 152 trials (preceded by 10 practice trials). Figure 2 panel B displays a time series of (within-person) standardized NA responses and gambling winnings for a single participant.

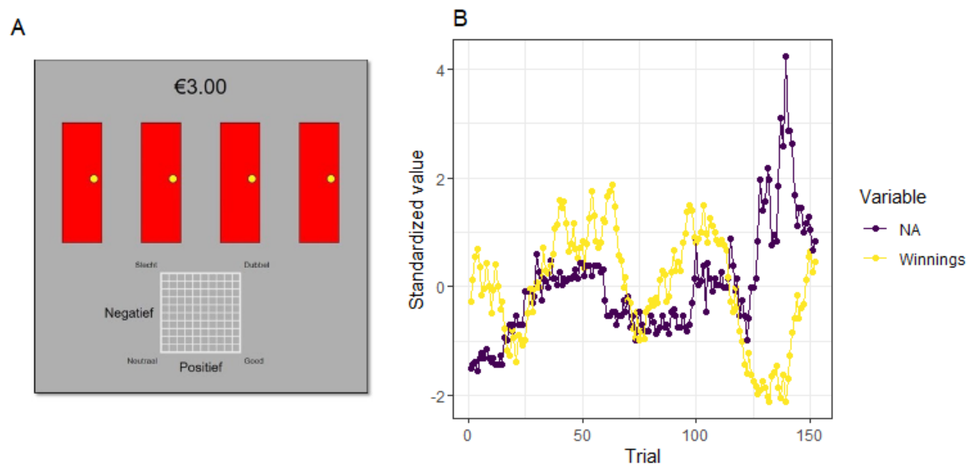

Figure 2: Panel A displays an example of a trial. Participants see four doors (middle), the evaluative space grid (bottom) and their cumulative gambling winnings (top). Upon selecting a door, a monetary win or loss was displayed behind the door, and participants asked to indicate their emotional state on the evaluative space grid. Panel B displays a time series of a single participant. Standardized gambling winnings and NA reports are displayed.

We are interested in the relationship between negative affect (NA) and the gambling winnings displayed on the screen, where NA functions as the criterion variable y_t and the gambling winnings as the covariate x_t. To this end, we extract the NA responses by using the $y-$coordinates of the evaluative space grid. Moreover, the practice trials are removed, and all variables are standardized to have mean zero and unit variance. For each participant, we then fit four models: (1) the LR, (2) the LCV (3) the ACLR, and (4) the ADL model, and compute their BIC values. All four models postulate within-subjects variation in NA and a contemporaneous effect of the displayed gambling winnings on NA, but they assume different forms of serial dependence, as previously explained. To again highlight the specificities of the ALCR model, recall that it assumes that only the variation in NA that is not due to the gambling winnings carries over across trials. The ADL and LCV in contrast allow also the effects of the experimental manipulation to be lasting in time.

We first inspect the autocorrelations in the residuals[4] of a simple LR model, displayed in Figure 3, panel A. As expected, serial dependence is present for many individuals. Moreover, BIC never favored the LR model. Rather, the ADL, ACLR, and LCV models were selected for respectively 72, 14, and 3 participants. AIC favored these models in 77, 9, and 3 participants, hence selecting the more general ADL model for 5 participants. This result pattern is in line with the pattern observed in section 5.4.2 where we saw that AIC tends to favor more general models than BIC for the sample sizes we consider. While we by no means have a guarantee that the ADL is the ”correct” model, and we should further verify the adequacy of the restrictions it invokes, these values suggest to refute both the CF specification, the LCV, and the LR for the majority of participants.

Next, Figure 3, panel B displays violin plots of the estimates of the four models over individuals. Overall, the estimates of the gambling stimuli effects $\hat{\beta}_x$ are negative, indicating that increased winnings are associated with decreased NA. The estimates of the serial dependence $\hat{\beta}_{Ly}$ are positive overall. Importantly, there are substantial differences between the approaches, since the violins display different characteristics (e.g. spread) depending on which approach is selected. For a small number of participants the LR even results in *positive* estimates $\hat{\beta}_x^{LR}$, which would rather paradoxically imply that larger winnings are associated with *increased* NA. The estimates of the error variance are on average largest for the LR model and smallest for the ADL model. The ADL model appears to explain most variability in the affective responses on average, over competitor models.

We then try to tie the size of these estimate differences to the extent that the restrictions appear violated, based on the ADL estimates. Concretely, we expect that the differences one obtains in the covariate effect ($\hat{\beta}_x$) when fitting an ADL model versus an ACLR model depend on how strongly the CF restriction appeared violated. The researchers conclusions regarding the covariate should thus differ as a function of the extent that $|\hat{\beta}_{Lx}^{ADL} + \hat{\beta}_x^{ADL}\hat{\beta}_{Ly}^{ADL}| > 0$, where $|.|$ denotes absolute value. For the LCV model, we expect conclusions to differ as a function of the extent that $|\hat{\beta}_{Lx}^{ADL}| > 0$. This is indeed what we see in Figure 3, panels C1 and C2. Finally, we also inspect the BIC values, expecting that BIC favors the ADL model over a particular competitor to the extent that the restriction imposed by the latter appears violated. This pattern is indeed found in Figure 3 panels D1 and D2.

7 Discussion

In this manuscript, we have endeavored to address the important issue of how to select an appropriate technique for handling serial dependence in the context of regression models of ILD. We compared the

[4]These residuals were extracted using the lm() function in R.

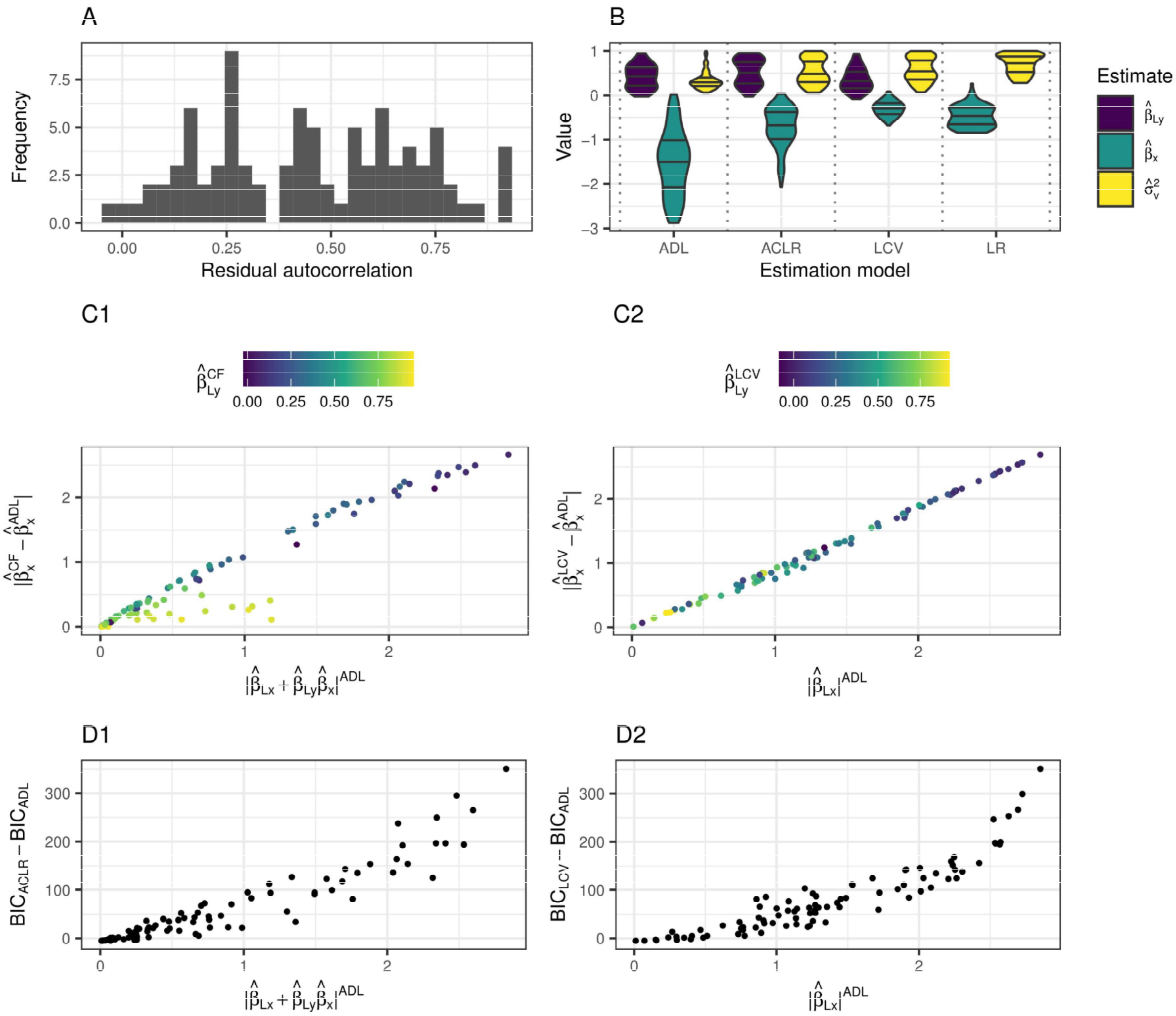

Figure 3: Panel A displays a histogram of the lag-1 autocorrelations in the residuals of a LR model fit to the data. Panel B displays violin plots of the estimates of the parameters over individuals. Colors distinguish the estimates of covariate effect ($\hat{\beta}_x$), the serial dependence ($\hat{\beta}_{Ly}$ or $\hat{\beta}_{Lu}$), and the error variance ($\hat{\sigma}_v^2$). The x-axis denotes which estimation model the estimates belong to. Dotted lines are included to facilitate discrimination. Panels C1 and C2 show how the estimates differ between the ACLR and ADL (C1), and LCV and ADL (C2). The y-axis displays the difference in estimates of β_x between the approaches in absolute value. The x-axis displays the the extent to which the corresponding restriction appeared violated based on the ADL estimates. Color is included to show the interaction between the extent of change and the estimated effect of β_{Ly}^{CF} (C1) and β_{Ly}^{LCV} (C2). Panels D1 and D2 show that the differences in BIC (y-axis) depend on the extent to which the corresponding restriction appeared violated.

practice of neglecting serial dependence with two common forms of dynamic specification, allowing for AR errors versus allowing for AR effects for the observed variables. We first elucidated the differences between the various approaches by considering a more general ADL model, where the restrictions implied under each approach can be scrutinized. The restrictions imposed under the LR and LCV modeling approaches are relatively straightforward, hence attention was paid to the nonlinear CF restriction imposed under the correction approach, and reference to the econometric literature was provided. The CF restriction provides a general condition by which a LR model with AR(1) errors can be written as an equivalent ADL model. It implies that the covariate only has a transient effect on the criterion variable, its influence does not carry forward or last in time.

As evidenced in our simulation studies and analytic example, when presented with serial dependence, neglecting this information (i.e., neglecting approach) and proceeding with inference based on the LR model parameter estimates can evidently be problematic. The estimates of the regression effects can be biased, sometimes to the point of implying nonsensical results. Furthermore, SE estimates can be biased, which can invalidate the results of hypothesis tests. These considerations become particularly important when covariates are autocorrelated. Importantly, misspecification bias also occurs when erroneously using the correcting and lagged criterion variable approaches. Furthermore, the size of the biases depends on the extent to which the restrictions are violated. Yet, the ADL was able to accurately recover the parameters in each case, since it relaxes the various restrictions. These considerations extend to multilevel models (i.e. DSEM and RDSEM), as demonstrated in appendix 3.

Regarding model selection, in large T situations (e.g. $T = 1000$), BIC and the likelihood-ratio tests performed well at selecting the true model from the competing approaches, whereas AIC often wrongly selected the more general ADL model. In smaller samples, (e.g. $T = 50$), the decision can become more difficult in some conditions. Particularly choosing between the ACLR and ADL models for data generated under a 'slight' violation of the CF restriction was troublesome. Likelihood-ratio tests suffered from larger type 2 errors in these conditions. AIC tended to favor the more general ADL model, both correctly and incorrectly, while BIC displayed a tendency to favor more restrictive alternatives. We conclude that for shorter time series AIC appears to be the more prudent index in the sense of avoiding misspecification whereas for longer time series BIC appears to be the better option.

In practice the number of available time points may be low, and uncertainty about these decisions high. In these cases, the researcher must contemplate the possibility of bias by constraining β_{Lx}, and the cost in terms of estimation precision due to estimating β_{Lx} freely (Hendry and Mizon, 1978). Nevertheless, we found that freely estimating lagged effects when these are present in the data can increase the precision of $\hat{\beta}_x$.

In our data analytic example, we showed that our findings are relevant for psychological science. Differences obtained between the approaches when applied to a data analytic problem displayed a direct correspondence with the extent to which the various restrictions appeared violated, and BIC favored the ADL model to the extent that the various restrictions appeared violated. Interestingly, we also found evidence that different model structures appeared appropriate for different individuals.

An at once exciting and unnerving conclusion is that whenever serial dependence is present, it should be included in the modeling strategy. Moreover, by simply 'correcting' for the dependence in the residuals, one is inadvertently implying a specific form of dynamics, which to the best of our knowledge never appears tested for in applied work, nor justified in the first place.

For applied researchers who have reason to believe that one of the models discussed in this manuscript is applicable, we therefore recommend that the analysis commences with an ADL model. From this ADL model, the appropriateness of imposing various specific restrictions corresponding to the more widely used and reported LCV, ACLR, or LR models can be tested empirically using model selection techniques.

The strategy of starting with fitting a general model and sequentially imposing restrictions and testing their adequacy is referred to as *general to specific modeling* in the econometric literature. This strategy is intended to safeguard applied researchers from the perils of misspecification (Sims, 1980; Campos et al., 2005). Our results suggest that the strategy should play an important role as well in the analysis of ILD in the behavioral sciences. An important disclaimer in this respect is that the ADL model we discussed in this manuscript needs not be the appropriate general model. Therefore, the restrictions it imposes can also be scrutinized in the context of yet more general models. For instance, one could include higher-order lags of the variables in the model to check whether the serial dependence is appropriately dealt with by conditioning on the previous time point.

8 Appendices

8.1 Appendix 1: Lasting effects of the covariate in the ADL and LCV models

In the manuscript we noted that the special case for which the ADL model implies only a transient effect of the covariate is when the CF restriction holds. To see that an ADL model in general, and also a LCV model, implies lasting effects of previous covariate scores on the criterion at time t, we present derivations analogue to the ones presented in the manuscript.

First, we express the general ADL model in terms of lag operators (cf. Hoover, 1988) as:

$$\begin{aligned} y_t &= \beta_{Ly} y_{t-1} + \beta_x x_t + \beta_{Lx} x_{t-1} + v_t \\ y_t - \beta_{Ly} y_{t-1} &= \beta_x x_t + \beta_{Lx} x_{t-1} + v_t \\ (1 - \beta_{Ly} L) y_t &= \beta_x (1 + \frac{\beta_{Lx}}{\beta_x} L) x_t + v_t. \end{aligned} \tag{9}$$

Now we can again divide by $(1 - \beta_{Ly} L)$:

$$\begin{aligned} (1 - \beta_{Ly} L) y_t &= \beta_x (1 + \frac{\beta_{Lx}}{\beta_x} L) x_t + v_t \\ (1 - \beta_{Ly} L) y_t &= (\beta_x + \beta_{Lx} L) x_t + v_t \\ (1 - \beta_{Ly} L)^{-1} (1 - \beta_{Ly} L) y_t &= (1 - \beta_{Ly} L)^{-1} (\beta_x + \beta_{Lx} L) x_t + (1 - \beta_{Ly} L)^{-1} v_t \qquad |\text{Divide by } (1 - \beta_{Ly} L) \\ y_t &= (\beta_x + \beta_{Lx} L) x_t + \beta_{Ly} (\beta_x + \beta_{Lx} L) x_{t-1} + \beta_{Ly}^2 (\beta_x + \beta_{Lx} L) x_{t-2} + ... \\ &\quad + v_t + \beta_{Ly} v_{t-1} + \beta_{Ly}^2 v_{t-2} + ... \\ &\qquad |(1 - \beta_{Ly} L)^{-1} = 1 + (\beta_{Ly} L) + (\beta_{Ly} L)^2 + ... \\ y_t &= \beta_x x_t + \beta_{Lx} x_{t-1} + \beta_{Ly} \beta_x x_{t-1} + \beta_{Ly} \beta_{Lx} x_{t-2} + \beta_{Ly}^2 \beta_x x_{t-2} + \beta_{Ly}^2 \beta_{Lx} x_{t-3} + ... \\ &\quad + v_t + \beta_{Ly} v_{t-1} + \beta_{Ly}^2 v_{t-2} + ... \\ y_t &= \beta_x x_t + \sum_{i=0}^{\infty} \beta_{Ly}^i (\beta_{Lx} + \beta_x \beta_{Ly}) x_{t-(i+1)} + \sum_{i=0}^{\infty} \beta_{Ly}^i v_{t-i}. \end{aligned}$$

.

In case of the LCV model, where $\beta_{Lx} = 0$, the obtained result simplifies to:

$$y_t = \beta_x x_t + \sum_{i=0}^{\infty} \beta_{Ly}^i (0 + \beta_x \beta_{Ly}) x_{t-(i+1)} + \sum_{i=0}^{\infty} \beta_{Ly}^i v_{t-i}$$
$$y_t = \beta_x x_t + \sum_{i=0}^{\infty} \beta_{Ly}^{i+1} \beta_x x_{t-(i+1)} + \sum_{i=0}^{\infty} \beta_{Ly}^i v_{t-i}$$
$$y_t = \sum_{i=0}^{\infty} \beta_{Ly}^i \beta_x x_{t-i} + \sum_{i=0}^{\infty} \beta_{Ly}^i v_{t-i}.$$

.

In both cases, the ADL and the LCV model, we see an accumulation of past effects of x_t on y_t because covariate influences carry over in time. This is in contrast to the CF constrained ADL (equation 8 in the manuscript), where the covariate only has a transient effect.

8.2 Appendix 2: Common factor restrictions in a VAR(1) model

In these appendices we investigate what a CF restriction implies if the covariate follows an AR(1) process, with AR effect γ, and covariate and criterion are represented jointly in terms of a VAR(1) process. Similarly to how we could observe the ACLR's CF restriction as a restriction on the parameters of an equivalent ADL, we now investigate the CF restriction on the ADL parameters as restrictions on the parameters of an equivalent VAR(1) model (see Mizon, 1995, McGuirk and Spanos, 2009).

In the manuscript we distinguished two cases depending on properties of the VAR(1) transition matrix. To make matters as digestible as possible, we discuss the two separately.

8.2.1 Appendix 2a: Common factors and equal AR effects in a diagonal transition matrix

Consider the following bivariate VAR(1) with a diagonal transition matrix where $\{|\beta_{Ly}|, |\gamma|\} < 1$:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \beta_{Ly} & 0 \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \zeta_t \end{bmatrix},$$
$$\begin{bmatrix} \epsilon_t \\ \zeta_t \end{bmatrix} \overset{iid}{\sim} MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon\zeta} \\ \sigma_{\epsilon\zeta} & \sigma_\zeta^2 \end{bmatrix} \right). \qquad (10)$$

Since the transition matrix is diagonal, the variables are 'mutually granger non-causal' (Spanos, 1987, Hamilton, 1994, p.303-304). There is a contemporaneous relationship, at each time point the error processes[5]

[5] The terminology 'innovation process' is usually used for error processes in time series contexts, we use the error terminology

covary with covariance $\sigma_{\epsilon\zeta}$. We specify a direction of this contemporaneous relationship by decomposing ϵ_t into a part of variability explained by ζ_t, and a part v_t which is unrelated to ζ_t:

$$\epsilon_t = \beta_x \zeta_t + v_t. \tag{11}$$

Here, $\beta_x = r_{\zeta\epsilon}\frac{\sigma_\epsilon}{\sigma_\zeta} = \frac{\sigma_{\epsilon\zeta}}{\sigma_\zeta^2}$. The variance of the remaining term v_t, σ_v^2, is restricted by this decomposition: $\sigma_v^2 = \sigma_\epsilon^2 - \beta_x^2 \sigma_\zeta^2$.[6]

By design, v_t is uncorrelated with ζ_t, so we may express the system as:

$$\begin{aligned} y_t &= \beta_{Ly} y_{t-1} + \beta_x \zeta_t + v_t, \\ x_t &= \gamma x_{t-1} + \zeta_t, \\ \begin{bmatrix} v_t \\ \zeta_t \end{bmatrix} &\overset{iid}{\sim} MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_\zeta^2 \end{bmatrix} \right). \end{aligned} \tag{13}$$

Since x_t, is generated by a stable AR(1) process with coefficient γ, it follows that $\zeta_t = (1 - \gamma L)x_t$. This implies that

$$\begin{aligned} y_t &= \beta_{Ly} y_{t-1} + \beta_x \zeta_t + v_t, \\ \zeta_t &= (1 - \gamma L)x_t, \\ \begin{bmatrix} v_t \\ \zeta_t \end{bmatrix} &\overset{iid}{\sim} MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_\zeta^2 \end{bmatrix} \right). \end{aligned} \tag{14}$$

We thus may express the equation for y_t as an ADL in lag operator notation:

$$(1 - \beta_{Ly} L) y_t = \beta_x (1 - \gamma L) x_t + v_t. \tag{15}$$

From this, it is evident that the system will common factor if $\beta_{Ly} = \gamma$, which we could define as β_{Ly}^*.

here due to its familiarity for researchers working with with linear regression models.

[6] In terms of error variances, we get:

$$\begin{aligned} \sigma_\epsilon^2 &= E(\epsilon_t^2) \\ \sigma_\epsilon^2 &= E((\beta_x \zeta_t + v_t)^2) \\ \sigma_\epsilon^2 &= E(\beta_x^2 \zeta_t^2 + 2\beta_x \zeta_t v_t + v_t^2) \\ \sigma_\epsilon^2 &= \beta_x^2 E(\zeta_t^2) + 2\beta_x E(\zeta_t v_t) + E(v_t^2) \\ \sigma_\epsilon^2 &= \beta_x^2 \sigma_\zeta^2 + \sigma_v^2. \end{aligned} \tag{12}$$

Hence, if the AR effects in the diagonal transition matrix are equal, there exists a common factor $(1-\beta^*_{Ly}L)$:

$$(1-\beta^*_{Ly}L)y_t = \beta_x(1-\beta^*_{Ly}L)x_t + v_t. \tag{16}$$

This implies that y_t and x_t have identical lag-one autocorrelations, which are in this case both equal to β^*_{Ly}: From Lütkepohl (2005), p.27, eq. 2.1.31, we know that the lag-one autocovariances implied by a VAR(1) model can be obtained via $B\,\Sigma_{xy}$, where B is the transition matrix and Σ_{xy} is the model-implied (lag-zero) covariance matrix. In our case, with the diagonal transition matrix, we get the following expressions for the lag-one autocovariances of y_t and x_t, σ_{yLy} and σ_{xLx}:

$$\begin{aligned}\sigma_{yLy} &= \beta^*_{Ly}\sigma^2_y,\\ \sigma_{xLx} &= \beta^*_{Ly}\sigma^2_x.\end{aligned} \tag{17}$$

Dividing by the respective variances σ^2_y and σ^2_x yields the model-implied autocorrelations ρ_{yLy} and ρ_{xLx}, which equal β^*_{Ly} in both cases.

8.2.2 Appendix 2b: Common factors and unequal AR effects in a triangular transition matrix

Consider the unrestricted ADL, where we specify an AR(1) model for the serial dependence in the covariate:

$$\begin{aligned} y_t &= \beta_x x_t + \beta_{Ly} y_{t-1} + \beta_{Lx} x_{t-1} + v_t,\\ x_t &= \gamma x_{t-1} + \zeta_t,\\ \begin{bmatrix} v_t \\ \zeta_t \end{bmatrix} &\overset{iid}{\sim} MVN\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma^2_v & 0\\ 0 & \sigma^2_\zeta\end{bmatrix}\right). \end{aligned} \tag{18}$$

We will write this system in reduced form (i.e. as a VAR, containing lagged predictors and covarying errors), and use the lag operator consistently:

$$
\begin{aligned}
y_t &= \beta_{Ly} L y_t + \beta_x x_t + \beta_{Lx} L x_t + v_t && |\text{ADL definition for } y_t \\
y_t &= \beta_{Ly} L y_t + \beta_x (\gamma L x_t + \zeta_t) + \beta_{Lx} L x_t + v_t && |x_t = \gamma L x_t + \zeta_t \\
y_t &= \beta_{Ly} L y_t + \beta_x \gamma L x_t + \beta_{Lx} L x_t + \beta_x \zeta_t + v_t && |\text{Distribute and rearrange} \\
y_t &= \beta_{Ly} L y_t + \beta_x \gamma L x_t + \beta_{Lx} L x_t + \epsilon_t && |\text{Define } \epsilon_t = \beta_x \zeta_t + v_t \\
y_t &= \beta_{Ly} L y_t + (\beta_x \gamma + \beta_{Lx}) L x_t + \epsilon_t. && |\text{Factorize}
\end{aligned}
$$

As in the previous section, defining $\epsilon_t = \beta_x \zeta_t + v_t$ implies that ϵ_t and ζ_t covary with covariance $\sigma_{\epsilon\zeta} = \beta_x \sigma_\zeta^2$, furthermore, $\sigma_\epsilon^2 = \sigma_v^2 + \beta_x^2 \sigma_\zeta^2$.

Now we impose a CF restriction: $\beta_{Lx} = -\beta_{Ly}\beta_x$:

$$
\begin{aligned}
y_t &= \beta_{Ly} L y_t + (\beta_x \gamma + \beta_{Lx}) L x_t + \epsilon_t && |\text{Previous equation} \\
y_t &= \beta_{Ly} L y_t + (\beta_x \gamma - \beta_x \beta_{Ly}) L x_t + \epsilon_t && |\beta_{Lx} = -\beta_x \beta_{Ly} \\
y_t &= \beta_{Ly} L y_t + \beta_x (\gamma - \beta_{Ly}) L x_t + \epsilon_t. && |\text{Factorize}
\end{aligned}
$$

And as such the system is:

$$
\begin{aligned}
y_t &= \beta_{Ly} L y_t + \beta_x (\gamma - \beta_{Ly}) L x_t + \epsilon_t, \\
x_t &= \gamma L x_t + \zeta_t, \\
\begin{bmatrix} \epsilon_t \\ \zeta_t \end{bmatrix} &\overset{iid}{\sim} MVN \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon\zeta} \\ \sigma_{\epsilon\zeta} & \sigma_\zeta^2 \end{bmatrix} \right).
\end{aligned} \tag{19}
$$

.

In typical VAR notation, and knowing that $\sigma_{\epsilon\zeta} = \beta_x \sigma_\zeta^2$ and $\sigma_\epsilon^2 = \beta_x^2 \sigma_\zeta^2 + \sigma_v^2$, we are left with the result found in McGuirk and Spanos, 2009 (equation 18), but for a single covariate:

$$
\begin{aligned}
\begin{bmatrix} y_t \\ x_t \end{bmatrix} &= \begin{bmatrix} \beta_{Ly} & \beta_x(\gamma - \beta_{Ly}) \\ 0 & \gamma \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \zeta_t \end{bmatrix}, \\
\begin{bmatrix} \epsilon_t \\ \zeta_t \end{bmatrix} &\overset{iid}{\sim} MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon\zeta} \\ \sigma_{\epsilon\zeta} & \sigma_\zeta^2 \end{bmatrix} \right).
\end{aligned} \tag{20}
$$

We can see that the CF restriction implies a crossregressive effect of x_{t-1} on y_t, $\beta_x(\gamma - \beta_{Ly})$, that is proportional to the difference in AR effects resulting in an upper triangular transition matrix. In the case of equal AR effects, said crossregressive effect becomes 0 and the transition matrix diagonal. This special case thus corresponds to the previously discussed VAR(1) system.

The lag-one autocorrelation of y_t implied by the model can be derived as (see again Lütkepohl, 2005, p.27, eq. 2.1.31):

$$
\begin{aligned}
\rho_{yLy} &= \frac{\sigma_{yLy}}{\sigma_y^2} \\
\rho_{yLy} &= \frac{\beta_{Ly}\sigma_y^2 + \beta_x(\gamma - \beta_{Ly})\sigma_{xy}}{\sigma_y^2} \\
\rho_{yLy} &= \beta_{Ly} + \beta_x(\gamma - \beta_{Ly})\frac{\sigma_{xy}}{\sigma_y^2}.
\end{aligned} \tag{21}
$$

The implied lag-one autocorrelation of x_t is:

$$
\begin{aligned}
\rho_{xLx} &= \frac{\sigma_{xLx}}{\sigma_x^2} \\
\rho_{xLx} &= \frac{\gamma\sigma_x^2}{\sigma_x^2} \\
\rho_{xLx} &= \gamma.
\end{aligned} \tag{22}
$$

Also at the level of the implied lag-one autocorrelations, we see that if $\gamma \neq \beta_{Ly}$, then the autocorrelation of y_t will differ from the one of x_t as a function of the difference between the AR effects.

8.3 Appendix 3: Multilevel (Hierarchical) models

8.3.1 Multilevel CF restrictions

ILD often consist of $N = many$ individual time series from different participants. In such designs, the regression models are often formulated as multilevel models and modeling the level 1 (i.e., within-subjects-level) residuals as AR(1) processes has also been considered (Asparouhov and Muthén, 2020). Consider as a simple example a multilevel ACLR model, where parameters β_x^{CF} and β_{Lu}^{CF} are fixed across individuals ($i \in \mathbb{N}$ indexes the individual):

$$\begin{aligned} y_{it} &= \beta_x^{CF} x_{it} + u_{it}, \\ u_{it} &= \beta_{Lu}^{CF} u_{i\,t-1} + v_{it}, \\ v_{it} &\overset{iid}{\sim} N(0, \sigma_v^2). \end{aligned} \tag{23}$$

It is evident that the same substitution procedure employed for the errors u_t in the $N = 1$ case can be employed for the within-subjects-level errors of the above multilevel model, u_{it}. Doing so, one finds that the multilevel ACLR model corresponds to a multilevel ADL model constrained by the CF restriction $-\beta_{Lu}^{CF}\beta_x^{CF}$:

$$\begin{aligned} y_{it} = \beta_x^{CF} x_{it} + \beta_{Lu}^{CF} y_{i\,t-1} - \beta_{Lu}^{CF}\beta_x^{CF} x_{i\,t-1} + v_{it}, \\ v_{it} \overset{iid}{\sim} N(0, \sigma_v^2). \end{aligned} \tag{24}$$

If β_{Lu}^{CF} and β_x^{CF} were allowed to vary across individuals by specifying them as random effects, then also the CF restriction would turn into a random effect. The take away message is that multilevel specifications with AR(1) residuals impose the analogue kinds of restrictions on the model parameters as those discussed for $N = 1$ models. These restrictions will therefore again lead to misspecification bias.

8.3.2 Reanalysis of Asparouhov and Muthén (2020)

Asparouhov and Muthén (2020) conducted a simulation study, pitting a multilevel LCV model (DSEM in their terminology) against a multilevel ACLR model (RDSEM in their terminology), finding biases in the estimates of the regression effects when the models were fit to data generated by the alternative. Specifically, they conducted simulations with the following parameter settings for the DSEM and RDSEM specification: $\beta_x = 1$, $\beta_{x_b} = -1$ (a between person-level effect of the covariate), β_{Ly} (DSEM) or β_{Lu} (RDSEM) $= 0.7$. The covariate $x_{i\,t}$ was generated according to an AR(1) model, with an AR effect of 0.7. The researchers reported biases for $\hat{\beta}_x$, $\hat{\beta}_{x_b}$, and $\hat{\beta}_{Ly}$ or $\hat{\beta}_{Lu}$ when the models were fit to data generated under the alternative.

It follows from the exposition we provide in this manuscript that fitting a multilevel ADL model to this data will lead to an accurate recovery of the parameters from data generated under either the DSEM or RDSEM specification. Specifically, data generated by the RDSEM model will, when estimated with the multilevel ADL model, result in a within-subjects effect $\beta_{Lx}^{ADL} = -\beta_x^{ADL}\beta_{Ly}^{ADL}$. Data generated under the DSEM model will simply result in the ADL recovering $\beta_{Lx}^{ADL} = 0$. For this analysis, we employed Bayesian estimation using Mplus as in Asparouhov and Muthén (2020). The only modification we made to their code was to include a lagged effect for the covariate in the DSEM estimation code.

8.3.3 Results

Simulation models	RDSEM		DSEM		True value
$\hat{\beta}_{Ly}$	.69	[-.01]	.70	[.00]	0.7
$\hat{\gamma}$	.70	[.00]	.70	[.00]	0.7
$\hat{\beta}_x$	1.00	[.00]	1.00	[.00]	1
$\hat{\beta}_{Lx}$	-.69	[-.01]	.00	[.00]	-0.7 or 0
$\hat{\beta}_b$	-1.01	[.01]	-.96	[-.04]	-1

Table 3: Results of the reanalysis of the simulation data reported in Asparouhov and Muthén (2020). We report the average ADL parameter estimates over replications, and bias in square brackets.

Table 3 shows that, as expected, the parameters are estimated accurately by the multilevel ADL model when data are generated under either the DSEM or RDSEM specification. Importantly, we obtain for the RDSEM condition an estimate of the effect of the lagged within-subjects covariate $\beta_{Lx}^{ADL} = -\beta_x^{ADL}\beta_{Ly}^{ADL}$, and for the DSEM condition we recover $\beta_{Lx}^{ADL} = 0$.

References

Aitken, A. C. (1936). Iv.—on least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, *55*, 42–48.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Ariens, S., Adolf, J. K., & Ceulemans, E. (2022). Collinearity issues in autoregressive models with time-varying serially dependent covariates. *Multivariate Behavioral Research*, 1–19.

Ariens, S., Ceulemans, E., & Adolf, J. K. (2020). Time series analysis of intensive longitudinal data in psychosomatic research: A methodological overview. *Journal of Psychosomatic Research*, *137*, 110191.

Asparouhov, T., & Muthén, B. (2020). Comparison of models for the analysis of intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(2), 275–297.

Asparouhov, T., & Muthén, B. (2021). Residual structural equation models.

Burridge, P. (1981). Testing for a common factor in a spatial autoregression model. *Environment and Planning A*, *13*(7), 795–800.

Campos, J., Ericsson, N. R., & Hendry, D. F. (2005). General-to-specific modeling: An overview and selected bibliography. *FRB International Finance Discussion Paper*, (838).

Cochrane, D., & Orcutt, G. H. (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American statistical association*, *44*(245), 32–61.

Dixon-Gordon, K. L., & Laws, H. (2021). Emotional variability and inertia in daily life: Links to borderline personality and depressive symptoms. *Journal of Personality Disorders*, *35*(Supplement A), 162–171.

Gard, D. E., Sanchez, A. H., Cooper, K., Fisher, M., Garrett, C., & Vinogradov, S. (2014). Do people with schizophrenia have difficulty anticipating pleasure, engaging in effortful behavior, or both? *Journal of abnormal psychology*, *123*(4), 771.

Garijo, A. M. A., & Lacambra, J. M. (2011). The likelihood ratio test of common factors under non-ideal conditions. *Investigaciones Regionales= Journal of Regional Research*, (21), 37–52.

Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the cogito study. *Multivariate Behavioral Research*, *53*(6), 820–841.

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton University Press.

Hendry, D. F., & Juselius, K. (2000). Explaining cointegration analysis: Part 1. *The Energy Journal*, *21*(1).

Hendry, D. F., & Mizon, G. E. (1978). Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the bank of england. *The Economic Journal*, *88*(351), 549–563.

Hendry, D. F., Pagan, A. R., & Sargan, J. D. (1984). Dynamic specification. *Handbook of econometrics*, *2*, 1023–1100.

Hoover, K. D. (1988). On the Pitfalls of Untested Common-Factor Restrictions: The Case of the Inverted Fisher Hypothesis. *Oxford Bulletin of Economics and Statistics*, *50*(2), 125–138. https://ideas.repec.org/a/bla/obuest/v50y1988i2p125-38.html

Johnson, A. R., Bank, S. R., Summers, M., Hyett, M. P., Erceg-Hurn, D. M., Kyron, M. J., & McEvoy, P. M. (2020). A longitudinal assessment of the bivalent fear of evaluation model with social interaction anxiety in social anxiety disorder. *Depression and Anxiety*, *37*(12), 1253–1260.

Krone, T., Albers, C. J., & Timmerman, M. E. (2017). A comparative simulation study of ar (1) estimators in short time series. *Quality & quantity*, *51*(1), 1–21.

Kung, Y.-T., Chi, S.-C., Chen, Y.-C., & Chang, C.-M. (2021). Using residual dynamic structural equation modeling to explore the relationships among employees' self-reported health, daily positive mood, and daily emotional exhaustion. *Healthcare*, *9*(1), 93.

Lafit, G., Adolf, J. K., Dejonckheere, E., Myin-Germeys, I., Viechtbauer, W., & Ceulemans, E. (2021). Selection of the number of participants in intensive longitudinal studies: A user-friendly shiny app and tutorial for performing power analysis in multilevel regression models that account for temporal dependencies. *Advances in methods and practices in psychological science*, *4*(1), 2515245920978738.

Larsen, J. T., Norris, C. J., McGraw, A. P., Hawkley, L. C., & Cacioppo, J. T. (2009). The evaluative space grid: A single-item measure of positivity and negativity. *Cognition and Emotion*, *23*(3), 453–480.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Maeshiro, A. (2000). An illustration of the bias of ols for yt= λ y t-1+ u t. *The Journal of Economic Education*, *31*(1), 76–80.

Mak, H. W., & Schneider, S. (2020). Individual differences in momentary pain-affect coupling and their associations with mental health in patients with chronic pain. *Journal of Psychosomatic Research*, *138*, 110227.

McGuirk, A., & Spanos, A. (2009). Revisiting error-autocorrelation correction: Common factor restrictions and granger non-causality. *Oxford Bulletin of Economics and Statistics*, *71*(2), 273–294.

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in mplus. *Psychological methods*, *25*(5), 610.

Michaelides, M., & Spanos, A. (2020). On modeling heterogeneity in linear models using trend polynomials. *Economic Modelling*, *85*, 74–86.

Mizon, G. E. (1995). A simple message for autocorrelation correctors: Don't. *Journal of Econometrics*, *69*(1), 267–288.

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201–218.

Mur, J., & Angulo, A. (2006). The spatial durbin model and the common factor tests. *Spatial Economic Analysis*, *1*(2), 207–226.

Muthén, L., & Muthén, B. (2017). Mplus user's guide. eighth edition.

Myin-Germeys, I., van Os, J., Schwartz, J. E., Stone, A. A., & Delespaul, P. A. (2001). Emotional reactivity to daily life stress in psychosis. *Archives of general psychiatry*, *58*(12), 1137–1144.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Ravindran, N., Hu, Y., McElwain, N. L., & Telzer, E. H. (2020). Dynamics of mother–adolescent and father–adolescent autonomy and control during a conflict discussion task. *Journal of Family Psychology*, *34*(3), 312.

Sargan, J. D. (1980). Some tests of dynamic specification for a single equation. *Econometrica: Journal of the econometric society*, 879–897.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461–464.

Silk, J. S., Steinberg, L., & Morris, A. S. (2003). Adolescents' emotion regulation in daily life: Links to depressive symptoms and problem behavior. *Child development*, *74*(6), 1869–1880.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48.

Spanos, A. (1987). Error autocorrelation revisited: The ar (1) case. *Econometric Reviews*, *6*(2), 285–294.

Stevenson, B. L., Kummerfeld, E., Merrill, J. E., Blevins, C., Abrantes, A. M., Kushner, M. G., & Lim, K. O. (2022). Quantifying heterogeneity in mood-alcohol relationships with idiographic causal models. *Alcoholism: Clinical and Experimental Research*.

Takano, K., Sakamoto, S., & Tanno, Y. (2014). Repetitive thought impairs sleep quality: An experience sampling study. *Behavior Therapy*, *45*(1), 67–82.

Talata, Z. (2005). Model selection via information criteria. *Periodica Mathematica Hungarica*, *51*(1), 99–117.

Vanhasbroeck, N., Loossens, T., Anarat, N., Ariens, S., Vanpaemel, W., Moors, A., & Tuerlinckx, F. (2022). Stimulus-driven affective change: Evaluating computational models of affect dynamics in conjunction with input. *Affective Science*, 1–18.

Whelen, M. L., & Strunk, D. R. (2021). Does cognitive behavioral therapy for depression target positive affect? examining affect and cognitive change session-to-session. *Journal of consulting and clinical psychology*, *89*(9), 742.

Wooldridge, J. M. (2012). *Introductory econometrics: A modern approach*. Cengage learning.