

How to Evaluate Causal Dominance Hypotheses in Lagged Effects Models

Chuenjai Sukpan and Rebecca M. Kuiper

Utrecht University

ABSTRACT

The (Random Intercept) Cross-Lagged Panel Model ((RI-)CLPM) is increasingly used in psychology and related fields to assess the longitudinal relationship of two or more variables on each other. Researchers are interested in the question which of the lagged effects is causally dominant receives considerable attention. However, currently used methods do not allow for the evaluation of causal dominance hypotheses. This paper will show the performance of the Generalized Order-Restricted Information Criterion Approximation (GORICA), an extension of Akaike's Information Criterion (AIC), in the context of causal dominance hypotheses using a simulation study. The GORICA proves to be an adequate method to evaluate causal dominance in lagged effects models.

KEYWORDS

Causal dominance;
informative hypotheses;
model selection; order
restrictions

1. Introduction

Structural equation modeling (SEM) has become an increasingly popular modeling technique in the social and behavioral sciences, because of its ability to analyze structural relationships (Bollen & Noble, 2011). Loosely speaking, SEM combines factor analysis and multiple regression into one multivariate statistical technique. Specifically, this paper will be tailored towards two related SEM models: the cross-lagged panel model (CLPM) and the random intercept cross-lagged panel model (RI-CLPM). The CLPM can be used to analyze panel data in which two or more variables are repeatedly measured at multiple time points. Hamaker et al. (2015) introduced the RI-CLPM as an extension of the traditional CLPM, to facilitate the separation of within-person variation from between-person variation. Both the CLPM and the RI-CLPM are utilized to unravel processes that develop over time.

The exploration of causal dominance is gaining attention across various fields. Next, we discuss three studies that indicate the growing interest among researchers in understanding causal dominance relationships. Schuurman et al. (2016) conducted a study using a multilevel multivariate autoregressive modeling approach to investigate cross-lagged associations between variables over time. The study underscores the significance of standardizing coefficients within individuals for valid comparisons and cautions against drawing misleading conclusions by disregarding individual differences. An empirical example focusing on competence and exhaustion in individuals diagnosed with burnout is presented. Overall, the study highlights the benefits of

employing a multilevel multivariate autoregressive modeling approach and emphasizes the importance of standardizing coefficients within individuals to accurately assess the strength of cross-lagged associations.

Another study, namely Kuiper (2021b) demonstrated the evaluation of causal dominance hypotheses using the Generalized Order-Restricted Information Criterion Approximation (GORICA; Altınışık, 2018) with CT meta-analyzed lagged regression estimates and illustrated its performance. However, the evaluation of the performance of the GORICA was limited to a small simulation study.

Scott (2021) investigated relationships among various cross-lagged models, including Autoregressive Latent Trajectory (ALT), General Cross-lagged Model (GCLM), Latent Growth Curve Model with Structured Residuals and Unspecified Growth Trajectory (LGCM-SR-UGT), Cross-lagged Panel Model (CLPM), Random Intercept Cross-lagged Panel Model (RI-CLPM), and Mean Stationary GCLM. Through simulations, the study addressed two key aspects: (1) evaluating fit indices to determine "good" fit and utilizing Bayes Factor to select the best from the set of various cross-lagged models, and (2) examining the impact of disregarding variability in trajectories on cross-lagged estimates. The findings of the study from Scott (2021) revealed that RMSEA identified models with "good" fit, and the Bayes Factor favored cross-lagged models closely aligned with the true model over less related models. Furthermore, various patterns of bias in path estimates and standard errors were observed. The presence of causal dominance, in combination with time-variant between-person variance and covariance, exerted a significant influence on the bias in

path estimates. Thus, it is important to select the right type of cross-lagged model. Our study primarily focuses on the concept of evaluating causal dominance hypotheses using GORICA in the context of RI-CLPM and CLPM.

When researchers selected the right cross-lagged model and standardized its estimates correctly, they can evaluate their expectations based on past experiences and theoretical considerations. To evaluate these expectations statistically, researchers can translate their ideas into what we will refer to as “informative hypotheses” (Hojtink, 2011). Informative hypotheses contain inequality constraints among the parameters of interest, that allow researcher to statistically evaluate the relative size of specific parameters. In the (RI-)CLPM-framework, the interest is often in assessing “causal dominance”, that is, assessing which of the cross-lagged effects between two variables is the strongest. Specifically, in the situation in which two variables mutually influence each other at a subsequent time point, the variable that exerts the strongest influence on the other or, better, has the most predictive strength is said to be causally dominant.

Consider the study by Masselink et al. (2018), who were interested in the effect between self-esteem and depressive symptoms. Although the existence of the effect between these variables was already known, they focused on the evaluation of causal dominance between the two. Their hypothesis of interest was that the cross-lagged effect of self-esteem on depressive symptoms, that is, the effect of self-esteem on time point $t - 1$ on depressive symptoms on time point t (γ_t) is stronger in absolute sense than the cross-lagged effect of depressive symptoms on self-esteem, that is, the effect of depressive symptoms on time point $t - 1$ on self-esteem on time point t (β_t), which can be formalized as $H_1 : |\beta_t| < |\gamma_t|$, where $|\cdot|$ denotes the absolute value and the parameters with subscript t are the standardized parameters between time point $t - 1$ and time point t . The absolute values of the cross-lagged effects ensures that the ordering of the absolute strengths of effects are compared. Ultimately, Masselink and colleagues expected that the cross-lagged effect of self-esteem on depressive symptoms tends to be stronger in absolute sense than the cross-lagged effect of depressive symptoms on self-esteem using the SEM model. Masselink et al. (2018) studied the effects between self-esteem and depressive symptoms in a RI-CLPM using Chi-square test and then inspected the estimates themselves. Even though their findings (w.r.t. the test and size of the estimates) were in favor of the hypothesis $|\beta_t| < |\gamma_t|$, their result does not quantify its support.

Altınışık (2018) introduced the GORICA, a methodology applicable not only to normal linear models but also to generalized linear models (GLMs), generalized linear mixed models (GLMMs), and structural equation models (SEMs), including (RI-)CLPMs. The (RI-)CLPM has gained popularity in the literature, as researchers from various fields have shown a keen interest in exploring the concept of causal dominance between variables. Hence, there is need for investigating hypotheses related to causal dominance. To address this gap, we are delighted to present the performance of the GORICA which enables the evaluation of causal

dominance hypotheses in (RI-)CLPMs. In this paper, we will investigate the performance of the GORICA to (RI-)CLPMs using a simulation study, and apply the method to the data from Masselink et al. (2018), who studied the longitudinal association between self-esteem and depressive symptoms.

The article is organized as follows. First, the CLPM and the RI-CLPM are explained in further depth. Second, informative hypotheses and causal dominance will be further illustrated using hands-on examples from the field of psychology. Third, we will consider model selection as it is often encountered in SEM, namely, using the AIC. We will add to this by elaborating on various extensions of the AIC: namely, the order-restricted information criterion (ORIC), the generalized order-restricted information criterion (GORIC), and the generalized order-restricted information criterion approximation (GORICA). Fourth, we will investigate the performance of the GORICA in simulations containing a bivariate and a tri-variate (RI-)CLPM, and evaluate its performance by the true hypothesis rate (i.e., the percentage of all simulations in which the true hypothesis is selected). Fifth, we apply the GORICA on the data obtained by Masselink et al. (2018). We will end with a discussion of the findings of the present study. Additionally, the [supplementary material](#) ([Supplementary material](#) for Sukpan & Kuiper) contains relevant and detailed information that is referenced in this paper. Furthermore, for those interested in evaluating causal dominance hypotheses on RI-CLPM, the R code is available from <https://github.com/Chuenjai/Causal-dominance>.

2. Lagged Effects Models

2.1. The Cross-Lagged Panel Model

The Cross-Lagged Panel Model (CLPM) is one of the most popular approaches to analyze effects between variables over time (Biesanz, 2012). CLPM, also referred to as a cross-lagged path model, is frequently used to model panel data or longitudinal data in social and behavioral data. Panel data consist of a relatively small number of repeated measurements from the same cases (e.g., individuals, dyads, or families). It can be used if two or more variables have been measured at two or more occasions. The models are considered “lagged” because they estimate effects between variables across different time points. They are considered “crossed” because they also estimate effects from one variable to another over time and vice-versa (Allen, 2017; Hamaker et al., 2015; Usami et al., 2019). To facilitate the explanation of the model, we will use the study of Masselink et al. (2018) as an example, in which the effects between self-esteem and depressive symptoms in young adolescents was investigated. The data of Masselink et al. (2018) consist of 2105 patients who were recruited from seven secondary schools (age range: 10–14 years) in the Netherlands at three time points (three waves) one year apart (namely 2013, 2014, and 2015). Masselink et al. (2018) applied a bivariate RI-CLPM (with three time points) to their data, as depicted in [Figure 1](#). Here, α and δ denote the autoregressive effect

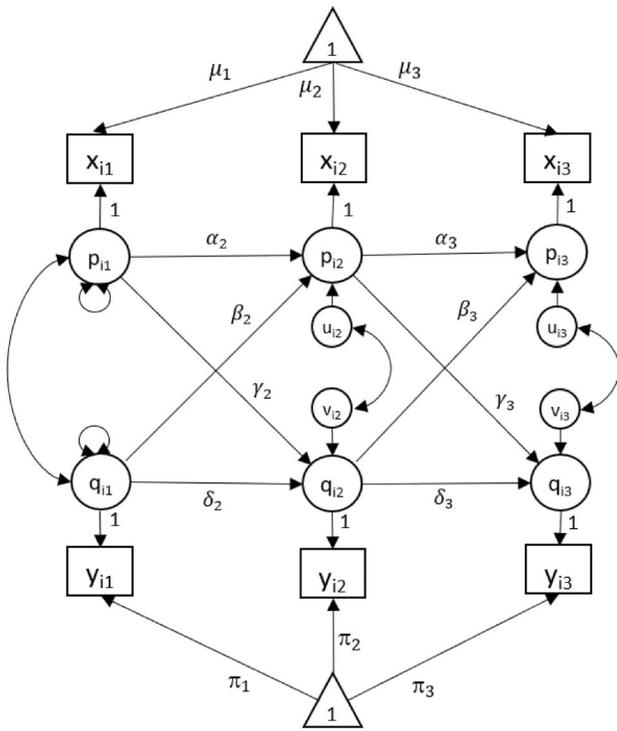


Figure 1. The bivariate CLPM for three waves and two observed variables; squares denote the observed variables, circles denote latent variables, triangles represent constants for the mean structure. The latent variables p and q are the grand-mean centered observed variables. Single-headed arrows indicate regression and double-headed arrows indicate correlations or (co)variances. The figure is based on Hamaker et al. (2015).

(same variable, different time points) and γ and β denote the cross-lagged ones (different variables and time points).

Let us consider a bivariate CLPM, where one with three time points is depicted in Figure 1. Let x_{it} and y_{it} denote the measurements for time-point t ($t = 1, \dots, T$) for individual i ($i = 1, \dots, N$). In a lagged effects model, the variables x_{it} and y_{it} are preferably centered to improve the interpretation of the mean/intercept parameters (Hamaker et al., 2015). In a CLPM, one can grand-mean center the variables:

$$x_{it} = \mu_t + p_{it} \tag{1}$$

$$y_{it} = \pi_t + q_{it} \tag{2}$$

For time-point t and individual i . Here, μ_t and π_t are the grand means at time point t for the two variables. The latent variables p_{it} and q_{it} , are the individual temporal deviation terms from these grand means. Notably, Equations (1) and (2) are not a true measurement model, because no measurement errors are specified. When no constraints are imposed on these means, they are allowed to change over time. Such changes may result from underlying developmental trajectories. However, this is typically not the focus in CLPM and thus the CLPM implicitly removes such group-level longitudinal trajectories from the cross-lagged relations (Usami et al., 2019).

When the data are centered, the structural equations for an individuals' temporal deviations p_{it} and q_{it} are modeled as exogenous variables for $t = 1$ (i.e., variances and covariance of variables are separately estimated), whereas, for

$t \geq 2$, p_{it} and q_{it} are modeled as

$$p_{it} = \alpha_t p_{i,t-1} + \beta_t q_{i,t-1} + u_{it} \tag{3}$$

$$q_{it} = \gamma_t p_{i,t-1} + \delta_t q_{i,t-1} + v_{it} \tag{4}$$

where α_t and δ_t are autoregressive parameters and β_t and γ_t are cross-lagged parameters. Note that these parameters have a subscript t , implying that they may change over time (Usami et al., 2019), that is, they may differ across waves, as can be seen in Figure 1.

The CLPM had a large impact on the analyses of psychological data in social and behavioral sciences. However, several authors have pointed out that the CLPM makes a number of assumptions that might not be met in applied contexts or that researchers might not be aware of when using the model (Allison, 2009; Berry & Willoughby, 2017; Hamaker et al., 2015; Mund & Nestler, 2019). First, the CLPM generally lacks explicit theories of change, it assumes that individuals fluctuate around a common grand mean in each of the involved variables over time. Hence, the CLPM assumes that there are no stable between-person differences in these variables. Stated otherwise, the CLPM does not consider that the average level of a variable across time is higher for some individuals than for others. As a consequence, if stable between-person differences in x or y are present, they are included in the estimated autoregressive and cross-lagged paths (Berry & Willoughby, 2017; Hamaker et al., 2015; Mund & Nestler, 2019). Second, although cross-lagged panel analysis can also be conducted using structural equation modeling (SEM), many cross-lagged panel models assume that variables are measured without error. For many variables in psychological research, this is clearly not the case, which leads to biased results. Some scholars have also argued that measurement error may also be misidentified as real change when models have only two time points. In these cases, measurement error could still influence the results of SEM (Allen, 2017). Third, the time interval can be of any length, though it must be contextually appropriate to have meaningful interpretations. If the interval is too short, measurement will occur before the effects can be observed. If the interval is too long, the effects will dissipate before the next measurement (Allen, 2017; Mund & Nestler, 2019).

As a result, the lagged parameters that are obtained with the CLPM do not represent the actual within-person effects over time, and this may lead to erroneous conclusions regarding the presence, dominance, and sign of predictive effects (Hamaker et al., 2015). The RI-CLPM can help to mitigate the first issue and is explained next.

2.2. The Random Intercept Cross-Lagged Panel Model

The RI-CLPM (Hamaker et al., 2015) is an extension of the CLPM that includes latent variables, which are referred to as random intercepts. A bivariate RI-CLPM (with three time-points) is displayed in Figure 2. A RI-CLPM allows to estimate the within-person autoregressive and cross-lagged effects. The basic idea behind the RI-CLPM is that, instead of all individuals varying around a common mean over

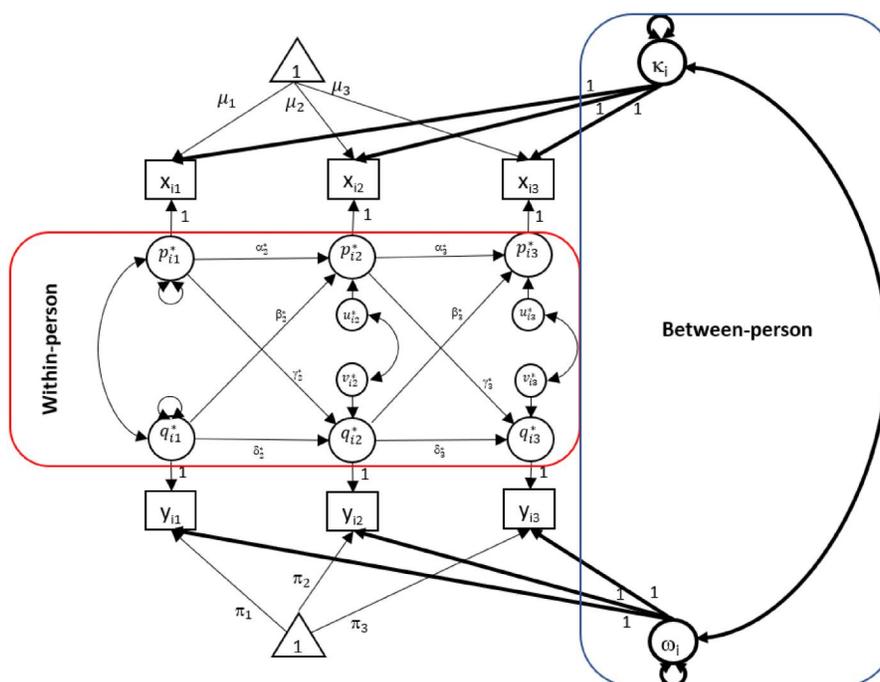


Figure 2. The bivariate RI-CLPM with the separation of within-person and between-person variation for three waves and two observed variables; squares denote the observed variables, circles denote latent variables, triangles represent constants for the mean structure. The variables p^* and q^* are the person-mean centered observed variables, and κ and ω are random intercepts. Single-headed arrows indicate regression and double-headed arrows indicate correlations or (co)variances. This figure is based on Hamaker et al. (2015).

time, each individual fluctuates around their own rather stable, person-specific mean level over time with respect to the constructs under investigation. This assumption is implemented by defining a latent intercept for each construct across all time points. The intercept for, for instance, effect satisfaction allows some individuals to be generally more satisfied with their effects than other individuals. Due to the separation of within-person and between-person variation, the autoregressive effects provide information on the within-person stability in x or y , instead of containing information on the (between-person) rank-order stability of these variables, where rank-order stability refers to the level to which the ordering of individuals on a construct remained unchanged over time. Similarly, the cross-lagged effects pertain to within-person associations. In a bivariate RI-CLPM, x_{it} and y_{it} are modeled as

$$x_{it} = \mu_t + \kappa_i + p_{it}^* \quad (5)$$

$$y_{it} = \pi_t + \omega_i + q_{it}^* \quad (6)$$

where μ_t and π_t are the temporal means at time point t ; κ_i and ω_i are random intercepts for individual i ; and p_{it}^* and q_{it}^* are residuals at time point t for individual i . The random intercepts κ_i and ω_i are time-invariant, stable trait factors that represent individual trait-like deviations from temporal means: They can be thought of as latent variables or factors whose factor loadings are all constrained to 1. The residuals p_{it}^* and q_{it}^* are temporal deviations from the expected scores of individual i at measurement t (i.e., temporal deviations from $\mu_t + \kappa_i$ and $\pi_t + \omega_i$). Consequently, p_{it}^* and q_{it}^* differ from p_{it} and q_{it} in Equations (1) and (2), which represent deviations from the group means. The RI-CLPM reduces to the CLPM if we set the variances of κ_i and ω_i to zero,

meaning there are no stable, trait-like between-person differences (Figure 2). Hence, the CLPM is nested within the RI-CLPM. The RI-CLPM is identified if two or more variables have been measured at three or more time points. Furthermore, the estimation of the RI-CLPM presupposes that the variables were measured without error, that is, the error variances of the observed indicators are constrained to be zero (Usami et al., 2019).

The structural equations are modeled as

$$p_{it}^* = \alpha_t^* p_{i,t-1}^* + \beta_t^* q_{i,t-1}^* + u_{it}^* \quad (7)$$

$$q_{it}^* = \gamma_t^* p_{i,t-1}^* + \delta_t^* q_{i,t-1}^* + v_{it}^* \quad (8)$$

Notably, these equations resemble Equations (3) and (4) from the CLPM, respectively. Nevertheless, the autoregressive and cross-lagged regression parameters differ. The autoregressive parameters α_t^* and δ_t^* do not represent the stability of the rank order of individuals from one occasion to the next, but rather the amount of within-person carry-over effect (Hamaker, 2012; Hamaker et al., 2015; Kuppens et al., 2010; Suls et al., 1998). The cross-lagged parameters β_t^* and γ_t^* refer to within-person effects due to the lagged relations pertain exclusively to within-person fluctuations, and thus should be interpreted as the extent to which a variable affects another variable at the following time point and vice versa. As an example, let us take the study of Masselink et al. (2018). Masselink et al. (2018) investigate the longitudinal association between self-esteem and depressive symptoms. In Figure 2, the autoregressive parameters α_t^* represents the within-person effect between previous self-esteem on current self-esteem and δ_t^* represents the within-person effect between previous depressive symptoms on current depressive symptoms. The cross-lagged parameters γ_t^*

denote the within-person cross-lagged effect between current depressive symptoms and previous self-esteem and β_t^* is the within-person cross-lagged effect between current self-esteem and previous depressive symptoms. For readability, in the remainder of the text, we will leave out the superscript *. From the text, it will be clear when we refer to RI-CLPM parameters.

In summary, RI-CLPM accounts for trait-like, time-invariant stability that separates the within-person process from stable between-person differences through the inclusion of random intercepts. This random intercept partials out between-person variance such that the lagged effects in the RI-CLPM actually pertain to within-person dynamics (Hamaker et al., 2015). However, using the RI-CLPM does not mitigate potential problems concerning measurement error, which may still be misidentified as real changes, or time intervals that are too short for any effects to occur or too long for effects to remain.

3. Informative Hypotheses

Scientists often have clear expectations about the order or the sign of the parameters in their statistical model when they formulate their hypotheses. For instance, in a (RI-)CLPM, researchers might have expectations about the predictive strength of the cross-lagged effects: Masselink et al. (2018) stated that self-esteem and depressive symptoms are associated with each other and the cross-lagged effect of self-esteem on depressive symptoms, that is, the effect of self-esteem on time point $t - 1$ on depressive symptoms on time point t (γ_t) is stronger in absolute sense than the cross-lagged effect of depressive symptoms on self-esteem, that is, the effect of depressive symptoms on time point $t - 1$ on self-esteem on time point t (β_t). This a-priori expectation is reflected by the following hypothesis:

$$H_1 : |\beta_t| < |\gamma_t|. \quad (9)$$

In cross-lagged longitudinal analysis, many researchers in the field of psychology want to examine the absolute strength of effects between two variables over time. The variable with the strongest cross-lagged effect is “causally dominant”. Often, researchers have their own experiences or some evidence based on their expertise or on previous research, researchers have a-priori hypothesis regarding which variable has more predictive strength. For example, Amtmann et al. (2020) studied the cross-lagged effect between pain intensity and sleep disturbance and stated that the lagged effect of sleep on pain (γ) tends to be stronger (in absolute sense) than the lagged effect of pain on sleep (β). The investigation of the causal dominance hypothesis can be written as $|\beta| < |\gamma|$. As another example, Van der Schuur et al. (2019) assessed whether the cross-lagged effect between current stress and previous media use (β) is stronger in absolute sense than cross-lagged effect between current media use and previous stress (γ). This is reflected by the hypothesis $|\beta| > |\gamma|$. In the case of wave-specific parameters, like in Masselink et al. (2018), the hypotheses involve more than two parameters and more than one order-

restriction, $|\beta_2| < |\gamma_2|$; $|\beta_3| < |\gamma_3|$. In case of a tri-variate model, hypotheses often involve more than two parameters and more than one order-restriction. One can, for instance, hypothesize the direction of the “causal dominance” for pairs of reciprocal cross-lagged effects. For example, in a tri-variate model, one can be interested in the hypothesis $|\phi_{12}| < |\phi_{21}|$, $|\phi_{13}| < |\phi_{31}|$, $|\phi_{23}| < |\phi_{32}|$, with ϕ_{12} the cross-lagged effect of the variable 1 on time point $t - 1$ on the variable 2 on time point t , ϕ_{21} the cross-lagged effect of the variable 2 on time point $t - 1$ on the variable 1 on time point t and so on. Note that such comparisons are only meaningful and fair when parameters are on a comparable scale, which oftentimes come down to inspecting standardized parameter estimates.

While researchers often have these type of informative hypotheses, their evaluation cannot be done with current model comparison techniques reported in SEM software. Fortunately, this can be done with GORICA, a recently developed information criterion. In the next section, we first discuss the currently reported methods and state why they do not suffice. Second, we elaborate on the GORICA and how it can evaluate the hypothesis/hypotheses of interest. Since the performance of the GORICA is not yet investigated for the CLPM and RI-CLPM, we will examine it using a simulation study, which will be discussed in the subsequent section.

4. Model Selection in SEM

One or more theories regarding the model parameters can be reflected by one or more hypotheses. These hypotheses, also referred to as models, can be compared. In case of model selection and model comparison, the models to be compared are different specifications of the same statistical model and can thus be seen as hypotheses on model parameters. In current SEM software, there exist two types of model comparison methods: The Chi-square difference test and model selection (using information criteria). The Chi-square difference test can compare two nested models (e.g., a model where all parameters are freely estimated and a nested model where some of the parameters are constrained). One then needs the difference between the Chi-square values and their degrees of freedom from the two nested models.¹ Since the Chi-square test cannot compare non-nested models, we will not further discuss this model comparison technique.

Another technique to compare models is model selection using information criteria.² Many information criteria have

¹In case of constraining variances to zero (e.g., the variances of the random intercepts in a RI-CLPM), the test should be adjusted, since variances are nonnegative values. Note that the reference value zero lies on the boundary of the parameter space of a variance. Consequently, the zero variance model should be tested against a model with a positive variance instead of a model with a freely-estimated variance. This can be done by using the Chi-bar-square difference test (Hamaker et al., 2015; Stoel et al., 2006), available in the R package Chibarsq.difftest (Kuiper, 2021a) for which example code can be found on: chi-bar-square test. Notably, the Chi-bar-square difference test cannot compare non-nested models nor can it evaluate inequality constraints other than testing positive variances.

²There is another framework to compare the relative evidence for one model over the other, namely the Bayesian framework based on Bayes Factors (BFs; Kass and Raftery (1995)). The BF quantifies the relative support provided by

been proposed, for instance, Akaike's information criterion (AIC; Akaike, 1973), Bayesian information criterion (BIC; Schwarz, 1978), and Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). These criteria are suited to compare both nested and non-nested models. Since the AIC is often reported in SEM software and has order-restricted extensions (as will be explained later on), we will focus on the AIC here.

4.1. The Akaike Information Criterion and Order-Restricted Extensions

The Akaike Information Criterion (AIC) is one of the most widely used methods for choosing a "best approximating" model from several competing models given a particular dataset. It was developed by Akaike in 1973. The basis of the AIC relies on several assumptions. It is assumed that the data is a realization of a random variable, which has an unknown probability distribution. Fortunately, one can draw inferences about the "true" distribution using the distribution of the data. Using this assumption, the best approximating model would be the model in which the "distance" between the estimated distribution and "true" distribution is as small as possible (Burnham & Anderson, 2002). Akaike showed that this selection of the best model can be done with the AIC:

$$AIC_m = -2l_m + 2K_m \quad (10)$$

where K_m is the number of distinct model parameters and l_m is the maximum log-likelihood under the restrictions in Model m .

The AIC can only evaluate hypotheses that contain equality restrictions (and freely estimated parameters). However, theories often involve orderings of parameters, which are expressed by inequality restrictions. Suppose we have independent random variables from each of K populations specified by scalar-valued, unknown parameters $\theta_1, \theta_2, \dots, \theta_K$ satisfying the simple-order restriction $\theta_1 \leq \theta_2 \leq \dots \leq \theta_K$. The AIC cannot evaluate these simple-order restricted hypotheses, whereas the order-restricted information criterion (ORIC) proposed by Anraku (1999), a modification of the AIC, can. ORIC can detect the configuration of the true parameters with the simple-order restriction. The ORIC is defined as follows:

$$ORIC_m = -2l_m + 2p_m, \quad (11)$$

where l_m is the maximum log-likelihood under the m th candidate model and p_m is the dimension for model m .

Kuiper et al. (2011, 2012) proposed the generalized order-restricted information criterion (GORIC) which is a generalization of the ORIC. While ORIC is applicable to simple-order restriction (e.g. $\theta_1 \leq \theta_2 \leq \theta_3$) and tree order restrictions (e.g. $\theta_2 \leq \theta_1, \theta_3 \leq \theta_1$), the GORIC can be used for the evaluation of (in)equality constrained hypotheses going beyond these hypotheses. For example, the GORIC can be used to evaluate the hypothesis $\frac{\theta_1 + \theta_2}{2} > \theta_3$ or $\theta_1 -$

$\theta_2 > \theta_3 - \theta_4$ or H_1 in Equation (9), the GORIC for Model m is calculated by

$$GORIC_m = -2l_m + 2PT_m \quad (12)$$

where the $GORIC_m$ incorporates the likelihood part, that is, the maximum order-restricted likelihood, thus under the constraint (l_m) and the penalty part (PT_m). The penalty again reflects the dimension of the model and can be seen as the expected number of parameters. Note that, in case of simple-order or tree-order restricted hypotheses, PT_m reduces to p_m , and, in case of equality-restricted hypotheses, PT_m reduces to K_m . The GORIC can be applied to univariate and multivariate normal linear models.

Altınışık (2018) proposed the generalized order-restricted information criterion approximation (GORICA), the generalization of the GORIC. The GORICA can be utilized to evaluate the same type of (in)equality constrained hypotheses as the GORIC. The advantage of the GORICA is that it is applicable to not only normal linear models, but also easily to generalized linear models (GLMs), generalized linear mixed models (GLMMs), and structural equation models (SEMs) and thus also to (RI-)CLPMs. The performance of the GORICA approximates the performance of the GORIC and extends its use to a broader range of models. The GORICA for Model m is defined as:

$$GORICA_m = -2L_m + 2PT_m \quad (13)$$

where the GORICA also incorporates a likelihood part (L_m), the asymptotic maximum order-restricted log-likelihood of the maximum likelihood estimation (MLEs) and a penalty part (PT_m), which is the same as the penalty in the GORIC.

For all these information criteria (ICs), it holds that the hypothesis with the smallest criterion value is the best from the set of hypotheses. Thus, IC values only denote the ordering of the candidate models and not the relative strength. The extent to which a hypothesis is better than the other can be quantified using IC weights, the relative likelihood of a hypothesis in a set of hypotheses (Burnham & Anderson, 2002; Kuiper et al., 2012; Wagenmakers & Farrell, 2004). The IC weights (w_m) can be obtained as follows:

$$w_m = \frac{\exp\{-\frac{1}{2} IC_m\}}{\sum_{m'=1}^M \exp\{-\frac{1}{2} IC_{m'}\}}. \quad (14)$$

The strength of evidence in favor of one model over another is obtained by dividing their weights. For example, the support for H_1 vs. H_2 is denoted by w_1/w_2 . If hypotheses H_1 and H_2 have IC weights of 0.75 and 0.25, respectively, H_1 has $0.75/0.25 = 3$ times more support than H_2 . Note that the GORICA weights will asymptotically equal the GORIC weights. Additionally, in case of equality-restricted hypotheses, the GORICA weights will asymptotically equal the AIC weights; but only in the case of equality-restrictions, since the AIC is not equipped to evaluate inequality-restrictions. The GORICA values and weights and the ratio of GORICA weights can be obtained with the goric function (Vanbrabant & Kuiper, 2020) of the R package restriktor (Vanbrabant & Rosseel, 2020).

Since the focus of the paper is on evaluating informative hypotheses (more specifically, causal-dominance hypotheses)

the data for two competing hypotheses. BFs can be calculated in R, for example, using the bain package (Gu et al., 2019, 2018; Hoijtink et al., 2019). For the comparison of the performance of GORICA and bain, the interested reader is referred to the [supplementary material](#).

Table 1. Specifications in simulations for bivariate (RI-)CLPM with a total of 144 conditions ($2 \times 4 \times 6 \times 3$).

Model	Number of waves	Sample size	Cross-lagged effects (β, γ)
RI-CLPM and CLPM	3, 4, 5, 6	25, 50, 75, 100, 200, 300	(0.1, 0.2) (0.1, 0.4) (-0.07, -0.07)

in (RI-)CLPMs, we will use the GORICA. Specifically, our goal is to investigate the performance of the GORICA to evaluate causal dominance hypotheses in lagged-effects models. Here, we will focus on two statistical models: CLPM and RI-CLPM, and we will inspect the performance of the GORICA in these models with a simulation study. The design of our simulation studies is described next.

5. Simulation study

To evaluate the performance of the GORICA, we will conduct a simulation study. We employ two types of simulations: (1) a bivariate (RI-)CLPM, like used in Masselink et al. (2018), and (2) a tri-variate (RI-)CLPM. In the simulation, we will thus inspect the CLPM as well. Namely, although the RI-CLPM is often more realistic, there may be processes that come from a CLPM data generating process. In the simulations, we will investigate the ability of the GORICA to correctly select the true hypothesis regarding causal dominance. The performance of the GORICA is measured by the true hypothesis rate (THR; i.e., the percentage of times the true hypothesis is selected).

5.1. Simulation I: Bivariate (RI-)CLPM

The first simulation aims to examine the performance of the GORICA in a bivariate (RI-)CLPM, as depicted in Figures 1 and 2. We simulated bivariate data, using R (R Core Team, 2020), with a varying number of waves, varying number of participants, and varying strengths of cross-lagged effects. Our choice of population values is based on the study by Masselink et al. (2018), who used three waves, which is the minimum number of waves that is required to fit a RI-CLPM, and a number of participants of $N=2105$. Therefore, we started varying the waves from three waves on up until 6 (i.e., $W \in \{3, 4, 5, 6\}$). Note that it is to be expected that the performance increases with number of wave. In case 6 waves do not show a good performance yet, we will increase the number of wave. For the number of participants, we started with a low value, namely 25. Then, we increased it with steps of 25 and later on steps of 100 (or more) up until 2100. As will become clear later on, for much lower number of participants, GORICA performed well and/or the trend of the performance was clear, this led to the decision to stop at $N=300$. Consequently, we inspect the following number of participants: $N \in \{25, 50, 75, 100, 200, 300\}$. Additionally, two types of cross-lagged population parameter specifications are considered: in one setting, the cross-lagged effects are unequal (β, γ) = (0.1, 0.2) and (β, γ) = (0.1, 0.4); in the other setting, the cross-lagged effects are equal (β, γ) = (-0.07, -0.07), using the parameter estimates in Masselink et al. (2018). The

autoregressive population parameters are held constant over the simulations: $(\alpha, \delta) = (0.22, 0.28)$, using the estimates found in Masselink et al. (2018).³ Table 1 gives an overview of the simulation design in a bivariate (RI-)CLPM.

In each generated dataset, we evaluate the causal dominance hypothesis as specified in Equation (9), that is, $H_1 : |\beta| < |\gamma|$, versus its complement $H_2 : |\beta| > |\gamma|$. Note that H_1 is true in the setting with unequal parameter specifications, while H_1 and H_2 are equally true in the equal parameter specification, since the truth is on the boundary of the hypotheses.

In the simulation, 1000 datasets will be generated per simulation condition. CLPM data is created by drawing data for each participant from the first-order vector autoregressive (VAR(1)), which can be regarded as a single-subject CLPM that generally has multiple observations (waves). Thus, the lagged-effect between current and previous observations for one participant can be described as a VAR(1) model (Hamilton, 1994).

$$Y_t = \mathbf{c} + \Phi Y_{t-1} + \mathbf{u}_t \quad (15)$$

where Y_t and Y_{t-1} represent the values of the two variables at time points t and $t-1$ (for $t = 1, \dots, T$), respectively, the vector \mathbf{c} is the intercept, the 2×2 matrix Φ contains the VAR(1) regression parameters (α, β, γ , and δ), and the vector \mathbf{u}_t contains the residuals, also referred to as innovations, at time t .

We sample $T = 100$ VAR(1) observations using the `VAR.sim()` function from the `tsDyn` package (Antonio et al., 2009) and use only the last W of those T timepoints, with $W \in \{3, 4, 5, 6\}$ the number of waves in that simulation. When there are three waves, the last $W = 3$ observations from each variable are used, and this process is repeated for each participant (as depicted in Figure 3 with $W = 3$ waves, 2 variables, and N participants).

To generate RI-CLPM data, the creation of CLPM data is extended by adding person-specific means to the data. These means are generated from a bivariate normal distribution with mean $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix $\boldsymbol{\Omega} = \mathbf{I}_2$, a 2×2 identity matrix. Resulting in an $N \times 2$ matrix of person-specific means for the N participants and 2 variables. Note that the means are, by definition, equal over the waves.

Each generated dataset will be analyzed with both the CLPM and the RI-CLPM (Figures 1 and 2 for a graphical representation) using the R package `lavaan` (Rosseel, 2012). Note that the autoregressive and cross-lagged paths are constrained to be the same across time in both models. Hence, the models do not contain wave-specific parameters.

³We used the estimates from the first to second wave of the RI-CLPM in Masselink et al. (2018) since they estimate wave-specific ones. Note that, for ease, we use a (RI-)CLPM where the parameters are not wave-specific. In the example section and the supplementary material, we will show how the GORICA can be applied to models with and without wave-specific parameters.

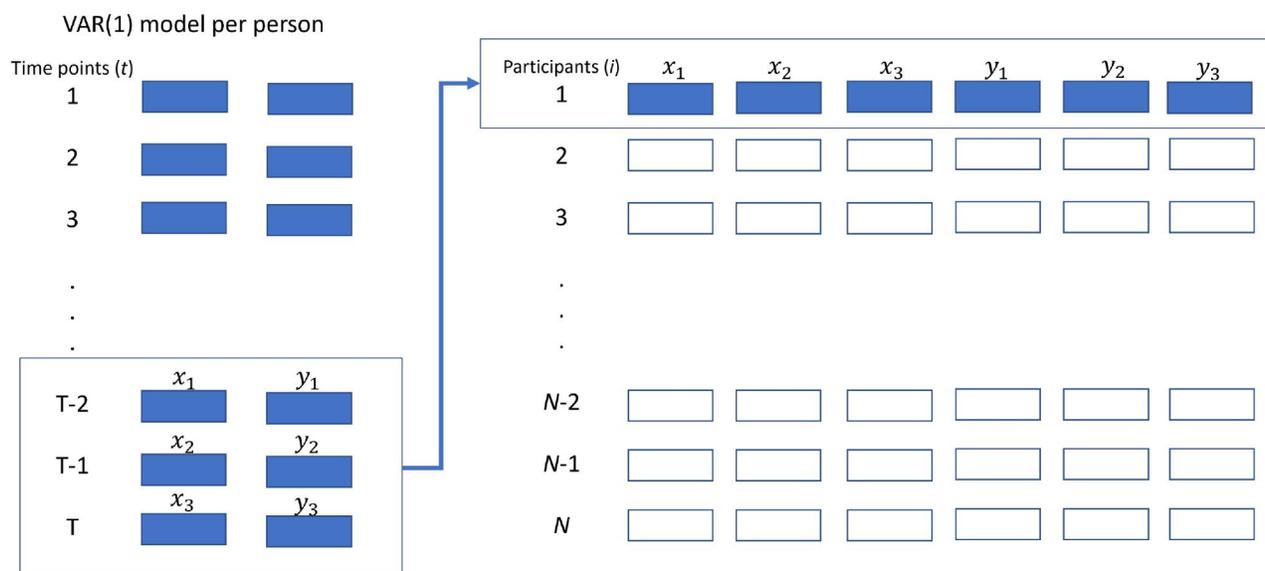


Figure 3. An illustration of how the three-wave data for the first participant ($i=1$) is obtained from data generated from the VAR(1) model. This is repeated N times, that is, for each individual.

Inspection of the mean parameter estimates of α , β , γ , and δ over the simulations (by checking whether they are close/asymptotically equal to the population parameter value) ensured the quality of the data generation.

Per simulation, both the standardized autoregressive estimates (α and δ), the standardized cross-lagged estimates (β and γ), and the covariance matrix of those estimates are extracted. This serves as the input for the calculation of the GORICA. The *goric* function (Vanbrabant & Kuiper, 2020) in the *restriktor* package (Vanbrabant & Rosseel, 2020) is used to calculate the GORICA values and the GORICA weights per simulation, which are used to calculate the percentage of times the true hypothesis is selected (THR).

When applying the RI-CLPM, we also use the GORICA to evaluate whether the random intercept variances are larger than zero ($\text{Var}_{[\kappa]} > 0$ and $\text{Var}_{[\omega]} > 0$, where $\text{Var}_{[\kappa]}$ and $\text{Var}_{[\omega]}$ are the variances of κ and ω , respectively, the random intercepts in a bivariate RI-CLPM). We did this to ensure that the model requires random intercepts. Especially, when the number of participants, the number of waves, and/or the difference between the cross-lagged effects are small, the random intercept variances may be non-positive. We also found this in our simulation. In practice, one would now conclude that the RI-CLPM is a poor fit, leading to the recommendation of using a CLPM instead of a RI-CLPM. In the simulation, this implies discarding the drawn sample, since our focus is on evaluating the GORICA within a RI-CLPM. Hence, we drew another sample such that in total we base our conclusions still on 1000 simulated data sets. As an example, in a bivariate RI-CLPM with $W = 3$ waves, $N = 25$ participants and $(\beta, \gamma) = (0.1, 0.2)$, the GORICA did not find support for positive random intercept variances in 3 out of 1003 iterations. Pages 1 and 9 (i.e., Tables 1 and 2) of the [supplementary material](#) (Supplementary material for Sukpan & Kuiper) provide more detail on how often we encountered negative variances (in which simulation condition). There, one can see that this happens in only 0.02% of the simulation cases when inspecting the bivariate model and

0.09% of the simulation cases when inspecting the tri-variate model. Notably, when the number of waves is larger than 4, the GORICA always found support for positive random intercept variances, irrespective of the number of participants. Furthermore, when the number of participants, the number of waves, and/or the difference between the cross-lagged effects are small, we came across the following warnings: “the covariance matrix of latent variables is not positive definite matrix” and “in sqrt (ETA2): NaNs produced. These warnings typically indicate a poorly fitted model as well, which can be expected in such situations; for example, situations with a small number of participants and only a few waves. Bear in mind that, in the bivariate RI-CLPM with three waves of data, 26 parameters need to be estimated (even though there are 7 equality constraints). Consequently, it may be hard to fit the model in the case of $N = 25$ participants. When this happened, we did not proceed with the next steps (e.g., applying the GORICA), we discarded the drawn data set. To still obtain 1000 data sets to base our conclusions on, we then sampled once more a data set (from the same population), as described in Figure 3. Next, we estimated the (RI-)CLPM using this new data set. If there were no negative variances or warnings, we proceeded with applying GORICA. In total, we thus simulated 1000 datasets, all of which exhibited neither negative variances nor warnings. We could have decided to not inspect situations where the number of participants, the number of waves, and/or the difference between the cross-lagged effects are small, but we are also interested in seeing the performance for such cases. Notably, when we would use 200 participants or more in our simulation, we would not have encountered this problem. The results of the performance of the GORICA, when the (RI-) CLPM is the data generating model, are discussed in the upcoming section.

5.1.1. The Performance of GORICA in a Bivariate (RI-) CLPM

Figure 4 shows the results when evaluating the true, informative hypothesis $H_1: |\beta| < |\gamma|$ versus its complement

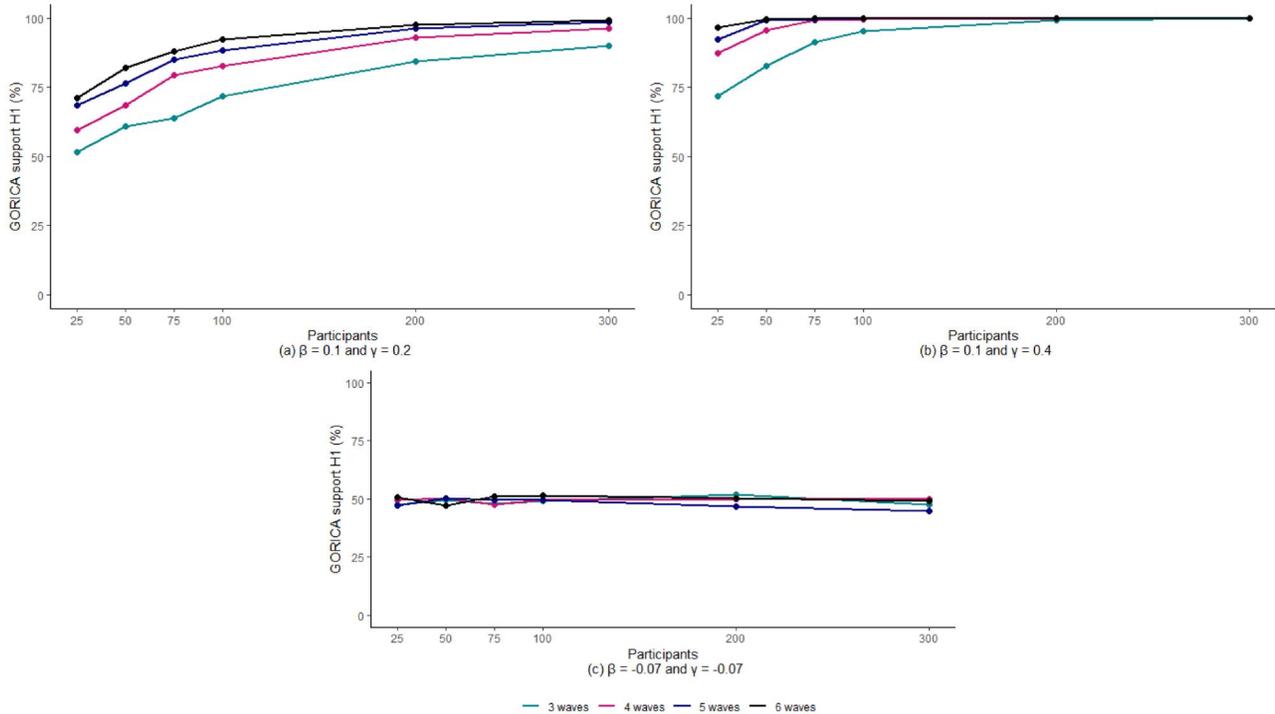


Figure 4. The (true) hypothesis rates when using the GORICA in a bivariate RI-CLPM for all simulation conditions, namely the number of participants (x-axis), the number of waves (different Colored lines), and the three different pairs of cross-lagged parameters (β and γ): (a) $(\beta, \gamma) = (0.1, 0.2)$, (b) $(\beta, \gamma) = (0.1, 0.4)$, and (c) $(\beta, \gamma) = (-0.07, -0.07)$. Note that in plots (a) and (b), H_1 is the true hypothesis, while in Plot (c) both $H_1 : |\beta| < |\gamma|$ and $H_2 : |\beta| > |\gamma|$ are true.

($H_2 : |\beta| > |\gamma|$) using the GORICA. The performance of the GORICA is similar under both the RI-CLPM and the CLPM. Hence, the results for the RI-CLPM are shown here, while the results and patterns for the CLPM are reported in the [supplementary material](#). In [Figure 4a](#), where $|\beta| = 0.1$ and $|\gamma| = 0.2$, the THR is larger than 50% and increases with the number of waves and the number of participants. When the difference between the cross-lagged effects larger ([Figure 4b](#) versus [Figure 4a](#)), the THR rises faster as well, and is actually close to 100% under all specifications concerning the number of observations and the number of waves. In [Figure 4c](#), these cross-lagged parameters (β and γ) based on [Masselink et al. \(2018\)](#) where both hypotheses are correct since $|\beta| = |\gamma| = 0.07$, both hypotheses obtain the same amount of support (i.e., the THR fluctuates around 50%). This behavior is to be expected, because the hypotheses under consideration are both correct and equally complex.⁴ The GORICA thus performs adequately when evaluating causal dominance in a (RI-)CLPM, because when there is causal dominance, this is generally detected, that is, the GORICA does quantify the support for the hypothesis of interest ($H_1 : |\beta| < |\gamma|$), while if there is no causal dominance (i.e., when the strength of the parameters is equally strong), both hypotheses obtained a similar amount of support. [Kuiper \(2021b\)](#) found the same trends, when looking at so-called CTmeta-analyzed estimates in a less comprehensive simulation study. Furthermore,

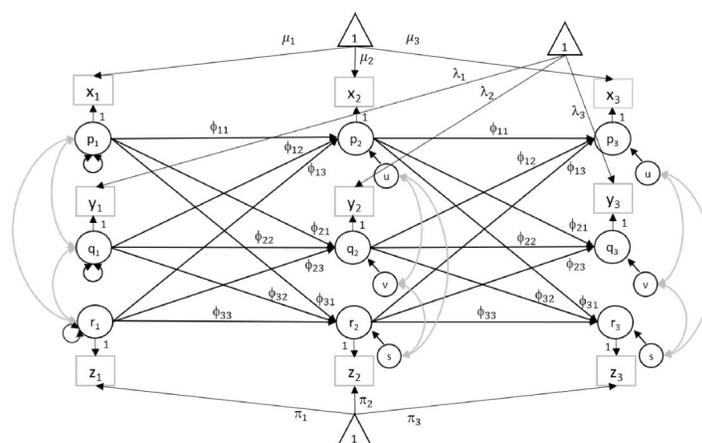
these results validate our expectation that GORICA has the capability to evaluate causal dominance in a bivariate (RI-)CLPM.

5.2. Simulation II: Tri-Variate (RI-)CLPM

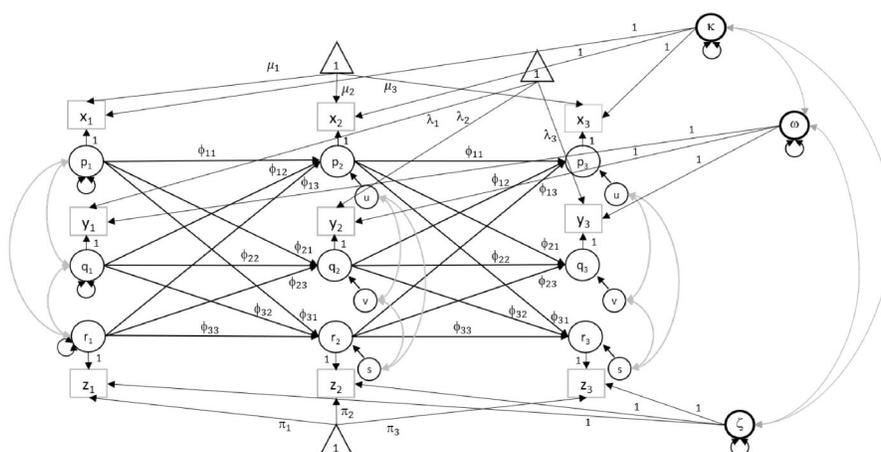
In the second type of simulation study, the previous simulations are extended to a tri-variate (RI-)CLPM. Compared to the bivariate (RI-)CLPM, in a tri-variate (RI-)CLPM three, rather than two, sets of variables are measured at multiple occasions and may affect one another. The tri-variate CLPM and RI-CLPM are path diagram displayed in [Figure 5](#). The tri-variate data is generated analogously to the bivariate simulation, with 1000 datasets per simulation condition but with three variables rather than two. The number of participants will again be varied with $N \in \{25, 50, 75, 100, 200, 300\}$. Here, only three waves will be considered, because in that case the performance of the GORICA is already good; and we already saw that increasing the number of waves increases the THR (as to be expected). Three sets of cross-lagged population parameter specifications will be considered which are based on the bivariate population values: $(\phi_{12}, \phi_{21}, \phi_{13}, \phi_{31}, \phi_{23}, \phi_{32}) = (0.1, 0.2, 0.1, 0.2, 0.1, 0.2)$, $(\phi_{12}, \phi_{21}, \phi_{13}, \phi_{31}, \phi_{23}, \phi_{32}) = (0.1, 0.4, 0.1, 0.4, 0.1, 0.4)$, and $(\phi_{12}, \phi_{21}, \phi_{13}, \phi_{31}, \phi_{23}, \phi_{32}) = (-0.07, -0.07, -0.07, -0.07, -0.07, -0.07)$. [Table 2](#) presents a summary of the simulation conditions.

The lagged effect parameters are again generated in such a way that they are the same across waves, and hence, no wave-specific parameters will be inspected. The autoregressive population parameters for all specifications are 0.22, 0.28, and 0.28 as in the previous simulation type. The causal

⁴In practice, one will find an equal fit, an equal complexity, and thus GORICA weights of 0.5 (which is the same as the weights based on the complexity/penalty). This then indicates that both hypotheses are equally supported, which reflects support for the border of the hypothesis, that is, support for equal strength of parameters.



(a) Tri-variate CLPM



(b) Tri-variate RI-CLPM

Figure 5. The tri-variate CLPM and tri-variate RI-CLPM; where squares denote the observed variables, circles denote latent variables and triangles represent constants for the mean structure. The latent variables p and q in Figure 5(a) are the grand-mean centered observed variables, while p , q , and r are the person-mean centered observed variables in Figure 5(b). Single-headed arrows indicate regression, double-headed arrows indicate correlations or (co)variances.

Table 2. Specifications in simulations for tri-variate (RI-)CLPM with a total of 36 conditions ($2 \times 1 \times 6 \times 3$).

Model	Number of wave	Sample size	Cross-lagged effects ($\phi_{12}, \phi_{21}, \phi_{13}, \phi_{31}, \phi_{23}, \phi_{32}$)
RI-CLPM and CLPM	3	25, 50, 75, 100, 200, 300	(0.1, 0.2, 0.1, 0.2, 0.1, 0.2) (0.1, 0.4, 0.1, 0.4, 0.1, 0.4) (-0.07, -0.07, -0.07, -0.07, -0.07, -0.07)

dominance hypothesis of interest in this simulation can be written as

$$H_3 : |\phi_{12}| < |\phi_{21}|, |\phi_{13}| < |\phi_{31}|, |\phi_{23}| < |\phi_{32}|. \quad (16)$$

H_3 will be compared with its complement, representing all orderings except the one in H_3 .⁵

Similarly, to previous simulations, the hypothesis of interest H_3 is true in the first two sets of cross-lagged parameter specifications, leading to the expectation that the THR is expected to converges to 100% in these simulations. In the third set of simulations, where the parameter are hypothesized to be

equal, H_3 and its complement are both true (that is, the truth is on their border). Because H_3 is less complex than its complement, and hence has a lower penalty, it should be selected on the basis of parsimony. Bear in mind that, due to sampling variation, the cross-lagged effects will not be estimated as exactly equal and the complement will sometimes be the best when the difference in likelihood outweighs the difference in penalty. Hence, it can be expected that the support for H_3 is lower than 100%, but higher than 50%. Similarly, to the bi-variate simulation study, the THR is calculated as the percentage of times H_3 is the preferred hypothesis.⁶

Note that, again, non-positive variances of the random intercept in the RI-CLPM where encountered. For the

⁵Notably, if there is a competing theory stating different orderings (perhaps even the opposite ordering of this one) or no causal dominance (stating that the reciprocal cross-lagged parameters are of equal strength), then this should be added to the set and then the unconstrained hypothesis should be included as safeguard. In this simulation, we thus assume that there is one theory, which is then compared against all other possible theories.

⁶In practice, one will find an equal fit and GORICA weights that equal the weights based on the complexity/penalty. This then indicates that both hypotheses are equally supported, which reflects support for their border.

simulations in a tri-variate RI-CLPM with 3 waves, the GORICA did not find support for positive variances of the random intercepts in some of the simulations with $N = 25, 50, 75, 100$ and 200 . In practice, one would now conclude that the RI-CLPM is not the correct model, but the CLPM is. Since we are interested in the performance of the GORICA in case of a RI-CLPM, we disregard these simulations and sampled new data (like in the simulation study for the bivariate model). For more details, refer to the [supplementary material](#) available at [Supplementary material](#) for Sukpan & Kuiper.

5.2.1. The performance of GORICA in a Tri-Variate (RI-)CLPM

Figure 6 shows the performance of the GORICA for the tri-variate RI-CLPM, the results for the CLPM can be found in the [supplementary material](#), where one can see the same patterns as for the RI-CLPM. In Figures 6a and b, H_3 is the true hypothesis, which is detected by the GORICA. The THR is structurally above 90% and increases quickly to 100% when the number of participants and/or the difference between the cross-lagged parameters increases. In Figure 6c, when the cross-lagged parameters are specified as equal, the less complex hypothesis H_3 receives most support, namely in about 87% of the simulations. Consequently, the complement of H_3 which is also true, but more complex is preferred in 13% of the simulations. This is due to equalities never being exactly true and sometimes the difference in likelihood will outweigh the difference in penalty.

In practice, one will find equal fit for both hypotheses. Then, the GORICA weights are solely determined on the penalty differences. This, then indicates support for the border of the hypothesis, that is, support for equal strength of parameters. Since H_3 is the least complex one, it will obtain higher support. The mean GORICA weight, in this scenario, converges to 0.77 which equals the weight based on solely the penalty values. Note that when H_3 is true, the mean GORICA weight (and THR) of H_3 converges to 1 (and 100%), see the mean weights of all simulations in the [supplementary material](#).

5.2.2. What if the Hypothesis of Interest is Not True and Thus its Complement is?

In the previous simulation set-ups, the hypothesis of interest was the true data generating hypothesis. What if the hypothesis of interest is not true and thus its complement is? In the bivariate (RI-)CLPM, either H_1 or its equally-sized complement H_2 is true ($H_1: |\beta| < |\gamma|$ and $H_2: |\beta| > |\gamma|$). Consequently, when H_2 is true, the THR for H_2 is the same as the THR for H_1 , resulting in the same plots. Bear in mind that, when H_2 is true, the selection rate of H_1 will be 100% minus the THR of H_2 .

In the tri-variate (RI-)CLPM, the informative causal dominance hypothesis has a different complexity than its complement since it consist of many orderings for examples the ones displayed in H_4 and H_5 in Equation (17). Consequently, when the complement of H_3 is true, its THR

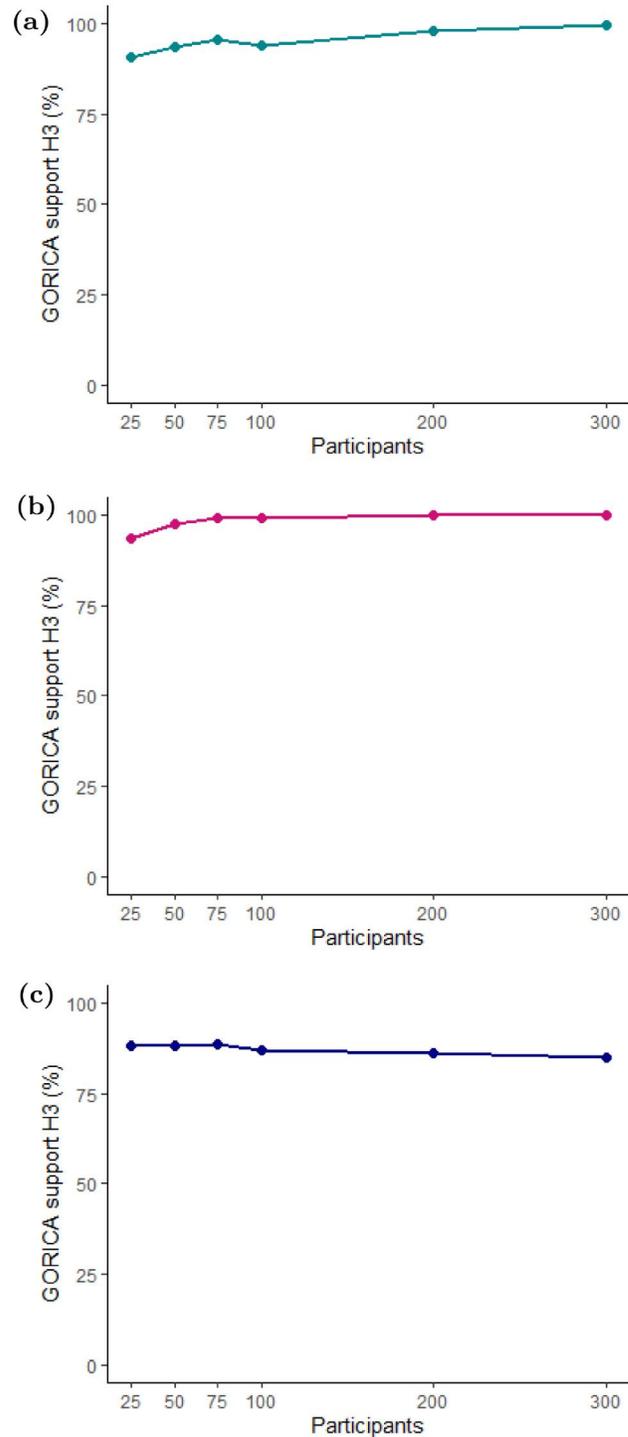


Figure 6. The (true) hypothesis rates when using the GORICA in a tri-variate RI-CLPM for all simulation conditions, namely the number of participants (x-axis) and the three different pairs of cross-lagged parameters ($\phi_{12}, \phi_{21}, \phi_{13}, \phi_{31}, \phi_{23}, \phi_{32}$): (a) $(\phi_{12}, \phi_{21}, \phi_{13}, \phi_{31}, \phi_{23}, \phi_{32}) = (0.1, 0.2, 0.1, 0.2, 0.1, 0.2)$, (b) $(\phi_{12}, \phi_{21}, \phi_{13}, \phi_{31}, \phi_{23}, \phi_{32}) = (0.1, 0.4, 0.1, 0.4, 0.1, 0.4)$, and (c) $(\phi_{12}, \phi_{21}, \phi_{13}, \phi_{31}, \phi_{23}, \phi_{32}) = (-0.07, -0.07, -0.07, -0.07, -0.07, -0.07)$. Note that in plots (a) and (b), H_3 is the true hypothesis, while in Plot (c) both H_3 and its complement are true but H_3 is the most parsimonious one.

is not the same as the THR for H_3 , that is, the plots will differ. Hence, it is of interest to inspect what happens if the hypothesis of interest in the tri-variate model is not true, but its complement is. For ease of generating the data and explaining the set-up, we will again use the set-up in which

H_3 is true. The difference is that we will now investigate what happens if we evaluate a hypothesis other than H_3 , which is incorrect, versus the complement of that hypothesis. This will be done for two types of incorrect hypotheses: One where all the order restrictions are in the complete opposite direction (H_4) and one where only one (namely the first) order restriction is reversed (H_5):

$$\begin{aligned} H_4 &: |\phi_{12}| > |\phi_{21}|, |\phi_{13}| > |\phi_{31}|, |\phi_{23}| > |\phi_{32}|, \\ H_5 &: |\phi_{12}| > |\phi_{21}|, |\phi_{13}| < |\phi_{31}|, |\phi_{23}| < |\phi_{32}|. \end{aligned} \quad (17)$$

We will run a simulation for each of these two hypotheses, using the same parameter specifications as in the previous simulations. Note that for the equal population cross-lagged parameter specification, the simulation results are not different from the previous simulations, because the complexity of H_4 and H_5 is (approximately) the same as of that H_3 . Hence, only the simulations with order-restricted parameters will be discussed. For these simulations, both H_4 and H_5 are incorrect and their complements are true. Hence, asymptotically, we expect no support for H_4 and H_5 . The performance of the GORICA in a RI-CLPM is presented in Figure 7. Similar results for the CLPM are shown in the supplementary material.

Using the GORICA to evaluate the hypotheses of interest in a (RI)-CLPM, it can be seen that, when the complement hypotheses are true, the support for the incorrect hypotheses of interest H_4 and H_5 decreases with the number of participants and with the difference between the cross-lagged parameters (Figure 7). When the number of participants N is 300 or more and the difference between the cross-lagged parameters is large, the incorrect hypotheses of interest H_4 receive almost no support anymore. Notably, for a small number of participants, it is hard to choose the true hypothesis from the incorrect ones (namely, the THRs are larger than 75% for $N = 25$ in Figures 7a and b), especially if the hypothesis of interest is close to the truth (like H_5 which differs in only one order restriction). Additionally, it can be seen that the support for H_4 , which contains solely incorrect order restrictions, is always lower than the support for H_5 , which contains only one incorrect order restriction.

6. Example

In this example section, we will demonstrate the use of the GORICA on the RI-CLPM using the panel data of Masselink et al. (2018), who assessed the longitudinal effect between self-esteem and depressive symptoms. We will inspect two types of RI-CLPMs. Namely, one where we have wave-independent parameters (i.e., where the parameters are the same across waves) and one where we do allow for wave-specific lagged-effects parameters. The first is comparable to the model we used in our simulations and is depicted in Figure 8. The second is the one Masselink et al. (2018) use and is displayed in Figure 9. In both figures, the γ represents the effect between self-esteem at time point t and depressive symptoms at time point $t + 1$, while the β denotes the effect between depressive symptoms at time point t and self-esteem at time point $t + 1$. The estimates from these two types of RI-CLPMs are obtained using the R

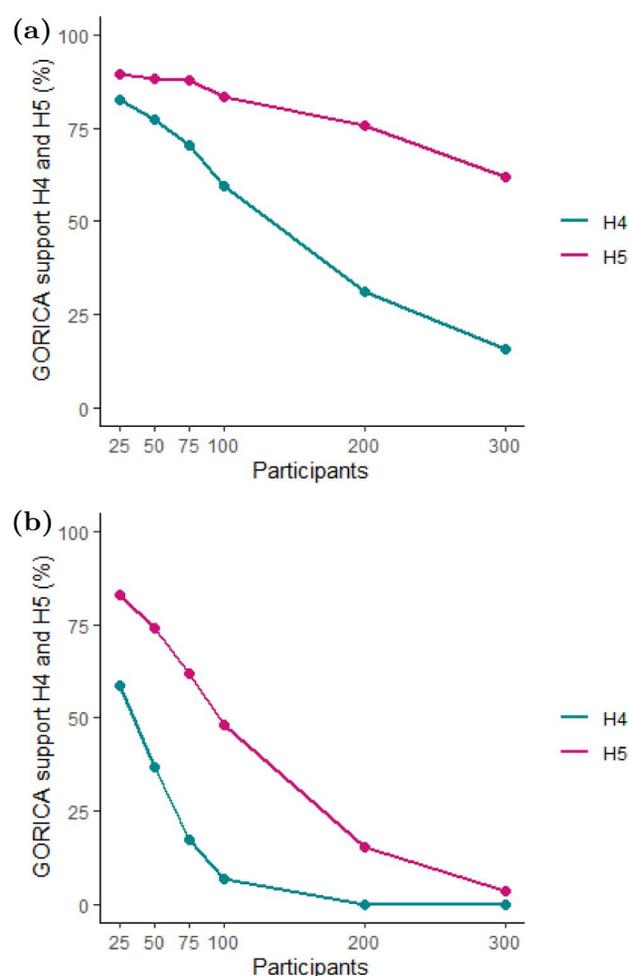


Figure 7. The number of times the incorrect hypotheses H_4 (Green) and H_5 (pink) are chosen in a RI-CLPM, for various participant numbers (x-axis) and two cross-lagged parameter specifications: (a) $(\phi_{12}, \phi_{21}, \phi_{13}, \phi_{31}, \phi_{23}, \phi_{32}) = (0.1, 0.2, 0.1, 0.2, 0.1, 0.2)$ and (b) $(\phi_{12}, \phi_{21}, \phi_{13}, \phi_{31}, \phi_{23}, \phi_{32}) = (0.1, 0.4, 0.1, 0.4, 0.1, 0.4)$.

package lavaan. The R code used in the analysis can be accessed at <https://github.com/Chuenjai/Causal-dominance>.

Many studies have shown that self-esteem and depressive symptoms are related (Orth et al., 2009; Steiger et al., 2014; Mlawer et al., 2021). Moreover, Masselink et al. (2018) have an expectation that adolescents with the cross-lagged effect of self-esteem on depressive symptoms is hypothesized to be stronger in absolute sense than the cross-lagged effect of depressive symptoms on self-esteem. Formalizing this theoretical expectation yields the following hypothesis for the model with wave-independent parameters (Figure 8): $H_{1a} : |\beta| < |\gamma|$, where $|\beta|$ and $|\gamma|$ are the absolute values of β and γ . In the case of wave-specific parameters model, the hypothesis of interest is reflected by $H_{1b} : |\beta_2| < |\gamma_2|; |\beta_3| < |\gamma_3|$. Bear in mind that, to compare the dominance of the effects, one should compare the absolute value of the strength of (standardized) effects. In this example, one can clearly see the impact if you do not use absolute values. Namely, -0.053 is smaller than -0.047 , while $|-0.053| = 0.053$ is higher than $|-0.047| = 0.047$. So, when the interest lies in the absolute strength, then one should specify this in the hypothesis. We will examine whether the hypothesis of interest (H_{1a} and H_{1b} in the wave-independent

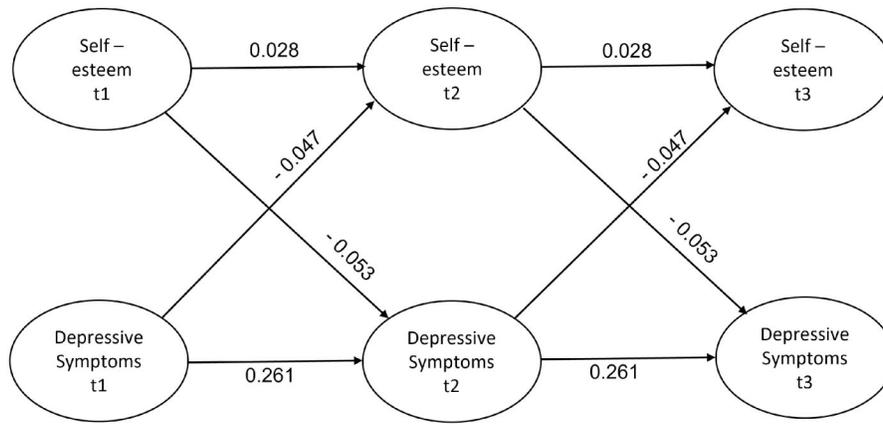


Figure 8. Simplified graph of the RI-CLPM with standardized estimates for the autoregressive parameter estimates and the cross-lagged parameter estimates at time point t and $t + 1$ ($t = 1, 2$) where the standardized cross-lagged paths are constrained to be the same across time.

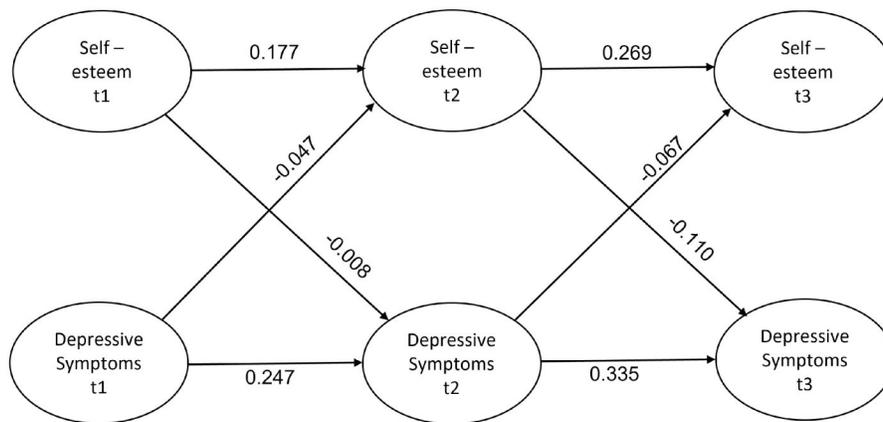


Figure 9. Simplified graph of the RI-CLPM with standardized coefficients for the autoregressive parameter estimates and the cross-lagged parameter estimates at time point t and $t + 1$ ($t = 1, 2$) in a wave-specific parameters model.

Table 3. GORICA values and weights for the hypotheses of interest (H_{1a} and its complement) in a wave-independent parameters model.

Hypothesis	log-likelihood (L_m)	penalty term (PT_m)	GORICA values ($GORICA_m$)	GORICA weights (w_m)
H_{1a}	34.238	12.5	-43.475	0.502
Complement	34.230	12.5	-43.460	0.498

Table 4. GORICA values and weights for the hypotheses of interest (H_{1b} and its complement) in a wave-specific parameters model.

Hypothesis	Log-likelihood (L_m)	penalty term (PT_m)	GORICA values ($GORICA_m$)	GORICA weights (w_m)
H_{1b}	8.410	2.977	-10.865	0.586
Complement	8.607	3.523	-10.168	0.414

parameters model and the wave-specific parameters model, respectively) is more likely than its complement (Tables 3 and 4, respectively). The GORICA evaluation of the hypotheses is done using the *goric* function from the R package *restrktor*. Next, we get the separate sections, a wave-independent parameters model and a wave-specific parameters model and the discussions.

6.1. Wave-Independent Parameters Model

In a wave-independent parameters model, the autoregressive parameters (α and δ) and the cross-lagged parameters

(β and γ) are constrained to be the same across time, like we did in the simulation. Figure 8 displays a simplified representation of the wave-independent parameter RI-CLPM together with the standardized estimates for the data of Masselink and colleagues (the *standardized* estimates $\hat{\beta} = -0.047$, $\hat{\gamma} = -0.053$, $\hat{\alpha} = 0.028$, and $\hat{\delta} = 0.261$). The evaluation of H_{1a} and its complement are shown in Table 3.

The log-likelihood value (L_m) for H_{1a} is only a bit larger than the log-likelihood value of its complement, while both hypotheses are equally complex and thus have equal penalty terms (Table 3). Therefore, the GORICA value for H_{1a} is only a bit smaller than the GORICA value of the complement hypothesis. The GORICA weights show that both hypotheses receive approximately the same amount of support. As the hypotheses are equally complex and do not overlap, the true parameter values are expected to be at the boundary of the hypotheses. That is, it is expected that the cross-lagged effects are equal. Thus, there does not seem to be a causally dominant effect, and the lagged effect between self-esteem and depressive symptoms seems equally strong as the reciprocal lagged effect. Hence, based on this result, one could conclude that the hypothesis stating equally strong (standardized) cross-lagged effects should be the hypothesis of interest in a future study. Alternatively, one could think about a maximum in absolute difference

between the two standardized cross-lagged effects such that the estimates can be seen as equal in practical sense - implying that, if the absolute difference is larger, the standardized estimates are truly different. For example, one could specify $H_{1a} : |\beta| - |\gamma| < .05$.

6.2. Wave-Specific Parameters Model

Masselink et al. (2018) studied the effect between self-esteem and depressive symptoms using a RI-CLPM with wave-specific parameters. Here, we analyze the data in the same as they did. Figure 9 depicts the wave-specific parameter RI-CLPM together with the standardized estimates between time point 1 and time point 2 (i.e., $\hat{\beta}_2 = -0.047$, $\hat{\gamma}_2 = -0.008$, $\hat{\alpha}_2 = 0.177$, and $\hat{\delta}_2 = 0.247$) and between time point 2 and time point 3 (i.e., $\hat{\beta}_3 = -0.067$, $\hat{\gamma}_3 = -0.110$, $\hat{\alpha}_3 = 0.269$, and $\hat{\delta}_3 = 0.335$). The GORICA results for the evaluation of H_{1b} and its complement are presented in Table 4.

The log-likelihood value (L_m) for H_{1b} is smaller than the log-likelihood value of its complement, indicating that the (unconstrained) estimates are not in agreement with H_{1b} . Nevertheless, the log-likelihood values do seem to be close. The penalty term for the H_{1b} is smaller than the penalty term for the complement of H_{1b} . Balancing fit and complexity, the GORICA value for H_{1b} is smaller than the GORICA value of its complement, denoting that H_{1b} is preferred over its complement. The ratio of GORICA weights shows that H_{1b} is 1.415 times ($0.586/0.414 = 1.415$) more likely than its complement. Thus, there is support for H_{1b} , but we feel it is not compelling evidence for H_{1b} . Hence, in the wave-specific parameters model, the GORICA is slightly supporting the informative hypothesis from Masselink et al. (2018), stating that the cross-lagged effect of self-esteem on depressive symptoms is stronger in absolute sense than the cross-lagged effect of depressive symptoms on self-esteem. In a future study, one could evaluate H_{1b} again. Alternatively, one could think about a substantial absolute difference between the standardized cross-lagged parameters and use that in the hypothesis. For example, one could specify $H_{1b} : |\beta_2| - |\gamma_2| > .1, |\beta_3| - |\gamma_3| > .1$. In that case, finding a ratio of GORICA (for H_{1b} versus its complement) of 1 and more implies that the standardized cross-lagged effects differ substantially.

7. Discussion

In the literature, several studies have highlighted the extensive usage of the (RI-)CLPM in various research domains, including Zhou et al. (2020), Van der Schuur et al. (2019), and Masselink et al. (2018) and the interest in causal dominance. Our study inspected the evaluation of causal dominance within those multivariate autoregressive models. We used standardized parameters (as they are comparable). We did not inspect multilevel models, but if one uses such a model to evaluate causal dominance, one should use person-mean-centered parameters, as suggested by Schuurman et al. (2016).

To evaluate causal dominance hypotheses, we made use of the GORICA (Altınışık, 2018). The GORICA is a tool specifically designed to investigate (in)equality constrained hypotheses in a wide range of models, including (RI-)CLPM. Additionally, the GORICA has been shown to be applicable to CTmeta-analyzed models (Kuiper, 2021b). The study by Kuiper (2021b) utilized the GORICA with so-called CTmeta-analyzed lagged regression estimates (using a in a continuous-time setting) to evaluate causal dominance hypotheses and demonstrated its performance. This study is limited to a small simulation study, focusing on the performance of the GORICA with CTmeta-analyzed aggregated CLPM estimates (thus based on estimates that come from multiple studies). The study of Kuiper was meant as a follow-up to this study and also refers to this on page 961. In our research, we examine the performance of the GORICA in various conditions (for a single study) with more elaborate simulation studies. Scott (2021) conducted a simulation study focusing on the impact of disregarding time-varying inter-individual differences on cross-lagged parameter estimates, revealing diverse patterns of bias in path estimates and standard errors. Notably, causal dominance, in conjunction with time-varying between-person variance and covariance, was found to exert significant influence on bias in path estimates. While Scott also explored causal dominance in cross-lagged models, his study had a different focus compared to our research. Our study focuses on assessing the capability of GORICA to evaluate the causal dominance hypothesis within the context of (RI-)CLPM, while Scott utilized Bayes Factor to select the best statistical model among different types of cross-lagged models. One could also have used the Bayes Factor to evaluate causal dominance. Both the Bayes Factor and GORICA showcase comparable performance, as evidenced in the supplementary material (pages 24–25). The study of Scott underscores the crucial role of using an accurate type of cross-lagged model. Moreover, our method proves valuable for assessing causal dominance within (RI-)CLPMs. To repeat, our study introduces an approach for evaluating causal dominance in lagged effects models (more specifically, in (RI-)CLPMs) and demonstrates the performance of the GORICA using a simulation study.

We based our study on the research conducted by Masselink et al. (2018), which stated a stronger cross-lagged effect of self-esteem on depressive symptoms compared to cross-lagged effect of depressive symptoms on self-esteem in a wave-specific parameters model. Our application of the GORICA to Masselink et al. (2018) data yielded no support for causal dominance in a wave-independent parameters model. Moreover, we discovered evidence suggesting equal standardized cross-lagged effects. In a wave-specific parameters model, we found that the standardized cross-lagged effect of self-esteem on depressive symptoms has more predictive strength than the standardized cross-lagged effect of depressive symptoms on self-esteem; although, one could doubt whether the support is compelling.

In our simulations, we solely inspected bivariate and trivariate (RI-)CLPMs, because most (RI-)CLPMs include two

or three variables, since we believe these are most often used. For example, Zhou et al. (2020) studied the association between problematic internet use and mental health issues among Chinese college students and Van der Schuur et al. (2019) studied the relationship between social media use, social media stress, and sleep. Although, we did not investigate the performance of the GORICA under more-variate models, we expect – based on the performance of the tri-variate model compared to the bivariate model – that the power to detect the true hypothesis to increase, when comparing pairs of standardized cross-lagged parameters.

In summary, the GORICA can be used to evaluate causal dominance in bi- and tri-variate (RI-)CLPMs. Even though the current study aimed at investigating causal dominance, a statistical method by itself is insufficient to make causal claims. Namely, only when the model includes all relevant variables, and alternative explanations can be excluded, the results can be interpreted causally. Nevertheless, evaluating hypotheses concerning the predictive strength of (cross-)lagged parameters is extremely insightful. To fully benefit from what the GORICA has to offer, we recommend a number of participants of at least 100 with at least three waves of data collection.

We have demonstrated that the GORICA opens doors for longitudinal data analysis and to facilitate the use of the GORICA, the *goric* function (Vanbrabant & Kuiper, 2020) in the *restriktor* package (Vanbrabant & Rosseel, 2020) is a user-friendly method to calculate the GORICA values and weights. Moreover, other types of hypothesis can be evaluated as well. To facilitate readers in gaining experience with the GORICA, we have provided R code that can be accessed at <https://github.com/Chuenjai/Causal-dominance>, to apply the GORICA to longitudinal data, more specifically, the code to run the two types of models we ran for the Masselink et al. data in the Example section.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike* (pp. 199–213). Springer.
- Allen, M. (2017). *The sage encyclopedia of communication research methods*. SAGE.
- Allison, P. (2009). *Fixed effects regression models*. SAGE.
- Altınışık, Y. (2018). *Evaluation of inequality constrained hypotheses using an Akaike-type information criterion* [PhD dissertation]. Utrecht University.
- Amtmann, D., Bamer, A., Askew, R., & Jensen, M. (2020). Cross-lagged longitudinal analysis of pain intensity and sleep disturbance. *Disability and Health Journal*, 13, 100908. <https://doi.org/10.1016/j.dhjo.2020.100908>
- Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika*, 86, 141–152. <https://doi.org/10.1093/biomet/86.1.141>
- Antonio, Narzo, F., Aznarte, J., Stigler, M. (2009). *tsDyn: Time series analysis based on dynamical systems theory* [computer software manual]. <https://cran.r-project.org/web/packages/tsDyn/vignettes/tsDyn.pdf> (R package version 0.7)
- Berry, D., & Willoughby, M. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, 88, 1186–1206. <https://doi.org/10.1111/cdev.12660>
- Biesanz, J. 2012. Autoregressive longitudinal models. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 459–471). Guilford Press.
- Bollen, K., & Noble, M. (2011). Structural equation models and the quantification of behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 15639–15646. <https://doi.org/10.1073/pnas.1010661108>
- Burnham, K., & Anderson, D. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer-Verlag.
- Gu, X., Hoijtink, H., Mulder, J., & Rosseel, Y. (2019). Bain: A program for Bayesian testing of order constrained hypotheses in structural equation models. *Journal of Statistical Computation and Simulation*, 89, 1526–1553. <https://doi.org/10.1080/00949655.2019.1590574>
- Gu, X., Mulder, J., & Hoijtink, H. (2018). Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. *The British Journal of Mathematical and Statistical Psychology*, 71, 229–261. <https://doi.org/10.1111/bmsp.12110>
- Hamaker, E. (2012). Why researchers should think “within-person”: A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). The Guilford Press.
- Hamaker, E., Kuiper, R., & Grasman, R. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20, 102–116. <https://doi.org/10.1037/a0038889>
- Hamilton, J. (1994). *Time series analysis*. Princeton University Press.
- Hoijtink, H. (2011). *Informative hypotheses: Theory and practice for behavioral and social scientists*. Chapman and Hall/CRC.
- Hoijtink, H., Mulder, J., Van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, 24, 539–556. <https://doi.org/10.1037/met0000201>
- Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kuiper, R. (2021a). ChiBarSq.DiffTest: Chi-bar-square difference test of the RI-CLPM versus the CLPM and more general. [Computer software manual]. (R package version 0.0.0.9000).
- Kuiper, R. M. (2021b). Evaluating causal dominance of CTmeta-analyzed lagged regression estimates. *Structural Equation Modeling*, 28, 951–963. <https://doi.org/10.1080/10705511.2020.1823228>
- Kuiper, R., Hoijtink, H., & Silvapulle, M. (2011). An Akaike-type information criterion for model selection under inequality constraints. *Biometrika*, 98, 495–501. <https://doi.org/10.1093/biomet/asr002>
- Kuiper, R., Hoijtink, H., & Silvapulle, M. (2012). Generalization of the order-restricted information criterion for multivariate normal linear models. *Journal of Statistical Planning and Inference*, 142, 2454–2463. <https://doi.org/10.1016/j.jspi.2012.03.007>
- Kuppens, P., Allen, N., & Sheeber, L. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, 21, 984–991. <https://doi.org/10.1177/0956797610372634>
- Masselink, M., Van Roekel, E., Hankin, B. L., Keijsers, L., Lodder, G. M. A., Vanhalst, J., Verhagen, M., Young, J. F., & Oldehinkel, A. J. (2018). The longitudinal association between self-esteem and depressive symptoms in adolescents: Separating between-person effects from within-person effects. *European Journal of Personality*, 32, 653–671. <https://doi.org/10.1002/per.2179>
- Masselink, M., Van Roekel, E., & Oldehinkel, A. (2018). Self-esteem in early adolescence as predictor of depressive symptoms in late adolescence and early adulthood: The mediating role of motivational

- and social factors. *Journal of Youth and Adolescence*, 47, 932–946. <https://doi.org/10.1007/s10964-017-0727-z>
- Mlawer, F., Hubbard, J. A., Bookhout, M. K., & Moore, C. C. (2021). Levels and instability of daily self-esteem in adolescents: Relations to depressive and anxious symptoms. *Research on Child and Adolescent Psychopathology*, 49, 1083–1095. <https://doi.org/10.1007/s10802-021-00802-3>
- Mund, M., & Nestler, S. (2019). Beyond the cross-lagged panel model: Next-generation tools for analyzing interdependencies across the life course. *Advances in Life Course Research*, 41, 100249. <https://doi.org/10.1016/j.alcr.2018.10.002>
- Orth, U., Robins, R. W., Trzesniewski, K. H., Maes, J., & Schmitt, M. (2009). Low self-esteem is a risk factor for depressive symptoms from young adulthood to old age. *Journal of Abnormal Psychology*, 118, 472–478. <https://doi.org/10.1037/a0015922>
- R Core Team. 2020. R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. <https://www.R-project.org/>.
- Rosseel, Y. (2012). Llavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. <http://www.jstatsoft.org/v48/i02/> <https://doi.org/10.18637/jss.v048.i02>
- Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M., & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multi-level autoregressive model. *Psychological Methods*, 21, 206–221. <https://doi.org/10.1037/met0000062>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464. <https://doi.org/10.1214/aos/1176344136>
- Scott, P. W. (2021). Accounting for time-varying inter-individual differences in trajectories when assessing cross-lagged models. *Structural Equation Modeling*, 28, 365–375. <https://doi.org/10.1080/10705511.2020.1819815>
- Spiegelhalter, D., Best, N., Carlin, B., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64, 583–639. <https://doi.org/10.1111/1467-9868.00353>
- Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014). Low and decreasing self-esteem during adolescence predict adult depression two decades later. *Journal of Personality and Social Psychology*, 106, 325–338. <https://doi.org/10.1037/a0035133>
- Stoel, R., Garre, F., Dolan, C., & Van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*, 11, 439–455. <https://doi.org/10.1037/1082-989X.11.4.439>
- Suls, J., Green, P., & Hillis, S. (1998). Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Personality and Social Psychology Bulletin*, 24, 127–136. <https://doi.org/10.1177/0146167298242002>
- Usami, S., Murayama, K., & Hamaker, E. A. (2019). Unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, 24, 637–657. <https://doi.org/10.1037/met0000210>
- Van der Schuur, W. A., Sumter, S. R., & Baumgartner, S. E. (2019). Social media use, social media stress, and sleep: Examining cross-sectional and longitudinal relationships in adolescents. *Health Communication*, 34, 552–559. <https://doi.org/10.1080/10410236.2017.1422101>
- Vanbrabant, L., & Kuiper, R. (2020). *Goric function in R package restriktor*. <https://CRAN.R-project.org/package=restriktor> (R package version 0.2-800)
- Vanbrabant, L., & Rosseel, Y. (2020). *Restriktor: Restricted statistical estimation and inference for linear models*. <https://CRAN.R-project.org/package=restriktor> (R package version 0.2-800)
- Wagenmakers, E., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11, 192–196. <https://doi.org/10.3758/BF03206482>
- Zhou, N., Cao, H., Liu, F., Wu, L., Liang, Y., Xu, J., Meng, H., Zang, N., Hao, R., An, Y., Ma, S., Fang, X., & Zhang, J. (2020). A four-wave, cross-lagged model of problematic internet use and mental health among Chinese college students: Disaggregation of within-person and between-person effects. *Developmental Psychology*, 56, 1009–1021. <https://doi.org/10.1037/dev0000907>