Routledge
Taylor & Francis Group

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

Evaluating Causal Dominance of CTmeta-Analyzed Lagged Regression Estimates

Rebecca Kuiper

To cite this article: Rebecca Kuiper (2021) Evaluating Causal Dominance of CTmeta-Analyzed Lagged Regression Estimates, Structural Equation Modeling: A Multidisciplinary Journal, 28:6, 951-963, DOI: 10.1080/10705511.2020.1823228

To link to this article: https://doi.org/10.1080/10705511.2020.1823228

Published online: 18 Jun 2021.

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
2021, VOL. 28, NO. 6, 951–963
https://doi.org/10.1080/10705511.2020.1823228

Routledge
Taylor & Francis Group

TEACHER'S CORNER

OPEN ACCESS Check for updates

Evaluating Causal Dominance of CTmeta-Analyzed Lagged Regression Estimates

Rebecca Kuiper

Universiteit Utrecht

ABSTRACT

Meta-analysis techniques allow researchers to aggregate effect sizes, like standardized regression estimates, of different studies. Recently, continuous-time meta-analysis (CTmeta) has been developed such that the time-interval dependent lagged-parameter estimates can be properly meta-analyzed. This leads to overall standardized lagged-parameter estimates and their multivariate confidence interval. Often, researchers are not only interested in these overall estimates but also in a specific ordering of them: Many researchers have an a priori expectation regarding the ordering of the predictive strength of the cross-lagged relationships; referred to as causal dominance. For example, a researcher might expect, based on literature or expertise, that the lagged relationship between burnout and work engagement is weaker than the reciprocal lagged relationship. Such a hypothesis can be evaluated with an AIC-type theory-based model selection criterion: GORICA. This paper introduces and illustrates how the GORICA can be applied to CTmeta-analyzed standardized lagged-parameter estimates and demonstrate its performance.

KEYWORDS

Meta-analysis; first-order vector autoregressive (VAR (1)) model; cross-lagged panel model (CLPM); model selection

Introduction

Meta-analysis aims to aggregate effect-size estimates, like standardized regression parameters, from multiple studies to come to an overall estimate of one or more population parameters (cf. Becker & Wu, 2007; Borenstein et al., 2009). In principle, it takes a *weighted* average of the effect-size estimates, where the contribution of each study is weighted by the standard error of the estimate (or the covariance matrix of the estimates), that is, the amount of information or certainty in the estimate(s).

Lagged-effects models, such as the first-order vector-autoregressive (VAR(1); Hamilton, 1994) and cross-lagged panel models (CLPMS; Bollen & Curran, 2006; Mayer, 1986) are increasingly the target for meta-analysis (for instance, Jacobson & Newman, 2017; Maricuțoiu et al., 2002; Masselink et al., 2018; Nohe et al., 2015). It is well-known that the lagged parameters are time-interval dependent (Gollob & Reichardt, 1987; Kuiper & Ryan, 2018; Pelz & Lew, 1970; Voelkle & Oud, 2013): Studies that use different uniform time intervals between observations (e.g., 1 hour vs 3 hours or 1 month vs 2 months) can come to very different parameter estimates, and seemingly contradictory conclusions, about the same underlying process. Averaging these incomparable estimates would lead to not-interpretable overall estimates. This problem can be overcome by using the recently developed continuous-time meta-analysis (CTmeta; Kuiper & Ryan, 2020), which assumes an underlying continuous-time process. Like the classical meta-analysis, CTmeta renders overall (standardized) parameter estimates. Nevertheless, this is often not sufficient information to answer the researcher's research question.

Typically, researchers use lagged-effects models to assess the Granger-causal relationships between pairs of variables, through the estimation of cross-lagged regression parameters (Granger, 1969; Rogosa, 1979, 1980). Generally, their primary aim is to compare the size and sign of the estimated cross-lagged parameters (ϕ_{jk} vs ϕ_{kj}). This relative strength is often referred to as "causal dominance" (Bentler & Speckart, 1981; Finkel, 1995; Hamaker et al., 2015; Rogosa, 1980). For instance, Moberly and Watkins (2008) examine which of momentary ruminative self-focus (RSF) and negative affect (NA) can be considered the 'driving force' of the pair, by comparing the size of the cross-lagged relationships of RSF and NA on each other at the next measurement occasion. The meta-analysis of Nohe et al. (2015) investigates the 'causal dominance' relationships between work-family conflict and strain by comparing the size of the overall cross-lagged parameter estimates. Their hypotheses of interest can be expressed as: $H_1 : \phi_{12} < \phi_{21}$ versus $H_2 : \phi_{12} > \phi_{21}$. Such hypotheses are often referred to as informative hypotheses (Hoijtink, 2012), inequality-constrained hypotheses, order-restricted hypotheses, or theory-based hypotheses.

While the overall (standardized) parameter estimates often do not answer the 'causal dominance' questions, as presented by H_1 and H_2, it possible to evaluate these hypotheses by using an AIC-type theory-based model selection criterion: GORICA (Altınışık et al., 2021; Kuiper et al., 2012, 2011). This paper demonstrates how this is done and what the advantages are. Furthermore, this paper will give insight into the performance of theory-based model selection using the GORICA on CTmeta-analyzed estimates.

Background

First, the core concepts of lagged-effects models, CTmeta-analysis, and model selection using the GORICA are briefly introduced.

CONTACT Rebecca Kuiper r.m.kuiper@uu.nl Methods and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, P.O. Box 80140, Netherlands.

Second, the remainder of this paper focuses on illustrating how causal dominance hypotheses can be evaluated and on demonstrating its performance by a small simulation study.

The discrete-time VAR(1) model

A popular choice of model to analyze multiple repeated measurements are (first-order) lagged-effects models, where lagged parameters describe the relationship between current observations and past observations. In the context of panel data (i.e., data for a large number of participants but with relatively few observations measured far apart in time), this is referred to as the cross-lagged panel model (CLPM). In the context of time series data (i.e., single-subjects data with many observations measured at a higher frequency), this is referred to as the (discrete-time) first-order vector autoregressive (DT-VAR(1)) model. Since both models are conceptually very similar, the DT-VAR(1) terminology will be used throughout. Figure 1 depicts a bivariate DT-VAR(1) model as a path model.

Let $y_{i,m}$ be the vector with q observed variables for individual i ($i = 1, \ldots, N$) at measurement occasion m, which is regressed on the preceding observation through

$$\boldsymbol{y}_{i,m} = c_i + \boldsymbol{\Phi}\boldsymbol{y}_{i,m-1} + \epsilon_{i,m}$$

where c_i is a q-vector of intercepts which is related to the person-specific mean of $y_{i,m}$ by $\boldsymbol{\mu}_i = (I - \boldsymbol{\Phi})^{-1}c_i$; $\epsilon_{i,m}$ represents a q-vector of errors for measurement m which are independent and identically distributed: $\epsilon_{i,m} \sim (\mathbf{0}, \Sigma_\epsilon)$; and $\boldsymbol{\Phi}$ is the $q \times q$ matrix of lagged regression parameters, that is, autoregressive (ϕ_{jj}) and cross-lagged ($\phi_{jk}, j \neq k$) parameters. The next subsection briefly discusses the time-interval dependency of $\boldsymbol{\Phi}$ and how its elements should be meta-analyzed using CTmeta.

Time-interval dependency and CTmeta

It is well known that autoregressive and cross-lagged parameter estimates are a function of the time interval (Chatfield, 2004; Dormann & Griffin, 2015; Gollob & Reichardt, 1987; Hamilton, 1994; Kuiper & Ryan, 2018; Oud, 2002). This refers to the phenomenon that estimates differ in sign, size, and/or relative ordering only due to the use of different spacing between measurements. Therefore, the lagged parameters will be denoted by $\boldsymbol{\Phi}(\Delta t)$ in the remainder. To illustrate the time-interval dependency, let us assume that measurements are taken every, say, 2 hours instead of every 1 hour, then a model is fitted on every second measurement wave in Figure 1. In that case, the estimate of $\boldsymbol{\Phi}(2) = \boldsymbol{\Phi}(1)^2$ is obtained instead of the estimate of $\boldsymbol{\Phi}(1)$. Note that this is a matrix exponential and, therefore, the elements in $\boldsymbol{\Phi}(2)$ are a sum of multiple different products of the parameters in the original matrix $\boldsymbol{\Phi}(1)$. For example, element (1,1) in $\boldsymbol{\Phi}(2)$, that is, the autoregressive parameter $\phi_{11}(2)$, does not equal the square of $\phi_{11}(1)$ but equates $\phi_{11}(1)\,\phi_{21}(1) + \phi_{21}(1)\,\phi_{22}(1)$. As a numerical example, take the lagged-parameter matrix

$$\boldsymbol{\Phi}(1) = \begin{bmatrix} 0.50 & 0.15 \\ 0.25 & 0.40 \end{bmatrix}.$$

When a twice as large time interval is used, an estimate of the following lagged-parameter matrix would be found:

$$\boldsymbol{\Phi}(2) = \begin{bmatrix} 0.29 & 0.14 \\ 0.23 & 0.20 \end{bmatrix}.$$

This shows the implications for meta-analysis: Taking a weighted average of estimates based on different time intervals results in a set of overall estimates which do not accurately reflect the true underlying process for any time interval. Furthermore, including the time interval as linear and/or quadratic moderator of the effect (cf. Card, 2019) fails to capture the exponential relationship (Kuiper & Ryan, 2020). Alternatively, when doing separate meta-analyses per time interval used in the primary studies (i.e, using dummy variables), an overall estimate per time interval is obtained. Moreover, the power is reduced because not all primary studies are used, only the ones that use that time interval. To overcome these problems, continuous-time meta-analysis (CTmeta) was developed, which is based on the following reasoning. Even though the elements in

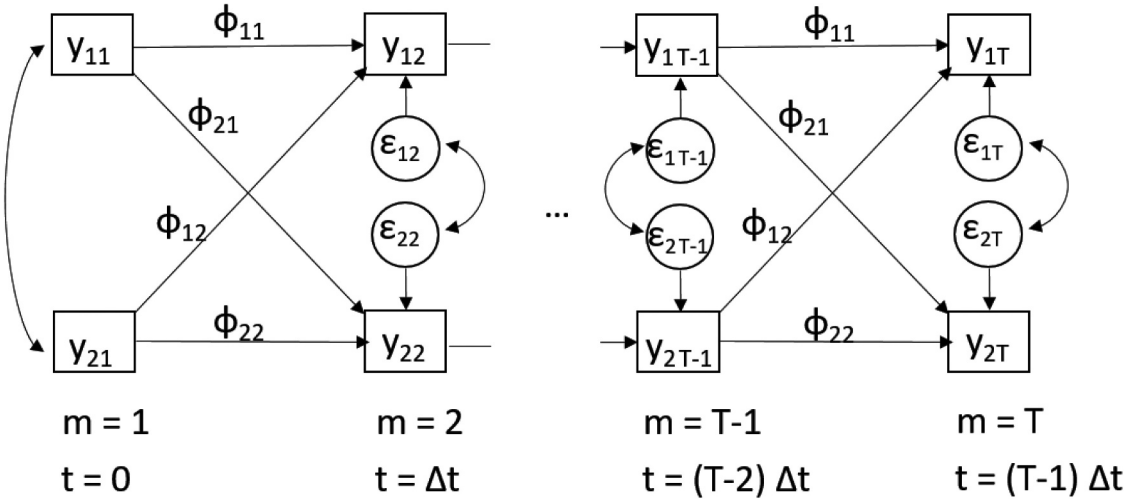


Figure 1. A graphical representation of a bivariate VAR(1) model.

$\boldsymbol{\Phi}(1)$ and $\boldsymbol{\Phi}(2)$ differ in size, they do describe the same process. Thus, results from studies using different time intervals are informative about one another, but need to be made comparable. Assuming that the underlying process is continuous in time (cf. Boker & Laurenceau, 2006; Oravecz et al., 2009, 2011; Oud & Delsing, 2010a; Voelkle & Oud, 2013), the results from studies using different time intervals (e.g., $\hat{\boldsymbol{\Phi}}(0.5)$, $\hat{\boldsymbol{\Phi}}(1)$, and $\hat{\boldsymbol{\Phi}}(1.7)$) can all be mapped back to one underlying effects matrix A, called the drift matrix, using

$$\boldsymbol{\Phi}(\Delta t) = e^{\boldsymbol{A}\Delta t}.$$

An elaborate discussion regarding the interpretation of the drift matrix can be found in Ryan et al. (2018). In the numerical example, the underlying drift matrix equals

$$\boldsymbol{A} = \begin{bmatrix} -0.79 & 0.36 \\ 0.60 & -1.03 \end{bmatrix}.$$

The underlying drift matrix $\boldsymbol{A}$ can be used to transform the (standardized) lagged relationships such that they are comparable:

$$\begin{aligned} \boldsymbol{\Phi}(\Delta t^*) &= (e^{\boldsymbol{A}\Delta t})^{\frac{\Delta t^*}{\Delta t}} \\ &= (\boldsymbol{\Phi}(\Delta t))^{\frac{\Delta t^*}{\Delta t}}. \end{aligned} \quad (3)$$

Using this formula, the lagged-parameter matrix estimates from studies using different time intervals, for instance, $\hat{\boldsymbol{\Phi}}(0.5)$, $\hat{\boldsymbol{\Phi}}(1)$, and $\hat{\boldsymbol{\Phi}}(1.7)$, can all be transformed to the estimate for a targeted time interval Δt^*, say, $\hat{\boldsymbol{\Phi}}(1)$. Since these transformed (standardized) estimates are comparable, these can be meta-analyzed, which is what CTmeta does.

Thus, CTmeta averages the transformed standardized lagged relationships, obtained using Equation 3, for one or more targeted time intervals Δt^* (where weighting is done based on the covariance matrices of the study-specific transformed standardized lagged relationships). The CTmeta method is implemented in i) the R-package *CTmeta* (Kuiper, 2020c), available from my GitHub page (https://github.com/rebeccakuiper/CTmeta) as demonstrated in the section "GORICA on CTmeta in R" below, and ii) an interactive (Shiny) web application (Kuiper, 2020b). CTmeta renders the overall standardized lagged-parameter estimates and their multivariate confidence intervals. The latter is based on the covariance matrix of the overall estimates instead of their single variances. By using the covariances as well, the multivariate structure is accounted for. These multivariate confidence intervals are sometimes refers to as elliptical confidence intervals.

The next subsection discusses how the GORICA can evaluate researchers' (causal dominance) theories regarding the overall CTmeta estimates.

GORICA

An information criterion selects the hypothesis that describes the data best (highest fit) with the smallest (least complex) hypothesis in terms of number of distinct parameters, out of a set of candidate hypotheses. An often used information criterion is the Akaike information criterion (AIC; Akaike, 1973):

$$AIC = -2\,\{\text{maximum log likelihood} - \text{penalty}\},$$

where the penalty equals the number of distinct model parameters: e.g., the number of distinct regression parameters, including the intercept, and the distinct error (co)variance(s). The formula shows the trade-off between the fit (likelihood) and the complexity (penalty) of the candidate hypotheses. The AIC is an estimate of the Kullback–Leibler (KL) discrepancy (Kullback & Leibler, 1951), the distance between a candidate hypothesis and the true unknown hypothesis. Therefore, the hypothesis with the smallest AIC value is the preferred one in the set of candidate hypotheses. The AIC can evaluate hypotheses where (some) parameters are set equal to zero or equal to each other; e.g., $\phi_{12} = \phi_{21}, \phi_{13} = \phi_{31}$.

By using the generalized order-restricted information criterion (GORIC; Kuiper et al., 2012, 2011) or its approximation (GORICA; Altınışık et al., shed), researchers' theories, like causal dominance theories, can directly be examined by evaluating theory-based hypotheses, like $\phi_{12} > \phi_{21}$, $\phi_{13} > \phi_{31}$. Thus, the GORIC and GORICA can evaluate theory-based hypotheses containing order restrictions on the parameters ("<" and/or ">") besides equality restrictions ("="). The GORIC is, like the AIC, an estimate of the KL discrepancy and is of the form

$$\begin{aligned} GORIC = -2\,\{&\text{maximum order-restricted log likelihood} \\ &- \text{penalty}\}. \end{aligned}$$

In comparison with the AIC, this expression is based on the order-restricted maximum likelihood (i.e., the maximum likelihood under the order restrictions in the hypothesis) and has a corrected penalty (using so-called chi-bar-square weights) such that the order restrictions are properly accounted for. The latter comes loosely speaking down to deriving the expected number of distinct parameters. For example, $\phi_{12} < \phi_{21}$ represents 1.5 distinct regression parameters and not 2, as would be the case in the AIC. If there are no order restrictions (i.e., only equality constraints ("=") and/or no constraints (",")), the GORIC reduces to the AIC. To ease the calculation of the GORIC for a broad range of models, the GORICA was derived using the fact that maximum likelihood estimates (mle's) are asymptotically normally distributed:

$$\begin{aligned} GORIC = -2\,\{&\text{maximum order-restricted log likelihood} \\ &\text{mle}'\text{s} - \text{penalty}\}. \end{aligned}$$

Here, the fit part is based on the mle's, which are a summary for the data, instead of the data themselves, which is used in the GORIC. Furthermore, the fit part of the GORICA is always based on the normal distribution even if the data do not follow one (like in a logistic regression). The fit values of the GORIC and GORICA differ in absolute sense but asymptotically not in relative sense when comparing candidate hypotheses. The penalty of the GORICA equates that of the GORIC.

Like AIC values, GORICA values denote the ordering of the candidate hypotheses and not their relative strength, but the following transformation does:

$$w_i = \frac{exp\left(-\frac{1}{2}\ GORICA_i\right)}{\sum_{m=1}^{M} exp\left(-\frac{1}{2}\ GORICA_m\right)}$$

for $i = 1, \ldots, M$, with M the total number of hypotheses in the set. These are called GORICA weights and reflect the relative likelihood of a hypothesis given the data and the set of hypotheses (Akaike, 1978; Burnham & Anderson, 2002; Kuiper et al., 2012; Wagenmakers & Farrell, 2004). For instance, GORICA weights for Hypothesis H_1 and a competing hypothesis H_2 of $w_1 = 0.875$ and $w_2 = 0.125$ mean that H_1 has $w_1/w_2 = 7$ times more support than the competing hypothesis H_2. Notably, the GORICA weights asymptotically equate the GORIC weights, which equal Akaike weights (i.e., AIC weights) in case there are no order restrictions.

The set of hypotheses of interest should consist of at least two hypotheses. When there are multiple theories, these can thus be included as competing hypotheses. Let us assume that the literature states two competing hypotheses: $\phi_{12} > \phi_{21}$, $\phi_{13} > \phi_{31}$ and $\phi_{12} < \phi_{21}$, $\phi_{13} < \phi_{31}$. Since these hypotheses do not cover all possible theories (e.g., $\phi_{12} < \phi_{21}$, $\phi_{13} > \phi_{31}$ is not included), GORICA selects the best out of a set of weak hypotheses when both hypotheses are weak. Therefore, in case the hypotheses do not cover the whole parameter space (as is also the case when there is only one theory of interest), a safeguard hypothesis should be included (Kuiper et al., 2012). There are two possibilities: the unconstrained hypothesis H_u, where none of the parameters are restricted and represents all possible theories, and the complement of the hypothesis/-es of interest, representing al other theories. The unconstrained hypothesis should be used to investigate whether the hypotheses of interest are weak or not. When at least one is not (e.g., $w_1 > w_u$, that is, $w_1/w_u > 1$), the relative support for the hypotheses of interest (e.g., w_1/w_2) can be inspected. Using the complement can be more powerful (Vanbrabant et al., 2020) and acts like another hypothesis of interest, but is in software only available for one theory-based hypothesis.

It is important to note that comparing parameters (e.g., $\phi_{12} < \phi_{21}$) is only meaningful if these parameters are measured on the same scale. Hence, (overall) standardized lagged parameters should be used; which should also be the target in CTmeta such that comparable estimates are averaged.

There are two R functions that can calculate GORICA values and weights: the gorica function in the gorica package (Kuiper et al., 2020) and the goric function (Vanbrabant & Kuiper, 2020) in the restriktor package (Vanbrabant & Rosseel, 2020). There are some differences in functionality (Kuiper, 2020d), but both functions render the same results of course. The goric function of the restriktor package is used in this paper.

The following sections demonstrate how the GORICA can be applied to CTmeta-analyzed lagged-parameter estimates and give insight into its advantages and performance. The paper concludes with a discussion.

Illustration

This section illustrates the application of GORICA to CTmeta. The illustration consists of simulated data which mimics the set-up of a published empirical meta-analysis, which was also used in the CTmeta article Kuiper and Ryan (2020). The code to reproduce all of the analyses shown below is available on my GitHub page (https://github.com/rebeccakuiper/CTmeta).

Dataset

To evaluate the use of the GORICA on CTmeta-analyzed estimates, the working example of Kuiper and Ryan (2020) is used, which is based on the design of Maricuțoiu et al., 2002. In the working example, there are 25 primary panel data studies which all study the cross-lagged relationships between work engagement and burnout (i.e., $q = 2$). These two variables are assumed to have a contemporaneous pairwise correlation of $cor(y_{1m}, y_{2m}) = 0.3$. The underlying dynamic relationships between these two variables is described by the one in Equation 2. Based on this underlying drift matrix, Figure 2 depicts the elements in $\mathbf{\Phi}(\Delta t)$ for a range of time intervals between 0 and 5 years, obtained by the Shiny web application Kuiper (2020a). In the 25 studies, the range of study-specific time intervals between measurement occasions is $\Delta t_s \in \left(\frac{1}{365}, 3\right)$ years $(s = 1, \ldots, 25)$, and the range of study-specific sample sizes is $T_s \in (67, 2897)$; the study-specific values can be found in Table 1.

For each primary study (i.e., for Study s), T_s measurements were generated from a VAR(1) model with standardized lagged parameters $\mathbf{\Phi}(\Delta t_s)$, based on the drift matrix in Equation 2, using the R-package *tsDyn* (Fabio Di Narzo et al., 2009); see Kuiper and Ryan (2020) for more details. Then, a DT-VAR(1) model is fitted to each dataset using the *vars* package (Pfaff, 2008). The standardized parameter estimates from each simulated dataset serve as input to CTmeta. This results in overall standardized lagged-parameter estimates and their covariance matrix, which is input for the GORICA.

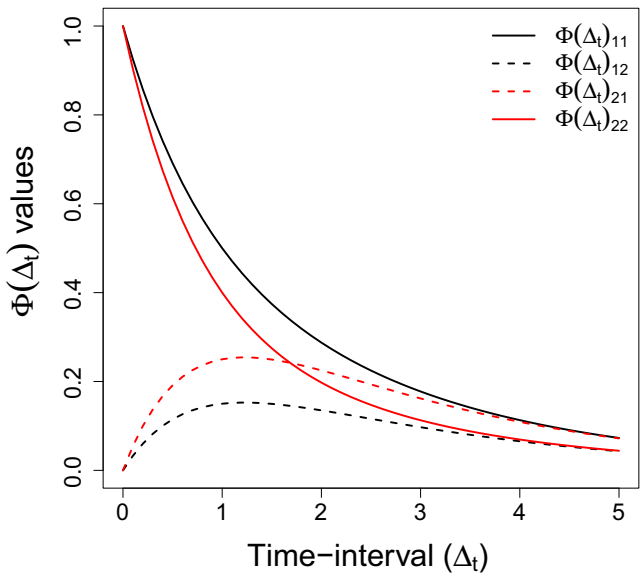


Figure 2. The bivariate lagged parameters $\phi(\Delta t)$ as a function of the time interval Δt.

Table 1. The study-specific sample sizes T_s and study-specific time intervals Δt_s (in years) used in the working example.

Study (s)	T_s	Reported Time Interval	Δt_s
1	643	12 months	1
2	651	12 months	1
3	473	12 months	1
4	387	4 months, 2.5 months and 6.5 months	$\frac{1}{3}$
5	187	9 months	$\frac{3}{4}$
6	209	12 months	1
7	2897	12 months	1
8	160	2 months	$\frac{1}{6}$
9	1964	36 months, 48 months	3
10	848	12 months	1
11	926	12 months	1
12	274	8 months	$\frac{2}{3}$
13	433	24 months	2
14	256	1 day	$\frac{1}{365}$
15	409	24 months	2
16	926	12 months	1
17	162	2 months	$\frac{1}{6}$
18	262	14 months	$\frac{14}{12}$
19	247	12 months	1
20	102	10 months	$\frac{10}{12}$
21	171	48 months	4
22	201	12 months	1
23	309	12 months	1
24	77	1 month	$\frac{1}{12}$
25	67	8 months	$\frac{2}{3}$

The theory-based causal-dominance hypothesis to be evaluated with the GORICA is one stating that the lagged relationship between work engagement and burnout is weaker that the reciprocal lagged relationship (H_1), which is evaluated versus its complement (H_c):

$$H_1 : \phi_{12} < \phi_{21}$$

$$H_c : \phi_{12} > \phi_{21}.$$

CTmeta

When applying CTmeta for a targeted interval, the resulting output gives, among other things, the overall lagged estimates (called "Overall_standPhi_DeltaTStar" in the R output) and their multivariate confidence intervals ("elliptical_CI"). For the working example, the resulting overall CTmeta estimates for a targeted interval of $\Delta t^* = 1$ year and their multivariate confidence intervals are displayed in Table 2. This process can be repeated for other targeted time intervals of interest as well.

Table 2. The CTmeta results of the working example: The resulting overall estimates of the lagged effects parameters for a targeted time interval of 1 year (i.e., $\phi_{11}(1)$, $\phi_{12}(1)$, $\phi_{21}(1)$, and $\phi_{22}(1)$) and the lower and upper bound of their multivariate confidence intervals.

	$\phi_{11}(1)$	$\phi_{12}(1)$	$\phi_{21}(1)$	$\phi_{22}(1)$
Overall estimate	0.491	0.148	0.249	0.399
Lower bound	0.477	0.133	0.233	0.384
Upper bound	0.506	0.162	0.264	0.415

GORICA on CTmeta

To evaluate H_1, the GORICA should be applied on the CTmeta-analyzed estimates. Here, this is done for all the time-intervals used in the 25 primary studies, assuming that all these intervals are of interest. In practice, this should be done for the time-interval(s) which are of interest (which are preferably within the range of the ones used in the primary studies). The code to replicate this analysis can be found at my GitHub page (https://github.com/rebeccakuiper/CTmeta). The section "GORICA on CTmeta in R" below also gives some example R code to do this for one targeted time-interval based on a subset of the example data (namely 3 out of the 25 studies) using the CTmeta package and restriktor package in R.

The resulting GORICA weights for the causal dominance hypothesis H_1 are given in the second column of Table 3. These are the GORICA weights for H_1 based on the overall standardized lagged-parameter estimates that resulted from CTmeta on the 25 primary studies, where H_1 is evaluated against its complement. From this column, it can be concluded that there is maximum support for H_1 (i.e., GORICA weights of 1) for almost all targeted time-intervals. For the smallest time-interval (one day), the support for H_1 is a bit higher but about equal to the support for its complement stating the opposite ordering of cross-lagged parameters. In that case, H_1 is only $0.584/(1-0.584) \approx 1.4$ times more supported than its complement. Thus, for that time interval, there is no compelling evidence for (or against) H_1. This makes sense since the effect of work engagement on burnout will not be substantial after one day. Similarly, the support for H_1 will decrease to 0.5 for higher targeted time intervals. Namely, the cross-lagged parameters will both go to 0, as can be seen from Figure 2, and, thus, both H_1 and H_c will be equally supported (since their border is the truth and because they are equally complex). This can also be seen from the GORICA weights for H_1, which decrease after $\Delta t^* = 36/12 = 3$ years, but are still very high in this case.

Since support is found for H_1 for all time-intervals and, for most, even compelling support, this increases the confidence in

Table 3. GORICA weights for H_1 (versus its complement H_c), for all 12 unique time intervals considered in the meta-analysis. In the second column, GORICA weights for the causal dominance hypothesis H_1 regarding the overall estimates. In columns 3 to 6, some descriptive statistics (minimum, maximum, mean, and standard deviation) for the study-specific GORICA weights for H_1 regarding the study-specific lagged-parameter estimates.

		Per study			
Δt^*	Overall	min	max	mean	s.d.
1/365	0.584	0.500	0.532	0.504	0.007
1/12	1.000	0.497	0.975	0.602	0.124
2/12	1.000	0.495	0.999	0.654	0.161
4/12	1.000	0.491	1.000	0.704	0.191
8/12	1.000	0.487	1.000	0.731	0.204
9/12	1.000	0.487	1.000	0.732	0.205
10/12	1.000	0.487	1.000	0.732	0.205
12/12	1.000	0.487	1.000	0.729	0.205
14/12	1.000	0.487	1.000	0.725	0.203
24/12	1.000	0.492	1.000	0.683	0.186
36/12	1.000	0.496	0.994	0.619	0.147
48/12	0.995	0.498	0.906	0.566	0.099

the finding that H_1 is the best hypothesis. Hence, the conclusion is that the lagged relationship of work engagement on burnout is weaker than its reciprocal relationship and that the support for this finding is prominent.

Advantages

This section describes two advantages of applying the GORICA to meta-analyzed lagged estimates. First, by using the CTmeta-analyzed estimates instead of the study-specific ones, the power to select the correct hypothesis is increased. Second, by using the GORICA instead of a classical criterion like the AIC or a confidence interval comparison approach, the hypothesis of interest is directly investigated.

Increased power to select correct hypothesis

The GORICA can also be applied to each primary study (per targeted time-interval). The 25 times 12 resulting GORICA weights for H_1 are plotted in Figure 3. Additionally, some descriptive statistics of the 25 study-specific GORICA weights for H_1 are depicted in columns 3 to 6 in Table 3, for each targeted time-interval. The plot and table show that not in all studies there was compelling evidence for H_1: There are studies indecisive (i.e., GORICA weights around 0.5) and, on average, there is support for H_1 but not overwhelming support: e.g., for $\Delta t^* = 1$, the support for H_1 is $0.729/(1-0.729) \approx 2.69$ times larger than for its complement.

While the evidence for H_1 is not convincing for each study, the evidence based on all 25 studies combined is (cf. column 2 in Table 3 and the black squares in Figure 3). Combining all the information from all the 25 studies does result in prominent (even maximum) support for H_1. This demonstrates the advantage of a meta-analysis and model selection on the resulting overall estimates: By using all information, the power to select the correct hypothesis increases. To obtain further insight in the behavior and performance of the GORICA weights in various CTmeta settings (e.g., regarding the size of the estimates), a small simulation study will be conducted in the next section. First, the preference of GORICA over AIC and a confidence interval comparison is demonstrated in the next subsection.

Direct evaluation of theory-based hypotheses

This subsection discusses two alternatives to the GORICA and demonstrates why the use of the GORICA is preferred. To investigate the hypothesis of interest H_1, the conservative method of inspecting the overlap of confidence intervals (CIs) can be used. Based on the eliptical/multivariate 95% CIs of the two cross-lagged estimates in Table 2, it is concluded that they significantly differ because the confidence intervals do not overlap.[1] Because of the size of the estimates, the conclusion is that $\phi_{12}(1)$ is smaller than $\phi_{21}(1)$, which is the ordering of interest. This, however, does not quantify the support for the

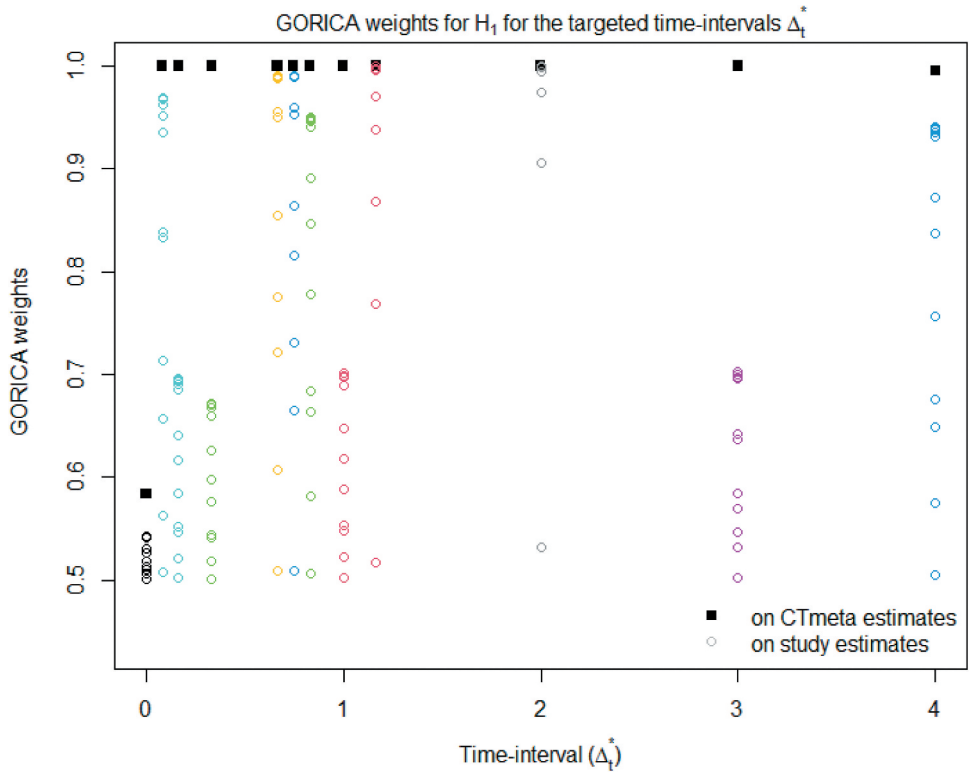


Figure 3. GORICA weights for H_1, for all 12 unique time intervals considered in the meta-analysis. The black squares are the ones based on the overall estimates; the colored rounds are the 25 times 12 study-specific GORICA weights for H_1.

[1]Note that the significance level is not 5% when comparing two confidence intervals. The confidence interval of the difference between the parameters should be inspected, which is not part of the output.

hypothesis of interest. Moreover, what if the CIs do overlap a bit and what if there are multiple pairs of cross-lagged relationships of interest and some of their CIs do overlap and others do not. To straightforwardly investigate the hypothesis of interest and quantify its support, model selection is needed.

Instead of the GORICA, the AIC can also be applied to the CTmeta results. In that case, the hypotheses of interest depicted in Equation 4 are represented by the following:

$$H_0 : \phi_{12} = \phi_{21}$$

$$H_u : \phi_{12},\ \phi_{21},$$

where H_0 is not of real interest but reflects 'no causal dominance' and H_u is the complement of H_0 and reflects both possibilities of causal dominance (i.e., H_1 and H_c in Equation 4). Since CTmeta does not render likelihood values, the AIC is not easily calculated. Fortunately, the GORICA weights asymptotically equal the AIC weights in case of no order restrictions. Therefore, the GORICA weights are calculated and denoted as AIC weights. The AIC weights for H_u are displayed in Table 4. When comparing this table to Table 3, it can be seen that the results per study differ (columns 3 to 6). The AIC weights have a lower minimum and mean and, therefore, a bit higher standard deviation. Nevertheless, the conclusion is the same: Some of the primary studies are underpowered and there is an advantage of taking studies together since the overall support for H_u (column 2) is very high. There is compelling overall support for H_u (second column), except for the smallest targeted time interval (one day). The difference between the AIC and GORICA is how these criteria address the theory-based hypothesis H_1. Using the AIC, it can be concluded that there is prominent support for unequal cross-lagged effects since there is overall support for H_u and not for H_0. Based on the sizes of the estimates, it can be concluded that this favors H_1. This is, however, not a quantification of the support for H_1. Furthermore, what if the interest lies in multiple pairs of cross-lagged parameters? Then, the evaluation is less straightforward and may even lead to inconclusive results, which is shown in Kuiper and Hoijtink (2010) who elaborately compare the AIC with the ORIC, the precursor of the GORIC and thus GORICA. In sum, like with using confidence intervals, the AIC does not offer a straightforward way to investigate the hypothesis of interest and quantify its support, while the GORICA does.

Table 4. AIC weights for H_u (versus H_0), for all 12 unique time intervals considered in the meta-analysis. In the second column, AIC weights for the causal dominance hypothesis H_u regarding the overall estimates. In columns 3 to 6, some descriptive statistics (minimum, maximum, mean, and standard deviation) for the study-specific AIC weights for H_u regarding the study-specific lagged-parameter estimates.

		Per study			
Δt^*	Overall	min	max	mean	s.d.
1/365	0.341	0.269	0.296	0.273	0.006
1/12	1.000	0.269	0.935	0.377	0.157
2/12	1.000	0.269	0.997	0.446	0.211
4/12	1.000	0.269	1.000	0.521	0.261
8/12	1.000	0.270	1.000	0.567	0.289
9/12	1.000	0.270	1.000	0.569	0.291
10/12	1.000	0.270	1.000	0.570	0.291
12/12	1.000	0.270	1.000	0.566	0.291
14/12	1.000	0.270	1.000	0.559	0.289
24/12	1.000	0.269	1.000	0.493	0.256
36/12	1.000	0.269	0.985	0.402	0.191
48/12	0.988	0.269	0.779	0.335	0.115

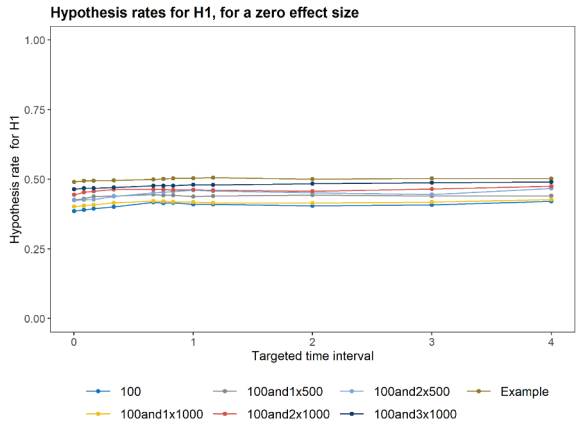
Performance of GORICA on CTmeta

The advantage of the meta-analysis and model selection on the resulting overall estimates is clear: an increased power to select the correct hypothesis and ability of evaluating the theory-based hypothesis. To examine how well the GORICA on CTmeta-analyzed estimates works, a small simulation study based on the working example is conducted. The procedure above is repeated for various settings for 1,000 simulated datasets.[2] Hence, each iteration in the simulation is a meta-analysis on 25 studies, where the GORICA is applied to the resulting overall lagged-parameter estimates, which is done for several targeted time intervals. The settings that are varied are the effect size in the population (large, medium, small, and zero) and the study-specific sample sizes (7 different sets). The other setting will remain the same: the study-specific time-intervals from Table 1, the 25 studies, and the 2 variables. The performance of the GORICA on CTmeta-analyzed estimates is measured by the true hypothesis rate (THR; i.e., the number of times the correct hypothesis is chosen) and the mean GORICA weights for the correct hypothesis.

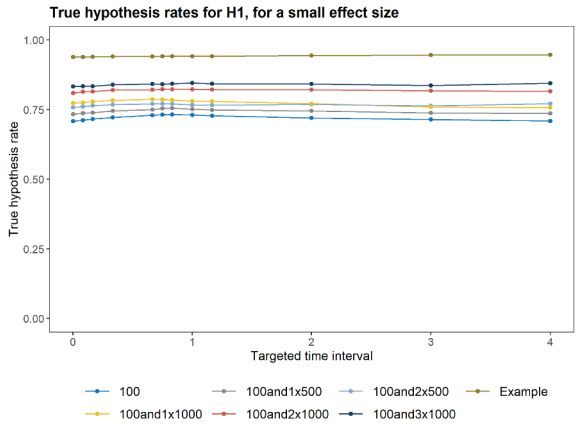
The four effect size specifications are based on the population values of $\mathbf{\Phi}(1)$ in Equation 1, were element (2,1) is varied. The specification in Equation 1, using 0.25, reflects a large effect size. For the effect size to be medium, small, and zero (implying that both cross-lagged relationships are equally strong), this element is set to 0.20, 0.17, and .15, respectively. At first, the study-specific sample sizes specifications as in Table 1 are used. Afterward, all of them are set to 100. Then, this is varied by setting the first 1 to 3 sample sizes to 500 or 1000. This leads to 7 specifications: "Example", "100", "100 and 1 x 500", "100 and 2 x 500", "100 and 1 x 1000", "100 and 2 x 1000", and "100 and 3 x 1000"; where the sum of the study-specific sample sizes are 13241, 2500, 2900, 3300, 3400, 4300, and 5200, respectively.

Based on these settings, datasets are generated for 25 primary studies where each primary study measures the lagged relationships between two variables. The same procedure as in the working example is used and evaluates H_1 and its complement H_c, given in Equation 4, with the GORICA. Note that H_1 is true for all settings, except for the zero effect size setting in which both H_1 and H_c are true (since the truth is on their border) and they have the same complexity. It is expected that the THRs and the mean GORICA

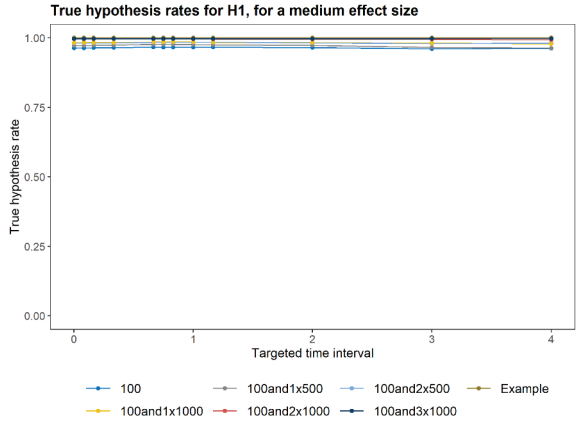
[2]When simulating data, some samples were discarded. Namely, the ones where the DT-VAR(1) lagged-parameter matrix had at least one negative eigenvalue (since in that case there does not exist an underlying drift matrix A) and the ones where the covariance matrices were not positive definite (comparable to negative variance; since in that case, CTmeta cannot be performed). For more details see Kuiper and Ryan (2020).



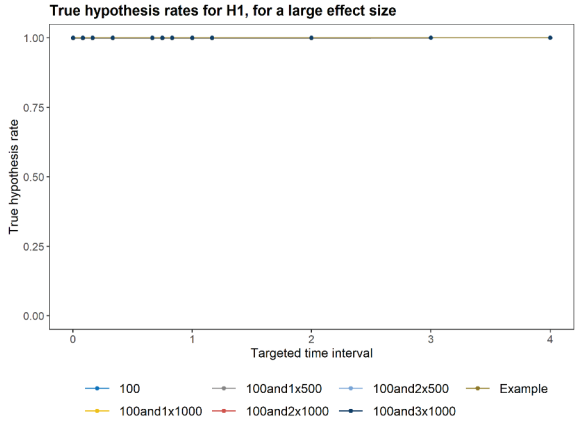
(a) Zero effect size



(b) Small effect size



(c) Medium effect size



(d) Large effect size

Figure 4. (True) hypothesis rates for H_1 (y-axis) for the 12 targeted time intervals (x-axis), 7 sample size sets (colored lines), and 4 effect size specifications (Plot a-d). Note that, for a zero effect size (Plot a), both H_1 and its complement are true; for the other three, H_1 is true.

weights for H_1 will go to one, when the effect size and/or combined sample size increases. In case of the zero effect size, it is expected that the THRs and the mean GORICA weights for H_1 is 0.5 (when the combined sample size is large enough) since the truth is on the border of H_1 and its complement and the hypotheses are of equal size (i.e., have the same penalty).

The THRs and mean GORICA weights for H_1 are plotted in Figures 4 and 5, respectively. These figures show that, for a zero effect size (Plots a), the hypothesis rate and mean weights for H_1 are around 0.5. Furthermore, the higher the combined sample size, the closer these values are to 0.5. Additionally, the figures show that the higher the effect size (going from Plots b to d), the higher the THRs and mean GORICA weights. Similarly, the higher the combined sample size (within Plots b, c, and d), the higher the THRs and mean GORICA weights. For a large effect size, the THRs are even 1 for all targeted time intervals, meaning full support for the true hypothesis.

When inspecting the results across targeted time intervals (i.e., various values on the x-axis), it can be seen that the THRs are (about) equal, while there is some variation in mean GORICA weights. Thus, the (average) support for H_1 varies over the targeted time intervals, which makes sense since the estimates of the parameters of interest do as well (cf. Figure 2). The pattern of the mean GORICA weights also resembles the pattern of the lagged-parameter curves in the **Φ**-plot in Figure 2, with the exception of the ceiling effect for the mean GORICA weight because it has a maximum value of 1. Since, here, the support is still over 0.5, the THR does not vary (that much). In case the targeted time intervals are that large that the corresponding cross-lagged estimates are both near zero, there is a zero effect size setting again and not only the mean GORICA weights go to 0.5 but eventually also the THR. To demonstrate this phenomenon for the GORICA weights, the targeted time-interval range is extended to 10 years (which is outside of the range within the primary studies); see Figure 6.

GORICA on CTmeta in R

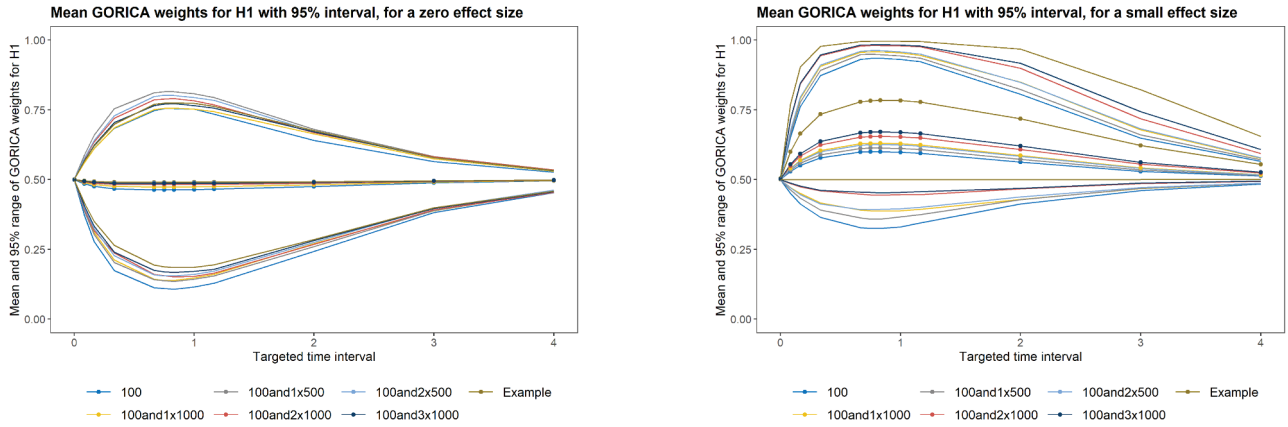
Example R code based on the first 3 out of the 25 studies of the work engagement and burnout working example is given by:

```
if (!require("restriktor")) install.packages("restriktor") # install this package first (once)
library(restriktor) # for goric function
#
library(devtools)
install_github("rebeccakuiper/CTmeta")
library(CTmeta)
?CTmeta
citation("CTmeta")

# Input needed in example below with q=2 variables and S=3 primary studies
#
N <- matrix(c(643, 651, 473))
DeltaT <- matrix(c(2, 3, 1))   # Time interval used in the 3 primary studies
DeltaTStar <- 1                # Targeted time interval
#
# Here, the example matrices stored in the package will be used.
# These contain the estimates for all 3 primary studies, stacked in one matrix.
Phi <- myPhi
SigmaVAR <- mySigmaVAR

## Example CTmeta: random effects model ("FEorRE = 2") without moderators ##
out_CTmeta <- CTmeta(N, DeltaT, DeltaTStar, Phi, SigmaVAR, FEorRE = 2)
#out_CTmeta

## Evaluate dominance of cross-lagged relationships ##
# Extract the vectorized overall standardized Phi matrix and its covariance matrix
est <- out_CTmeta$Overall_vecStandPhi_DeltaTStar
VCOV <- out_CTmeta$CovMx_OverallPhi_DeltaTStar
# Specify hypothesis
H1 <- "overallPhi12 < overallPhi21"
# Evaluate dominance of cross-lagged relationships via GORICA
goricaResult <- goric(est, VCOV = VCOV, H1, type = "gorica", comparison = "complement")
# In this example, the complement is "overallPhi12 > overallPhi21"
summary(goricaResult)
```

When using GORICA, only the parameters addressed in the theory-based hypotheses, the structural parameters, and their covariance matrix are needed. In contrast, the estimates of the residual covariance matrix $\boldsymbol{\Sigma}_{\epsilon}$ are so-called nuisance parameters and are not needed. In this example, the autoregressive relationships, which are also nuisance parameters here, could have been left out. Of course, including or excluding nuisance parameters does not effect the resulting GORICA weights.

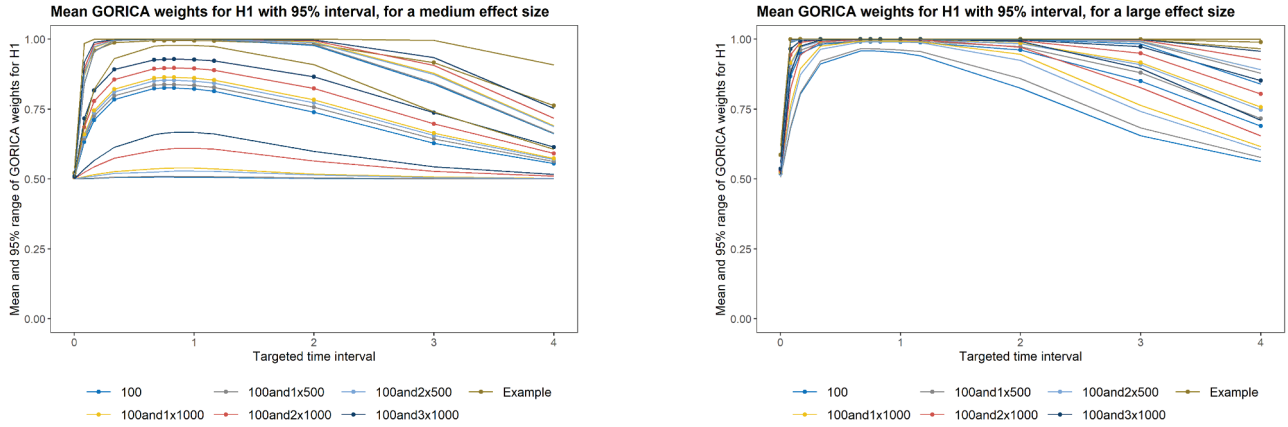
The code above renders the following output in R:

```
restriktor: generalized order-restriced information criterion approximation:

Results:
model  loglik  penalty   gorica  gorica.weights
1          H1  10.007    3.500  -13.013           0.934
2  complement   7.364    3.500   -7.728           0.066
---

Relative GORICA-weights:
vs. H1  vs. complement
H1          1.000   14.047
complement  0.071   1.000
---
The order-restricted hypothesis "H1" has 14.047 times more support than its complement.
```



(a) Zero effect size

(b) Small effect size



(c) Medium effect size

(d) Large effect size

Figure 5. Mean GORICA weights for H_1 (lines with dots) with the 5th percentile (lower lines) and 95th percentile (upper lines) for the 12 targeted time intervals (x-axis), 7 sample size sets (colored lines), and 4 effect size specifications (Plot a-d). Note that, for a zero effect size (Plot a), both H_1 and its complement are true; for the other three, H_1 is true.

From this, it is concluded that the lagged relationship of work engagement on burnout is smaller than its reciprocal effect; and that the support for this hypothesis is 14 times stronger than for its complement (i.e., any other theory) which states the opposite ordering in a two-parameter case. When inspecting the CTmeta results, this makes sense since the overall standardized lagged-parameter estimates in this 3-studies example are 0.08 and 0.15, respectively.

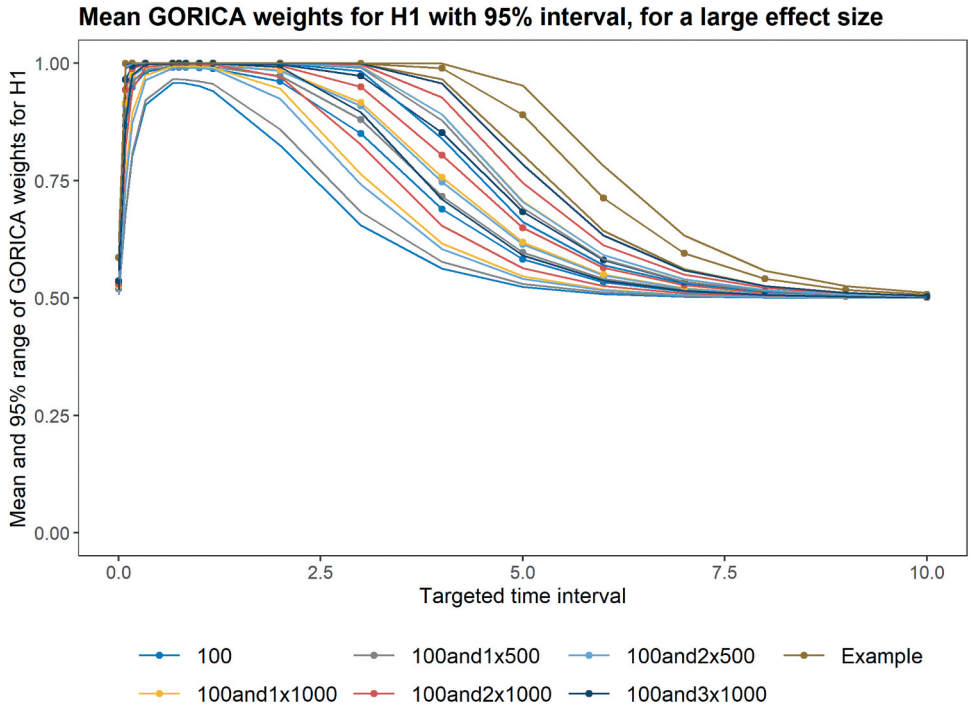


Figure 6. Mean GORICA weights for H_1 (lines with dots) with the 5th percentile (lower lines) and 95th percentile (upper lines) for a large effect size; for the targeted time intervals between 0 and 10 (x-axis), and 7 sample size sets (colored lines).

Discussion

This paper demonstrates how causal dominance theories regarding cross-lagged relationships based on multiple studies can be evaluated using the GORICA. A small simulation study shows that the properties of this method is good: The support for the correct hypothesis (reflected by the THR and the mean GORICA weight) increases with combined sample size and with effect size and will asymptotically reflect full support. Furthermore, the mean GORICA weights are a function of the targeted time intervals, like the estimates themselves are.

Only a small simulation study is conducted to show the performance of the GORICA on CTmeta-analyzed estimates. Since the results resemble that of the GORICA for a single CLPM and a single random-intercept CLPM (RI-CLPM; Hamaker et al., 2015),[3] it is to be expected that the performance will also be good for other settings, like a trivariate model. A more extensive simulation of the performance of the GORICA on a single (RI-)CLPM can be found in Sukpan and Kuiper (unpublished). They also investigated the performance for various choices of waves (in combination with different choices of the number of persons). Additionally, it is also possible to evaluate order-restricted theory-based hypotheses on autoregressive (and cross-lagged) relationships. Since most researchers are interested in the comparison of reciprocal cross-lagged relationships, only these were investigated. Nevertheless, it is to be expected that the performance is the same for other types of hypotheses.

Like in the CTmeta article, the key assumption is that the underlying process is continuous-time in nature and, thus, the same two limitations hold. First, the sampling frequency in the primary studies should be sufficiently high to capture the dynamics of interest (cf. Shannon, 1984). Second, when oscillating behavior is present in the system of interest, as indicated by complex eigenvalues of the lagged-parameter matrix, CTmeta breaks down (cf. Hamerle et al., 1991). In that case, there is no unique mapping from the discrete-time to the continuous-time parameter matrices.

Last but not least, while the interest of researchers is in 'causal dominance' relations, additional assumptions are required for the overall lagged parameters to reflect causal relationships (cf. Usami et al., 2019). Nonetheless, the availability of a method to evaluate the ordering of CTmeta-analyzed (cross-)lagged parameters is a necessary step on the road to establishing any reliable conclusions regarding 'causal dominance' relationships.

ORCID

Rebecca Kuiper http://orcid.org/0000-0002-0267-5197

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Akademiai Kiado.

Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, *30*, 9–14. https://doi.org/10.1007/BF02480194

[3]In these models, random intercepts are used to separate the between from the within-persons variance components by decomposing observations into a stable person-specific mean and a deviation from that mean at a given time point. The DT-VAR(1) model is then applied to the latter deviations. Notably, this RI-CLPM model is only identified for $m \geq 3$.

Altınışık, Y., Van Lissa, C. J., Hoijtink, H., Oldehinkel, A. J., & Kuiper, R. M. (2021). Evaluation of inequality constrained hypotheses using a generalization of the AIC. *Psychological Methods*.

Becker, B. J., & Wu, M.-J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science*, *22*, 414–429. https://doi.org/10.1214/07-STS243

Bentler, P. M., & Speckart, G. (1981). Attitudes "cause" behaviors: A structural equation analysis. *Journal of Personality and Social Psychology*, *40*, 226. https://doi.org/10.1037/0022-3514.40.2.226

Boker, S. M., & Laurenceau, J.-P. (2006). *Dynamic system modeling: An application to the regulation of intimicy and disclosure in marriage*. Oxford University Press.

Bollen, K. A., & Curran, P. (2006). *Latent curve models: A structural equation approach*. Wiley and Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Card, N. A. (2019). Lag as moderator meta-analysis: A methodological approach for synthesizing longitudinal data. *International Journal of Behavioral Development*, *43*, 80–89. https://doi.org/10.1177/0165025418773461

Chatfield, C. (2004). *The analysis of time series: An introduction* (6th ed.). Chapman and Hall.

Dormann, C., & Griffin, M. A. (2015). Optimal time lags in panel studies. *Psychological Methods*, *20*, 489. https://doi.org/10.1037/met0000041

Fabio Di Narzo, A., Aznarte, J. L., & Stigler, M. (2009). tsDyn: Time series analysis based on dynamical systems theory. R package version 0.7.

Finkel, S. E. (1995). *Causal analysis with panel data, sage university papers series on quantitative applications in the social sciences*. Sage.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development*, *58*, 80–92. https://doi.org/10.2307/1130293

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 424–438. https://doi.org/10.2307/1912791

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116. https://doi.org/10.1037/a0038889

Hamerle, A., Nagl, W., & Singer, H. (1991). Problems with the estimation of stochastic differential equations using structural equations models. *The Journal of Mathematical Sociology*, *16*, 201–220. https://doi.org/10.1080/0022250X.1991.9990088

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

Hoijtink, H. (2012). *Informative hypotheses: Theory and practice for behavioral and social scientists*. CRC Press.

Jacobson, N. C., & Newman, M. G. (2017). Anxiety and depression as bidirectional risk factors for one another: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *1*, 1–46. https://doi.org/10.1037/bul0000111

Kuiper, R. M. (2020a). Phi-and-Psi-Plots and find deltaT [Shiny app software]. https://www.uu.nl/staff/RMKuiper/Websites%20%2F%20Shiny%20apps

Kuiper, R. M. (2020b). CT meta-analysis on lagged effects [shiny app software]. https://www.uu.nl/staff/RMKuiper/Websites%20%2F%20Shiny%20apps

Kuiper, R. M. (2020c). CTmeta: Continuous-time meta-analysis on lagged effects [R package]. https://github.com/rebeccakuiper/CTmeta

Kuiper, R. M. (2020d). AIC-type theory-based model selection for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*.

Kuiper, R. M., & Hoijtink, H. (2010). Comparisons of means using exploratory and confirmatory approaches. *Psychological Methods*, *15*, 69–86. https://doi.org/10.1037/a0018720

Kuiper, R. M., Hoijtink, H., & Silvapulle, M. J. (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. M., Hoijtink, H., & Silvapulle, M. J. (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

Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. *Structural Equation Modeling*, *25*, 809–823. https://doi.org/10.1080/10705511.2018.1431046

Kuiper, R. M., & Ryan, O. (2020). Meta-analysis of lagged regression models: A continuous-time approach. *Structural Equation Modeling*, *27*, 396–413. https://doi.org/10.1080/10705511.2019.1652613

Kuiper, R. M., Yasin, A., & Van Lissa, C. J. (2020). gorica: Evaluation of inequality constrained hypotheses using GORICA. R package version 0.1.0.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79–86. https://doi.org/10.1214/aoms/1177729694

Maricuțoiu, L. P., Sulea, C., & Iancu, A. (2002). Work engagement or burnout: Which comes first? A meta-analysis of longitudinal evidence. *Burnout Research*, *5*, 35–43. https://doi.org/10.1016/j.burn.2017.05.001

Masselink, M., Van Roekel, E., Hankin, B., Keijsers, L., Lodder, G., Vanhalst, J., Verhagen, M., Young, J., & Oldehinkel, A. (2018). The longitudinal association between self-esteem and depressive symptoms in adolescents: Separating between-person effects from within-person effects. *European Journal of Personality*, 1–19. https://doi.org/10.1002/per.2179

Mayer, L. (1986). On cross-lagged panel models with serially correlated errors. *Journal of Business and Economic Statistics*, *4*, 347–357. https://www.tandfonline.com/doi/abs/10.1080/07350015.1986.10509531

Moberly, N. J., & Watkins, E. R. (2008). Ruminative self-focus and negative affect: An experience sampling study. *Journal of Abnormal Psychology*, *117*, 314–323. https://doi.org/10.1037/0021-843X.117.2.314

Nohe, C., Meier, L. L., Sonntag, K., & Michel, A. (2015). The chicken or the egg? A meta-analysis of panel studies of the relationship between work–family conflict and strain. *Journal of Applied Psychology*, *100*, 522–536. https://doi.org/10.1037/a0038012

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical Ornstein–Uhlenbeck model for continuous repeated measurement data. *Psychometrika*, *74*, 395–418. https://doi.org/10.1007/s11336-008-9106-8

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods*, *16*, 468–490. https://doi.org/10.1037/a0024375

Oud, J. (2002). Continuous time modeling of the cross-lagged panel design. *Kwantitatieve Methoden*, *69*, 1–27. https://psycnet.apa.org/record/2007-07934-004

Oud, J., & Delsing, M. (2010a). *Longitudinal research with latent variables, chapter continuous time modeling of panel data by means of SEM*. Springer.

Pelz, D., & Lew, R. (1970). Heise's causal model applied. *Sociological Methodology*, *2*, 28.

Pfaff, B. (2008). VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software*, *27*, 1–32. https://doi.org/10.18637/jss.v027.i04

Rogosa, D. (1979). Causal models in longitudinal research: Rationale, formulation, and interpretation. *Longitudinal Research in the Study of Behavior and Development*, 263–302. https://doi.org/10.1037/0033-2909.88.2.245

Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, *88*, 245. https://doi.org/10.1037/0033-2909.88.2.245

Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). *Continuous time modeling in the behavioral and related sciences, chapter A continuous time approach to intensive longitudinal data: What, why and how?*. Springer.

Shannon, C. E. (1984). Communication in the presence of noise. *Proceedings of the IEEE*, *72*, 1192–1201. https://doi.org/10.1109/PROC.1984.12998

Sukpan, C., & Kuiper, R. M. (unpublished). Evaluating causal dominance hypotheses using an AIC-type criterion.

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*, 637–657. https://doi.org/10.1037/met0000210

Vanbrabant, L., & Kuiper, R. M. (2020). goric function in R package restriktor [R package version 0.2–800]. https://CRAN.R-project.org/package=restriktor

Vanbrabant, L., & Rosseel, Y. (2020). restriktor: Restricted statistical estimation and inference for linear models. R package version 0.2–800.

Vanbrabant, L., Van Loey, N., & Kuiper, R. M. (2020). Evaluating a theory-based hypothesis against its complement using an AIC-type information criterion with an application to facial burn injury. *Psychological Methods*, *25*, 129–142. https://doi.org/10.1037/met0000238

Voelkle, M. C., & Oud, J. H. L. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. *British Journal of Mathematical and Statistical Psychology*, *66*, 103–126. https://doi.org/10.1111/j.2044-8317.2012.02043.x

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*, 192–196. https://doi.org/10.3758/BF03206482