

From Data to Causes I: Building A General Cross- Lagged Panel Model (GCLM)

Organizational Research Methods
1-37

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1094428119847278

journals.sagepub.com/home/orm



**Michael J. Zyphur¹, Paul D. Allison², Louis Tay³,
Manuel C. Voelkle^{4,5}, Kristopher J. Preacher⁶,
Zhen Zhang⁷, Ellen L. Hamaker⁸,
Ali Shamsollahi⁹, Dean C. Pierides¹⁰,
Peter Koval¹¹, and Ed Diener^{12,13}**

Abstract

This is the first paper in a series of two that synthesizes, compares, and extends methods for causal inference with longitudinal panel data in a structural equation modeling (SEM) framework. Starting with a cross-lagged approach, this paper builds a general cross-lagged panel model (GCLM) with parameters to account for stable factors while increasing the range of dynamic processes that can be modeled. We illustrate the GCLM by examining the relationship between national income and subjective well-being (SWB), showing how to examine hypotheses about short-run (via Granger-Sims tests) versus long-run effects (via impulse responses). When controlling for stable factors, we find no short-run or long-run effects among these variables, showing national SWB to be relatively stable, whereas income is less so. Our second paper addresses the differences between the GCLM and other methods. Online Supplementary Materials offer an Excel file automating GCLM input for Mplus (with an example also for Lavaan in R) and analyses using additional data sets and all program input/output. We also offer an introductory GCLM presentation at <https://youtu.be/tHnnaRNPbXs>. We conclude with a discussion of issues surrounding causal inference.

¹Department of Management & Marketing, Business & Economics, University of Melbourne, Melbourne, Australia

²Department of Sociology, University of Pennsylvania, PA, USA

³Department of Psychology, Purdue University, IN, USA

⁴Institut für Psychologie, Humboldt University Berlin, Berlin, Germany

⁵Max Planck Institute for Human Development, Berlin, Germany

⁶Department of Psychology & Human Development, Vanderbilt University, TN, USA

⁷Department of Management, W. P. Carey School of Business, Arizona State University, AZ, USA

⁸Department of Methods and Statistics, Utrecht University, Netherlands

⁹ESSEC Business School, Cergy-Pontoise, France

¹⁰Department of Management Work and Organisation, University of Stirling, Stirling, UK

¹¹Department of Psychology, University of Melbourne, Melbourne, Australia

¹²Department of Psychology, University of Utah, UT, USA

¹³Department of Psychology, University of Virginia, VA, USA

Corresponding Author:

Michael J. Zyphur, Department of Management and Marketing, Parkville, VIC 3010, Australia.

Email: mzyphur@unimelb.edu.au

Keywords

panel data model, cross-lagged panel model, causal inference, Granger causality, structural equation model, vector autoregressive VAR model, autoregression, moving average, ARMA, VARMA, panel VAR

Causal inference is a core part of scientific research and policy formation. There are multiple pathways to causal inference (Cartwright, 2007, 2011), but a popular approach uses longitudinal panel data made up of multiple units measured at multiple occasions. Such data are useful for causal inference by helping to control for confounds and modeling lagged relationships as units of analysis change over time (Hausman & Taylor, 1981; Kessler & Greenberg, 1981; Liker, Augustyniak, & Duncan, 1985). With this approach, organizational researchers regularly use panel data to infer causality, often with cross-lagged panel models.

For example, at an individual level of analysis, Meier and Spector (2013) studied 663 people at five occasions, finding reciprocal effects among counterproductive work behaviors work stressors, inferring “a vicious cycle with negative consequences for all parties involved” (p. 537). At a higher level of analysis, Van Iddekinge et al. (2009) studied 861 locations of an organization at six occasions, showing reciprocal effects for human resources (HR) factors and performance, thus offering the advice that “human capital investments . . . can yield a high return” (p. 840). At a national level of analysis, Diener, Tay, and Oishi (2013) studied 135 countries at six occasions, finding reciprocal effects for income and subjective well-being (SWB), inferring that in terms of SWB, “people did not adapt to income increases” (p. 275).

By using such observational data, this work has the potential to show real-world evidence of effects that may otherwise be difficult to uncover. As medical researchers note, such evidence may be useful due to “its potential for complementing the knowledge gained from traditional clinical trials, whose well-known limitations make it difficult to generalize findings” (Sherman et al., 2016, p. 2293; see also Booth & Tannock, 2014). However, given this potential, many studies often fail to capitalize on the unique opportunities that panel data offer, including strengthening causal inferences by controlling for stable factors and testing hypotheses about the long-run behavior of the systems being studied. This occurs due to the lack of integration across fields in the tools used for longitudinal data modeling (contrast: Bollen & Curran, 2006; Box, Jenkins, & Reinsel, 2008; Hsiao, 2014; Lütkepohl, 2005; McArdle & Nesselroade, 2014). The result is that organizational researchers often fail to examine a range of theoretically relevant processes and effects when modeling panel data.

For example, many researchers use latent curve models separately from lagged effects models, perhaps due to a belief that modeling curves precludes lagged effects (e.g., Rogosa & Willett, 1985) or that econometric tools “are usually less applicable for the kinds of data psychologists and micro HR/OB scholars have,” often with few measured occasions T (Ployhart & Ward, 2011, p. 414). Yet, accounting for curves (i.e., trends) is crucial for lagged effects models (Box et al., 2008; Lütkepohl, 2005), and many econometric tools are designed specifically for the “small T ” case (Arellano, 2003; Baltagi, 2014; Hsiao, 2014).

To help researchers overcome the limitations of current panel data modeling methods, we synthesize, compare, and extend approaches to panel data modeling in two papers. Our primary goals are to: (a) show how panel data can help test hypotheses (or infer processes) in more powerful and useful ways than are typically found in the organizational literature and for this purpose, (b) introduce methods from disciplines that may be foreign to many readers.

We tackle these by starting with a typical cross-lagged panel model to build a more general structural equation model (SEM), which we call a **general cross-lagged panel model (GCLM)**, that controls for

stable factors and increases the range of dynamic processes that can be modeled. Our approach is designed for the typical panel data case where $T < 20$ (and usually $T < 10$), but most of what we discuss can be applied to larger T cases by using dynamic structural equation modeling (DSEM; see Asparouhov, Hamaker, & Muthén, 2018). Our second paper compares our approach to others, including multilevel panel data models. Across both papers, we offer an integrative overview drawn from multiple traditions, resulting in powerful new conceptual and statistical tools for modeling panel data.

In what follows, we first conceptually treat GCLM parameters. Then, we treat tests of short-run effects as *direct* effects among variables, versus long-run “impulse responses” that capture all *indirect* effects of one variable on another over time. We then describe a general SEM for estimation and hypothesis testing. To illustrate a GCLM, we reexamine the income-SWB relationship at the national level (from Diener et al., 2013), failing to support effects among these variables. We also model individual and organizational effects from Meier and Spector (2013) and Van Iddekinge et al. (2009) to exemplify our points—we reanalyze their data and present GCLM findings in Appendix A in the Supplementary Materials (available in the online version of the journal).

All input/output for the Mplus program are available in the online Supplementary Materials, along with an Excel file to automate Mplus input for a GCLM and its variants. We also include an example of the GCLM in Lavaan for R and note that the Mplus2lavaan (2019) program for R can help translate most Mplus input to Lavaan. All Supplementary Material can be cited and is available for download at <https://doi.org/10.26188/5c9ec7295fef>. To help the reader, we also offer a presentation on the GCLM and the processes it captures at <https://youtu.be/tHnnaRNPbXs>. We conclude by discussing issues in causal inference under uncertainty, including threats to causality due to trends and regime changes (i.e., parameter changes over time).

Before proceeding, we emphasize that our goal is to offer a practical framework for modeling panel data based on the idea that “it pays to experiment with the . . . techniques that panel data make available” (Halaby, 2004, p. 541). In the end, we agree that “there is no such thing as *the* methodology for analyzing panel data, but a collection of . . . techniques that have accumulated from a series of heterogeneous motivations” (Arellano, 2003, p. 2). Our goal is to explore these techniques and expand the toolkits of researchers who regularly use panel data to make causal inferences. In this tradition, we seek to improve current practices.

Building a General Cross-Lagged Panel Model

There are many useful introductions to longitudinal data models (e.g., Allison, 2005, 2009; Baltagi, 2013; Bollen & Brand, 2010; Bond, 2002; Cole, 2012; Enders, 2014; Halaby, 2004; Hamaker, Kuiper, & Grasman, 2015; Hsiao, 2007; Lütkepohl, 2006, 2013). We draw on this work to build a GCLM while focusing on its conceptual logic and tools for hypothesis testing that follow from it (see YouTube). Although the GCLM may seem complex, any subset of its parameters (in Table 1) can be used to build a panel data model, and our methods for hypothesis testing will both clarify and simplify causal inference using the GCLM.

To begin in a familiar way, we first introduce a cross-lagged panel model and treat the conceptual underpinnings of its parameters. With this structure in place, we then offer several ways to extend the model, proposing a GCLM that includes additional parameters to expand the range of dynamic processes that can be modeled and then used for hypothesis testing.

A Cross-Lagged Panel Model

We start with a cross-lagged panel model where all variables are a function of the past (see Figure 1). Throughout, our figures use SEM notation as follows: Observed variables are squares, latent variables are circles, single-headed arrows show dependence, and double-headed arrows are

Table 1. Parameters, Their Purposes, and SEM Specifications (for Observed Variables x_{it} and y_{it}).

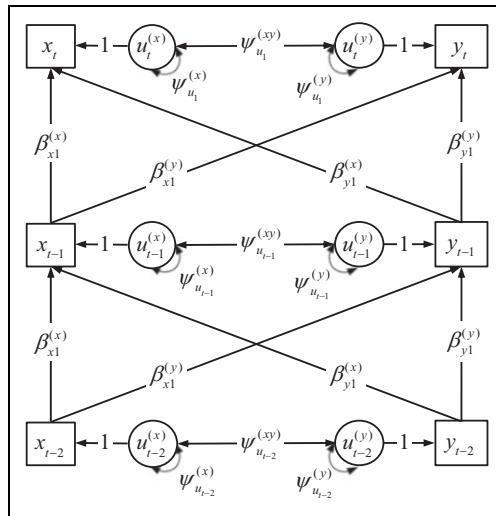
$\alpha_t^{(x)}$ or $\alpha_t^{(y)}$	<p>Name: Occasion effect (also called global shock or cross-sectional effect)</p> <p>Purpose: Controls for time-specific effects common to all observed units at t</p> <p>Specification: Unrestricted intercept for each observed variable</p>
$\beta_{x1}^{(x)}$ or $\beta_{y1}^{(y)}$	<p>Name: Autoregression or autoregressive effect (AR)</p> <p>Purpose: Models (in)stability by allowing past impulses to persist over time, controlling for past impulses when assessing other effects</p> <p>Specification: Regress the future on the past for the same variable at a lag h (e.g., $\beta_{xh}^{(x)}$), which can be done at some lag order p for an AR(p) model</p>
$\beta_x^{(x)}$ or $\beta_y^{(y)}$	<p>Name: Total AR effect</p> <p>Purpose: Summarizes all direct AR effects when $p > 1$; if less than 1.0, a mean-reverting process is implied (regressing to η_i); if equal to 1.0, a random walk is implied wherein past impulses persist over time; if greater than 1.0, a more chaotic is implied wherein past impulses are amplified over time</p> <p>Specification: Sum of all direct AR effects for a given variable</p>
$\beta_{y1}^{(x)}$ or $\beta_{x1}^{(y)}$	<p>Name: Cross-lagged (CL) effect</p> <p>Purpose: Models causal effect of past impulses from one variable on future realizations of another variable and controls for past impulses across variables</p> <p>Specification: Regress the future on the past for different variables at a lag h (e.g., $\beta_{yh}^{(x)}$), which can be done at some lag order c for a CL(c) model</p>
$\beta_y^{(x)}$ or $\beta_x^{(y)}$	<p>Name: Total CL term</p> <p>Purpose: Summarizes all direct CL effects when $q > 1$</p> <p>Specification: Sum of all direct CL effects linking one variable to another</p>
$u_{it}^{(x)}$ or $u_{it}^{(y)}$	<p>Name: Impulse (also sometimes called a shock, innovation, perturbation)</p> <p>Purpose: Mimics random assignment along an observed variable at a time t, similar to an intervention at a time t that is uncorrelated with other variables</p> <p>Specification: Latent variable with “factor loading” of 1.0 on observed variable</p>
$\psi_{u_t}^{(x)}$ or $\psi_{u_t}^{(y)}$	<p>Name: Impulse variance</p> <p>Purpose: Models differences across units in their standings along $u_{it}^{(x)}$ and $u_{it}^{(y)}$</p> <p>Specification: Unrestricted variance for latent impulse variables</p>
$\psi_{u_t}^{(xy)}$	<p>Name: Co-movement (or impulse co-movement)</p> <p>Purpose: Models impulses that are common to multiple variables at a time t, allows controlling for the past when assessing effects over time</p> <p>Specification: Unrestricted covariance among latent impulse variables</p>
$\eta_i^{(x)}$ or $\eta_i^{(y)}$	<p>Name: Unit effect (or time-invariant effect, fixed effect)</p> <p>Purpose: Controls for consistency over time, eliminating stable confounds</p> <p>Specification: Latent variable with “factor loading” on all T occasions</p>
$\lambda_t^{(x)}$ or $\lambda_t^{(y)}$	<p>Name: Time-varying unit effect (often called a factor loading)</p> <p>Purpose: Allows unit effects to affect observed variables differently at each t</p> <p>Specification: Unrestricted factor loadings (except for $t = T$ at 1.0 to scale η_i)</p>
$\psi_\eta^{(x)}$ or $\psi_\eta^{(y)}$	<p>Name: Unit effect variance</p> <p>Purpose: Models differences across units in their standings on η_i</p> <p>Specification: Unrestricted variance for η_i</p>
$\psi_\eta^{(xy)}$	<p>Name: Unit effect covariance</p> <p>Purpose: Models covariance in unit effects, allowing them to be controlled when assessing other modeled effects (similar to a “fixed effects” model)</p> <p>Specification: Unrestricted covariance among latent variables η_i</p>
$\delta_{x1}^{(x)}$ or $\delta_{y1}^{(y)}$	<p>Name: Moving average (MA)</p> <p>Purpose: Allows temporary effects of the past on the future for the same variable, typically in order to temporarily increase or decrease AR effects</p> <p>Specification: Regress the future on a past impulse for the same variable at a lag h (e.g., $\delta_{xh}^{(x)}$), which can be done at some lag order q for an MA(q) model</p>

(continued)

Table 1. (continued)

$\delta_x^{(x)}$ or $\delta_y^{(y)}$	<p>Name: Total MA effect</p> <p>Purpose: Summarizes all direct MA effects when $q > 1$</p> <p>Specification: Sum of all direct MA effects for a given variable</p>
$\beta_x^{(x)} + \delta_x^{(x)}$ or $\beta_y^{(y)} + \delta_y^{(y)}$	<p>Name: Total AR and MA effect</p> <p>Purpose: Summarizes all direct effects of a past impulse on the same variable</p> <p>Specification: Sum of all direct AR and MA effect for a given variable</p>
$\delta_{y1}^{(x)}$ or $\delta_{x1}^{(y)}$	<p>Name: Cross-lagged moving average (CLMA)</p> <p>Purpose: Allows temporary effects of the past on the future for different variables, typically in order to temporarily increase or decrease CL effects</p> <p>Specification: Regress the future on a past impulse of a different variable at a lag h (e.g., $\delta_{yh}^{(x)}$), which can be done at a lag order l for a CLMA(l) model</p>
$\delta_y^{(x)}$ or $\delta_x^{(y)}$	<p>Name: Total CLMA term</p> <p>Purpose: Summarizes all direct CLMA effects when $l > 1$</p> <p>Specification: Sum of all direct CLMA effects of a past impulse</p>
$\beta_y^{(x)} + \delta_y^{(x)}$ or $\beta_x^{(y)} + \delta_x^{(y)}$	<p>Name: Total CL and CLMA effect</p> <p>Purpose: Summarizes all direct effects of one variable on another</p> <p>Specification: Sum of all CL and CLMA terms linking one variable to another</p>

Note: SEM = structural equation model.

**Figure 1.** An AR(1)CL(1) model.

(co)variances (we omit intercepts/means). For simplicity, we use two variables, x_{it} and y_{it} , for a unit i at a time t , for N units at T occasions (where $t = 1, 2, \dots, T$). Lagged occasions are indicated by a lag h , so if $h = 1$ in y_{it-h} , then y is an observation for the occasion before t , or y_{it-1} . We initially assume x_{it} and y_{it} are observed, error-free, normal, linearly related, and measured at similar occasions across N with similar spacing across T . We later note that many of these assumptions can be relaxed using various approaches, which we also discuss at greater length in our online appendices.

We start with a cross-lagged panel model using some specialized notation as follows:

$$x_{it} = \alpha_t^{(x)} + \beta_{x1}^{(x)} x_{it-1} + \beta_{y1}^{(x)} y_{it-1} + u_{it}^{(x)} \quad (1)$$

$$y_{it} = \alpha_t^{(y)} + \beta_{y1}^{(y)} y_{it-1} + \beta_{x1}^{(y)} x_{it-1} + u_{it}^{(y)} \quad (2)$$

Here, α_t is an occasion-specific intercept, which we call an *occasion effect*; $\beta_{x1}^{(x)}$ and $\beta_{y1}^{(y)}$ are autoregressive (AR) effects; $\beta_{x1}^{(x)}$ and $\beta_{x1}^{(y)}$ are cross-lagged (CL) effects; and u_{it} is a time-specific residual, which we call an *impulse*. We treat these in more detail in the following but for now point out that **terms have a superscript in parentheses to indicate outcome variables**, whereas each regression coefficient β has a subscript that refers to a predictor and a time lag. Thus, $\beta_{x1}^{(y)}$ is the effect of x on y at lag $h = 1$, with the past as the subscript x that affects the future as the superscript y (i.e., $\beta_{x1}^{(y)}$ is a y AR effect, whereas $\beta_{x1}^{(y)}$ is an $x \rightarrow y$ CL effect).

Occasion effects α_t . To model causal effects in panel data, it is important to account for overall changes in a sample across occasions, which may be due to a variety of aggregate factors that are unrelated to lagged effects (i.e., AR and CL terms). We account for these with an occasion effect α_t . For example, if y_{it} were work stressors, α_t could capture aggregate change in work demands such as seasonal effects for all employees; if y_{it} were organization performance, aggregate change in industry-related factors could be captured by α_t ; if y_{it} were national income, the global financial crisis could be captured by α_t . Such effects are often controlled using $T - 1$ dummy variables that code for occasion of measurement, but as we show later, an SEM automatically accounts for α_t with observed-variable intercepts.

Autoregressive (AR) effects $\beta_{y1}^{(y)}$ and $\beta_{x1}^{(x)}$. A key part of the cross-lagged model are AR effects that link the past and future (see Figure 1). With this approach, a unit's current state is a function of its past, so x_{it} and y_{it} depend on x_{it-1} and y_{it-1} with AR terms $\beta_{x1}^{(x)}$ and $\beta_{y1}^{(y)}$, respectively; we show a lagged path from a past y_{it-h} to a future y_{it} as $\beta_{yh}^{(y)}$. These effects show how a process unfolds, reflecting the fact that the current states of a system do not spontaneously arise. Instead, they depend on past states, such as current national income depending on past income, current organization performance depending on past performance, or current individual counterproductive behaviors depending on past behaviors.

An AR term $\beta_{y1}^{(y)}$ captures this as a proportion of the past that persists directly to the next occasion (y_{it-1} is multiplied by $\beta_{y1}^{(y)}$). Thus, if $\beta_{y1}^{(y)} = 0$, it implies current levels of national income, organization performance, or individual behaviors are independent of past levels, but as $\beta_{y1}^{(y)} \rightarrow 1$, the **present increasingly resembles the past**. This can be seen in Figure 1 by tracing AR paths as indirect effects of the past, which can be thought of as a kind of persistence, momentum, inertia, memory, or carryover. Alternatively, **inverse meanings (e.g., $1 - \beta_{y1}^{(y)}$) can be adopted, such as forgetting, decay, entropy, regulatory strength, or regression to the mean, because as $\beta_{y1}^{(y)} \rightarrow 0$, past states fade more quickly**. On the other hand, **negative AR terms can indicate a pendulum-like process, wherein values on a variable tend to alternate between positive and negative over time**—a type of “anti-dependence”—such as in a system that responds by counteracting past states (e.g., cycles of activity vs. respite).

We return to AR terms when treating long-run effects, and Online Appendix B treats the special case of $AR \geq 1$, but for now, we lay a foundation for seeing CL effects as causal by noting that AR terms help control for some confounds. For example, employees may engage in counterproductive work behaviors as a matter of habit rather than due to increases in work stressors, so controlling for past counterproductive work behaviors with AR terms is relevant. Similarly, organizations may experience high performance due to persistent market forces rather than changes to HR practices, so again, performance AR terms may be useful. Also, nations may experience low well-being that persists for reasons that may be unrelated to decreases in national income. Thus, **AR terms reflect persistence, but they also control for a variable's past levels** to help avoid drawing erroneous causal conclusions using CL terms.

This understanding of AR terms motivates a discussion of CL effects, but before this, it is important to note that **some processes cannot be modeled by a single AR term, such as lagged**

effects that take longer than a lag $h = 1$ to appear or complex processes that can be modeled by both a positive and negative AR term at different lags. As Figure 1 shows, AR terms recursively link the past to the future (e.g., y_{it-2} impacts y_{it} indirectly via y_{it-1}), but some processes may require not only an indirect effect from y_{it-2} to y_{it} via y_{it-1} but also a direct effect $\beta_{y2}^{(y)}$ of y_{it-2} on y_{it} . The latter is a higher-order lag, which would be a model with an AR “lag order” of 2, noted as p in an AR(p) model—an AR(2) model here. In this case, researchers may combine all p AR terms that impact a future y (e.g., $\beta_y^{(y)} = \beta_{y1}^{(y)} + \beta_{y2}^{(y)}$), which is equivalent to combining all p AR terms that emanate from a past y and allows a single AR term to be used as a kind of shorthand to infer about the AR dynamics affecting a future observation of y (e.g., on the whole persistent with a positive $\beta_y^{(y)}$ or anti-persistent with a negative value). We discuss issues related to higher-order lags later, and our online Excel file can automate Mplus input for models with varying lag orders.

Cross-lagged effects $\beta_{y1}^{(x)}$ and $\beta_{x1}^{(y)}$. By including AR terms, it becomes possible to use the past of one variable to uniquely predict the future of another. Such CL effects imply that each unit’s current state is a function of its past on other variables; so for example, x_{it} is predicted by y_{it-1} with a CL effect $\beta_{y1}^{(x)}$, and y_{it} is predicted by x_{it-1} with $\beta_{x1}^{(y)}$; a path from x_{it-h} to y_{it} is $\beta_{xh}^{(y)}$. These effects enact a temporal order that is consistent with causation, which is partly used to justify treating CL terms as causal. Often called Granger causality, the idea is that if a predictor uniquely accounts for the future along a variable, then this can serve as provisional evidence of causation (Granger, 1969, 1980, 1986, 1988).

With this logic, CL terms are used to infer causality, but as Figure 1 shows, they only imply a “short-run” effect as a direct effect of the past on the future. Just like AR terms, these depict a system’s short-run behavior, with implications for CL terms ≥ 1 , as noted in Online Appendix B. Yet, investigating long-run behavior requires examining how the past indirectly affects the future along all AR and CL paths simultaneously (e.g., the total effect of an initial x_{i1} on y_{i2}, y_{i3} , etc.). As we show later, such long-run effects map onto key hypotheses such as “a vicious cycle” among work stressors and behavior (Meier & Spector, 2013, p. 537), or if “human capital investments . . . yield a high return” (Van Iddekinge et al., 2009, p. 840), or if “people [do] not adapt to income increases” (Diener et al., 2013, p. 275).

We will cover long-run hypothesis tests when we treat impulse responses. For now, we note that just like AR effects, higher-order CL terms may be needed for some processes. For example, work stressors may have delayed or complex effects on counterproductive work behaviors, requiring a second lag $c = 2$ in a CL(2) model, such as an effect $\beta_{x1}^{(y)}$ of x_{it-1} on y_{it} and a direct effect $\beta_{x2}^{(y)}$ of x_{it-2} . Here, a single CL term can be used as a shorthand for all CL effects associated with a future y or similarly, all CL effects emanating from a past x ($\beta_x^{(y)} = \beta_{x1}^{(y)} + \beta_{x2}^{(y)}$). We later show how to specify and select models with higher-order CL lags and how to test effects among variables in more nuanced ways using impulse responses.

Impulses u_{it} . The model also includes a residual term to allow units to differ over time due to random inputs (Denrell, Fang, & Liu, 2014). Although residuals are often taken for granted in regression, in cross-lagged models, they actually have an important substantive role that requires some theoretical prefacing. For example, consider that rules and routines guide social entities but behavior and events are never predictable as novelties emerge over time (Becker, Knudsen, & March, 2006; Weick, 1998). The same is true for larger economic changes (Lütkepohl, 2015), which are typically unpredictable or even a priori unexplainable (Cochrane, 1994). This is echoed by research efforts in social science that fail to explain substantial variation because of the stochastic nature of many phenomena (Abelson, 1985).

To capture such random inputs for each unit i at a given occasion t , we include a random term u_{it} , so x_{it} and y_{it} are functions of $u_{it}^{(x)}$ and $u_{it}^{(y)}$ with variances $\psi_{u_i}^{(x)}$ and $\psi_{u_i}^{(y)}$, respectively. Put simply, u_{it}

represents random events that make observations unique, such as events that raise/lower a person's work stressors by 1 point, or organization performance, or national income at a time t (i.e., $u_{it} = \pm 1$). Thus, although u_{it} is often called a residual or a disturbance and $\psi_{u_{it}}$ is a residual variance, we call u_{it} an impulse and $\psi_{u_{it}}$ an impulse variance because they do not imply error. Instead, they capture a unit-specific "shock" or an unpredictable "surprise" in the system being modeled over time (Stock & Watson, 2001). This conceptualization will later be used to facilitate an understanding of causality in our model.

Before this, however, we note that an impulse u_{it} may be nonindependent if a shock affects multiple variables. Such common causes can still be random, such as unpredictable changes in work assignments causing simultaneous change in work stressors and behavior, or unpredictable changes in organization leadership causing change in both HR practices and performance, or unpredictable events in a nation causing change to SWB and income. We call such effects a *co-movement*, which can be modeled in various ways (Bai & Ng, 2008; Bernanke, Boivin, & Elias, 2005; Stock & Watson, 2005, 2011). As Figure 1 shows, we use a covariance $\psi_{u_{it}}^{(xy)}$ among impulses $u_{it}^{(x)}$ and $u_{it}^{(y)}$, but again, $\psi_{u_{it}}^{(xy)}$ is not a residual covariance and instead represents something unique in a unit i at a time t that is shared by two variables.

The *co-movement* $\psi_{u_{it}}^{(xy)}$ aids in interpreting CL effects as causal by accounting for predictor covariance—this is how regression accounts for confounds. To understand this, we make the crucial point that AR and CL terms can be seen as modeling the persistence of past impulses (Hamaker, 2005), essentially reflecting how long these last as they propagate into the future. For example, how long does a random change to work stressors, organization performance, or national income last—conversely, how long until it fades? Figure 1 shows this by path tracing from an impulse to future observations on AR and CL paths. Impulses persist longer if AR and CL terms are larger (persisting indefinitely if an AR term $\beta_{y1}^{(y)} = 1$); conversely, regression to the mean occurs more quickly if they are smaller. We treat this further in Online Appendix B but for now emphasize that a co-movement $\psi_{u_{it}}^{(xy)}$ can help in understanding AR and CL terms as capturing the independent effects of past impulses.

Summary and limitations. The cross-lagged model has many useful properties. It controls for occasion effects α_t while modeling system dynamics with AR effects $\beta_{y1}^{(y)}$ and $\beta_{x1}^{(x)}$ and CL effects $\beta_{y1}^{(x)}$ and $\beta_{x1}^{(y)}$. Here, AR terms are meant to indicate the proportion of the past that directly persists to the future on a variable, which is meant to serve as a control when assessing direct effects of other variables with CL terms. What remains should be random impulses $u_{it}^{(x)}$ and $u_{it}^{(y)}$ with variances $\psi_{u_{it}}^{(x)}$ and $\psi_{u_{it}}^{(y)}$ and a co-movement $\psi_{u_{it}}^{(xy)}$, all of which are meant to reflect factors uniquely affecting a unit i at a time t —again, we assume that x and y are error free, but we later note that this can be easily addressed in SEM.

However, there are two limitations of this approach that motivate a GCLM. First, all units are treated as if they were the same in the long run—Figure 1 does not reflect any stable between-unit differences. This is anathema to organization research in which individual and organizational differences such as personality or culture are well recognized and units differ systematically over time. By failing to model stable factors, they will be confounded with the system dynamics that should be reflected by AR and CL terms (Hamaker et al., 2015). Thus, a more general model is needed to account for stable factors, which we will call *unit effects*.

The second limitation is that the dynamic process linking the past and the future via AR and CL terms is assumed to follow a simple, indirect-effects structure. As we noted, AR and CL terms depict persistence (or regression to the mean) of a past impulse, but this might persist (or fade) in complex ways. Thus, a more general model may help to overcome the indirect-effects structures associated with AR and CL terms, which we will treat in the next section using *moving average (MA)* and *cross-lagged moving average (CLMA)* terms.

A General Cross-Lagged Model

To generalize the cross-lagged model, we now sequentially introduce unit effects as well as MA and CLMA terms. In doing so and in what follows, we draw on three modeling traditions: (a) vector autoregressive (VAR) models (Canova & Ciccarelli, 2013; Lütkepohl, 2005; Sims, 1980), (b) vector autoregressive moving average (VARMA) models (Box et al., 2008; Browne & Nesselroade, 2005), and (c) dynamic panel data models (Arellano, 2003; Arellano & Bond, 1991; Baltagi, 2014; Hsiao, 2014). From this work, we take the idea that processes and effects may be more complex than AR and CL terms imply. Furthermore, there may be stable factors that differentiate units of analysis over time, to which we now turn.

Unit Effects λ, η_i . Researchers often seek to explain two distinct causes of variation in people, organizations, and larger entities. On the one hand there is variation *within* units as each one changes relative to itself over time; AR, CL, and impulse terms capture these dynamics as units experience random shocks that persist via AR and CL paths. On the other hand, units may systematically differ from each other, producing variation *between* units of analysis due to factors that create stability rather than occasion-specific change.

To elaborate, if a unit i is a person, psychological factors can explain stability over time, including stable patterns of embodied cognition (Barsalou, 2008), social roles and norms (Andersen & Chen, 2002; Fournier, Moskowitz, & Zuroff, 2008), cognitive ability (Deary, Pattie, & Starr, 2013), personality or affective traits (Matthews, Deary, & Whiteman, 2003), and habits of thought/action that emerge in stabilized person-environment interactions (Fleeson, 2001; Mischel, & Shoda, 2008; Neal, Wood, & Quinn, 2006). Alternatively, if i is a group, organization, or a nation, substantial scholarship treats how collectives emerge as stable entities, such as by the formation of institutions (March & Olsen, 1989) and collective routines to guide social and material processes (Feldman & Orlikowski, 2011; Winter, 2013).

Such causes of between-unit differences are not the same as causal effects among variables as they change over time (Allison, 2005; Hamaker et al., 2015). Instead, between-unit differences are akin to unit-specific trends (e.g., long-run averages) that systematically differentiate units over time (i.e., between-unit differences). These should not confound the AR, CL, and impulse terms that represent perturbations around any such trends (see Online Appendix B) because stable factors are constant by definition and thus do not have a clear role in models of causality *over time*. To account for this, we treat each unit i as a function of unit-specific factors that are constant or nearly constant over T , modeled as a unit effect η_i , so x_{it} and y_{it} are functions of $\eta_i^{(x)}$ and $\eta_i^{(y)}$ with variances $\psi_{\eta}^{(x)}$ and $\psi_{\eta}^{(y)}$, respectively.

Unlike u_{it} , η_i captures factors that make a unit similar to itself (rather than different) over time. This makes η_i similar to a familiar common factor, which some researchers may think of as a “random intercept” in a multilevel model (see Hamaker et al., 2015). However, by including a covariance $\psi_{\eta}^{(xy)}$ for unit effects (see Figure 2), stable factors are controlled when estimating AR, CL, and impulse terms; in econometrics, this is often called a “fixed effects” model because covariance among lagged predictors and unit effects is accounted for. Indeed, this covariance is crucial for holding stable factors constant across occasions.

Although everything changes with time, we model η_i because ignoring it assumes that no stable between-unit differences exist (Bond, 2002; Shrout, 2011), which may be difficult to justify (Roberts & DelVecchio, 2000). This said, there are reasons to allow the effect of η_i to vary over time. For example, situations differ in their effects on individual behavior, with constraining situations altering trait expressions (Cooper & Withey, 2009; Fleeson, 2004; see also Kanfer & Ackerman, 1989). Larger systems also change in how enduring characteristics are expressed (Gersick & Hackman,

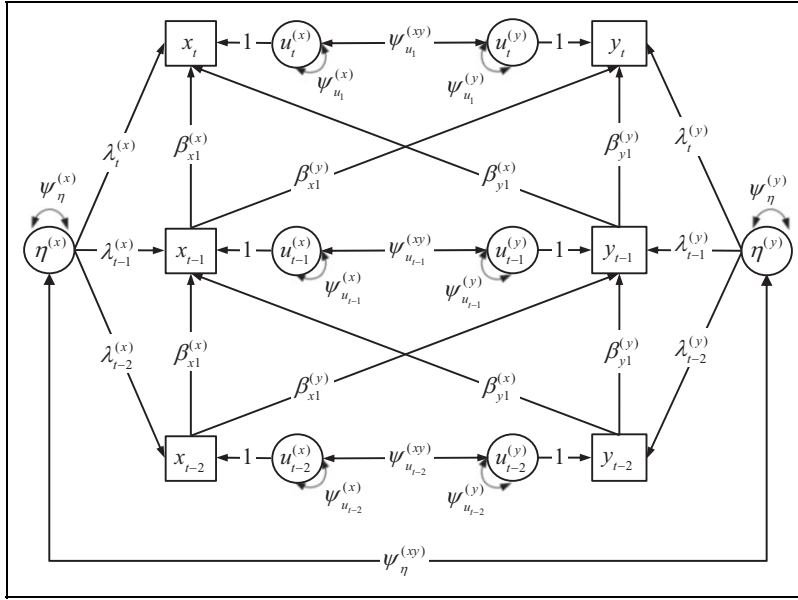


Figure 2. An AR(1)CL(1) model with unit effects.

1990; Kirkman & Rosen, 1999). We model such time-varying effects with occasion-specific “factors loadings” $\lambda_t^{(x)}$ and $\lambda_t^{(y)}$:

$$x_{it} = \alpha_t^{(x)} + \lambda_t^{(x)} \eta_i^{(x)} + \beta_{x1}^{(x)} x_{it-1} + \beta_{y1}^{(x)} y_{it-1} + u_{it}^{(x)} \quad (3)$$

$$y_{it} = \alpha_t^{(y)} + \lambda_t^{(y)} \eta_i^{(y)} + \beta_{y1}^{(y)} y_{it-1} + \beta_{x1}^{(y)} x_{it-1} + u_{it}^{(y)} \quad (4)$$

This can be seen as a $\lambda_t \times \eta_i$ interaction, with either “an aggregate shock $[\lambda_t]$ that has individual-specific effects $[\eta_i]$, or a permanent characteristic $[\eta_i]$ that has changing effects over time $[\lambda_t]$ ” (Arellano, 2003, p. 63; Bai, 2009, 2013; Moral-Benito, 2013). The common scenario where $\lambda_t = 1$ has been called *mean stationarity* (Bun & Sarafidis, 2015), with various ways to proceed if it does not hold (see Ahn, Lee, & Schmidt, 2001, 2013; Nauges & Thomas, 2003). We avoid this assumption, but as we note later, it can be imposed by fixing $\lambda_t = 1$ (except at a first occasion $\lambda_{t=1}$, which as we note later is freely estimated to reflect that η_i is actually a unit-specific average adjusted by AR/CL terms; Hamaker, 2005).

Moving average effects $\delta_{x1}^{(x)}$ and $\delta_{y1}^{(y)}$. To generalize model dynamics, we now introduce MA and CLMA terms. The idea motivating these is that **long-run and short-run dynamics may be different as impulses persist/fade over time, but AR and CL terms imply equivalent long- and short-run dynamics as a single set of parameters linking the past to the future**. Because AR and CL terms imply impulse persistence, relying on only them to capture system dynamics is akin to assuming that unexpected changes persist or fade multiplicatively vis-à-vis AR and CL terms. This can be modified by making the future a *direct* function of past impulses, which is how MA and CLMA terms modify the typical cross-lagged model.

We begin with **MA terms, which modify AR paths by making observations a direct function of past impulses** (Box et al., 2008; Hamaker, Dolan, & Molenaar, 2002). This allows MA terms to modify the short-run persistence of an impulse, whereas AR (and CL) terms still reflect long-run

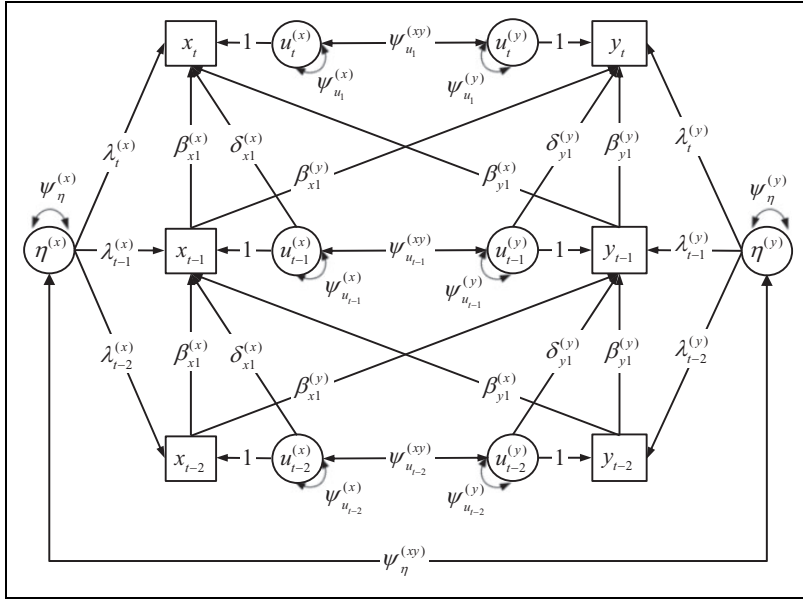


Figure 3. An AR(1)CL(1)MA(1) model with unit effects.

dynamics. As Figure 3 shows, x_{it} is a function of $u_{it-1}^{(x)}$ with an MA term $\delta_{x1}^{(x)}$ and y_{it} a function of $u_{it-1}^{(y)}$ with an MA term $\delta_{y1}^{(y)}$; a path from $u_{it-h}^{(y)}$ to y_{it} is $\delta_{yh}^{(y)}$:

$$x_{it} = \alpha_t^{(x)} + \lambda_t^{(x)} \eta_t^{(x)} + \beta_{x1}^{(x)} x_{it-1} + \beta_{y1}^{(x)} y_{it-1} + \delta_{x1}^{(x)} u_{it-1}^{(x)} + u_{it}^{(x)} \quad (5)$$

$$y_{it} = \alpha_t^{(y)} + \lambda_t^{(y)} \eta_t^{(y)} + \beta_{y1}^{(y)} y_{it-1} + \beta_{x1}^{(y)} x_{it-1} + \delta_{y1}^{(y)} u_{it-1}^{(y)} + u_{it}^{(y)} \quad (6)$$

By including MA terms, generality is added to the way that dynamic processes can be modeled—specifically, by allowing MA terms to modify the way AR terms imply short-run persistence of impulses. This is seen by path tracing in Figure 3, where **short-run persistence of an impulse is a sum of MA and AR terms (i.e., a total effect of $u_{it-1}^{(y)}$ on y_{it} is $\beta_{y1}^{(y)} + \delta_{y1}^{(y)}$)**, but AR (and CL) terms determine long-run dynamics as an impulse persists beyond the MA term into the future. The key to understanding how this works is by recognizing that **an impulse can only travel along an MA term once, whereas AR and CL terms repeatedly link the past to the future over time**. To show the value of differentiating short-run and long-run dynamics, we elaborate on two possible cases: (a) After an unexpected change in the form of an impulse, a system may stay perturbed in the short run but then rapidly adapt later versus (b) a system that adapts rapidly in the short run but then adapts very slowly later on.

To explain the first case, consider that **as MA terms become more positive ($\delta_{y1}^{(y)} > 0$), it allows for short-run persistence of a past impulse while still allowing small AR terms to imply very weak long-run persistence**. In the extreme case that AR terms equal 0, MA terms can reflect the persistence of an impulse to the next occasion $t + 1$ only, such as if changes in work stressors, HR practices, or national SWB persist to the next occasion but regress to the mean by the second occasion $t + 2$. This case could be common because individuals and organizations can fully adapt to the unexpected in some cases (Becker et al., 2006; Frederick & Loewenstein, 1999; Levitt & March, 1988), as can entire nations when agents respond in real-time to policy changes (see classic thought in Lucas, 1976).

In the second case, adaptation may occur rapidly at first but then slow over time. For example, there may be short-run adaptive responses to changes in counterproductive work behaviors (e.g., management interventions), organizational performance (e.g., increased competition), or national income (e.g., less stringent budget controls), but if these responses fade or become ineffective, then what remains of the initial change may persist. This is made possible by MA terms because as they become more negative (i.e., $\delta_{y1}^{(y)} < 0$), the short-run persistence of an impulse is reduced while positive AR terms allow what remains to persist. Consider if an AR term $\beta_{y1}^{(y)} = .95$ were combined with an MA term $\delta_{y1}^{(y)} = -.75$, allowing the persistence of an impulse $u_{it-2}^{(y)}$ to be a small $\beta_{y1}^{(y)} + \delta_{y1}^{(y)} = .2$ at the next occasion y_{it-1} . Yet, this would persist due to the AR term, with an effect of $u_{it-2}^{(y)}$ on y_{it} being $(\beta_{y1}^{(y)} + \delta_{y1}^{(y)})\beta_{y1}^{(y)} = .2 \times .95 = .19$. Alternatively, if an AR term $\beta_{y1}^{(y)} = 1$, the past would never fade in a system that is permanently impacted by its past (see Online Appendix B for insights).

To add additional generality to the model, higher-order MA lags may be included for q MA effects in an MA(q) model. Here, the sum of all MA terms is a shorthand for how MA effects from a single past u modify short-run persistence ($\delta_y^{(y)} = \delta_{y1}^{(y)} + \delta_{y2}^{(y)}$), with the total short-run persistence of an impulse then being $\beta_y^{(y)} + \delta_y^{(y)}$. As noted in the following, choosing an AR(p) and MA(q) model can be done by model checking, but theory and past findings are also key (Allen & Fildes, 2001, 2005; Armstrong, 2001, 2005, 2007, 2012; Armstrong, Green, & Graefe, 2015). Often, simpler models are better (Green & Armstrong, 2015), and “we rarely seem to meet situations for which either $p \dots$ or q need to be greater than 2. Frequently, values of zero or unity will be appropriate” (Box et al., 2008, p. 102).

Cross-lagged moving average effect $\delta_y^{(x)}$ and $\delta_x^{(y)}$. Just as AR and MA terms allow modeling a separate short-run and long-run dynamic structure, Figure 4 shows that the structure associated with CL terms can be extended analogously by making each unit's standing on an observed variable a direct function of other variables' past impulses. We call these CLMA terms, which arise when x_{it} is regressed on $u_{it-1}^{(y)}$ to estimate $\delta_{y1}^{(x)}$ and y_{it} is regressed on $u_{it-1}^{(x)}$ to estimate $\delta_{x1}^{(y)}$; a path from $u_{it-h}^{(x)}$ to y_{it} is $\delta_{xh}^{(y)}$. With single-order lags for all terms, we refer to this full GCLM as an AR(1)MA(1)CL(1)CLMA(1) model:

$$x_{it} = \alpha_t^{(x)} + \lambda_t^{(x)} \eta_i^{(x)} + \beta_{x1}^{(x)} x_{it-1} + \delta_{x1}^{(x)} u_{it-1}^{(x)} + \beta_{y1}^{(x)} y_{it-1} + \delta_{y1}^{(x)} u_{it-1}^{(y)} + u_{it}^{(x)} \quad (7)$$

$$y_{it} = \alpha_t^{(y)} + \lambda_t^{(y)} \eta_i^{(y)} + \beta_{y1}^{(y)} y_{it-1} + \delta_{y1}^{(y)} u_{it-1}^{(y)} + \beta_{x1}^{(y)} x_{it-1} + \delta_{x1}^{(y)} u_{it-1}^{(x)} + u_{it}^{(y)} \quad (8)$$

By including CLMA terms, the model changes how causal effects can be understood. As noted previously, lagged effects can be seen as implying an effect of past impulses on future observed variables. In turn, just as the short-run persistence for a variable becomes AR+MA, the short-run effect of one variable on another becomes CL+CLMA. As Figure 4 shows, $u_{it-1}^{(x)}$ impacts y_{it} via CL and CLMA paths, so this is akin to estimating an effect of $u_{it-1}^{(x)}$ on y_{it} as the short-run effect $\beta_{x1}^{(y)} + \delta_{x1}^{(y)}$. The idea here is that impulses can be treated as causes, which has been called Granger-Sims causality (Kuersteiner, 2010; Lechner, 2010).

This type of causality uses an interesting through experiment to ground it: Consider that if impulses are random, then it is as if a natural experiment were done at each occasion by randomly assigning units to a new level on a variable (e.g., $u_{it} = +1$). Thus, impulses act as a conceptual randomization device, with CL+CLMA terms showing the short-run effects of random interventions (see Dufour, Pelletier, & Renault, 2006; Dufour & Renault, 1998; Dufour & Tessier, 1993; Lütkepohl, 1993; Sims, 1980). Indeed, longitudinal methods that use planned or natural experiments can rely on this logic by using treatment variables as predictors of random impulses (i.e., putting a

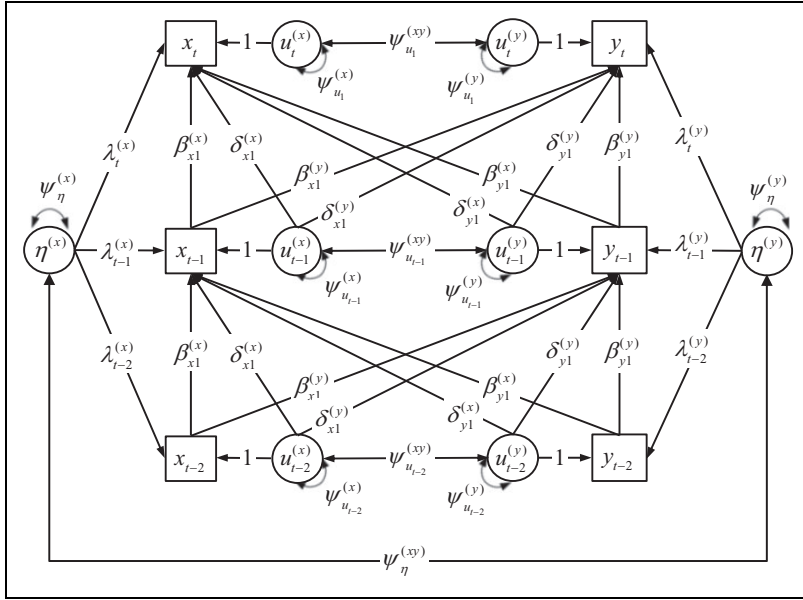


Figure 4. A full GCLM, AR(1)MA(1)CL(1)CLMA(1) model with unit effects.

time-varying treatment variable “behind” each random impulse; Angrist & Kuersteiner, 2011; Bojinov & Shephard, 2017; Stock & Watson, 2018). Of course, not everyone endorses the idea that impulses approximate randomization, but the fact that cross-lagged models are common and can be shown to rely on past impulses (see Online Appendix B) may help readers appreciate this kind of thought experiment.

The issue of causality aside, CLMA terms offer pragmatic value by allowing complex forms of dependence among variables. We treat this by analogizing the two previous MA cases. The first involved delayed adaptation, such that an unexpected change in a variable has an effect on the future of another, but adaptation then limits the duration of effects. This could be a case of short-lived effects of work stressors on counterproductive behaviors, HR practices on performance, or national income on SWB as each system adapts to the change. Here, an impulse $u_{it-1}^{(x)}$ may have a large positive CLMA term but a small CL term associated with y_{it} , so the early effect does not “repeat” at each future occasion of x . Indeed, as Figure 4 shows, if AR and CL terms were zero, MA and CLMA terms can model a system that fully adapts to previous impulses in all modeled variables (i.e., regression to the mean at $t + 2$).

The second case involved a small short-run effect that is highly persistent, such as a reverse-causal case of counterproductive behaviors affecting work stressors, organizational performance affecting HR practices, or national SWB affecting income, but with each effect being small yet long-lived over time. For this, an impulse $u_{it-1}^{(x)}$ may have a negative CLMA term $\delta_{x1}^{(y)} = -.75$ and a large positive CL term $\beta_{x1}^{(y)} = .95$ associated with the outcome y_{it} . In this case, the short-run effect is positive and small $\beta_{x1}^{(y)} + \delta_{x1}^{(y)} = .20$, but this small effect can persist via CL (and AR) terms due to the large $\beta_{x1}^{(y)} = .95$ (see also Online Appendix B).

The point is that CLMA terms add generality to the kinds of dynamics that can be modeled. For this purpose, researchers may include l higher-order CLMA terms in a CLMA(l) model, such as if $l = 2$ for a CLMA(2) model. Again, the CLMA effects from a single past u can act as a kind of shorthand indicating how CLMA terms modify short-run effects ($\delta_{x1}^{(y)} = \delta_{x1}^{(y)} + \delta_{x2}^{(y)}$), which in turn implies that an overall short-run effect of x ’s impulses on y is $\beta_{x1}^{(y)} + \delta_{x1}^{(y)}$. Given

the potential complexity of models with such lags, we now offer ways to simplify hypothesis testing with the GCLM.

Hypothesis Testing With the GCLM

To facilitate testing hypotheses with the GCLM, there are methods that can be easily implemented even if the models are very complex (e.g., many higher-order lags). As we now describe, short-run effects can be evaluated with Granger-Sims causality tests, whereas long-run effects can be evaluated with impulse responses that indirectly link past impulses to future observed variables over time. We now treat each of these in turn.

Short-Run Effects: Granger-Sims Tests

To facilitate hypothesis testing with the GCLM, we offer a four-step process that is easy to use in SEM software (inspired by Granger, 1969; Sanggyun & Brown, 2010; Sims, 1980, 1986). The method maps onto the Granger-Sims logic that impulses on one variable can be understood as causes of future observations on another. This is a test for short-run effects because our four steps only assess the direct effects of past impulses—short-run effects are direct effects; long-run effects involve indirect effects. For this, null hypothesis significance tests can be used, but we use fit criteria to balance parsimony and statistical fit.

Step 1: Estimate a panel data model of interest, such as the full GCLM in Figure 4 and Equations 7 and 8, and obtain model fit information such as information criteria (e.g., Akaike Information Criterion or Bayesian Information Criterion).

Step 2: Test an $x \rightarrow y$ effect by constraining CL and CLMA terms linking x 's impulse $u^{(x)}$ to y , such as $\beta_{x1}^{(y)} = \delta_{x1}^{(y)} = 0$. In the presence of higher-order lags, only constrain effects on variables at occasions that are subject to all lagged effects (i.e., only constrain effects "entering" occasions that are endogenous to all AR, CL, MA, and CLMA paths). Effects prior to this are not the full model being tested and may be due to unmodeled effects before $t = 1$ (of note is that our online Excel worksheet facilitates this by only labeling parameters that should be constrained to zero). After instituting the appropriate constraints, evaluate model fit and compare it to Step 1, such as larger information criteria indicating that the model in Step 2 is less justified, in turn implying $x \rightarrow y$ causation. Any inferred short-run effect of $u^{(x)}$ on y is then the sum of CL and CLMA terms (e.g., $\beta_{x1}^{(y)} + \delta_{x1}^{(y)}$).

Step 3: Test a $y \rightarrow x$ effect with the same approach, comparing results to Step 1.

Step 4: Test $x \rightarrow y$ and $y \rightarrow x$ "feedback" or "reciprocal effects" with all constraints from Steps 2 and 3 and compare to Step 1. If feedback exists, then intervening to change $u^{(x)}$ or $u^{(y)}$ may impact y , x , or both via feedback. This test may help provide evidence, for example, of "a vicious cycle" of effects among x and y over time (Meier & Spector, 2013, p. 537).

However, these four steps only offer a picture of short-run effects rather than the form effects take over time (Dufour et al., 2006; Dufour & Renault, 1998; Hsiao, 1982; Lütkepohl, 1993). Consider that with more than two variables such as x , m , and y , there may be a direct effect $x \rightarrow y$ and an indirect effect such as $x \rightarrow m \rightarrow y$ over time, but only the former is tested. To tackle these issues, we now treat long-run effects using the logic of impulse responses.

Long-Run Effects: Impulse Responses

Although tests for short-run effects are common, their results may not be useful for planning interventions, which requires predicting the results of actions over time (Cartwright & Hardie,

2012). For this, we use impulse responses, which we treat as total effects of a past impulse on future observations over time, including all indirect effects via AR and CL paths (Lütkepohl, 2005; Sims, 1980; Stock & Watson, 2005). Impulses are the focus because, as Figure 4 shows, “changes in the variables are induced by non-zero residuals, that is, by shocks Hence, to study the relations between the variables, the effects of . . . shocks are traced through the system” (Lütkepohl, 2013, p. 154). Indeed, methods to account for natural or planned experiments can adopt this logic by modeling impulses via time-varying treatment variables (Angrist & Kuersteiner, 2011; Bojinov & Shephard, 2017; Stock & Watson, 2018).

By conceptualizing u_{it} as random, it is as if a person, organization, or a nation were randomly assigned to a level of an observed variable at each t (e.g., $u_{it} = +1$), allowing a thought experiment that traces the effects of this as a set of “impulse responses” over time. These are estimable in various ways (see Box et al., 2008; Lütkepohl, 2005; Stock & Watson, 2005), but a familiar way is to estimate total effects (direct + indirect) of an early impulse on future observations. For example, Figure 4 implies four impulse responses: from $u_{i1}^{(x)}$ and $u_{i1}^{(y)}$ to future x and y at each future occasion. Plots of the effects offer a simple way to see how interventions may work (e.g., Sims, 1992, pp. 982–994; the reader may skip to Figures 5a–5d).

By estimating and plotting these effects and their confidence intervals (CIs), researchers can test hypotheses that map more directly onto research questions such as if “human capital investments . . . can yield a high return” (Van Iddekinge et al., 2009, p. 840) or if, in terms of SWB, “people [do] not adapt to income increases” (Diener et al., 2013, p. 275). Impulse responses can show such effects across all paths modeled in a GCLM. Indeed, in the case that effects do not fade due to AR or CL terms = 1 (see Online Appendix B), impulse response analysis offers a simple way to show how all lagged parameters may imply persistent effects in a studied time frame.

This said, impulse response analysis has limitations. Some of these we treat later, but for now we note that the earliest impulse that can be used has a t equal to a model’s highest lag order. This is because higher-order lags involve missing MA and CLMA terms in early occasions (as we note in our next section). Thus, impulse responses must begin at the first impulse with all modeled effects “leaving” the impulse. Also, as is well known for mediation analysis, indirect effect estimates are not normally distributed, so testing can be done using bootstrapped CIs or Bayesian analogues (Dufour et al., 2006; Kilian, 1999; Wright, 2000).

SEM Specification and Estimation

To model panel data, SEMs are useful because of their flexibility (Allison, 2005; Bollen & Brand, 2010). Due to its generality and stable algorithms, we use the approach found in Mplus (see Asparouhov & Muthén, 2014; Muthén, 2002; Muthén & Asparouhov, 2009; L. K. Muthén & Muthén, 1998–2018). As a special case of this, we show an SEM as:

$$\mathbf{y}_i = \mathbf{v} + \mathbf{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i \quad (9)$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i \quad (10)$$

with all terms typically understood as follows: \mathbf{y}_i is a k -length vector of observed variables, \mathbf{v} is a k -length vector of intercepts, $\mathbf{\Lambda}$ is a $k \times m$ matrix of factor loadings for m latent variables, $\boldsymbol{\epsilon}_i$ is a k -length vector of residuals with $k \times k$ covariance matrix Θ , $\boldsymbol{\eta}_i$ is an m -length vector of latent variables, $\boldsymbol{\alpha}$ is an m -length vector of intercepts or means, \mathbf{B} is an $m \times m$ matrix of regression coefficients, and $\boldsymbol{\zeta}_i$ is an m -length vector of residuals with an $m \times m$ covariance matrix Ψ . This structure is usually used to correct for measurement error by estimating terms in \mathbf{v} , $\mathbf{\Lambda}$, and Θ , with error-free latent terms in $\boldsymbol{\eta}_i$ and Ψ .

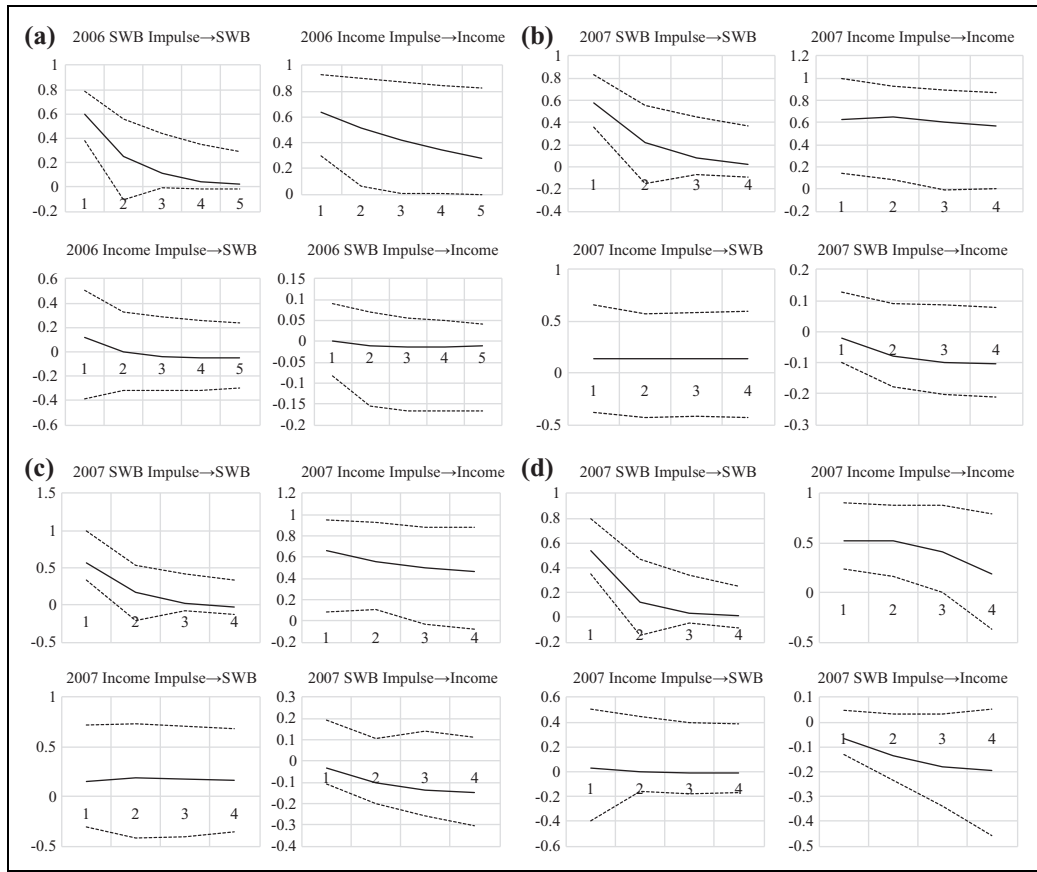


Figure 5. (a) Impulse response functions for AR(1)MA(1) model. Note: The y-axis is effect estimates, and the x-axis is the response horizon in years so that the plotted lines indicate the effect of a 1-unit impulse in 2006 over the next 5 years. Solid lines represent effect estimates; dotted lines represent 97.5% and 2.5% confidence intervals obtained using a nonparametric bootstrap with roughly 15,000 replications. Impulse responses begin at the first occasion $t = 1$ because the highest lag order in the model = 1. (b) Impulse response functions for AR(1)MA(2) model. Note: See the Note for Figure 5a, except the impulse begins in 2007 at $t = 2$ (rather than 2006 at $t = 1$) because the highest lag order in the model = 2, so the first occasion is “lost” when estimating effects. Thus, we show the effect of a 1-unit impulse in 2007 over 4 years. (c) Impulse response functions for AR(2)MA(1) model. Note: See the Note for Figure 5b. (d) Impulse response functions for AR(2)MA(2) model. Note: See the Note for Figure 5b.

For concision we assume error-free measures that reduce Equation 9 to $\mathbf{y}_i = \mathbf{A}\boldsymbol{\eta}_i$ (but measurement error can be addressed by multiple-indicator models using Equation 9) so that \mathbf{A} maps each observed variable in \mathbf{y}_i onto an analog in $\boldsymbol{\eta}_i$. This means that $\boldsymbol{\eta}_i$ stacks all variables in our model, with occasion effects in $\boldsymbol{\alpha}$, regression paths and each “factor loading” (e.g., λ_t) in \mathbf{B} , and impulse and unit effect variances and covariance in $\boldsymbol{\Psi}$. Online Appendix C describes the model in greater detail, including an expansion of Equations 9 and 10 for a GCLM.

To estimate a GCLM, any unit effect $\boldsymbol{\eta}_i$ should be treated as latent with loadings estimated at a first occasion $t = 1$ (Allison, 2005; Bai, 2013; Moral-Benito, 2013). As our second paper shows, treating $\boldsymbol{\eta}_i$ as observed—by “within-group” centering or via dummy variables—causes “dynamic panel bias” in lagged effects (Bond, 2002; Hamaker & Grasman, 2015; Nickell, 1981). To avoid this, maximum likelihood (ML) or Bayes estimators treat $\boldsymbol{\eta}_i$ as missing (e.g., Dempster, Laird, & Rubin,

1977). Here, we use an ML estimator robust to data missing at random and non-normality (although these can be modeled; see Asparouhov & Muthén, 2016; Muthén & Asparouhov, 2015). Conveniently, ML performs well even in the presence of AR or CL terms that imply highly persistent processes (i.e., “unit roots” or “integrated” processes such as $\beta_{y1}^{(y)} \geq 1$; see Allison, Williams, & Moral-Benito, 2017; Binder, Hsiao, & Pesaran, 2005; Moral-Benito, 2013; Williams, Allison, & Moral-Benito, 2018).

For the sake of concision, we describe model identification conditions in Online Appendix C but note that many combinations of AR, MA, CL, and CLMA lags are possible (i.e., different p , q , c , and l , respectively) and each will have unique identification conditions. Our online Excel worksheet automates Mplus input for models with different lag orders for different observed variables, but researchers should be aware of constraints on identification as lag orders increase. A basic GCLM with single lag orders is identified with $T \geq 4$, but even complex models will often be identified if $T \geq 6$ (for general insight, see Bollen, 1989).

Also, there are special considerations for model with higher-order lags, which become interpretable at the first occasion t that is subject to all lagged effects (i.e., when t equals the highest lag order in a model $+1$; see Online Appendix C). Thus, **the highest lag order is equal to the number of early occasions that are “lost” because they cannot be predicted by occasions before $t = 1$** . In these cases, the GCLM includes freely estimated AR and CL terms in early occasions to account for unmodeled effects prior to $t = 1$ (see Online Appendix C).

Finally, we offer a few comments about λ_t . *the impulse.* First, for the last occasion, $\lambda_{t=T} = 1$ is fixed to provide η_t an observed-variable scale. Second, given the structure of our model, the first $\lambda_{t=1}$ will capture unit effects due to unobserved past occasions $t < 1$ (see Hamaker, 2005). Thus, the first term $\lambda_{t=1}$ may not be of interest when interpreting effects. Due to this, **if researchers want a model that excludes time-varying terms (i.e., if $\lambda_t = 1$ is desired across all T), then only terms after the first occasion should be fixed, so that $\lambda_{t=1}$ is free and $\lambda_{t>1} = 1$** . Again, our online Excel worksheet automates this for Mplus program input.

Income and Subjective Well-Being

To illustrate model estimation and interpretation, we reanalyze data from Diener et al. (2013), who used Gallup World Poll data to study the relationship between SWB and income at the national level (other examples are in Online Appendix A). SWB was measured by self-rated life evaluations on a 0 to 10 scale; income was equivalized, log-transformed, and then multiplied by 2 to stabilize model estimation. With $N = 135$ nations from 2006 to 2011 ($T = 6$) and roughly 1,000 people responding for each country i at each year t , the data represent about 95% of the world’s adult population. The mean for each country i at each year t was computed to represent average income x_{it} and SWB y_{it} . Given the Gallup sampling technique, missing data are assumed missing at random. Descriptive statistics are in Table 2.

These data are useful for studying causal effects because income and SWB cannot be easily manipulated and methods with observed proxies for this can have strong assumptions (e.g., Ettner, 1996; Lindahl, 2005; Meer, Miller, & Rosen, 2003). Also, diverse causes can explain covariance in well-being and income. Deaton (2002) notes three possible cases for SWB or health: “[1] Income might cause health, [2] health might cause income, or [3] both might be correlated with other factors; indeed, all three possibilities might be operating” (p. 15; Deaton, 2003; see also Diener & Biswas-Diener, 2002). The GCLM addresses these issues as follows.

First, income may lead to SWB by reducing monetary stressors and increasing access to positive environments (Diener & Biswas-Diener, 2002). Such effects can be understood in relation to life circumstances and the relative comparisons that they allow (Clark, Frijters, & Shields, 2008; Frijters, Haiksen-DeNew, & Shields, 2005). Yet, second, some “literature has been skeptical about *any*

Table 2. Descriptive Statistics.

Variable	Mean	σ^2		Correlations											
SWB 2006	5.261	1.090													
SWB 2007	5.414	1.096	.942												
SWB 2008	5.394	1.095	.930	.948											
SWB 2009	5.425	1.072	.914	.899	.892										
SWB 2010	5.424	1.123	.922	.898	.888	.963									
SWB 2011	5.426	1.083	.887	.887	.889	.887	.923								
INC 2006	7.698	.870	.811	.778	.764	.680	.724	.710							
INC 2007	7.661	.914	.828	.824	.803	.751	.756	.752	.956						
INC 2008	7.728	.927	.824	.807	.791	.718	.742	.743	.939	.969					
INC 2009	7.727	.922	.832	.807	.788	.732	.763	.753	.952	.963	.984				
INC 2010	7.750	.901	.838	.809	.791	.744	.784	.782	.932	.963	.973	.978			
INC 2011	7.751	.898	.831	.803	.788	.734	.775	.779	.914	.953	.959	.971	.988		

Note: SWB = average subjective well-being; INC = average income logged; σ^2 = estimated population variance.

causal link from income . . . and instead tends to emphasize causality in the opposite direction” (Deaton, 2003, pp. 118-119). In terms of well-being, some research shows no lasting effect of income (Easterlin, Morgan, Switek, & Wang, 2012) but an effect of well-being on income via employment and other factors (Binder & Coad, 2010; De Neve & Oswald, 2012; Michaud & Van Soest, 2008; Oswald, Proto, & Sgroi, 2015). Still other studies find bidirectional causality or “feedback” effects (e.g., Chen, Clarke, & Roy, 2014; Devlin & Hansen, 2001; Erdil & Yetkiner, 2009; French, 2012), which many researchers propose should exist for various reasons (e.g., Deaton, 2003; Diener, 2012).

Third, in terms of confounding factors, our model controls for occasion effects (α_t) and other factors. For example, time-varying events such as changes in job status or family demands can impact income and SWB simultaneously (Lyubomirsky, King, & Diener, 2005). Such factors cause co-movement in impulses (i.e., $\psi_{u_t}^{(xy)}$), as found in a GCLM. Also, time-invariant factors produce stable covariance in SWB and income, such as persistent attitudes, behaviors, institutions, and cultures that can cause both income and SWB (Diener & Biswas-Diener, 2002; Ferrer-i-Carbonell & Frijters, 2004). The GCLM accounts for this with unit effects $\eta_i^{(x)}$ and $\eta_i^{(y)}$ that are controlled by including their covariance $\psi_{\eta}^{(xy)}$ while also allowing time-varying effects of such stable factors with factor loadings λ_t .

In sum, a GCLM helps in studying variables like income and SWB or health because researchers want to make causal inferences about them (e.g., Sacks, Stevenson, & Wolfers, 2012). However, weak methods often require admitting that “we shall have little to say about a causal interpretation” (Sacks et al., 2013, p. 8). By way of example, we now explore the process of GCLM specification and checking on the road to causal inference.

Model Specification

Causal inference with the GCLM requires choosing lag orders and some number of unit effects. To make this choice, alternative models can be compared, but this requires first choosing which models to specify for comparison. To guide this, **conservative models are typically best for out-of-sample generalizations, wherein conservatism means simpler models that rely on theory, past findings, and contextual information** (Allen & Fildes, 2001, 2005; Armstrong et al., 2015). We now motivate four such models for comparison.

Past research shows that SWB (y_{it}) is adaptive, implying impulses that revert to stable unit means (Clark et al., 2008; Ferrer-i-Carbonell & Frijters, 2004). This implies $AR < 1$ and a unit effect $\eta_i^{(y)}$. Yet, adaptation may be faster than AR terms allow (Binder & Coad, 2010; Di Tella, Haisken-De New, & MacCulloch, 2010; Stutzer & Frey, 2006). An MA term can accommodate this without requiring higher-order AR lags. Also, it may be able to assist with potentially complex dynamics caused by the global financial crisis (GFC), so we estimate one AR and MA parameter for an AR(1)MA(1) model for SWB y_{it} .

National income x_{it} is related to GDP, which is highly stochastic with countries having different GDP trends (Cogley, 1990; Fleissig & Strauss, 2001; Mankiw & Shapiro, 1985; Stock & Watson, 1988). This implies large AR terms and a unit effect $\eta_i^{(x)}$. Yet, there is debate about GDP dynamics (Murray & Nelson, 2000), which the GFC complicates. We tackle this with a time-varying unit effect $\lambda_t^{(x)} \eta_i^{(x)}$ and MA terms. Yet, higher-order AR and MA lags may be needed to account for the GFC, so we estimate four models: AR(1)MA(1), AR(1)MA(2), AR(2)MA(1), and AR(2)MA(2). Given that the model for SWB y_{it} is always AR(1)MA(1), when reporting results, we refer to the AR and MA lag orders for income x_{it} .

For the effects among income x_{it} and SWB y_{it} , past findings suggest only one CL and CLMA term may be needed (e.g., Diener et al., 2013). Thus, we model one of each for the effects among income x_{it} and SWB y_{it} (as a CL[1]CLMA[1] model for both variables).

Results for all models are in Table 3, with occasion effects omitted for concision and impulse/unit effect covariances standardized as correlations. Impulse responses for all models are in Figures 5a through 5d (generated as indirect effects from an initial impulse to future observed occasions using Mplus's "MODEL INDIRECT" command), with 95% bootstrapped CIs using 20,000 draws, with missing data in early periods reducing convergence to roughly 15,000. All Mplus input and output is available in our online materials, including an Excel worksheet used to create Mplus input and impulse responses for these four specific models.

Model Selection

Model selection can be done by substantive and statistical checking. We first offer a substantive interpretation of results by checking estimates for consistency with theory and contextual knowledge (see Table 3). For this, we rely on impulse responses because they simplify model comparisons in the presence of varying lag orders (see Figures 5a-5d). We then discuss the use of model fit indices for model selection.

Substantive checking. We first examine the SWB dynamics, with an AR(1)MA(1) structure in all four models. As expected, the persistence of impulses quickly falls (top-left of Figures 5a-5d), with impulses almost entirely faded by the fourth future year. Also, 95% CIs include zero by the second year, so statistical significance exists only for the direct effect of a past impulse. In Table 3, this is the combined AR and MA term $\beta_{y1}^{(y)} + \delta_{y1}^{(y)}$, which ranges from .535 to .595, so only 53.5% to 59.5% of a previous year's impulse persists to the next year. Also, AR parameters $\beta_{y1}^{(y)}$ range from .226 to .423, implying mean-reversion ($\beta_{y1}^{(y)} < 1$). Finally, the time-varying unit effects $\lambda_t^{(y)}$ are in a range consistent with stable unit effects, with the first $\lambda_{t=1}^{(y)}$ being large, as expected, and those in future occasions ranging from .455 to .765. In sum, all results show an expected mean-reverting process, and meaningful unit effects imply stability over time. Therefore, substantively, all models seem appropriate.

On the other hand, income dynamics tell a different story (see top-right of Figures 5a-5d). The AR(1)MA(1) and AR(2)MA(2) models in Figures 5a and 5d imply mean-reversion, with Table 3 showing the AR(1)MA(1) model's AR effect $\beta_{x1}^{(x)} = .813$ and the AR(2)MA(2) model's total AR effect $\beta_{x1}^{(x)} = .792$. However, time-varying unit effects $\lambda_t^{(x)}$ in Table 3 are small after the first $\lambda_{t=1}^{(x)}$,

Table 3. Model Results (models are referred to using the lag specification for income x_{it}).

Parameter	Model Estimates (SE)			
	AR(1)MA(1)	AR(1)MA(2)	AR(2)MA(1)	AR(2)MA(2)
SWB→SWB AR/MA Terms $\beta_{y1}^{(y)}$ and $\delta_{y1}^{(y)}$				
$\beta_{y1}^{(y)}$.423 (.391)	.390 (.359)	.301 (.330)	.226 (.223)
$\delta_{y1}^{(y)}$.171 (.339)	.191 (.319)	.261 (.282)	.309 (.203)
$\beta_{y1}^{(y)} + \delta_{y1}^{(y)}$.595** (.100)	.580** (.094)	.561** (.096)	.535** (.084)
Income→Income AR/MA Terms $\beta_{x1}^{(x)}$ and $\delta_{x1}^{(x)}$				
$\beta_{x1}^{(x)}$.813** (.096)	.958** (.127)	1.274** (.388)	1.920** (.255)
$\beta_{x2}^{(x)}$	—	—	-.275 (.373)	-1.129** (.387)
$\beta_{x.}^{(x)}$.813** (.096)	.958** (.127)	.998** (.082)	.792** (.207)
$\delta_{x1}^{(x)}$	-.175 (.149)	-.326 (.253)	-.612* (.295)	-1.394** (.352)
$\delta_{x2}^{(x)}$	—	.055 (.086)	—	.639 (.512)
$\delta_{x.}^{(x)}$	-.175 (.149)	-.271 (.192)	-.612* (.295)	-.755** (.271)
$\beta_{x.}^{(x)} + \delta_{x.}^{(x)}$.638** (.124)	.686** (.197)	.386 (.296)	.037 (.132)
Income→SWB CL/CLMA Terms $\beta_{x1}^{(y)}$ and $\delta_{x1}^{(y)}$				
$\beta_{x1}^{(y)}$	-.078 (.116)	.131 (.315)	.217 (.325)	-.016 (.135)
$\delta_{x1}^{(y)}$.194 (.274)	.009 (.249)	-.064 (.233)	.046 (.289)
$\beta_{x1}^{(y)} + \delta_{x1}^{(y)}$.116 (.250)	.139 (.164)	.153 (.150)	.029 (.194)
SWB→Income CL/CLMA Terms $\beta_{y1}^{(x)}$ and $\delta_{y1}^{(x)}$				
$\beta_{y1}^{(x)}$	-.021 (.042)	-.103 (.072)	-.102 (.068)	-.003 (.019)
$\delta_{y1}^{(x)}$.023 (.049)	.080 (.073)	.066 (.089)	-.066* (.033)
$\beta_{y1}^{(x)} + \delta_{y1}^{(x)}$.002 (.040)	-.023 (.040)	-.036 (.042)	-.069* (.027)
Co-Movement in Impulses $\psi_{u_t}^{(xy)}$ as Correlations				
$\psi_{u_1}^{(xy)}$.003 (.321)	.643 (.593)	.726* (.346)	.499 (.272)
$\psi_{u_2}^{(xy)}$.537* (.243)	.446* (.214)	.430* (.190)	.381 (.232)
$\psi_{u_3}^{(xy)}$.007 (.120)	.003 (.134)	.029 (.137)	.024 (.126)
$\psi_{u_4}^{(xy)}$.015 (.125)	-.023 (.123)	-.028 (.118)	-.114 (.124)
$\psi_{u_5}^{(xy)}$.384* (.151)	.321* (.137)	.304* (.128)	.305* (.140)
$\psi_{u_6}^{(xy)}$.168 (.134)	.107 (.133)	.059 (.151)	-.082 (.568)
Unit Effect Variances $\psi_{\eta}^{(y)}$ and $\psi_{\eta}^{(x)}$, and Covariance $\psi_{\eta}^{(xy)}$ as a Correlation				
$\psi_{\eta}^{(y)}$.376	.246	.286	.612
$\psi_{\eta}^{(x)}$.034	.021	.014	.041
$\psi_{\eta}^{(xy)}$.903** (.042)	.961** (.059)	.956** (.040)	.846** (.037)
Time-Varying Unit Effects ("factor loadings") $\lambda_t^{(y)}$ and $\lambda_t^{(x)}$ as Correlations				
$\lambda_1^{(y)}$	0.974** (0.013)	0.956** (0.063)	0.936** (0.083)	0.966** (0.011)
$\lambda_2^{(y)}$	0.578 (0.441)	0.478 (0.321)	0.520* (0.253)	0.732** (0.212)
$\lambda_3^{(y)}$	0.593 (0.436)	0.482 (0.319)	0.514* (0.252)	0.742** (0.209)

(continued)

Table 3. (continued)

Parameter	Model Estimates (SE)			
	AR(1)MA(1)	AR(1)MA(2)	AR(2)MA(1)	AR(2)MA(2)
$\lambda_4^{(y)}$	0.559 (0.436)	0.458 (0.303)	0.494* (0.242)	0.712** (0.212)
$\lambda_5^{(y)}$	0.625 (0.412)	0.52 (0.296)	0.549* (0.234)	0.765** (0.2)
$\lambda_6^{(y)}$	0.563 (0.446)	0.455 (0.327)	0.491 (0.258)	0.72** (0.215)
$\lambda_1^{(x)}$	0.916** (0.061)	0.734** (0.254)	0.677** (0.253)	0.936** (0.071)
$\lambda_2^{(x)}$	0.259* (0.109)	−0.031 (0.222)	−0.037 (0.148)	0.412 (0.223)
$\lambda_3^{(x)}$	0.207 (0.112)	0.152 (0.089)	0.091 (0.086)	0.107 (0.208)
$\lambda_4^{(x)}$	0.204* (0.103)	0.162* (0.074)	0.133* (0.054)	0.195 (0.19)
$\lambda_5^{(x)}$	0.187 (0.117)	0.150* (0.076)	0.115* (0.052)	0.203 (0.174)
$\lambda_6^{(x)}$	0.205 (0.108)	0.162* (0.082)	0.132* (0.056)	0.225 (0.206)
Fit Indices				
χ^2	84.750	78.648	76.981	58.937
df / k	39 / 51	36 / 54	36 / 54	35 / 55
CFI / TLI	.976 / .960	.978 / .959	.979 / .961	.988 / .977
RMSEA / SRMR	.093 / .019	.094 / .026	.092 / .031	.071 / .030
AIC / BIC	844.77 / 992.94	845.94 / 1002.82	845.32 / 1002.20	836.66 / 996.45
aAIC / aBIC	908.67 / 831.61	920.19 / 832.00	919.57 / 831.38	914.63 / 822.47

Note: Columns are named after the AR/MA specification for income. SWB = subjective well-being; AR = autoregressive; MA = moving average; CL = cross-lagged; CLMA = cross-lagged moving average; CFI = Confirmatory Fit Index; TLI = Tucker-Lewis Index; RMSEA = root mean squared error of approximation; SRMR = standardized root mean squared residual; AIC = Akaike Information Criterion; BIC = Bayes Information Criterion; aAIC = sample-size adjusted AIC; aBIC = sample-size adjusted BIC.

* $p < .05$. ** $p < .01$.

ranging from $-.031$ to $.162$. This indicates that unit effects account for little observed variation in income and therefore that countries regress to very similar means for income over time. This does not seem plausible given cross-national differences.

Alternatively, the AR(1)MA(2) and AR(2)MA(1) models in Figures 5b and 5c imply expected persistence in income, with Table 3 showing the AR(1)MA(2) model having an AR effect $\rho_{x1}^{(x)} = .958$ and the AR(2)MA(1) an overall AR effect $\beta_x^{(x)} = .998$ (Wald tests cannot distinguish these from 1). These effects are consistent with theory and past findings about the random-walk nature of income, but with an interesting twist: MA terms are negative, with the AR(1)MA(2) model having an overall $\delta_x^{(x)} = -.271$ and the AR(2)MA(1) model $\delta_{x1}^{(x)} = -.612$. This implies past impulses have an effect on the next year of roughly .6 (i.e., only 60% of an impulse carries over; see Figures 5b and 5c). Yet, large AR terms allow this to persist, so an impulse is not forgotten. This may be due to the GFC: Income is persistent as expected, but past impulses during the GFC do not fully carry over. Also, time-varying unit effects in Table 3 are modest after the first $\lambda_{t=1}^{(x)}$, ranging from $-.031$ to $.412$, arguing against mean reversion as AR terms ≥ 1 imply that unit effects may capture stable change (see Online Appendix B). In sum, AR(1)MA(2) and AR(2)MA(1) models seem appropriate. Given its high persistence shown in Figure 5b, we favor the AR(1)MA(2) model to capture income dynamics.

For the income→SWB effect (bottom-left of Figures 5a-5d), all impulse responses include zero in 95% CIs. The short-run effect is positive, with a CL+CLMA term $\beta_{x1}^{(y)} + \delta_{x1}^{(y)}$ ranging from $.029$ to $.153$ and SEs range from $.15$ to $.25$ ($p > .05$). Yet, this weak effect is still useful for model selection

when examining the AR(1)MA(2) model (Figure 5b), which shows an interesting result of income's persistence: An increase in income has a weak but persistent effect on SWB due to income's large AR term $\beta_{x1}^{(x)} = .958$, thus further affecting SWB via the CL term $\beta_{x1}^{(y)} = .131$. This shows how income's large AR term implies a lasting effect on SWB, even if SWB is mean-reverting, leading us to prefer the AR(1)MA(2) model.

Finally, the SWB→income effect shows 95% CIs include zero at all time horizons (bottom-right of Figures 5a-5d). Yet, unlike the income→SWB effect, the SWB→income effect tends to be negative, with the short-run effect $\beta_{y1}^{(x)} + \delta_{y1}^{(x)}$ ranging from $-.069$ to $.002$ in the four models (in Table 3). Although this effect is weak, our preferred AR(1)MA(2) model for income still illustrates an interesting implication of SWB and income dynamics: An SWB impulse has an initial negative effect on income that persists due to income's AR process, which becomes stronger due to SWB's AR process and income-SWB feedback. This shows how a mean-reverting process such as SWB can have lasting effects on a highly persistent process such as income, leading us to prefer the AR(1)MA(2) model.

Statistical checking. Many researchers agree that model selection should use indices balancing statistical fit with model parsimony (Allen & Fildes, 2001, 2005; Armstrong, 2007; Armstrong et al., 2015; Burnham & Anderson, 2004; Hu & Bentler, 1998, 1999; Lütkepohl, 2005). However, different communities use fit indices differently. SEM researchers typically make recommendations based on simulations (e.g., Hu & Bentler, 1998, 1999). This often results in recommending fit index cutoffs that are not specific to panel data or predicting the results of interventions. Researchers from other fields do not always appreciate this approach.

For example, **forecasters empirically examine fit index performance for out-of-sample predictions with real data** (Fildes & Ord, 2002; Makridakis & Hibon, 2000), showing that **accurate prediction can be less a function of fit indices than substantive checking and other factors** (see Allen & Fildes, 2001, 2005; Armstrong et al., 2015; Green & Armstrong, 2015). Economists agree, noting that “statistical fit is overemphasized as a criterion As a policymaker, I want to use models to help evaluate the effects of out-of-sample changes in policies” (Kocherlakota, 2010, p.17), which requires substantive and contextual reasoning. Therefore, we do not unconditionally endorse the use of cutoff criteria often found in the SEM community—at least until such cutoffs are examined for use with panel data.

Here, we advocate balancing concerns about fit with substantive checking and an interest in parsimony. **If SEM fit indices show serious problems, this may be cause for concern, but modest differences in fit or poor fit for a model that accurately depicts a known process seem acceptable.** When in doubt, “you should probably aim towards simplicity at the expense of good specification” (Allen & Fildes, 2001, p. 21). However, “[o]f course, any simple model may sometimes be too simple” (Bernanke & Blinder, 1988, p. 1), and therefore theoretical and contextual knowledge of the processes being modeled should always be used.

To illustrate model selection by statistical checking, **we use the following fit indices: standardized root mean square error of approximation (RMSEA), Tucker-Lewis index (TLI), Comparative Fit Index (CFI), Akaike Information Criterion (AIC), Bayesian or Schwarz Information Criterion (BIC), and sample-size adjusted version of the AIC and BIC.** We also report the standardized root mean square residual (SRMR) but emphasize the former indices for their balance of parsimony and fit. Examining these indices in Table 3 shows no serious problems with any single model and very modest differences in terms of CFI, TLI, RMSEA, and SRMR. The AIC favors the more complex AR(2)MA(2) model, and the BIC favors the more parsimonious AR(1)MA(1) model, which is expected (see Burnham & Anderson, 2004; Lütkepohl, 2005). This is reversed for the sample-size adjusted AIC and BIC.

Table 4. Granger-Sims Tests.

χ^2	CFI / TLI	Model Estimates		
		RMSEA / SRMR	AIC / BIC	aAIC / aBIC
Step 1: Derive Fit of Full Model				
78.65** (36)	.978 / .959	.094 / .026	845.94 / 1002.82	920.19 / 832.00
Step 2: Constraint All Income → SWB Effects				
75.35** (38)	.981 / .966	.085 / .021	843.66 / 994.73	907.56 / 830.24
Step 3: Constrain All SWB → Income Effects				
86.04** (38)	.975 / .957	.097 / .021	844.71 / 995.79	908.61 / 831.29
Step 4: Constraining all CL/CLMA Terms				
84.90** (40)	.977 / .961	.091 / .021	843.71 / 988.98	901.36 / 830.81

Note: Parentheses after χ^2 values are degrees of freedom. SWB = subjective well-being; CFI = Confirmatory Fit Index; TLI = Tucker-Lewis Index; RMSEA = root mean squared error of approximation; SRMR = standardized root mean squared residual; AIC = Akaike Information Criterion; BIC = Bayes Information Criterion; aAIC = sample-size adjusted AIC; aBIC = sample-size adjusted BIC.

** $p < .01$.

Importantly, our preferred AR(1)MA(2) model shows acceptable levels of fit using typical SEM indices (e.g., CFI = .978; TLI = .959; SRMR = .026; RMSEA = .094), but it is the worst model in terms of AIC and BIC indices. However, the differences are inconsistent across models and are often minor. Therefore, we favor a AR(1)MA(2) model because of its acceptable fit and because the substantive relationships it shows are consistent with theory.

It is notable that other procedures can be used for model checking, such as for nonlinearity and local misfit using modification indices, covariance residuals, and residual plots (Asparouhov & Muthén, 2014). This is often considered obligatory, so we do not treat it here.

Model Interpretation and Hypothesis Testing

Using the AR(1)MA(2) model for inference (Figure 5b), we do not expect our results to conform to past studies given the sizable unit effects for SWB and the standardized $\psi_{\eta}^{(xy)} = .961$. In terms of the income→SWB effect, Table 4 shows Granger-Sims tests for a short-run effect, illustrating its small magnitude with CFI, TLI, RMSEA, and all AIC and BIC values improving by eliminating the effect. Impulse responses show a weak but positive long-run effect, with an impulse on income persisting into the future but with the 95% CI always containing zero. In sum, we find no meaningful effect of income on SWB.

More interesting is the weak, negative effect of SWB on income, which is opposite of what is often found (Deaton, 2003; Diener et al., 2013). Supporting this effect in the short-run, Table 4 shows that removing it reduces fit via CFI, TLI, and RMSEA. Yet, AIC and BIC terms show improved fit. In terms of the long-run effect, all impulse response CIs encompass zero. In sum, the weak nature of the effect implies it is untrustworthy, but were it present, then it could be explained. For example, some research shows that positive psychological states can negatively affect motivation and resource allocation for goal pursuit (Vancouver, More, & Yoder, 2008). The effect is sensible if increasing SWB demotivates seeking economic welfare or reductions in SWB orient people toward economic welfare.

In sum, by including AR, MA, CL, and CLMA terms, as well as time-varying unit effects and occasion effects, we do not find strong associations between income and SWB, and the SWB→income effect we find is negative, which runs counter to results using typical cross-lagged models (e.g., Diener et al., 2013). As we show in our second paper, this may be due to uncontrolled unit effects and/or a need for MA and CLMA terms in such past studies.

Discussion

This is the first of two papers in which we synthesize, compare, and extend panel data methods using SEM. In this first paper, we proposed a new panel data model, the GCLM, to incorporate stable factors in the form of unit effects while expanding the range of dynamic processes that can be modeled by using MA and CLMA terms. We treated these parameters and their application, covering model specification, checking, and interpretation by studying income-SWB dynamics, which did not support previous findings of positive effects among these variables (e.g., Diener et al., 2013). This suggests reappraising the sign and magnitude of income-SWB effects (Easterlin, 1995, 2001). We now conclude with thoughts on causal inference, starting with threats to this inference; Online Appendix D treats ways to modify the GCLM, including interactions, random slopes, and nonstandard measurement occasions.

Threats to Causal Inference: Trends and Regime Changes

To interpret GCLM results, it is important to address two threats to causal inference (Clements & Mizon, 1991): trends, including seasonal or cyclical effects, and changes in how a system functions, or regime changes (Granger & Newbold, 1974; Lütkepohl, 2005; Sims, Stock, & Watson, 1990). Grappling with these is important because if they exist, they may drive observed relationships rather than the random impulses that are meant to justify causal inference (Hendry, 2004). To raise awareness of these threats, we discuss each in turn.

Concerns over trends have generated substantial work (Harvey, 1985, 1997; Stock & Watson, 1988, 1999), covering unique types of trends: long-run trends due to things like maturation, periodic trends such as seasonal effects or circadian rhythms, cycles that wax and wane unpredictably (e.g., business cycles or depressive states), and random or stochastic trends caused by persistent impulses. Our model accounts for these in five ways: (a) An occasion effect α_t allows for aggregate trends; (b) a time-varying unit effect $\lambda_t \eta_i$ accounts for unit-specific trends; (c) persistent impulses are modeled by AR/CL terms; (d) impulse terms $\psi_{u_i}^{(y)}$, $\psi_{u_i}^{(x)}$, and $\psi_{u_i}^{(xy)}$ are free to reflect trending variances; and (e) MA/CLMA terms can account for some short-lived trends or cycles (Box et al., 2008; Granger & Morris, 1976).

However, additional tools may be required. For example, periodic trends like seasons or times of day can be modeled with latent variables (similar to “common methods factors”), or latent variables can act as additional unit effects to model unit-specific cyclical trends (e.g., a term $\lambda_{2t}^{(y)} \eta_{2i}^{(y)}$; see Bollen & Curran, 2006). Alternatively, trends due to persistent and complex dynamics and can be treated with higher-order AR, MA, CL, and CLMA terms.

This said, certainty about the existence of trends is often impossible (Heckman, 1991; Stock & Watson, 1999). Although de-trending data is often recommended (e.g., Curran & Bauer, 2011; Curran, Lee, Howard, Lane, & MacCallum, 2012; Hoffman & Stawski, 2009), there is no single way to do this, and tests for trends are often ambiguous (Davidson, 2013; Haldrup, Kruse, Teräsvirta, & Verneskov, 2013). The fact is that the evolution of any system involves mixtures of multiple processes, leading some to say that “no one really understands trends, even though most of us see trends [in] data” (Phillips, 2003, p. C35; Heckman, 1991). Also, visual inspections and detrending methods may be useful for $N = 1$ cases (see Jebb & Tay, 2016; Jebb, Tay, Wang, & Huang, 2015), but this is impractical with larger N . In the face of uncertainty, unit effects automatically de-trend data, but theoretical and contextual knowledge about a process can also be used (Allen & Fildes, 2001; Armstrong et al., 2015).

Next, regime changes refer to changes in the way a system functions over time—such as when water turns to ice, a person gets a new job, or an organization changes strategy. The idea is that there is a threshold beyond which a system functions differently, complicating prediction and causal

inference (see Bak, 1996; D'Souza & Nagler, 2015). To investigate this, increased variances and large AR terms may be observed due to chaotic behavior that occurs during a change (Carpenter et al., 2011; Dakos, van Nes, D'Odorico, & Scheffer, 2012; although see Hastings & Wysham, 2010). This may be part of a “critical slowing” in a system’s ability to recover from impulses (Scheffer, Carpenter, Dakos, & van Nes, 2015; Scheffer et al., 2009). The idea is that feedback mechanisms can become coupled in a system, causing it to become chaotic (Brock & Carpenter, 2010), wherein impulses are amplified or “accelerated” (similar to Bernanke & Mihov, 1998; Kiyotaki & Moore, 1997, 2002).

For example, consider people who experience multiple impulses in succession, such as job loss and a spouse’s death. Variability in emotions may increase as people try to cope, and AR effects may increase as emotions are no longer mean-reverting and people slip into depression (Van de Leemput et al., 2014). Such regime changes complicate causal inference and can be expected in complex systems subjected to random events in the form of impulses (Clements & Hendry, 2001; Hendry & Mizon, 2005; Stock & Watson, 1996).

There are multiple ways to handle regime changes, such as with time-varying AR, MA, CL, and CLMA terms to reflect parameter changes (Bringmann et al., 2016), while keeping in mind that this makes a model sensitive to noise (Boldea & Hall, 2013; Perron, 2006; Stock & Watson, 2009). As with trends, there is no magic bullet for regime changes, and their existence is often uncertain (Badagían, Kaiser, & Peña, 2015). Our model can account for some regime changes with an occasion effect α_t , a time-varying unit effect $\lambda_t \eta_i$ (and covariance $\psi_{\eta}^{(xy)}$), and impulse terms $\psi_{u_t}^{(y)}$, $\psi_{u_t}^{(x)}$, and $\psi_{u_t}^{(xy)}$ that are free to vary. Theory and contextualized knowledge of a process can be used to address additional concerns (Allen & Fildes, 2001; Armstrong et al., 2015).

Causal Inference Under Uncertainty

Even when tackling trends and regime changes, our approach is not without criticism, typically because it does not model the effects of randomly assigned interventions (Holland, 1986; Rubin, 2011). Without this, we theorize impulses as being akin to random assignment (see Lütkepohl, 2013; Sims, 1980, 1992; Stock & Watson, 2005, 2011). Yet, the validity of this theorizing is debatable, as in economics where GDP impulses are said to be due to improved technology (Christiano, Eichenbaum, & Evans, 1999). Also, interpreting impulse responses is complicated by correlated impulses (i.e., $\psi_{u_t}^{(xy)}$) because a “thought experiment of changing one error while holding the others constant makes most sense when the errors are uncorrelated” (Stock & Watson, 2001, p. 106). If this is false, an “analysis of the evolution of the system caused just by an innovation in one variable may not be appropriate” (Swanson & Granger, 1997, p. 357). To justify interpreting impulse responses, Table 3 shows modest co-movements (from $-.114$ to $.384$), but this cannot assuage more fundamental concerns.

However, such concerns should not derail using models like the GCLM. Consider that many researchers use cross-sectional regression methods, which in our model can be done by regressions among impulses $u_{it}^{(x)}$ and $u_{it}^{(y)}$ or unit effects $\eta_i^{(x)}$ and $\eta_i^{(y)}$. Yet, this requires assuming a single direction of causality, which models like the GCLM avoid while including lags that are consistent with causality (Sims, 1980; Uhlig, 2005). Although cross-sectional regressions are common, they lack a temporal ordering that defines causality, and “[a]lthough the phrase ‘instantaneous causality’ is somewhat useful on occasion, the concept is a weak one” (Granger, 1980, p. 340). Indeed, if factors like $\eta_i^{(x)}$ and $\eta_i^{(y)}$ are stable, they have no link to causality vis-à-vis a temporal ordering or interventions (Freedman, 2004; Holland, 1986, 2008; Winship & Morgan, 1999; although see Greiner & Rubin, 2011).

For these reasons, we focus on where change seems possible (as in Hamaker, 2012; Molenaar, 2004). For this, we emphasize impulses, which is useful for variables subject to random variation

and difficult to be experimented on (Aalen, Røysland, Gran, & Ledergerber, 2012; Dominici, Greenstone, & Sunstein, 2014; Granger, 1980, 1986, 1988, 2003). Of course, this approach has assumptions, but all methods have assumptions that must be balanced with their uses (Cartwright, 2007, 2009; Freedman, 2004; Sekhon, 2009). Even experiments have been criticized because they do not describe how to translate effects into interventions across contexts (Cartwright, 2011, 2012; Cartwright & Munro, 2010; Deaton & Cartwright, 2016).

The problem with all methods for causal inference is that their aim is to guide real-world action, but the consequences of action can never be predicted with certainty (Schön, 1995; Stone, 1989). Thus, even idealized methods such as randomized controlled trials (RCT) cannot enable unconditional inference because there are always gaps between evidence, action, and its consequences (Deaton & Cartwright, 2016). As Cartwright and Hardie (2012) explain:

You want evidence that a policy will work here, where you are. Randomized controlled trials do not tell you that. They do not even tell you that a policy works. What they tell you is that a policy worked there, where the trial was carried out Our argument is that the changes in tense—from “worked” to “work” to “will work”—are not just a matter of grammatical detail. To move from one to the other requires hard intellectual and practical effort. The fact that it worked there is indeed fact. But for that fact to be evidence that it will work here, it needs to be relevant to that conclusion. To make RCTs relevant you need a lot more information. (p. ix)

The point is that although RCTs are often seen as a gold standard for causality, the kind of relationships that they establish may be situated in contexts that do not help plan an intervention elsewhere. Furthermore, RCT findings may not be useful if trying to understand phenomena over time. Panel data models like ours may not offer the benefits of RCTs, but they have other virtues. In the end, models and experiments cannot predict the future with certainty. In the face of uncertainty, the GCLM is a useful complement to other methods, allowing researchers to assess short-run and long-run effects in dynamic processes.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by Australian Research Council's Future Fellowship scheme (project FT140100629).

Supplemental material

Supplemental material for this article is available online at <http://journals.sagepub.com/doi/suppl/10.1177/1094428119847278> and <https://doi.org/10.26188/5c9ec7295fed>.

References

- Aalen, O. O., Røysland, K., Gran, J. M., & Ledergerber, B. (2012). Causality, mediation and time: A dynamic viewpoint. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 175(4), 831-861. doi:10.1111/j.1467-985X.2011.01030.x
- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, 97(1), 129.
- Ahn, S. C., Lee, Y. H., & Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101(2), 219-255. doi:10.1016/S0304-4076(00)00083-X

- Ahn, S. C., Lee, Y. H., & Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, 174(1), 1-14. doi:10.1016/j.jeconom.2012.12.002
- Allen, P. G., & Fildes, R. (2001). Econometric forecasting. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 308-362). Norwell, MA: Kluwer Academic Publishers.
- Allen, P. G., & Fildes, R. (2005). Levels, differences and ECMs—Principles for improved econometric forecasting. *Oxford Bulletin of Economics and Statistics*, 67, 881-904. doi:10.1111/j.1468-0084.2005.00144.x
- Allison, P. D. (2005). *Fixed effects regression methods for longitudinal data using SAS*. Cary, NC: SAS Institute.
- Allison, P. D. (2009). *Fixed effects regression models*. Thousand Oaks, CA: Sage.
- Allison, P. D., Williams, R., & Moral-Benito, E. (2017). Maximum likelihood for cross-lagged panel models with fixed effects. *Socius*, 3, 1-17. doi:10.1177/2378023117710578
- Andersen, S. M., & Chen, S. (2002). The relational self: An interpersonal social-cognitive theory. *Psychological Review*, 109(4), 619-645.
- Angrist, J. D., & Kuersteiner, G. M. (2011). Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score. *Review of Economics and Statistics*, 93(3), 725-747. doi:10.1162/REST_a_00109
- Arellano, M. (2003). *Panel data econometrics*. Oxford: Oxford University Press.
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2), 277-297. doi:10.2307/2297968
- Armstrong, J. S. (2001). Role playing: A method to forecast decisions. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 15-30). New York, NY: Springer.
- Armstrong, J. S. (2005). The forecasting canon: Nine generalizations to improve forecast accuracy. *International Journal of Applied Forecasting*, 1, 29-35.
- Armstrong, J. S. (2007). Significance tests harm progress in forecasting. *International Journal of Forecasting*, 23(2), 321-327. doi:10.1016/j.ijforecast.2007.03.004
- Armstrong, J. S. (2012). Illusions in regression analysis. *International Journal of Forecasting*, 28, 689-694. doi:10.2139/ssrn.1969740
- Armstrong, J. S., Green, K. C., & Graefe, A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*, 68(8), 1717-1731. doi:10.1016/j.jbusres.2015.03.031
- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling*, 25, 359-388. doi:10.1080/10705511.2017.1406803
- Asparouhov, T., & Muthén, B. (2014). *Using Mplus individual residual plots for diagnostic and model evaluation in SEM*. Retrieved from <http://www.statmodel.com/examples/webnotes/webnote%2020.pdf>
- Asparouhov, T., & Muthén, B. (2016). Structural equation models and mixture models with continuous nonnormal skewed distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 1-19.
- Badagián, A. L., Kaiser, R., & Peña, D. (2015). Time series segmentation procedures to detect, locate and estimate change-points. In J. Beran, Y. Feng, & H. Hebbel (Eds.), *Empirical economic and financial research: Theory, methods, and practice* (pp. 45-59). Heidelberg: Springer.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229-1279. doi:10.3982/ECTA6135
- Bai, J. (2013). *Likelihood approach to dynamic panel models with interactive effects* (MRPA Paper No. 50267). Retrieved from: <http://mrpa.ub.uni-muenchen.de/50267/>
- Bai, J., & Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics*, 3(2), 89-163. doi:10.1561/0800000002
- Bak, P. (1996). *How nature works: The science of self-organized criticality*. New York, NY: Springer.

- Baltagi, B. H. (2013). Dynamic panel data models. In N. Hashimzade & M. Thornton (Eds.), *Handbook of research methods and applications on empirical macroeconomics* (pp. 229-248). Cheltenham, UK: Edward Elgar.
- Baltagi, B. H. (2014). *Econometric analysis of panel data* (5th ed.). New York, NY: Wiley.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645. doi:10.1146/annurev.psych.59.103006.093639
- Becker, M. C., Knudsen, T., & March, J. G. (2006). Schumpeter, Winter, and the sources of novelty. *Industrial and Corporate Change*, 15(2), 353-371. doi:10.1093/icc/dtl003
- Bernanke, B. S., & Blinder, A. S. (1988). *Credit, money, and aggregate demand* (No. w2534). Cambridge, MA: National Bureau of Economic Research.
- Bernanke, B. S., Boivin, J., & Eliasziw, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*, 120(1), 387-422. doi:10.1162/0033553053327452
- Bernanke, B. S., & Mihov, I. (1998). Measuring monetary policy. *Quarterly Journal of Economics*, 113(3), 869-902. doi:10.1162/003355398555775
- Binder, M., & Coad, A. (2010). An examination of the dynamics of well-being and life events using vector autoregressions. *Journal of Economic Behavior & Organization*, 76(2), 352-371. doi:10.1016/j.jebo.2010.06.006
- Binder, M., Hsiao, C., & Pesaran, M. H. (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory*, 21(4), 795-837. doi:10.1017/S0266466605050413
- Bojinov, I., & Shephard, N. (2017). Time series experiments and causal estimands: Exact randomization tests and trading. *arXiv preprint arXiv:1706.07840*.
- Boldea, O., & Hall, A. R. (2013). Testing structural stability in macroeconomic models. In N. Hashimzade & M. Thornton (Eds.), *Handbook of research methods and applications on empirical macroeconomics* (pp. 206-228). Cheltenham, UK: Edward Elgar.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Bollen, K. A., & Brand, J. E. (2010). A general panel model with random and fixed effects: A structural equations approach. *Social Forces*, 89(1), 1-34. doi:10.1353/sof.2010.0072
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. New York, NY: Wiley.
- Bond, S. R. (2002). Dynamic panel data models: A guide to micro data methods and practice. *Portuguese Economic Journal*, 1(2), 141-162. doi:10.1007/s10258-002-0009-9
- Booth, C. M., & Tannock, I. F. (2014). Randomised controlled trials and population-based observational research: Partners in the evolution of medical evidence. *British Journal of Cancer*, 110(3), 551-555. doi:10.1038/bjc.2013.725
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis: Forecasting and control* (4th ed.). New York, NY: Wiley.
- Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2016). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, 22(3), 409-425.
- Brock, W. A., & Carpenter, S. R. (2010). Interacting regime shifts in ecosystems: Implication for early warnings. *Ecological Monographs*, 80(3), 353-367. doi:10.1890/09-1824.1
- Browne, M., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models. In A. Madeau & J. J. McArdle (Eds.), *Advances in psychometrics: A festschrift for R. P. McDonald* (pp. 415-452). Mahwah, NJ: Erlbaum.
- Bun, M. J. G., & Sarafidis, V. (2015). Dynamic panel data models. In B. H. Baltagi (Ed.), *The Oxford handbook of panel data* (pp. 76-110). Oxford: Oxford University Press.

- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261-304. doi:10.1177/0049124104268644
- Canova, F., & Ciccarelli, M. (2013). Panel vector autoregressive models: A survey. In *VAR models in macroeconomics—New developments and applications: Essays in honor of Christopher A. Sims* (pp. 205-246). Bingley, UK: Emerald Group Publishing Limited.
- Carpenter, S. R., Cole, J. J., Pace, M. L., Batt, R., Brock, W. A., Cline, T., . . . Weidel, B. (2011). Early warnings of regime shifts: A whole-ecosystem experiment. *Science*, 332(6033), 1079-1082. doi:0.1126/science.1203672
- Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge: Cambridge University Press.
- Cartwright, N. (2009). Causality, invariance, and policy. In H. Kincaid & D. Ross (Eds.), *The Oxford handbook of philosophy of economics* (pp. 410-421). Oxford: Oxford University Press.
- Cartwright, N. (2011). Predicting what will happen when we act. What counts for warrant? *Preventive Medicine*, 53(4-5), 221-224. doi:10.1016/j.ypmed.2011.08.011
- Cartwright, N. (2012). Presidential address: Will this policy work for you? Predicting effectiveness better: How philosophy helps. *Philosophy of Science*, 79(5), 973-989. doi:10.1086/668041
- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford: Oxford University Press.
- Cartwright, N., & Munro, E. (2010). The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, 16(2), 260-266. doi:10.1111/j.1365-2753.2010.01382.x
- Chen, W., Clarke, J. A., & Roy, N. (2014). Health and wealth: Short panel Granger causality tests for developing countries. *The Journal of International Trade & Economic Development*, 23(6), 755-784. doi:10.1080/09638199.2013.783093
- Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? In J. B. Taylor & M. Woodford (Eds.), *Handbook of macroeconomics* (Vol. 1, pp. 65-148). Amsterdam: Elsevier.
- Clark, A. E., Frijters, P., & Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, 46(1), 95-144. doi:10.1257/jel.46.1.95
- Clements, M. P., & Hendry, D. F. (2001). An historical perspective on forecast errors. *National Institute Economic Review*, 177(1), 100-112. doi:10.1177/002795010117700109
- Clements, M. P., & Mizon, G. E. (1991). Empirical analysis of macroeconomic time series: VAR and structural models. *European Economic Review*, 35(4), 887-917. doi:10.1016/0014-2921(91)90042-H
- Cochrane, J. H. (1994, December). Shocks. In *Carnegie-Rochester Conference series on public policy* (Vol. 41, pp. 295-364). Amsterdam: North-Holland.
- Cogley, T. (1990). International evidence on the size of the random walk in output. *Journal of Political Economy*, 98(3), 501-518. doi:10.1086/261692
- Cole, D. A. (2012). Latent trait-state models. In R. H. Hoyle (Ed.), *The handbook of structural equation modeling* (pp. 585-600). New York, NY: Guilford.
- Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social Psychology Review*, 13(1), 62-72. doi:10.1177/1088868308329378
- Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, 62, 583-619. doi:10.1146/annurev.psych.093008.100356
- Curran, P. J., Lee, T., Howard, A. L., Lane, S., & MacCallum, R. A. (2012). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. In J. R. Harring & G. R. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 217-253). Charlotte, NC: Information Age.

- Dakos, V., van Nes, E. H., D'Odorico, P., & Scheffer, M. (2012). Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology*, 93(2), 264-271. doi:10.1890/11-0889.1
- Davidson, J. (2013). Cointegration and error correction. In N. Hashimzade & M. Thornton (Eds.), *Handbook of research methods and applications on empirical macroeconomics* (pp. 165-188). Cheltenham, UK: Edward Elgar.
- De Neve, J. E., & Oswald, A. J. (2012). Estimating the influence of life satisfaction and positive affect on later income using sibling fixed effects. *Proceedings of the National Academy of Sciences*, 109(49), 19953-19958.
- Deary, I. J., Pattie, A., & Starr, J. M. (2013). The stability of intelligence from age 11 to age 90 years: The Lothian birth cohort of 1921. *Psychological Science*, 24(12), 2361-2368. doi:10.1177/0956797613486487
- Deaton, A. (2002). Policy implications of the gradient of health and wealth. *Health Affairs*, 21(2), 13-30. doi:10.1377/hlthaff.21.2.13
- Deaton, A. (2003). Health, inequality, and economic development. *Journal of Economic Literature*, 41, 113-158. doi:10.1257/002205103321544710
- Deaton, A., & Cartwright, N. (2016). *Understanding and misunderstanding randomized controlled trials* (Working Paper). Durham, UK: Centre for Humanities Engaging Science and Society (CHESS).
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1-38.
- Denrell, J., Fang, C., & Liu, C. (2014). Perspective—Chance explanations in the management sciences. *Organization Science*, 26(3), 923-940. doi:10.1287/orsc.2014.0946
- Devlin, N., & Hansen, P. (2001). Health care spending and economic output: Granger causality. *Applied Economics Letters*, 8(8), 561-564.
- Di Tella, R., Haisken-De New, J., & MacCulloch, R. (2010). Happiness adaptation to income and to status in an individual panel. *Journal of Economic Behavior & Organization*, 76(3), 834-852. doi:10.1016/j.jebo.2010.09.016
- Diener, E. (2012). New findings and future directions for subjective well-being research. *American Psychologist*, 67(8), 590-597. doi:10.1037/a0029541
- Diener, E., & Biswas-Diener, R. (2002). Will money increase subjective well-being? *Social Indicators Research*, 57(2), 119-169. doi:10.1023/A:1014411319119
- Diener, E., Tay, L., & Oishi, S. (2013). Rising income and the subjective well-being of nations. *Journal of Personality & Social Psychology*, 104(2), 267-276.
- Dominici, F., Greenstone, M., & Sunstein, C. R. (2014). Particulate matter matters: Quasi-experimental evidence is needed on the relations between human health and airborne particulate matter. *Science*, 344(6181), 257-259. doi:10.1126/science.1247348
- D'Souza, R. M., & Nagler, J. (2015). Anomalous critical and supercritical phenomena in explosive percolation. *Nature Physics*, 11(7), 531-538. doi:10.1038/nphys3378
- Dufour, J. M., Pelletier, D., & Renault, E. (2006). Short run and long run causality in time series: Inference. *Journal of Econometrics*, 132(2), 337-362. doi:10.1016/j.jeconom.2005.02.003
- Dufour, J. M., & Renault, E. (1998). Short run and long run causality in time series: Theory. *Econometrica*, 66(5), 1099-1125. doi:10.2307/2999631
- Dufour, J. M., & Tessier, D. (1993). On the relationship between impulse response analysis, innovation accounting and Granger causality. *Economics Letters*, 42(4), 327-333.
- Easterlin, R. A. (1995). Will raising the incomes of all increase the happiness of all? *Journal of Economic Behavior & Organization*, 27(1), 35-47. doi:10.1016/0167-2681(95)00003-B
- Easterlin, R. A. (2001). Income and happiness: Towards a unified theory. *Economic Journal*, 111(473), 465-484. doi:10.1111/1468-0297.00646
- Easterlin, R. A., Morgan, R., Switek, M., & Wang, F. (2012). China's life satisfaction, 1990-2010. *Proceedings of the National Academy of Sciences*, 109, 9775-9780.
- Enders, W. (2014). *Applied econometric time series*. New York, NY: John Wiley & Sons.

- Erdil, E., & Yetkiner, I. H. (2009). The Granger-causality between health care expenditure and output: A panel data approach. *Applied Economics*, 41(4), 511-518. doi:10.1080/00036840601019083
- Ettner, S. L. (1996). New evidence on the relationship between income and health. *Journal of Health Economics*, 15(1), 67-85.
- Feldman, M., & Orlikowski, W. (2011). Theorizing practice and practicing theory. *Organization Science*, 22, 1240-1253. doi:10.1016/0167-6296(95)00032 -1
- Ferrer-i-Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114(497), 641-659. doi:10.1111/j.1468-0297.2004.00235.x
- Fildes, R., & Ord, J. K. (2002). Forecasting competitions—Their role in improving forecasting practice and research. In M. Clements & D. Hendry (Eds.). *A companion to economic forecasting* (pp. 322-353). Oxford: Blackwell.
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, 80(6), 1011-1027. doi:10.1037/0022-3514.80.6.1011
- Fleeson, W. (2004). Moving personality beyond the person-situation debate the challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, 13(2), 83-87. doi:10.1111/j.0963-7214.2004.00280.x
- Fleissig, A., & Strauss, J. (2001). Panel unit-root tests of OECD stochastic convergence. *Review of International Economics*, 9(1), 153-162. doi:10.1111/1467-9396.00270
- Fournier, M. A., Moskowitz, D. S., & Zuroff, D. C. (2008). Integrating dispositions, signatures, and the interpersonal domain. *Journal of Personality and Social Psychology*, 94(3), 531-545. doi:10.1037/0022-3514.94.3.531
- Frederick, S., & Loewenstein, G. (1999). Hedonic adaptation. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 302-329). New York, NY: Sage.
- Freedman, D. A. (2004). Graphical models for causation, and the identification problem. *Evaluation Review*, 28(4), 267-293. doi:10.1177/0193841X04266432
- French, D. (2012). Causation between health and income: A need to panic. *Empirical Economics*, 42(2), 583-601. doi:10.1007/s00181-011-0541-5
- Frijters, P., Haiken-DeNew, J. P., & Shields, M. A. (2005). The causal effect of income on health: Evidence from German reunification. *Journal of Health Economics*, 24(5), 997-1017. doi:10.1016/j.jhealeco.2005.01.004
- Gersick, C. J., & Hackman, J. R. (1990). Habitual routines in task-performing groups. *Organizational Behavior and Human Decision Processes*, 47(1), 65-97.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral method. *Econometrica*, 37(3), 424-438. doi:10.2307/1912791
- Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2, 329-352. doi:10.1016/0165-1889(80)90069-X
- Granger, C. W. J. (1986). Comment on statistics and causal inference by P. Holland. *Journal of the American Statistical Association*, 81(396), 967-968.
- Granger, C. W. J. (1988). Some recent developments in a concept of causality. *Journal of Econometrics*, 39(1-2), 199-211. doi:10.1016/0304-4076(88)90045-0
- Granger, C. W. J. (2003). Some aspects of causal relationships. *Journal of Econometrics*, 112(1), 69-69.
- Granger, C. W. J., & Morris, M. J. (1976). Time series modeling and interpretation. *Journal of the Royal Statistical Society, A*, 139, 246-257.
- Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2), 111-120.
- Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, 68(8), 1678-1685. doi:10.1016/j.jbusres.2015.03.026

- Greiner, D. J., & Rubin, D. B. (2011). Causal effects of perceived immutable characteristics. *Review of Economics and Statistics*, 93(3), 775-785. doi:10.1162/REST_a_00110
- Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology*, 30, 507-544. doi:10.1146/annurev.soc.30.012703.110629
- Haldrup, N., Kruse, R., Teräsvirta, T., & Verneskov, R. T. (2013). Unit roots, non-linearities and structural breaks. In N. Hashimzade & M. Thornton (Eds.), *Handbook of research methods and applications on empirical macroeconomics* (pp. 61-94). Cheltenham, UK: Edward Elgar.
- Hamaker, E. L. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods & Research*, 33(3), 404-416. doi:10.1177/0049124104270220
- Hamaker, E. L. (2012). Why researchers should think “within-person”: A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of methods for studying daily life* (pp. 43-61). New York, NY: Guilford Publications.
- Hamaker, E. L., & Grasman, R. P. (2015). To center or not to center? Investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, 5, 1492. doi:10.3389/fpsyg.2014.01492
- Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. (2002). On the nature of SEM estimates of ARMA parameters. *Structural Equation Modeling*, 9(3), 347-368. doi: 10.1207/S15328007SEM0903_3
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20(1), 102-116. doi:10.1037/a0038889
- Harvey, A. C. (1985). Trends and cycles in macroeconomic time series. *Journal of Business & Economic Statistics*, 3(3), 216-227.
- Harvey, A. (1997). Trends, cycles and autoregressions. *The Economic Journal*, 107(440), 192-201.
- Hastings, A., & Wysham, D. B. (2010). Regime shifts in ecological systems can occur with no warning. *Ecology Letters*, 13(4), 464-472. doi:10.1111/j.1461-0248.2010.01439.x
- Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica: Journal of the Econometric Society*, 49, 1377-1398. doi:10.2307/1911406
- Heckman, J. J. (1991). Identifying the hand of past: Distinguishing state dependence from heterogeneity. *The American Economic Review*, 81(2), 75-79.
- Hendry, D. F. (2004). Causality and exogeneity in non-stationary economic time series. In *New directions in macromodelling* (pp. 21-48). Bingley, UK: Emerald Group Publishing Limited.
- Hendry, D. F., & Mizon, G. E. (2005). Forecasting in the presence of structural breaks and policy regime shifts. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models* (pp. 480-502). New York, NY: Cambridge University Press.
- Hendry, D. F., Pagan, A. R., & Sargan, J. D. (1984). Dynamic specification. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics* (Vol. II, pp. 1023-1100). Amsterdam: Elsevier.
- Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development*, 6(2-3), 97-120. doi:10.1080/15427600902911189
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945-960.
- Holland, P. W. (2008). Causation and race. In T. Zuberi & E. Bonilla-Silva (Eds.), *White logic, white methods: Racism and methodology* (pp. 93-109). New York, NY: Rowman & Littlefield.
- Hsiao, C. (1982). Autoregressive modeling and causal ordering of economic variables. *Journal of Economic Dynamics and Control*, 4(1), 243-259.
- Hsiao, C. (2007). Panel data analysis—Advantages and challenges. *Test*, 16(1), 1-22. doi:10.1007/s11749-007-0046-x
- Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). Cambridge: Cambridge University Press.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424-453.

- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. doi:10.1080/10705519909540118
- Jebb, A. T., & Tay, L. (2016). Introduction to time series analysis for organizational research methods for longitudinal analyses. *Organizational Research Methods*, 20(1), 61-94. doi:10.1177/1094428116668035
- Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in Psychology*, 6, 727. doi:10.3389/fpsyg.2015.00727
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, 74(4), 657-690.
- Kessler, R. C., & Greenberg, D. F. (1981). *Linear panel analysis: Models of quantitative change*. New York, NY: Academic Press.
- Kilian, L. (1999). Finite-sample properties of percentile and percentile-*t* bootstrap confidence intervals for impulse responses. *Review of Economics and Statistics*, 81(4), 652-660. doi:10.1162/003465399558517
- Kirkman, B. L., & Rosen, B. (1999). Beyond self-management: Antecedents and consequences of team empowerment. *Academy of Management Journal*, 42(1), 58-74. doi:10.5465/256874
- Kiyotaki, N., & Moore, J. (1997). Credit cycles. *The Journal of Political Economy*, 105(2), 211-248. doi:10.1086/262072
- Kiyotaki, N., & Moore, J. (2002). Balance-sheet contagion. *American Economic Review*, 92(2), 46-50. doi:10.1257/000282802320188989
- Kocherlakota, N. R. (2010). Modern macroeconomic models as tools for economic policy. *The Region, Federal Reserve Bank of Minneapolis*, 4(May), 5-21.
- Kuersteiner, G. M. (2010). Granger-Sims causality. In S. N. Durlauf & L. E. Blume (Eds.), *Macroeconometrics and time series analysis* (pp. 119-134). London: Palgrave Macmillan.
- Lechner, M. (2010). The relation of different concepts of causality used in time series and microeconometrics. *Econometric Reviews*, 30(1), 109-127. doi:10.1080/07474938.2011.520571
- Levitt, B., & March, J. G. (1988). Organizational Learning. *Annual Review of Sociology*, 14(1), 319-338. doi:10.1146/annurev.so.14.080188.001535
- Liker, J. K., Augustyniak, S., & Duncan, G. J. (1985). Panel data and models of change: A comparison of first-different and conventional two-wave models. *Social Science Research*, 14, 80-101. doi:10.1016/0049-089X(85)90013-4
- Lindahl, M. (2005). Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. *Journal of Human Resources*, 40(1), 144-168. doi:10.3368/jhr.XL.1.144
- Lucas, R. E. (1976). Economic policy evaluation: A critique. *Journal of Monetary Economics*, 1, 19-46. doi:10.1016/S0167-2231(76)80003-6
- Lütkepohl, H. (1993). Testing for causation between two variables in higher-dimensional VAR models. In H. Schneeweib & K. F. Zimmermann (Eds.), *Studies in applied econometrics* (pp. 75-91). Heidelberg: Physica-Verlag HD.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: Springer-Verlag.
- Lütkepohl, H. (2006). Forecasting with VARMA models. In G. Elliot, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting (Vol. 1, pp. 287-325)*. Amsterdam: Elsevier.
- Lütkepohl, H. (2013). Vector autoregressive models. In N. Hashimzade & M. Thornton (Eds.), *Handbook of research methods and applications on empirical macroeconomics* (pp.139-164). Cheltenham, UK: Edward Elgar.
- Lütkepohl, H. (2015). Forecasting unpredictable variables. In J. Beran, Y. Feng, & H. Hebbel (Eds.), *Empirical economic and financial research: Theory, methods, and practice* (pp. 287-304). Heidelberg: Springer.
- Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin*, 131(6), 803-855.
- Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications. *International Journal of Forecasting*, 16(4), 451-476. doi:10.1016/S0169-2070(00)00057 -1

- Mankiw, N. G., & Shapiro, M. D. (1985). Trends, random walks, and tests of the permanent income hypothesis. *Journal of Monetary Economics*, 16(2), 165-174.
- March, J. G., & Olsen, J. P. (1989). *Rediscovering institutions: The organizational basis of politics*. New York, NY: The Free Press.
- Matthews, G., Deary, I. J., & Whiteman, M. C. (2003). *Personality traits* (2nd ed.). Cambridge: Cambridge University Press.
- McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models*. Washington, DC: American Psychological Association.
- Meer, J., Miller, D. L., & Rosen, H. S. (2003). Exploring the health-wealth nexus. *Journal of Health Economics*, 22(5), 713-730. doi:10.1016/S0167-6296(03)00059-6
- Meier, L. L., & Spector, P. E. (2013). Reciprocal effects of work stressors and counterproductive work behavior: A five-wave longitudinal study. *Journal of Applied Psychology*, 98(3), 529-539. doi:10.1037/a0031732
- Michaud, P. C., & Van Soest, A. (2008). Health and wealth of elderly couples: Causality tests using dynamic panel data models. *Journal of Health Economics*, 27(5), 1312-1325. doi:10.1016/j.jhealeco.2008.04.002
- Mischel, W., & Shoda, Y. (2008). Toward a unified theory of personality: Integrating dispositions and processing dynamics within the cognitive-affective processing system (CAPS). In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality* (3rd ed., pp. 208-241). New York, NY: Guilford.
- Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, 2(4), 201-218. doi:10.1207/s15366359mea0204_1
- Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics*, 31(4), 451-472. doi:10.1080/07350015.2013.818003
- Mplus2lavaan. (2019). Retrieved from <https://rdrr.io/cran/lavaan/man/mplus2lavaan.modelSyntax.html>
- Murray, C. J., & Nelson, C. R. (2000). The uncertain trend in US GDP. *Journal of Monetary Economics*, 46(1), 79-95. doi:10.1016/S0304-3932(00)00018-0
- Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29, 81-117. doi:10.2333/bhmk.29.81
- Muthén, B. O., & Asparouhov, T. (2003). Modeling interactions between latent and observed continuous variables using maximum-likelihood estimation in Mplus. *Mplus Web Notes*, 6, 1-9. Retrieved from <http://www.statmodel2.com/download/webnotes/webnote6.pdf>
- Muthén, B. O., & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143-165). Boca Raton, FL: Chapman & Hall/CRC Press.
- Muthén, B. O., & Asparouhov, T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 12-23. doi:10.1080/10705511.2014.935843
- Muthén, L. K., & Muthén, B. O. (1998-2018). *Mplus: Statistical analysis with latent variables* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Nauges, C., & Thomas, A. (2003). Consistent estimation of dynamic panel data models with time-varying individual effects. *Annales d'Economie et de Statistique*, 70, 53-75.
- Neal, D. T., Wood, W., & Quinn, J. M. (2006). Habits: A repeat performance. *Current Directions in Psychological Science*, 15(4), 198-202. doi:10.1111/j.1467-8721.2006.00435.x
- Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49, 1417-1426. doi:10.2307/1911408
- Oswald, A. J., Proto, E., & Sgroi, D. (2015). Happiness and productivity. *Journal of Labor Economics*, 33(4), 789-822. doi:10.1086/681096
- Perron, P. (2006). Dealing with structural breaks. In T. C. Mills & K. Patterson (Eds.), *Palgrave handbook of econometrics: Volume 1, econometric theory* (pp. 278-352). Basingstoke, UK: Palgrave Macmillan.

- Phillips, P. C. B. (2003). Laws and limits of econometrics. *Economic Journal*, 113, C26-C52. doi:10.1111/1468-0297.00114
- Ployhart, R. E., & Ward, A. K. (2011). The “quick start guide” for conducting and publishing longitudinal research. *Journal of Business and Psychology*, 26(4), 413-422. doi:10.1007/s10869-011-9209-6
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 126(1), 3-25.
- Rogosa, D., & Willett, J. B. (1985). Satisfying a simplex structure is simpler than it should be. *Journal of Educational and Behavioral Statistics*, 10(2), 99-107. doi:10.3102/10769986010002099
- Rubin, D. B. (2011). For objective causal inference, design trumps analysis. *Annals of Applied Statistics*, 2(3), 808-840. doi:10.1214/08-AOAS187
- Sacks, D. W., Stevenson, B., & Wolfers, J. (2012). The new stylized facts about income and subjective well-being. *Emotion*, 12(6), 1181-1187.
- Sacks, D. W., Stevenson, B., & Wolfers, J. (2013). *Growth in income and subjective well-being over time*. Unpublished Manuscript, University of Michigan.
- Sanggyun, K., & Brown, E. N. (2010). A general statistical framework for assessing granger causality. *Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, IEEE, 2222-2225.
- Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., . . . Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, 461, 53-59. doi:10.1038/nature08227
- Scheffer, M., Carpenter, S. R., Dakos, V., & van Nes, E. (2015). Generic indicators of ecological resilience: Inferring the chance of a critical transition. *Annual Review of Ecology, Evolution, and Systematics*, 46, 145-167. doi:10.1146/annurev-ecolsys-112414-054242
- Schön, D. A. (1995). Causality and causal inference in the study of organizations. In R. F. Goodman & W. R. Fisher (Eds.), *Rethinking knowledge: Reflections across the disciplines* (pp. 69-101). New York, NY: State University of New York Press.
- Sekhon, J. S. (2009). Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science*, 12, 487-508. doi:10.1146/annurev.polisci.11.060606.135444
- Sherman, R. E., Anderson, S. A., Dal Pan, G. J., Gray, G. W., Gross, T., Hunter, N. L., . . . Shuren, J. (2016). Real-world evidence—What is it and what can it tell us. *New England Journal of Medicine*, 375(23), 2293-2297. doi:10.1056/NEJMs1609216
- Shrout, P. E. (2011). Commentary: Mediation analysis, causal process, and cross-sectional data. *Multivariate Behavioral Research*, 46(5), 852-860. doi:10.1080/00273171.2011.606718
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1), 1-48. doi:10.2307/1912017
- Sims, C. A. (1986). Are forecasting models usable for policy analysis? *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 2-16.
- Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36(5), 975-1000. doi:10.1016/0014-2921(92)90041-T
- Sims, C. A., Stock, J. H., & Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, 58, 113-144. doi:10.2307/2938337
- Stock, J. H., & Watson, M. W. (1988). Variable trends in economic time series. *The Journal of Economic Perspectives*, 2(3), 147-174. doi:10.1257/jep.2.3.147
- Stock, J. H., & Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14(1), 11-30.
- Stock, J. H., & Watson, M. W. (1999). Business cycle fluctuations in US macroeconomic time series. In J. B. Taylor & M. Woodford (Eds.), *Handbook of macroeconomics* (Vol. 1A, pp. 3-64). Amsterdam: North-Holland.
- Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *The Journal of Economic Perspectives*, 15(4), 101-115. doi:10.1257/jep.15.4.101

- Stock, J. H., & Watson, M. W. (2005). *Implications of dynamic factor models for VAR analysis* (No. w11467). Cambridge, MA: National Bureau of Economic Research.
- Stock, J. H., & Watson, M. W. (2009). Forecasting in dynamic factor models subject to structural instability. In N. Shephard & J. Castle (Eds.), *The methodology and practice of econometrics: Festschrift in honor of D. F. Hendry* (pp. 173-205). Oxford: Oxford University.
- Stock, J. H., & Watson, M. W. (2011). Dynamic factor models. In M. P. Clements & D. F. Hendry (Eds.), *Oxford handbook of economic forecasting* (pp. 35-60). Oxford: Oxford University Press.
- Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, 128(610), 917-948. doi:10.1111/eoj.12593
- Stone, D. A. (1989). Causal stories and the formation of policy agendas. *Political Science Quarterly*, 104(2), 281-300. doi:10.2307/2151585
- Stutzer, A., & Frey, B.S. (2006). Does marriage make people happy, or do happy people get married? *Journal of Socio-Economics*, 35(2), 326-347. doi:10.1016/j.socec.2005.11.043
- Swanson, N. R., & Granger, C. W. (1997). Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. *Journal of the American Statistical Association*, 92(437), 357-367. doi:10.1080/01621459.1997.10473634
- Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2), 381-419. doi:10.1016/j.jmoneco.2004.05.007
- Van de Leemput, I. A., Wichers, M., Cramer, A. O., Borsboom, D., Tuerlinckx, F., Kuppens, P., . . . Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences*, 111(1), 87-92. doi:0.1073/pnas.1312114110
- Van Iddekinge, C. H., Ferris, G. R., Perrewé, P. L., Perryman, A. A., Blass, F. R., & Heetderks, T. D. (2009). Effects of selection and training on unit-level performance over time: A latent growth modeling approach. *Journal of Applied Psychology*, 94(4), 829-843. doi:10.1037/a0014453
- Vancouver, J. B., More, K. M., & Yoder, R. J. (2008). Self-efficacy and resource allocation: support for a nonmonotonic, discontinuous model. *Journal of Applied Psychology*, 93(1), 35-47.
- Weick, K. E. (1998). Improvisation as a mindset for organizational analysis. *Organization Science*, 9, 543-555.
- Williams, R., Allison, P. D., & Moral-Benito, E. (2018). Linear dynamic panel-data estimation using maximum likelihood and structural equation modeling. *The Stata Journal*, 18, 293-326.
- Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, 25(1), 659-706. doi:10.1146/annurev.soc.25.1.659
- Winter, S. G. (2013). Habit, deliberation, and action: Strengthening the microfoundations of routines and capabilities. *Academy of Management Perspectives*, 27(2), 120-137. doi:10.5465/amp.2012.0124
- Wright, J. H. (2000). Confidence intervals for univariate impulse responses with a near unit root. *Journal of Business and Economic Statistics*, 18(3), 368-373.

Author Biographies

Michael J. Zyphur is an associate professor of Management at the University of Melbourne. His interests lie in quantitative research methods, including their history and philosophy, as well as the study of human wellbeing and ethics.

Paul D. Allison is a sociologist and professor emeritus of the University of Pennsylvania where he taught graduate courses in methods and statistics for more than 35 years. He is widely recognized for his work on, and teaching of, statistical methods. His current methodological research is on the analysis of longitudinal data, especially with determining the causes and consequences of events, and on methods for handling missing data.

Louis Tay is an associate professor in Industrial-Organizational Psychology at Purdue University. His research interests are in well-being, vocational interests, measurement, taxometrics, latent class modeling, and data science.

Manuel C. Voelkle is a professor for psychological research methods at the Humboldt-Universität zu Berlin and an adjunct researcher at the Max Planck Institute for Human Development. His research interests revolve around the design and analysis of multivariate empirical studies with an emphasis on the use of structural equation models for the analysis of longitudinal data.

Kristopher J. Preacher is a Professor in Quantitative Methods at Vanderbilt University. His research concerns the use and combination of structural equation modeling and multilevel modeling to model correlational and longitudinal data. Other interests include developing techniques to test mediation and moderation hypotheses, bridging the gap between substantive theory and statistical practice, and studying model evaluation and model selection in the application of multivariate methods to social science questions.

Zhen Zhang is a professor of Management and Dean's Council of 100 Distinguished Scholar in the W. P. Carey School of Business at Arizona State University. He obtained his Ph.D. in Human Resources and Industrial Relations in 2008 from the University of Minnesota. His research focuses on leadership process and development, biological basis of work behavior, the intersection between organizational behavior and entrepreneurship, and advanced research methods.

Ellen L. Hamaker is a full Professor at the Utrecht University. She received her PhD in 2004 from the University of Amsterdam and spent a year as a post doctoral fellow at the University of Virginia. Her research interests include time series analysis, longitudinal analysis, multilevel analysis, structural equation modeling, and Bayesian statistics.

Ali Shamsollahi is an assistant professor of Marketing, ESSEC Business School. His research interests revolve around inter-organizational relationships, marketing channels, and dynamic research methods suitable for the study of longitudinal data.

Dean C. Pierides is a lecturer in Business and Management at the University of Stirling and Executive Board member for the Research Committee on the Sociology of Organizations in the International Sociological Association. He was awarded a BA from the University of Pennsylvania, a DipEd, MEd and PhD from the University of Melbourne. His research is about how organizations and institutions deal with uncertainty, focusing on government agencies that are responsible for emergencies and disasters, whether, natural, financial, medical, or otherwise.

Peter Koval completed a PhD in Psychology at KU Leuven (2014), followed by a postdoctoral fellowship at the Australian Catholic University (2014-2017). He is currently a lecturer in the Melbourne School of Psychological Sciences, University of Melbourne. Peter's research focuses on understanding daily emotional functioning and dynamics and their relation with well-being and psychopathology.

Ed Diener is an alumni distinguished professor of Psychology (Emeritus) at the University of Illinois, where he has been a faculty member since 1974. He is currently a professor of psychology at the University of Utah and the University of Virginia. He joined Gallup as a Senior Scientist in 1999 and advises them on research in psychological well-being. His current research focuses on the theories and measurement of well-being; temperament and personality influences on well-being; income and well-being; and cultural influences on well-being; and how employee well-being enhances organizational performance.