Estimating the parameters of system dynamics models using indirect inference

Niyousha Hosseinichimeh,[a]* Hazhir Rahmandad,[b] Mohammad S. Jalali[b] and Andrea K. Wittenborn[c]

Abstract

There is limited methodological guidance for estimating system dynamics (SD) models using datasets common to social sciences that include few data points over time for many units under analysis. Here, we introduce indirect inference, a simulation-based estimation method that can be applied to common datasets and is applicable to SD models that often include intractable likelihood functions. In this method, the model parameters are found by ensuring that simulated data from the model and available empirical data produce similar auxiliary statistics. The method requires few assumptions about the structure of the model and error-generating processes and thus can be used in a variety of applications. We demonstrate the method in estimating an SD model of depression and rumination using a panel dataset. The overall results suggest that indirect inference can extend the application of SD models to new topics and leverage common panel datasets to provide unique insights.

Syst. Dyn. Rev. (2016)

Additional supporting information may be found in the online version of this article at the publisher's web site.

Background

Most system dynamics (SD) models use a single case study and apply traditional estimation methods (e.g. mean squared error, mean absolute percentage error) to time series data of that case to specify unknown parameter values. However, more flexible methods of estimation are needed in both theoretical and practical applications to leverage data structures beyond single-case time series. With the increasing availability of data on various research subjects—from individual level to company- and country-level phenomena—formal model calibration has become a requisite step in producing credible model-based analyses that are trusted by various academic audiences. However, three major challenges exist in estimating SD models. First, SD models are often complex and nonlinear and the likelihood functions are intractable; thus many conventional statistical methods

[a] Virginia Tech, Department of Industrial and Systems Engineering, Blacksburg, Virginia US
[b] MIT, Sloan School of Management, Cambridge, Massachusetts US
[c] Michigan State University, Department of Human Development and Family Studies, East Lansing, Michigan US

* Correspondence to: Niyousha Hosseinichimeh. E-mail: niyousha@vt.edu

Accepted by Laura Black, *Received 17 August 2015; Revised 27 January 2016, 30 August 2016, 9 September 2016; Accepted 13 September 2016*

System Dynamics Review
System Dynamics Review (2016)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/sdr.1558

do not directly apply. Second, due to the structure of many datasets, even the heuristic calibration methods commonly used in SD practice that minimize the differences between empirical time series and simulated counterparts may not apply. For example, many "panel" datasets include data at only a few points in time, but in many units under analysis (e.g. individuals, organizations or countries), complicating the matching of the simulations to data using traditional methods that require many data points over time for each unit. For the same reason, other methods such as Kalman filtering (Kalman, 1960) or extended Kalman filtering (Smith *et al.*, 1962), which adjust state variables based on measured system behaviors, or partial model calibration (Homer, 2012; Hosseinichimeh *et al.*, 2015) cannot be used effectively when very few data points are available over time. Third, in many applications, randomness, which is exogenous to a model's boundaries, has a significant role in the behavior of the system; therefore, noise should be considered explicitly in the estimation of the model. These complications call for the introduction of more rigorous, simulation-based estimation methods to the SD literature.

Simulation-based estimation methods have been introduced with the increasing computational power of computers and have made it possible to run many numerical simulations of large datasets in short periods of time. The basic idea behind these methods is to match properties of the simulated data to those of the empirical data. These methods include the method of simulated moments (Duffie and Singleton, 1993; Jalali *et al.*, 2015; Mcfadden, 1989; Pakes and Pollard, 1989), the efficient method of moments (Durlauf and Blume, 2008) and indirect inference (Gourieroux *et al.*, 1993; Gouriéroux *et al.*, 2010; Smith, 1993), among others. These methods are mostly useful for models with intractable likelihood function such as nonlinear dynamic models and models with missing, incomplete or noisy data.

In this article, we provide an introduction to indirect inference, one of the most flexible methods available in this space, and explain how it can be applied in SD modeling. This method has been applied in various fields for estimating different types of models including nonlinear ecological dynamic systems (Wood, 2010), dynamic panel models with intractable likelihood function (Gouriéroux *et al.*, 2010), continuous time models (Monfort, 1996) and stochastic volatility models (Monfardini, 1998), among others. However, to our knowledge, no study has implemented indirect inference for calibrating SD models distinguished by a focus on endogenous feedback relationships among variables. We first present an SD model of major depressive disorder (MDD) and discuss the challenges in estimating the unknown parameters of the model. We then introduce the indirect inference method and explain the steps needed to estimate unknown parameters of a model. We apply this method to our SD model of MDD to demonstrate the estimation procedures with an empirical dataset. Finally, we discuss the conditions under which SD studies can benefit from indirect inference.

DOI: 10.1002/sdr

An applied example

Here we present a simple SD model of MDD, a disabling condition that causes feelings of sadness and loss of interest. Different mechanisms including genetic, cognitive, environmental and biological factors contribute to the disorder. To keep the applied example simple, we focus only on the impact of rumination (a cognitive factor) and stressful life events (an environmental factor). Rumination is defined as repetitive thinking about the causes and consequences of a stressor without focusing on coping strategies or engaging in problem solving (Nolen-Hoeksema *et al.*, 2007). As shown in Figure 1, the model has two reinforcing loops (R1 and R2). Loop R1 captures the idea that when an individual with a ruminative response style experiences a stressful event, she spends time thinking about that stressor, keeping the stressor active, and thus increasing the chance of even more rumination (Ruscio *et al.*, 2015). In other words, engaging in rumination increases the duration of recalling a stressor (i.e., *memory time*) and thus increases the accumulation of *stressor memory*, causing an even higher level of *rumination* (loop R1). Current *rumination* is formulated as a stock adjusting with a time constant towards *indicated rumination*, which is a linear function of *stressor memory*

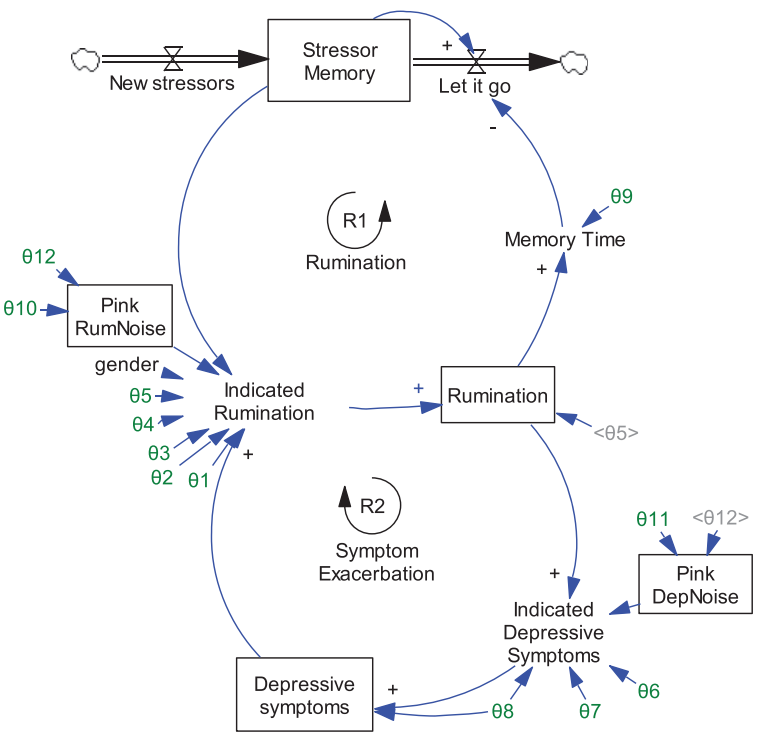

Fig. 1. The depression–rumination conceptual model. Boxes depict stock (or state) variables and arrows with values represent flows into/out of those stocks. Single-line arrows indicate causal relationships hypothesized among variables (the strength of which is estimated below). A stock variable is the accumulation of the difference between its inflows and outflows and, mathematically, is represented as an integral

DOI: 10.1002/sdr

(Michl *et al.*, 2013), current *depressive symptoms* (Nolen-Hoeksema *et al.*, 2007) and *gender* (Nolen-Hoeksema *et al.*, 1999).

Loop R2 demonstrates that more *rumination* elevates *depressive symptoms* and higher *depressive symptoms* lead to more *rumination* (Nolen-Hoeksema *et al.*, 2007). The stock of *depressive symptoms* is the smooth of *indicated depressive symptoms*, which is a function of *rumination*. Moreover, random events outside the model boundaries affect *rumination* and *depressive symptoms*. Randomness recognizes that *indicated depressive symptoms* and *indicated rumination* are not deterministic and vary by factors outside the model boundary based on a probability distribution; however, there is some autocorrelation in how those chance events unfold. Therefore, normally distributed pink noises are added to the *indicated rumination* and *indicated depressive symptoms*, $RumNoise \sim N(0, \sigma_r^2)$ and $DepNoise \sim N(0, \sigma_d^2)$ with correlation time ρ, respectively. We assume the same correlation time for rumination and depression because they are both generated by stressful events and other exogenous shocks that are similar in nature. All equations are presented in the online Appendix (supporting information). The model has 12 unknown parameters ($p = 12$), listed in Table 1, that need to be estimated.

The available dataset for estimating the SD model parameters reports all three variables of interest (i.e. depressive symptoms, rumination and stressful life events) for 661 adolescents from two middle schools in Connecticut (Michl *et al.*, 2013). The main challenge is that these variables are reported in few points in time. The tendency to ruminate was assessed at three points in time (T_1, T_2, and T_3), while the questionnaires related to depressive symptoms (Children's Depression Inventory) and stressful life events (Life Events Scale for Children) were completed only at T_1 and T_3. The time between the first and second assessments and the second and third

Table 1. Unknown parameters in the model

Unknown parameters ($\boldsymbol{\theta}$)	Unit
Rumination constant (θ_1)	RumScore
Effect of depressive symptoms on rumination (θ_2)	RumScore/DepScore
Gender coefficient (θ_3)	RumScore
Effect of stressors on rumination (θ_4)	RumScore/Disruption
Rumination coefficient (θ_5)	Dmnl[1]
Depression constant (θ_6)	DepScore
Effect of rumination on depressive symptoms (θ_7)	DepScore/RumScore
Depression Coefficient (θ_8)	Dmnl
Effect of rumination on memory time (θ_9)	1/RumScore
Standard deviation of rumination noise (θ_{10})	Dmnl
Standard deviation of depression noise (θ_{11})	Dmnl
Correlation time (θ_{12})	Month

[1]Dimensionless.

DOI: 10.1002/sdr

assessments are 4 and 3 months, respectively. Table 2 summarizes the means and standard deviations of variables each time data were collected.

Conventional estimation methods in SD compare data over time for each individual against the simulation model of that individual to estimate an individual-level model. For example, Croson and colleagues use data over 48 weeks for individual decisions in the beer game to estimate each person's four-decision rule parameters across many settings (Croson *et al.*, 2014). Yet in this example we only have seven data points for each individual; thus separately estimating a dozen parameters for each participant is infeasible. The richness of the current dataset is in the large number of available participants. If we assume the core model parameters represent basic biological processes that are similar across individuals (a common assumption in typical statistical models), we should be able to leverage the large sample size to estimate the dozen unknown model parameters. Yet conventional time-series estimation methods in the SD literature do not provide a recipe for such a scenario. To resolve such estimation challenges, we next introduce indirect inference.

Indirect inference method

General properties and historical background

The main idea behind the indirect inference method is to match properties of empirical and simulated data in order to estimate the unknown parameters of the model of interest. This method was developed to overcome the challenges of estimating complex model parameters for which the likelihood function is intractable. In the indirect inference method, the simulated data are generated by simulating the model of interest; then, an "auxiliary model" that typically consists of simple regression(s) is selected and parameters of the auxiliary model are estimated by using both the empirical and the simulated data. The difference between these two sets of parameters of the auxiliary model is minimized to estimate the parameters of the model of interest.

The indirect inference method has several advantages. First, there are few limitations to the types of models to which it can be applied. The only requirement is that the model of interest can be simulated for different values

Table 2. Means and standard deviations* of variables

Variable	Time 1, T_1	Time 2, T_2	Time 3, T_3
Depressive symptoms	9.48 (6.28)	—	9.78 (7.64)
Rumination	11.59 (7.52)	10.85 (7.62)	9.95 (7.95)
Stressful life events	4.96 (3.32)	—	4.20 (3.70)

*Standard deviations are in parentheses.

DOI: 10.1002/sdr

of its parameters. Second, although this method is a simulation-based technique, it can be relatively inexpensive to compute when the auxiliary model uses a maximum likelihood estimator, and thus the auxiliary model parameters have small variance and could be matched reliably with few simulations (Gouriéroux *et al.*, 2010). Third, the indirect inference method inherits the beneficial properties of the estimation method used for the auxiliary model (Gouriéroux *et al.*, 2010). For instance, if the maximum likelihood method is used to find the parameters of the auxiliary model, the estimated parameters resulting from indirect inference would also have small variance. Fourth, it can be used for both estimating and validating a model. The validation step allows the modeler to decide if the model's outputs are indistinguishable from empirical data or if notable differences exist after estimation that could inform further model refinement. In this article, we discuss one such validation test. We also investigate the method's validity using a separate approach in which indirect inference is applied to a synthetic dataset generated by simulation of the calibrated model, and we evaluate the method's ability to recover correct parameters from a structurally precise model.

The method of simulated moments (MSM; Mcfadden, 1989) is one of the first rigorous simulation-based estimation methods. It is the workhorse of modern econometrics and motivates the idea of indirect inference. In this method, model parameters are estimated by minimizing the difference between selected moments (e.g. mean and variance) of empirical data and corresponding moments of model-generated simulated data. Only a few studies exist that have implemented MSM to calibrate SD models. Rahmandad and Sabounchi (2011) calibrated a dynamic model of obesity at both individual and population levels by using MSM, and Jalali *et al.* (2013) discussed the application of MSM to SD models. The indirect inference method, proposed independently by Gourieroux *et al.* (1993) and Smith (1993), is very similar to MSM in matching some functions of empirical data against the same function calculated on simulated data. However, it is more general because, rather than only the statistical moments, a wider set of functions of the empirical and simulated data can be matched to estimate the unknown parameters. These functions are created using auxiliary models. Although the auxiliary model is typically a separate estimation, it does not need to capture the true data-generating process. The auxiliary model serves only as a lens through which we view the empirical data and calculate functions we then match to their simulated counterparts. The parameters of the model are set in a way that both empirical and simulated data produce very similar images as they pass through this lens.

Other methods exist that follow a similar logic. Structural equation modeling (SEM) is based on matching the observed covariance matrix and model-generated covariance matrix (Anderson and Gerbing, 1988). In actor-based network models, the statistical properties (e.g. degree distribution, centrality and clustering) of empirical networks are compared with those of

DOI: 10.1002/sdr

the simulated networks to estimate the model parameters (Snijders, 2001). Overall, indirect inference and its derivatives are among the most flexible econometric methods for estimating complex models using various data structures. Given its moderate computational costs, the method could be applied easily to models of modest size when a handful of model parameters require estimation. However, estimating a large number of parameters (e.g. hundreds) could be much more challenging because the underlying optimization problem is non-convex. Currently, there are no studies in the literature that apply the indirect inference method to SD models. In the next section, we introduce the method formally and present a step-by-step guide for applying it.

Description of the method

Consider a general dynamic model with stock (state) variables $\mathbf{z}$, the dynamics of which are described as $\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t} = f(\boldsymbol{\theta}_{\mathrm{D}}, \mathbf{z}, \mathbf{u}, \boldsymbol{\varepsilon}_1)$, and a set of exogenous variables, $\mathbf{u}$, and observable variables, $\mathbf{x}$, which are a function of $\mathbf{z}$:

$$\mathbf{x} = g(\boldsymbol{\theta}_{\mathrm{O}}, \mathbf{z}, \mathbf{u}, \boldsymbol{\varepsilon}_2) \tag{1}$$

Here, function f describes the dynamics of the system and function g defines the measurement process. It is assumed that the structure of both of these functions is known. A vector of random errors with a known distribution[1] ($\boldsymbol{\varepsilon} = \cup(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2)$) adds uncertainty to the dynamics and measurements. Finally, a set of parameters quantifying both the dynamics of the system (f) and the observation function (g), $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\mathrm{D}}, \boldsymbol{\theta}_{\mathrm{O}})$,[2] is unknown and the goal of the estimation process is to find these parameters. Note that the model and measurement functions may apply to a single case or multiple units of the phenomenon of interest. For example, a panel dataset includes measures on dynamics of several parallel units (e.g. people, firms, or countries) over time. Figure 2 summarizes the steps needed to estimate the model parameters, $\boldsymbol{\theta}$, by using the indirect inference method.

The intuition behind indirect inference is simple: if we calculate a set of empirical statistics from the data ($\mathbf{S}$), a good model of the data-generating process should be able to closely simulate these empirical statistics. Therefore, we search for a set of $\boldsymbol{\theta}$ parameters, which, when used to simulate the SD model, lead to simulated statistics that closely correspond to empirical statistics we have calculated using the empirical data. Consider a very simple example in which individuals' weight change over time is modeled in an SD

[1]The distribution of ε does not need to be known. ε can be a function of a random process with a known distribution and an unknown parameter of the model of interest ($\boldsymbol{\theta}$) (Gourieroux *et al.*, 1993). Moreover, if there is uncertainty in the initial conditions of stock variables, that uncertainty could be incorporated into ε.
[2]Note that $\boldsymbol{\theta}_{\mathrm{D}}$ and $\boldsymbol{\theta}_{\mathrm{O}}$ could each include multiple parameters describing a detailed SD model and various functions relating the stock variables to observable variables in an empirical setting.

DOI: 10.1002/sdr

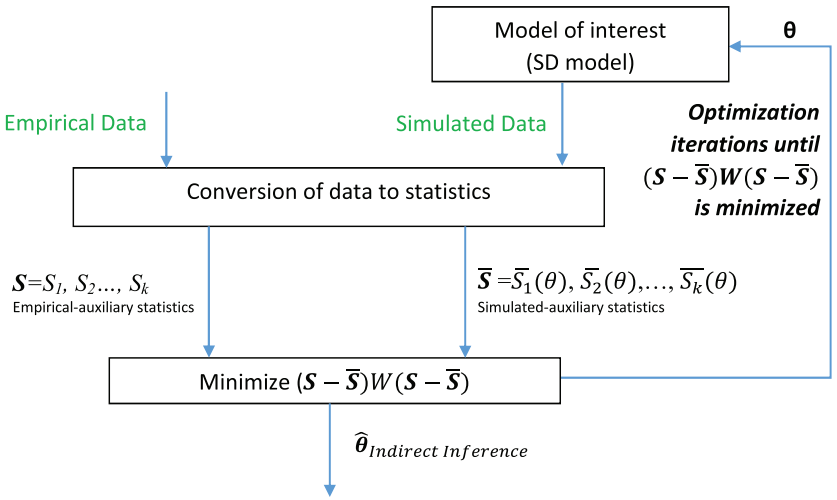

Fig. 2. The required steps for estimating parameters of a dynamic model

model with a single stock of **W** and the equations: d**W**/dt = **E** * (θ_1 + θ_2 * **W**) and **E** ~ Normal(0, θ_3). Assume that weight measurements at two points in time, $\mathbf{W}_1$ and $\mathbf{W}_2$, are available for 100 individuals, and we want to estimate the population-level **θ** parameters using these data. Note that we are assuming **θ** parameters are the same for all individuals, and individual differences are only coming from differences in initial weight and the randomness in the normally distributed **E** variable. Indirect inference instructs us to generate some empirical statistics, **S**, using this data, which should then be matched by simulations of the model. For example, **S** could include the average weight at time two, the variance of weight at time two, and the regression coefficients predicting $\mathbf{W}_2$ as a linear function of $\mathbf{W}_1$ and a constant (the regression has two coefficients), for a total of four statistics. We then simulate the body weight change of the 100 individuals, each starting from their actual $\mathbf{W}_1$ and growing based on the system's dynamics using a set of values of **θ**. We calculate the four elements of **S** for each simulated population, and change the **θ** until the four simulated statistics closely match the empirical ones, at which point we have found good estimates for the true **θ** values. The formal steps to accomplish this idea follow.

First, suitable statistics of empirical data, **x**, are generated **S** = (S_1, S_2, …, S_k). These statistics could include coefficients of an auxiliary model (e.g., a regression that estimates some elements of **x** based on other elements or lagged values) or they could be any statistics of a dataset such as mean and standard deviation (Wood, 2010). Second, the corresponding simulated statistics **S**(θ) = $S_1(\theta)$, $S_2(\theta)$, …, $S_k(\theta)$ are calculated/estimated. To gain better accuracy, for a given value of **θ**, the model of interest (SD model) is simulated H times by using H different streams of noise over time,

DOI: 10.1002/sdr

$\varepsilon_t\big(= \varepsilon_1^h, \ \ldots, \varepsilon_T^h\big), \mathrm{h} = 1, \ \ldots \mathrm{H}$. As a result, H replications of empirical x are generated and $\mathbf{S}(\theta) = (S_1(\theta), S_2(\theta), \ \ldots, S_k(\theta))$ is estimated for each replication of x. Third, the average of these estimators is found $\overline{S_k}(\theta) = \frac{1}{H}\sum_{h=1}^{H} S_k(\theta)^h)$ and θ is estimated by minimizing the difference between empirical-auxiliary statistics $(S_1, S_2, \ldots, S_k)$ and the average of corresponding simulated statistics $\overline{\mathbf{S}}(\theta) = \big(\overline{\mathbf{S}_1}(\theta), \overline{\mathbf{S}_2}(\theta), \ \ldots \ , \overline{\mathbf{S}_k}(\theta)\big)$ (Gourieroux *et al.*, 1993). The number of statistics, k, should be equal to or larger than the number of unknown parameters. A thorough explanation of each step is provided in the following sections.

1. **Define and estimate a set of empirical-auxiliary statistics**. The first step is to select a set of statistics, which, when matched in simulation, allow us to find the model parameters. There is substantial flexibility in terms of defining these statistics. Common statistics include mean, standard deviation, autocorrelation and correlation matrices of observed variables. Although these statistics are typically calculated across different units of analysis for cross-sectional and panel data (e.g. the mean weight in a group of subjects), they could also be calculated over time for a single case. Besides simple statistics, more complex auxiliary models could be defined that relate some of the observed variables to the other ones or to the lagged values of the same variable. The coefficients of these models (i.e. regression coefficients) could then be appropriate statistics to include in the statistics vector. We will refer to these statistics estimated from the empirical data as empirical-auxiliary statistics (S_1, S_2, …, S_k). Note that auxiliary models do not need to be accurate (i.e. the density function may not accurately describe the conditional distribution of x for the element of x being estimated (Durlauf and Blume, 2008)). It is an approximate model, which, unlike the model of interest, can be easily estimated with limited computational costs (e.g. using a simple linear regression). If the data-generating process (the model of interest, i.e. the SD model) is identical to the real-world data-generating process, we would then expect the replication of simulated auxiliary statistics to be close to the empirical ones. As a result, rather than trying to remove the biases in the auxiliary models (which may not be feasible because of limited data; for example, where we do not have continuous data but only irregular samples with measurement error), we find the data-generating model that gives the same results as those biased models using similar operations, which ensures that the model of data-generating process is fairly accurate. Note that in indirect inference the coefficients of auxiliary models are not the final output but rather a means to estimate the true model of the data-generating process. If they are inaccurate/biased, both the empirical- and simulated-auxiliary statistics would be biased for the same reasons and therefore should match when

DOI: 10.1002/sdr

the model of data-generating process matches the true process. In other words, the true model of the data-generating process can be accurately estimated by matching biased empirical and simulated auxiliary statistics. Consider a tangible example: you are given a warped picture of a scene and asked to find the actual scene from a line-up of undistorted images. The picture you have is distorted because it was taken with a warped lens that changed the shape and coloring of the actual image. Think of the characteristics of this warped picture as empirical-auxiliary statistics. Think of the search for the true scene as finding the parameters of the SD model that best matches the true data-generating process. Indirect inference uses the *same* warped lens to look at each candidate's true image, and picks the one that, once transformed with this warped lens, produces the closest fit to the initial (warped) picture. In this analogy, each candidate picture is generated when the SD model creates simulated data, and the warped lens is applied when simulated auxiliary statistics are calculated for those data. The best SD model is found when the simulated auxiliary statistics match the empirical ones.

While many different auxiliary models could be beneficial, the estimation would be more efficient if the auxiliary models were defined as precisely as possible; i.e. the auxiliary model is a good approximation of some aspects of f and g functions that are reflected in the estimated relationship (Guvenen and Smith, 2014). A more precise model reduces the variance of estimated regression coefficients (elements of S_1, S_2, …, S_k) and thus enables a reliable estimation with a smaller number of simulations, H. In general, good empirical-auxiliary statistics have information about the model parameters in question, and bound those parameters from multiple angles, so that the empirical evidence embedded in the statistics puts strong constraints on the value of unknown parameters.A couple of examples help to illustrate the idea of auxiliary models. Consider a nonlinear model: $y_i = \exp(x_i'\gamma) + u_i, u_i \sim \mathcal{N}[0, \sigma^2]$.The auxiliary model could be $y_i = x_i'\beta + \varepsilon_i,\ \varepsilon_i \sim \mathcal{N}[0, \sigma_\varepsilon^2]$ (Cameron and Trivedi, 2005). Another example is a two-level logistic model, $x_{ik} = \text{logit}^{-1}(p_{ik}) + e_{ik}$, in which $\text{logit}(p_{ik}) = \theta_0 + \theta_1 z_{ik} + u_k$ and x_{ik} is the ith observation in the kth group. This model has an intractable likelihood function and conventional estimation methods cannot estimate it. An auxiliary model for implementing indirect inference could be $x_{ik}^* = \beta_0 + \beta 1 z_{ik} + u_k + e_{ik}^*$ (Mealli and Rampichini, 1999).A good empirical-auxiliary statistic has four key characteristics. First, it should be relatively stable; i.e. its value should not be very sensitive to the process and measurement noise streams ($\boldsymbol{\varepsilon}$). The empirical value of a noise-sensitive statistic is not reliable and, as such, does not have much information to guide the identification of model parameters. Second, good statistics are sensitive to change in at least one of the parameters in $\boldsymbol{\theta}$. In the extreme, if a statistic does not vary with changes in any of the model parameters, there is no way to backtrack the value of any parameter

DOI: 10.1002/sdr

based on the information in that statistic. Both of these conditions could be partially tested using simulations. One could simulate a model in the range of considered parameters and measure the sensitivity of the simulated statistics with respect to model parameters $\left(\frac{\partial S_k}{\partial \theta}\right)$ and their sensitivity to different noise streams. Third, empirical statistics should be inexpensive to calculate; otherwise, the multiple iterations needed to solve the optimization problem may become infeasible. Therefore, simple linear regressions are preferred over regression models that require non-convex optimizations. Fourth, the number of statistics should be equal to or greater than the number of parameters that need to be estimated. In other words, $k \geq p$, where p and k are the number of elements in the vector $\boldsymbol{\theta}$ and the statistics vector $(S_1, S_2, \ldots, S_k)$, respectively. If $k < p$, unknown coefficients cannot be estimated and more statistics should be added to the vector of statistics to make k equal to or greater than p. After choosing the appropriate statistics, including the auxiliary model(s), the empirical-auxiliary statistics $(S_1, S_2, \ldots, S_k)$ are estimated or calculated using the empirical data x.

2. **Generate the simulated data using the SD model.** First, H independent drawings of $\varepsilon_t\left(\varepsilon_1^h, \ldots, \varepsilon_T^h\right)$ are generated. These streams of random numbers are generated only once and kept constant for the rest of the process. Then, for a given $\boldsymbol{\theta}$, the SD model is simulated H times (H replications using the independent drawings above). This process creates the simulated data that contains H paths $\left(x_1^h, \ldots, x_T^h\right)$, where $h = 1, \ldots, H$. The number of observations in each path should be equal to the number of observations in the empirical data. It should also be noted that the same $\varepsilon_t\left(\varepsilon_1^h, \ldots, \varepsilon_T^h\right)$ are used for each value of $\boldsymbol{\theta}$ simulated during optimization (i.e. we use the same noise seed values for each iteration of the optimization).

3. **Estimate the simulated-auxiliary statistics using the auxiliary model and simulated data.** For each of the H paths, the simulated statistics are estimated in the same fashion used in the calculation of the empirical-auxiliary statistics. The only difference is that, instead of using empirical data, simulated data are used to estimate those statistics, which we call simulated-auxiliary statistics. The key point is to generate the same statistics using the empirical and simulated data (they are both $\langle k \times 1 \rangle$ vectors). After finding the simulated statistics for each path, the average of these H simulated-auxiliary parameters is found as

$$\overline{S_k}(\theta) = \frac{1}{H} \sum_{h=1}^{H} S_k(\theta)^h \tag{2}$$

Typical values of H could range between a handful and hundreds, depending on the variance of the simulated auxiliary statistics. If that variance is high, a larger H is recommended to reduce error resulting from the simulation of

DOI: 10.1002/sdr

statistics. However, note that computational costs scale linearly with H and the incremental value of increasing H is limited because for the empirical statistics we only have a single path available; thus the total sampling error approximately scales with $(1 + 1/H)$.

4. **Minimize the difference between the empirical-auxiliary statistics and the simulated-auxiliary statistics.** The unknown parameters ($\boldsymbol{\theta}$) are estimated by minimizing the weighted differences between the empirical-auxiliary statistics $(S_1, S_2, \ldots, S_k)$ and the average of the simulated-auxiliary statistics $\left(\overline{S_1}(\theta), \overline{S_2}(\theta), \ldots, \overline{S_k}(\theta)\right)$. In other words, the parameters of the model of interest are estimated as

$$\widehat{\boldsymbol{\theta}} = \text{argmin}\left(\mathbf{S} - \frac{1}{H}\sum_{h=1}^{H}\mathbf{S}(\theta)^h\right)' \mathbf{W}\left(\mathbf{S} - \frac{1}{H}\sum_{h=1}^{H}\mathbf{S}(\theta)^h\right) \tag{3}$$

The weighting matrix ($\mathbf{W}$) can be any positive definite matrix in theory, but good choices of $\mathbf{W}$ are critical for obtaining reliable parameter estimates. Therefore, the calibration of the indirect inference is usually performed in a two-step procedure with two different values of $\mathbf{W}$. In the first step, $\mathbf{W}$ can be chosen to be a diagonal matrix in which the diagonal element i of the matrix is the inverse of square of the ith empirical statistic $(1/S_i^2)$ and the non-diagonal elements are zero (let us call this matrix $\mathbf{W_1}$). $\mathbf{W_1}$ ensures that some statistics do not dominate the optimization if their magnitude is much larger than the others. However, $\mathbf{W_1}$ is not theoretically optimal in the sense that it does not provide the lowest standard deviation for the estimated parameters. After performing optimization using the initial $\mathbf{W_1}$ and obtaining estimates of the model parameters $\left(\widehat{\boldsymbol{\theta}}_1\right)$, we switch to another $\mathbf{W}$, the inverse of the variance-covariance matrix of the simulated statistics (using $\widehat{\theta}_1$ to estimate this matrix) and repeat the estimation process. It is important to note that in calculating the variance-covariance matrix a large number of simulations (e.g. thousands)[3] using distinct noise streams will be needed. However, this step is done only once and not repeated during optimizations, so computational costs are not a concern. The intuition behind using the inverse of the variance–covariance matrix is that those statistics with a large variance (i.e. they are sensitive to the choice of random noise) should get lower weights. Although stopping after the second estimation gives reliable results in many applications, $\mathbf{W}$ can be recalculated (based on $\widehat{\boldsymbol{\theta}}$ achieved in the second step) to estimate a new set of parameters. This process can be iterated through until the estimated parameters converge across successive iterations.The initial assumed values for $\boldsymbol{\theta}$ could impact the speed of

[3]Note that number of simulations in steps 2 and 4 are different. In step 2, fewer simulation paths are needed, while for estimating the weight matrix consistently a much larger number of simulations should be used, because estimates of the covariance matrix require large numbers of simulations.

DOI: 10.1002/sdr

convergence in the optimization process or trap the optimization in a local optimum. If the coefficients of the auxiliary model and the unknown parameters in the main model are similar in their meaning, the initial values for model parameters could be chosen to equal the corresponding empirical-auxiliary statistics. If they are not similar, qualitative information on the appropriate range of such parameters or rough initial estimates using a relevant estimation method could help initialize the model from a more promising point in the parameter space. Even with good initial points, however, the optimization may become stuck in a local optimum, so the optimization algorithm should include multiple start points to increase the chances of finding the global optimum.

5. **Model assessment test.** When $k > p$, the model is said to be over-identified. Since in this method we minimize the distance between empirical-auxiliary statistics and their simulated counterparts, over-identification does not change the estimation approach. In fact, over-identification is helpful for further restricting the estimated parameters and offering tighter confidence intervals. It can also help further evaluate the model's overall goodness of fit, such that the optimal value of the objective function can be used to test how well the model has been specified. The following statistic (ξ_T) is distributed asymptotically as a chi-square with $k - p$ degrees of freedom. The null hypothesis is that the model of interest (our SD model) is not different from the true data-generating process. If the test statistic is larger than the threshold for chi-square distribution with the desired precision, we then reject the null hypothesis, inferring that the model's structure could be improved further:

$$\xi_T = \frac{H}{1+H} \min\left(\mathbf{S} - \frac{1}{H}\sum_{h=1}^{H}\mathbf{S}(\theta)^h\right)' \mathbf{W}\left(\mathbf{S} - \frac{1}{H}\sum_{h=1}^{H}\mathbf{S}(\theta)^h\right) \quad (4)$$

The estimation would be more efficient if the auxiliary model were defined as precisely as possible; i.e. the auxiliary model is a good approximation of some aspects of f and g functions which are reflected in the estimated relationship (Guvenen and Smith, 2014). A more precise model reduces the variance of estimated regression coefficients (elements of S_1, S_2, …, S_k) and thus enables a reliable estimation with a smaller number of simulations, H.

Calibration of the applied example

Here we demonstrate the application of indirect inference for estimating the SD model of depression depicted in Figure 1 using the panel dataset described in Table 1.

DOI: 10.1002/sdr

The steps to estimate the parameters of the depression–rumination model are as follows.

1. **Define and estimate a set of empirical-auxiliary statistics.** There is no simple rule to identify the best auxiliary model, and the modeler's judgment and insight guide this selection. However, estimation benefits generally from auxiliary models whose coefficients are informative about the parameters we want to estimate, e.g. those capturing similar types of information and relationships. Thus our auxiliary models include three regressions that are similar to the formulas to be estimated in the main model for *indicated rumination*, *indicated depressive symptoms* and *stressor memory* in the SD model (see Table 3). The first regression (presented in the first row, third column of Table 3) relates to rumination and includes coefficients informative about the *indicated rumination* formulation (first row, second column of Table 3, which quantifies the impact of gender, stressful life events and depression on rumination). Note that the SD model formulations are simulated in continuous time, whereas the auxiliary models use empirical data collected at discrete points in time (and are subscripted accordingly). In the SD model, we hypothesized that *indicated depressive symptoms* were a function of *rumination* (second row, second column of Table 3). As a result, we included rumination in the second regression (second row, third column of Table 3). The previous values of rumination and depressive symptoms were included in the first and second regressions, respectively, because the predictive power of the auxiliary models improved by adding them. In addition, incorporating previous values accounts for the inertia observed in those variables and encodes information about some of the time constants in our SD model. The third regression is an approximation of the change in stressor memory per month (third row, third column of Table 3). The change in stressor memory was divided by 7 months (the time between the two measurements) to get the stressor memory change per month.Note that one could come up with other auxiliary models and statistics, such as autocorrelations and correlations across empirical variables over time. By getting the auxiliary statistics from the above regression models, we tap into that more complex correlation information, while controlling for multiple factors. This allows us to use more informative signals than raw correlations and reduce the noise in the auxiliary statistics.

To estimate the auxiliary-empirical statistics, we ran the three regressions (third column of Table 3) using the ordinary least squares (OLS) technique (all statistics related to regressions are shown in the Appendix). In addition, we included the mean of depressive symptoms at T_3 and rumination at T_2 and T_3 as statistics. The resulting empirical statistics (S_1, S_2, …, S_{14}) are listed in Table 4. Because the number of statistics ($k = 14$) is greater than the number

DOI: 10.1002/sdr

Table 3. Formulae used in the SD model of depression and corresponding auxiliary models used in the indirect inference method

Concept	Formula used in the SD model	Corresponding auxiliary models
Rumination	$\text{indicated rumination}_t = (\theta_1 + \theta_2 \times \text{depressive symptoms}_t + \theta_3 \times \text{gender} + \theta_4 \times \text{stressor memory}_t)/(1-\theta_5)$	$\text{rumination}_3 = b_1 + b_2\ \text{depressive symptoms}_3 + b_3\ \text{gender} + b_4\ \text{stressor memory}_3 + b_5\ \text{rumination}_2 + b_6\ \text{rumination}_1 + \varepsilon_1$
Depressive symptoms	$\text{indicated depressive symptoms}_t = (\theta_6 + \theta_7 \times \text{rumination}_t)/(1-\theta_8)$	$\text{depressive symptoms}_3 = a_1 + a_2\ \text{rum}_3 + a_3\ \text{depressive symptoms}_1 + \varepsilon_2$
Stressor memory	$\text{stressor memory}_t - \text{stressor memory}_{t0} = \int_{t0}^{t} [\text{new stressors(s)} - \text{let it go}(s)]\mathrm{d}s$	$(\text{stressor memory}_3 - \text{stressor memory}_1)/7 = c_1 - c_2 \frac{\text{stressor memory}_1}{\text{rum}_1} + \varepsilon_3$

of unknown parameters (p = 12), we have enough degrees of freedom to test also the model's specification quality after estimation.

2. **Generate the simulated data using the SD model.** For generating a simulated data path, we first set the value of stocks to the corresponding empirical values (e.g. depressive symptoms$_0$ = depressive symptoms at T_1). We then generate H = 200 paths by adding random noise to the indicated rumination and indicated depression for each individual (the resulting noise matrix has 200 columns and 661 rows, with two noise values for each cell corresponding to the depressive symptoms and rumination noise). We repeat this procedure

Table 4. Value of the empirical auxiliary statistics

Regression	Statistic	Empirical auxiliary statistic
First row–third column of Table 3	b_1	−0.4663
	b_2	0.2313
	b_3	1.2021
	b_4	0.1316
	b_5	0.4548
	b_6	0.1749
Second row–third column of Table 3	a_1	2.0012
	a_2	0.2526
	a_3	0.5559
Third row–third column of Table 3	c_1	−0.0201
	c_2	−0.1222
Mean	Mean of depressive symptoms at T_3	9.7852
	Mean of rumination at T_2	10.8487
	Mean of rumination at T_3	9.9546

DOI: 10.1002/sdr

at every time step as we simulate each of the individuals over 7 months.

3. **Estimate the simulated-auxiliary statistics using the auxiliary model and simulated data.** After generating the simulated data, for a given **θ**, the simulated-auxiliary statistics are estimated in a similar fashion to the empirical-auxiliary statistics for each path. In this case, we run three regressions presented in the third column of Table 3, and include the other statistics to create a vector of simulated auxiliary statistics ($S_1(\theta)^h$, $S_2(\theta)^h$, …, $S_{14}(\theta)^h$) for each path. The average of these H simulated-auxiliary statistics is then found as $\frac{1}{H}\sum_{h=1}^{H} S_k(\theta)^h$.

4. **Minimize the difference between the empirical-auxiliary statistics and the simulated-auxiliary statistics.** A good estimate for the initial value of parameters can be found by running regressions on equations of indicated rumination and indicated depressive symptoms. The initial value for standard deviations of RumNoise and DepNoise are the residuals of these two regressions. The initial values of other parameters, effect of rumination on time constant (θ_9) and correlation time (θ_{12}) are arbitrarily selected. The initial values are summarized in the first column of Table 5. The unknown parameters (**θ**) are estimated by using the fmincon solver in MATLAB combined with its global search algorithm, which attempts smartly chosen start-points in the parameter space to increase the likelihood of finding the global optima. The same set of noise matrices are used in each iteration of

Table 5. Estimated parameters in the first and tenth rounds of optimization

Unknown parameters	Initial value (θ_0)	First round of optimization	10th round of optimization
Rumination constant (θ_1)	0.3320	−0.5064	−1.2504[−3.1920, 0.6911]
Effect of depressive symptoms on rumination (θ_2)	0.2490	0.1187	0.4236[−0.1661, 1.0132]
Gender coefficient (θ_3)	1.3540	0.7883	2.5152[0.5518, 4.4787]
Effect of stressors on rumination (θ_4)	0.1240	0.0824	0.2518[0.0227, 0.4809]
Rumination coefficient (θ_5)	0.5470	0.6202	0.1639[-0.8064, 1.1342]
Depression constant (θ_6)	2.0010	0.3207	0.3730[0.2968, 0.4491]
Effect of rumination on depressive symptoms (θ_7)	0.2520	0.0530	0.0699[0.0638, 0.0759]
Depression coefficient (θ_8)	0.5560	0.9102	0.8894[0.8822, 0.8967]
Effect of rumination on memory time (θ_9)	1.0000	2.1865	1.4741[1.3735, 1.5747]
Standard deviation of rumination noise (θ_{10})	5.8000	2.8678	7.8735[−0.1391, 15.8861]
Standard deviation of depression noise (θ_{11})	6.0500	0.0016	0.0002[−0.0307, 0.0311]
Correlation time (θ_{12})	1.0000	0.4266	1.6008[0.0456, 3.1559]

95% confidence interval is presented in parentheses.

DOI: 10.1002/sdr

the optimization to ensure comparability. The weight matrix (**W**) is estimated by using a large number of simulations (for this case we used 2000 simulations). The estimated parameters are shown in Table 5. All materials for estimating the parameters of the model are provided in the online Appendix.

Results

Table 5 shows the estimated parameters of the SD model, including the estimated parameters found in the first round of optimization and the estimated parameters found after 10 rounds of optimizations. In the first round of optimization the weighting matrix is $\mathbf{W_1}$, defined above. In the next rounds of optimizations, the weighting matrix is the inverse of the variance–covariance matrix of the statistics based on parameters estimated in the previous round of optimization. We ran 2000 simulations to estimate the weight matrices. The parameters fully converged after seven rounds of optimization.

Some of the parameters in Table 5 have tight confidence intervals, e.g. suggesting the strong impact of gender and stressors on rumination, of rumination and past depression on depressive symptoms, and of rumination on memory time. These are indeed the more relevant theoretical findings we may be seeking in this model. On the other hand, parameters specifying the pink noise characteristics are less reliable. It is possible that our data are not able to fully constrain the model parameters, or better model structures could be devised that fit these data more closely and with less variance in parameters. Moreover, the use of the analytical method used to calculate the confidence intervals, which relies on the normality of estimated parameters, may be imprecise. Indirect inference estimators are asymptotically normally distributed when auxiliary-empirical statistics are normally distributed. When that assumption is not reliable the use of bootstrapping methods (Dogan, 2007), while computationally more expensive, is more robust in finding the confidence intervals.

Figure 3 compares the results of the first round of optimization and the final optimization. The circles represent the simulated-auxiliary statistics and the bars depict the 95 percent confidence interval of empirical-auxiliary statistics (in which such confidence intervals are available from auxiliary model estimations). The estimated parameters from the first round of optimization generate a few simulated-auxiliary statistics that are far away from the 95 percent confidence interval of the empirical-auxiliary statistics (Figure 3A). After 10 rounds of optimization almost all of the simulated-auxiliary statistics are within the 95 percent confidence intervals of the empirical-auxiliary statistics (Figure 3B). Table 6 presents the values of the simulated and empirical auxiliary statistics shown in Figure 3.

DOI: 10.1002/sdr

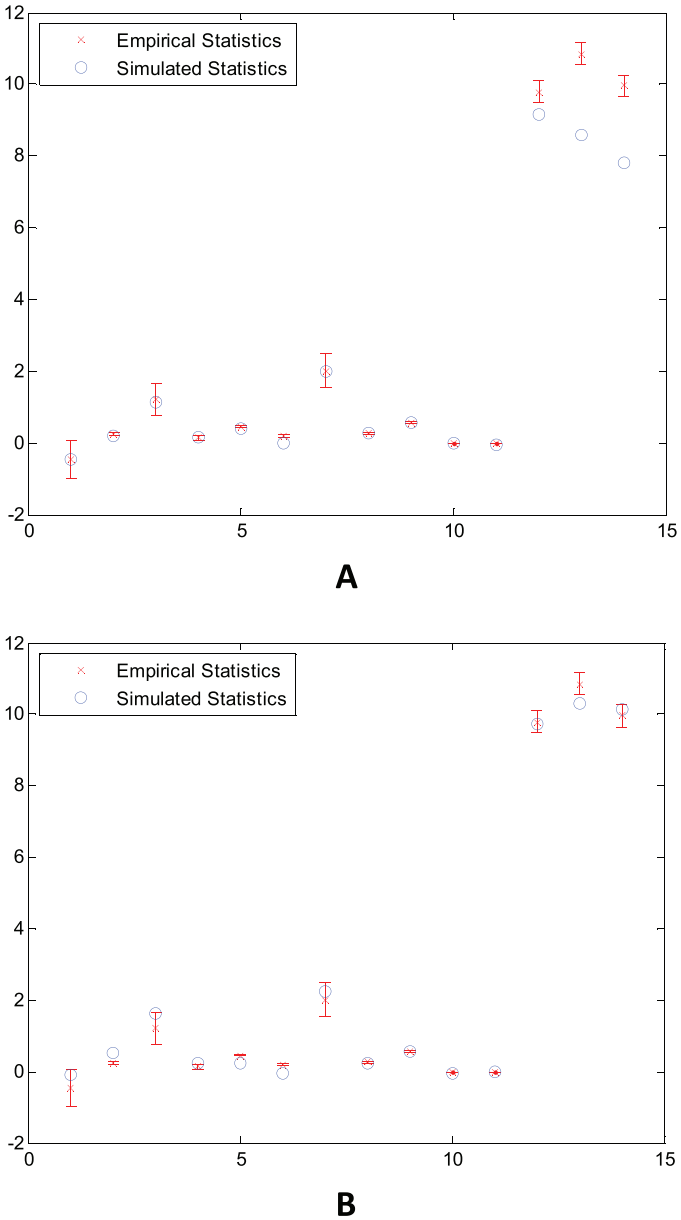

Fig. 3. Empirical-auxiliary statistics and simulated-auxiliary statistics generated using the estimated parameters from the first (A) and the tenth rounds of optimization (B)

Model specification and refinement

As explained in the fifth step of the method, when $k > p$ a test can be used to assess how well a model has been specified. Using Eq. (4), the test statistic ξ_T

DOI: 10.1002/sdr

was found to be 220. The 99 percent cut-off value for a chi-square distribution with 2 degrees of freedom (14 − 12 = 2) was 9.2. Although almost all simulated-auxiliary statistics are within the confidence interval of the empirical-auxiliary statistics, our ξ_T is still higher than the cut-off value $\left(\xi_T = 220 > \chi^2_2 = 9.2\right)$; hence the model can be further refined.[i]

Internal validity of this method can be tested using simulations. This is a good general practice when using complex models and estimation methods for which proofs of consistency and efficiency do not fully apply because of the divergence between theoretical requirements for those proofs and practical features of the problem at hand. Specifically, once a model is estimated, it can be used to generate a synthetic dataset, which is then used to re-estimate the model to find out if the new parameter estimates correspond closely to the empirical estimates. We thus check whether the parameters estimated by applying the indirect inference to a synthetic dataset—generated by simulating the calibrated model with coefficients reported in Table 5—are similar to true values (used for creating the synthetic data). The main idea behind this test is that in this case the data-generating process is perfectly modeled and true parameter values are known; thus any errors in parameter estimates can be attributed to the estimation method. The parameters in the third column of Table 5 are used to simulate the model and generate a synthetic dataset. All steps explained in the description of the model are then applied to the synthetic data to find the indirect inference estimates. As shown in Table 7, the true parameters that are used to generate the synthetic data (first column) are within the 95 percent confidence interval of the estimated parameters using the synthetic data (second column); however, the estimated confidence intervals are rather wide, suggesting the potential benefit of incorporating additional auxiliary statistics into the estimation process. In this synthetic estimation exercise the overall goodness of fit statistic, ξ_T, is 7.48, which is lower than the cut-off value (9.2), and thus the test does not reject the overall fit of the model to synthetic data, which is reassuring. Moreover, repeating this synthetic estimation a few hundred times would provide the bootstrapped confidence intervals for the model parameters, which are typically more reliable than the asymptotic estimates reported in Table 7.

Finally, building confidence in the estimation results calls for the inspection of individual parameter values and simulating the model extensively to ensure its behavior is robust and does not violate trends in data or physics of the problem. To demonstrate, we simulated the model using three different sets of parameters reported in Table 5 (i.e. initial parameters, parameters estimated in the first round of optimization and parameters estimated in the 10th round of optimization). For each set of parameters, we ran the model for 2500 female adolescents by changing the rumination and depression noise seeds and setting initial depressive symptoms, rumination and stressors at their mean. Results, reported in the online Appendix, show consistency among

DOI: 10.1002/sdr

Table 6. Values of empirical auxiliary statistics and simulated auxiliary statistics generated using the estimated parameters from the first and higher rounds of optimization

Regression	Statistics	Empirical auxiliary statistics	Simulated auxiliary statistics (first round of optimization)	Simulated auxiliary statistics (10th round of optimization)
	b_1	−0.4663[−1.4883, 0.5557][1]	−0.48496	−0.09359
	b_2	0.2313[0.1680, 0.2947]	0.196941	0.511783
	b_3	1.2021[0.3046, 2.0996]	1.128222	1.609224
	b_4	0.1316[0.0064, 0.2569]	0.134688	0.221293
Equation	b_5	0.4548[0.3819, 0.5276]	0.38188	0.225195
(5)	b_6	0.1749[0.1028, 0.2470]	−0.01602	−0.07331
	a_1	2.0012[1.0910, 2.9113]	2.004988	2.234688
Equation	a_2	0.2526[0.1894, 0.3157]	0.263558	0.234232
(6)	a_3	0.5559[0.4760, 0.6358]	0.538416	0.5394
Equation	c_1	−0.0201[−0.06850, 0.0282]	−0.02015	−0.06472
(7)	c_2	−0.1222[−0.1588, −0.0857]	−0.04511	−0.00205
	Mean of depressive symptoms at T_3	9.7852 [9.2013, 10.3690]	9.169462	9.722089
	Mean of rumination at T_2	10.8487 [10.2665, 11.4309]	8.563383	10.31732
Mean	Mean of rumination at T_3	9.9546 [9.3475, 10.5617]	7.8181	10.13648

[1]95% confidence interval are presented in brackets.

Table 7. Estimated parameters using empirical data and synthetic data

Unknown parameters	Parameters used to generate synthetic data	Estimated parameters using synthetic data
Rumination constant (θ_1)	−1.2504	−0.0915 [−3.83, 3.65][1]
Effect of depressive symptoms on rumination (θ_2)	0.4236	0.3111 [−0.03, 0.66]
Gender coefficient (θ_3)	2.5152	2.8423 [−1.58, 7.27]
Effect of stressors on rumination (θ_4)	0.2518	0.2411 [−0.19, 0.67]
Rumination coefficient (θ_5)	0.1639	0.1722 [−1.19, 1.54]
Depression constant (θ_6)	0.3730	−0.4226 [−7.00, 6.15]
Effect of rumination on depressive symptoms (θ_7)	0.0699	0.0948 [0.05, 0.14]
Depression coefficient (θ_8)	0.8894	0.8923 [0.77, 1.01]
Effect of rumination on memory time (θ_9)	1.4741	1.4920 [1.39, 1.60]
Standard deviation of rumination noise (θ_{10})	7.8735	7.1009 [0.90, 13.30]
Standard deviation of depression noise (θ_{11})	0.0002	17.9914 [−77.01, 113.00]
Correlation time (θ_{12})	1.6008	2.7767 [−1.22, 6.77]

[1]95% confidence interval are presented in brackets.

DOI: 10.1002/sdr

simulated trends and empirical ones, which is strongest for the 10th round of optimization.

Furthermore, polarity and magnitude of the estimated parameters should be considered to ensure that they are intuitively sound. For example, the negative polarity of θ_1—which was positive in the base case but estimated to be negative— calls for closer attention because rumination should not become negative. Further inspection suggests this negative constant is compensated for by the average impact of depression, gender and stressor memory (which are all positive), so that calculated values for indicated rumination remain non-negative.

Conclusion

This article provides a step-by-step introduction to the indirect inference method for estimating unknown parameters of dynamic models. In this method, the unknown parameters of the model of interest are estimated by matching the properties of empirical data and simulated data. In many applications, there are few empirical data points available over time; as a result, it is not feasible to use conventional estimation methods such as the least squared error. In addition, unlike traditional methods, indirect inference does not require calculation of the likelihood function, which may well be intractable for complex models. The indirect inference method extends the method of simulated moments by removing the requirement that the matching statistics be a set of valid moments. They can be parameters of an auxiliary model, which is not an accurate description of the data-generating process, but it can be estimated easily by conventional estimation methods. This extension opens the door to utilizing a large range of auxiliary statistics that may be more informative than regular moments, capture dynamic features of the data not included in regular moments, and include better signal-to-noise ratio. When the dynamic model is too complicated with intractable likelihood function, when very few empirical data points exist over time, or when the number of available valid moments are smaller than the number of parameters of the model, indirect inference might be one of the few feasible options to estimate the unknown parameters of an SD model.

Introducing the indirect inference approach to the SD community has two distinct benefits. First, it can make the contribution of SD to other fields more salient. For example, previous models of MDD have not explicitly incorporated the feedback mechanisms we discussed in our model (Wittenborn *et al.*, 2016). Our modeling and estimation results suggest that these feedback mechanisms are indeed important and may be central to understanding MDD. While the previous literature has speculated about such feedback mechanisms, it did not quantify them properly. More generally, qualitative data often establish the existence of a feedback mechanism,

DOI: 10.1002/sdr

and the really challenging task is estimating the strength of those mechanisms, which is central to both theory and practice. This estimation practice also contributes beyond SD by shedding light on the mechanisms through which stress and rumination contribute to depression. Although previous studies showed that rumination mediates the relations between stressful events and depression (Michl *et al.*, 2013), the mechanism of rumination's influence was not known; only one study hypothesized that rumination contributes to depression by keeping stressors alive without testing it (Ruscio *et al.*, 2015). We examined the validity of the hypothesized pathway by capturing it in the first loop of Figure 1 and estimating the significance of related parameters. In addition, in another article (Hosseinichimeh *et al.*, 2016), we estimate the time to forget a stressor separately for boys and girls, which is a useful measure for clinicians, and we simulate the model for diverse patients under different conditions and investigate the progression of depressive symptoms for them.

Second, this article also contributes to the SD discipline. In the absence of indirect inference, traditional calibration methods in the SD literature would not allow for the use of common data structures available in this field (e.g., with 2–3 data points per person) to estimate a feedback-rich model. Many empirical datasets in psychology, medicine, organization studies, economics and sociology share a similar structure. Thus a gap has emerged between the focus of SD modeling (which often focuses on building feedback-rich models) and the focus of mainstream research in social and behavioral fields (which often attempts to estimate simple models). Our results show the potential synergies between SD and indirect inference that could be explored well beyond MDD research.

Many advances in statistics have enabled researchers to estimate increasingly complex and realistic models with diverse types of data over the past three decades. We believe that for SD to contribute to mainstream disciplinary research across various fields of social and behavioral sciences, modelers must be able to draw on the best available methods in order to estimate feedback-rich, mechanism-based models using quantitative data. We hope that the introduction of indirect inference extends the toolbox of SD researchers and allows them to combine the benefits of a broad model boundary and feedback richness—which traditional SD brings to understanding various phenomena—with the quantitative rigor of modern econometrics.

Note

i. In applications of indirect inference, it is common that the chi-square test rejects the hypothesis that the model is indistinguishable from the true data-generating process. This is a very high bar to set for any simulation model: to generate data in a way that the resulting auxiliary statistics are indistinguishable from the true data-generating process. Various unobserved external influences often exist on the system's dynamics that make it hard, if not impossible, to satisfy such criteria. Moreover, the chi-square test is

DOI: 10.1002/sdr

potentially over-sensitive because it penalizes non-normal error terms with a normality assumption. Because normal distribution has thin tails, the statistic penalizes large errors very significantly, compared to what might be more appropriate for common fat-tailed distributions for the error. Thus it is more appropriate, and more often used, as a continuous metric to assess the quality of fit rather than a binary rejection/acceptance test.

Acknowledgements

We would like to thank Dr Kate McLaughlin for generously sharing her dataset. This research was funded by NIH/NIMH grant R21MH100515 (A.K. W., PI).

Biographies

Niyousha Hosseinichimeh is an assistant professor at the department of Industrial and Systems Engineering at Virginia Tech. Her research focuses on health policy and healthcare systems. She applies system dynamics approach and statistical methods to study health issues and to examine policy interventions.

Hazhir Rahmandad is an Associate Professor of System Dynamics at the MIT Sloan School of Management. His research aims at expanding the dynamic modeling and estimation tools and applying them to complex organizational and health problems.

Mohammad S. Jalali (PhD) is a research faculty at MIT Sloan School of Management. He is interested in simulation and model estimation methodologies, and the applications of dynamic modeling for complex sociotechnical problems. Mohammad is a former consultant at the World Bank and a former researcher at the U.S. Department of Energy. He is also the recipient of the 2015 Dana Meadows Award, the 2015 WINFORMS Excellence Award, and the 2014 Lupina Young Researcher Award.

Dr. Andrea Wittenborn is Associate Professor of Human Development and Family Studies at Michigan State University. Her research aims to improve the detection, continuous assessment, and treatment of depression.

References

Anderson JC, Gerbing DW. 1988. Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin* **103**(3): 411–423.

Cameron AC, Trivedi PK. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press: Cambridge, UK.

DOI: 10.1002/sdr

Croson R, Donohue K, Katok E, Sterman J. 2014. Order stability in supply chains: coordination risk and the role of coordination stock. *Production and Operations Management* **23**(2): 176–196.

Dogan G. 2007. Bootstrapping for confidence interval estimation and hypothesis testing for parameters of system dynamics models. *System Dynamics Review* **23**(4): 415–436.

Duffie D, Singleton KJ. 1993. Simulated moments estimation of Markov models of asset prices. *Econometrica* **61**(4): 929–952.

Durlauf SN, Blume LE. 2008. *The New Palgrave Dictionary of Economics*. Palgrave Macmillan: Basingstoke, UK.

Gourieroux C, Monfort A, Renault E. 1993. Indirect inference. *Journal of Applied Econometrics* **8**: S85–S118.

Gouriéroux C, Phillips PCB, Yu J. 2010. Indirect inference for dynamic panel models. *Journal of Econometrics* **157**(1): 68–77.

Guvenen F, Smith AA. 2014. Inferring labor income risk and partial insurance from economic choices. *Econometrica* **82**(6): 2085–2129.

Homer JB. 2012. Partial-model testing as a validation tool for system dynamics (1983). *System Dynamics Review* **28**(3): 281–294.

Hosseinichimeh N, Rahmandad H, Wittenborn AK. 2015. Modeling the hypothalamus–pituitary–adrenal axis: a review and extension. *Mathematical Biosciences* **268**: 52–65.

Hosseinichimeh N, Wittenborn AK, Rick J, McLaughlin KA, Jalali MS, Rahmandad H. 2016. *Rumination as a mediator exacerbating depression by keeping a stressful event alive*. Department of Industrial and Systems Engineering at Virginia Tech. Working paper.

Jalali MS, Rahmandad H, Ghoddusi H. 2013. Estimation of unknown parameters in system dynamics models using the method of simulated moments. In *Simulation Conference (WSC)*, Winter 2013.

Jalali MS, Rahmandad H, Ghoddusi H. 2015. Using the method of simulated moments for identification of dynamic models. In *Analytical Methods for Dynamic Modelers*. MIT Press: Cambridge, MA; 39–69.

Kalman RE. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* **82**(1): 35–45.

Mcfadden D. 1989. A method of simulated moments for estimation of discrete response models without numerical-integration. *Econometrica* **57**(5): 995–1026.

Mealli F, Rampichini C. 1999. Estimating binary multilevel models through indirect inference. *Computational Statistics and Data Analysis* **29**(3): 313–324.

Michl LC, McLaughlin KA, Shepherd K, Nolen-Hoeksema S. 2013. Rumination as a mechanism linking stressful life events to symptoms of depression and anxiety: longitudinal evidence in early adolescents and adults. *Journal of Abnormal Psychology* **122**(2): 339–352.

Monfardini C. 1998. Estimating stochastic volatility models through indirect inference. *Econometrics Journal* **1**(1): 113–128.

Monfort A. 1996. A reappraisal of misspecified econometric models. *Econometric Theory* **12**(4): 597–619.

Nolen-Hoeksema S, Larson J, Grayson C. 1999. Explaining the gender difference in depressive symptoms. *Journal of Personality and Social Psychology* **77**(5): 1061–1072.

Nolen-Hoeksema S, Stice E, Wade E, Bohon C. 2007. Reciprocal relations between rumination and bulimic, substance abuse, and depressive symptoms in female adolescents. *Journal of Abnormal Psychology* **116**(1): 198–207.

DOI: 10.1002/sdr

Pakes A, Pollard D. 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* **57**(5): 1027–1057.
Rahmandad H, Sabounchi N. 2011. Building and estimating a dynamic model of weight gain and loss for individuals and populations. In *29th International Conference of the System Dynamics Society*. DC: Washington.
Ruscio AM, Gentes EL, Jones JD, Hallion LS, Coleman ES, Swendsen J. 2015. Rumination predicts heightened responding to stressful life events in major depressive disorder and generalized anxiety disorder. *Journal of Abnormal Psychology* **124**(1): 17–26.
Smith AA. 1993. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* **8**(S1): S63–S84.
Smith GL, Schmidt SF, McGee LA. 1962. *Application of Statistical Filter Theory to the Optimal Estimation of Position and Velocity on Board a Circumlunar Vehicle*. National Aeronautics and Space Administration: Washington, DC.
Snijders TAB. 2001. The statistical evaluation of social network dynamics. *Sociological Methodology* **31**(1): 361–395.
Wittenborn AK, Rahmandad H, Rick J, Hosseinichimeh N. 2016. Depression as a systemic syndrome: mapping the feedback loops of major depressive disorder. *Psychological Medicine* **46**(3): 551–562.
Wood SN. 2010. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature* **466**(7310): 1102–1104.

Supporting information

Additional supporting information may be found in the online version at the publisher's web-site.

DOI: 10.1002/sdr