

AN EMPIRICAL ESTIMATION OF STATISTICAL INFERENCES FOR SYSTEM DYNAMICS MODEL PARAMETERS

Mohammed Mesabbah
Wael Rashwan
Amr Arisha

3S Group – College of Business,
Dublin Institute of Technology (DIT)
Aungier Street
Dublin 2, IRELAND

ABSTRACT

For system dynamics simulation (SD) models, an estimation of statistical distributions for uncertain parameters is crucial. These distributions could be used for testing models sensitivity, quality of policies, and/or estimating confidence intervals for these parameters. Assumptions related to normality, independence and constant variation are often misapplied in dynamic simulation. Bootstrapping holds a considerable theoretical advantage when used with non-Gaussian data for estimating empirical distributions for unknown parameters. Although it is a widely acceptable approach, it has had only limited use in system dynamics applications. This paper introduces an application of Direct Residual Bootstrapping (DRBS) for statistical inference in system dynamic model. DRBS has been applied successfully to ‘The Irish Elderly Patient Delayed Discharge’ dynamic model to estimate empirical distribution for some unknown parameters with a minimal computation effort. The computational results show that bootstrapping offers an efficient performance in cases of no availability of prior information of model parameters.

1 INTRODUCTION

Numerical parameter estimation is one of the pestering issues in System Dynamics (SD) modeling. One of the challenges is that the model validation depends highly on the accuracy of the parameter estimations (Meadows 1980). Failure to set uncertain parameters during the design phase can impact on the dynamic behavior, which in turn can lead to undesirable and unforeseen consequences (Ng, Sy, and Lee 2012). Lack of the availability of historical data often makes it difficult to estimate all model parameters. Hence, estimating unknown parameters based on expert judgments (hand-calibration) can be a plausible solution, and is a common practice for assigning parameter values (Lyneis and Pugh 1996). But this method has been criticized for few reasons: (1) it is often difficult, and relies mainly on the experience of the experts concerned; (2) the results are not replicable; (3) it is not optimal or even best estimate; (4) it makes sensitivity analyses and building confidence bands more cumbersome and less robust. In practical terms, statistical and judgment methods can work together to enhance the estimation of parameters, where former can be used to estimate the possible ranges (the distribution) of parameter values which can then be used to check the judgmental estimates (Sterman 2000). When formal parameter estimation procedures are used, modelers typically compare time series data against the same variables in their model, and minimize the weighted sum of a function of the error terms, using a nonlinear optimization algorithm to adjust uncertain parameter values until ‘best fit’ estimates are found (Oliva 2003; Jun and

Ng 2013). Rahmandad et al (2013) introduced a ‘simulated moments’ method as a non-parametric method of estimating the parameters of dynamic models. A number of statistically-based or optimization-based techniques can be employed to overcome the shortcomings of hand-calibration. Lyneis and Pugh (1996) concluded that combining priori estimations and qualitative information with statistical approaches improves the efficiency and reliability of system dynamics practice. Econometric approaches are the most commonly used methods, in particular multiple regression based on ordinary least squares (OLS), although early studies (Senge 1977), concluded that econometric approaches were not useful because of the propensity of SD models to violate the ordinary least squares estimation assumptions. OLS can estimate accurate parameters when there is: no co-linearity (correlation among the independent variables); no autocorrelation (where the dependent variable depends on its own previous values); no heteroskedastic (non-stationary variance thorough the sample), and normality assumptions of the residuals (Sterman 2000).

Sensitivity analyses are often used to judge the sensitivity of model behavior to uncertain assumptions about model formulations and parameter values (Moxnes 2005). Moxnes stated that, in addition to the importance of sensitivity analysis for model testing, learning and validating, it would be also useful to observe how the optimal policy changes in line with variations in model parameters. For these reasons, estimating model parameter(s) distribution is required in case some parameters are uncertain, and sufficient information about their variations is unavailable. Hypothesis testing and confidence interval estimation are useful to modelers to build confidence about parameter estimates. The literature reports confidence interval estimation as being used for steady state simulations such as discrete event simulation models and queuing systems (Alexopoulos, Goldsman, and Tang 2013; Calvin and Nakayama 2013). For system dynamics models, the methods reported confidence interval estimation in literature impose assumptions about the data, such as it being normally distributed, independent, and not auto correlated or heteroskedastic, whereas system dynamics models usually violate these assumptions (Dogan 2007).

In his work, Dogan (2007; 2004) used residual bootstrapping (RBS) for estimating confidence intervals for parameters in system dynamics models. Dogan (2004) makes a comparison between residual bootstrapping and likelihood ratio method using synthetic data, concluding that the former is more appropriate for dynamic models, as it does not require strong assumptions and is valid for small samples. This study uses this method to investigate the statistical inferences of system dynamics model parameters. Residual resampling is a robust method for data analysis, and Dogan (2007) proposed the residual bootstrapping method as suitable to estimate parameters’ confidence intervals, applying it to two small system dynamics models with synthetic data. The originality in this study is threefold: (1) it introduces double residual bootstrapping (DRBS) in order to reduce the required computations and to overcome the large number of parameter fittings; (2) it investigates the utility of the bootstrap method in cases where there is or is not prior information about the parameters; (3) it applies this modified method to large real system dynamic model with real data.

Section 2 of this article presents the research methodology, and provides a brief overview of the bootstrapping method, and its appropriateness and challenges for the study of dynamic systems. Followed by Section 3, which describes the proposed framework, Section 4 reports the empirical results and analysis of the healthcare case study.

2 METHODOLOGY

The bootstrap method is an approach for estimating the distribution of an estimator or test statistic by resampling data under conditions that hold in a wide variety of econometric applications. It provides approximations to distributions of statistics, coverage probabilities of confidence intervals, and rejection probabilities of hypothesis tests that are more accurate than the approximations of first-order asymptotic distribution theory. The technique, introduced by (Efron and Tibshirani 1994), is growing in popularity in the statistics and econometrics literatures as computation becomes cheaper and faster, and because it is applicable to general problems. In outline, the bootstrapping method involves resampling from a sample - that is, generating many new data sets resampled from the original data in order to estimate

parameters (θ) for each new generated data set, which allows an empirical distribution to be estimated for that parameter ($\hat{\theta}$). This empirical distribution can then be used to estimate relevant characteristics of the population by statistical inference. Resampling methods can be parametric or nonparametric (Chernick 2011; Efron and Tibshirani 1). Many ways in which bootstrapping can be used, including direct or residual bootstrapping, the latter being more appropriate for dynamic models (Dogan 2004). Helpful reviews and details of bootstrapping methods are provided by Efron and Tibshirani (1994) and DiCiccio and Efron (1996), and Hongyi Li and Maddala (1996) provide an extensive survey of the approach.

2.1 Automatic Model Parameters Calibration

Assume different multiple actual data variables P are used to estimate the parameter vector θ . Let $\{D_{it}, t = T_0, T_0 + 1, T_0 + 2, \dots, T\}$ be a series of the actual data variable D_i , where t is the subscript for time, T_0 the initial time, and T the final time. Assume $\tilde{\theta}$ is the information available about the parameters (i.e., estimated judgments, expert guesses, or previous estimates). Let $n = (T - T_0 + 1)$ be the total number of simulation runs, and $\{Y_{it}, t = T_0, T_0 + 1, T_0 + 2, \dots, T\}$ the corresponding simulated output endogenous variable Y_i that has been generated using a simulation model of a non-linear function that represents the dynamics of the simulation model Γ ; where i represents i^{th} time series and $i \in \{1, 2, \dots, P\}$. The function Γ represents a system of differential equations. The output(s) variables (Y_{it}) of the model are non-linear functions (Γ) of system state variables (s_t), known model inputs (v_t), and unknown model parameters (θ), where $Y_{it} = \Gamma(s_t, v_t, \theta)$. Model calibration is the process of estimating the unknown model parameters, which minimizes the sum of the non-linear function of the residuals ε_{it} . Generally, the residual is the difference between actual data (D_{it}) and simulated data (Y_{it}). Formally, the error function is given by equation (1). Kleijnen and Sargent (2000), Sterman (2000) and Reichelt, Lyeins, and Bespolka (1996) provide and compare ranges of error functions.

$$\mathfrak{F}_i(\varepsilon_{it}) = \sum_{t=T_0}^T \mathcal{F}(Y_{it} - D_{it}), \forall i = 1, 2, \dots, P \tag{1}$$

his study uses the root mean square errors (RMSE) metric as an error function which can be defined as $RMSE_i = \mathcal{F}(Y_{it} - D_{it}) = (n^{-1} \sum_{t=T_0}^T (Y_{it} - D_{it})^2)^{1/2}$. The calibration process mainly attempts to match the behavior of the structure (Oliva 2003). Several methods can be employed to calibrate or to estimate model parameters that optimize an arbitrary chosen error function $\mathfrak{F}_i(\varepsilon_{it})$ without losing generality. The vector of residuals term ε_i is assumed to be iid with mean zero. The calibration process in this study is described as a non-linear optimization problem, where decision variables are the parameters (θ) that should be selected in order to minimize a weighted average objective function of the residuals (ε_{it}). The weights (ω_i) reflect the importance of each time series data i . The optimization model is constrained by the lower and upper limits of the parameter(s) ($\ell_{\theta}, u_{\theta}$) and a system of differential equations represented by (Γ). Mathematically, the optimization model of the parameter estimation is provided here:

$$\min_{\theta} \mathfrak{F}_i(\varepsilon_{it}) = \sum_{i=1}^P \omega_i \sum_{t=T_0}^T \mathcal{F}(Y_{it} - D_{it}) \tag{2}$$

$$\text{subject to: } Y_{it} = \Gamma(s_t, v_t, \theta), \text{ and } \ell_{\theta} \leq \theta \leq u_{\theta}$$

where T_0 is the initial simulation time, and T is the final simulation time. Solving this non-linear optimization problem is beyond the scope of this study. However, a variety of solvers and algorithms are available, from commercial or open sources, that can be used to obtain the optimal (or near to optimal) solution. Miller (1998) reported briefly on optimization techniques (i.e., genetic algorithm, simulated annealing, gradient methods).

2.2 Residuals Bootstrapping Procedure (RBS)

Let B is number of bootstraps, and b the current bootstrap iteration where $b \in \{1, 2, 3, \dots, B\}$. Assume that B is sufficiently large to obtain a representative bootstrap sample. The algorithm steps are:

Step 1: Set the bootstrap iteration number $b = 0$.

Step 2: Fit the model data Y_{it} to the actual data D_{it} by solving the optimizing problem provided in equation (2) to obtain a first optimal estimate of the parameter vector ($\hat{\theta}$).

Step 3: Compute the error terms ($\hat{\varepsilon}_{it}$) for the optimal parameter values. Centering the residual terms ($\hat{\varepsilon}_{it}$) at zero mean is given by equation (3).

$$\hat{\varepsilon}_{it} = (Y_{it} - D_{it}) - \mu_{\hat{\varepsilon}_i} \quad (3)$$

where $\mu_{\hat{\varepsilon}_i} = (n)^{-1} \sum_{t=1}^T \hat{\varepsilon}_{it}$

Note: there are two ways to compute residuals:

(3a) where prior information about the parameters $\tilde{\theta}$ is available, compute (Y_{it}) via simulation using the original parameter values ($\tilde{\theta}$), where ($Y_{it} = \Gamma(s_t, v_t, \tilde{\theta})$),

(3b) where there is no such prior information, compute Y_{it} via simulation using the first estimate ($\hat{\theta}$), where ($Y_{it} = \Gamma(s_t, v_t, \hat{\theta})$).

Step 4: Set $b = b + 1$.

Step 5: Resample the residual terms using a parametric or nonparametric technique to obtain new residual values (ε_{it}^{*b}), where ε_{it}^{*b} represents the new residuals value of the i^{th} variable at time t in bootstrap iteration b .

Step 6: Generate new data sets (D_{it}^{*b}) by adding the re-sampled error terms to the model output. The new generated data are given by equation (4).

$$D_{it}^{*b} = Y_{it}^b + \varepsilon_{it}^{*b} \quad (4)$$

Step 7: Use the new generated data set (D_{it}^{*b}) to estimate the parameter value(s) ($\hat{\theta}^{*b}$) with automated calibration as provided by equation (2).

Step 8: Repeat steps 4, 5, 6, and 7 until $b = B$.

At the end of running this procedure, B estimates of the parameter vector (θ) will have been obtained. Formally, the parameter bootstrap estimates, represented by $\hat{\theta}^* = \{\hat{\theta}^{*b}, b = 1, 2, 3, \dots, B\}$. $\hat{\theta}^*$, can be used to construct estimate empirical distribution for the parameter vector (θ).

2.3 Parametric and Nonparametric Resampling

The simplest sampling technique is nonparametric sampling, because it does not require any distributional assumptions, but it requires a large sample of data to be sufficiently representative. In this method, each new set of residuals is created by resampling randomly, with replacement, from the set of residuals ($\hat{\varepsilon}_{it}$) associated with the best fit parameters ($\hat{\theta}$) or the original parameters values ($\tilde{\theta}$). No such assumption needs be made in the parametric method. The empirical distribution function is given as:

$$F(x) = n^{-1} \sum I(\theta \leq x) \quad (5)$$

where $I(\cdot)$ is the indicator function. On the other hand, parametric sampling is the process of sampling from a parametric probability function that can be obtained by fitting residual terms to a specific theoretical probability distribution. This method is efficient if the residual term fits a parametric distribution as well, as it does not require large data set. Once the distribution parameters are estimated, the distribution can be used to generate a new data set by sampling error terms from the parametric distribution. However, distribution assumptions must be respected in this method.

2.4 Autoregressive Model

As noted earlier, the advantage of bootstrapping is that it does not require statistical assumption for residuals as do the other traditional methods for estimating parameter distributions, especially given that most dynamic models violate these assumptions. Exploring and studying the residuals obtained from

fitting actual data to simulated data is most essential in order to investigate sources of variability and bias in parameter estimation problems (Oliva 2003; Sterman 2000). An autocorrelation spectrum can be used to detect the auto-dependency of the residuals for different time lags.

The assumption that the residuals are not correlated is most likely to be violated in dynamics models, which could lead to a bias in bootstrapping estimation when resampling the residuals ignoring that they may be auto-correlated. Formal tests can be conducted by computing the autocorrelation function values and their variances (Barlas 1989; Barlas, Topaloglu, and Yilankaya 1997). If the error terms are auto-correlated, they should be removed before resampling.

Formally, the centered residual terms ($\hat{\varepsilon}_{it}$) are computed using equation (3) in step 3 for each output variable and then autocorrelation coefficients (r_k^i) in the residual terms are computed for different lags $k \in \{0, 1, 2, \dots, \mathcal{K}\}$, where \mathcal{K} is the maximum number of lags and $\mathcal{K} < n - 1$. Mathematically, the equation of sample autocorrelation spectrum for lags $k \in \{0, 1, 2, \dots, \mathcal{K}\}$ is given by:

$$r_k^i = \frac{\sum_{t=1}^{n-k} \hat{\varepsilon}_{it} \cdot \hat{\varepsilon}_{i(t+k)}}{\sum_{t=1}^n \hat{\varepsilon}_{it}^2}, \forall i \leq P \tag{6}$$

Then the hypothesis as to whether $\{r_k^i: \forall k \leq \mathcal{K}\}$ are significantly different than zero can be tested. Hence, the null and alternative hypothesis can be formulated as following:

$$H_0: r_k^i = 0, \forall k \leq \mathcal{K}$$

$$H_1: \exists r_k^i \neq 0, \forall k \leq \mathcal{K}$$

The test statistic $\tau_i(k)$ for each individual r_k^i is provided by equation (7), using the autocorrelation function variance estimator developed by Barlas (1994).

$$\tau_i(k) = \frac{r_k^i}{\hat{\sigma}_{r_k^i}} \tag{7}$$

$$\hat{\sigma}_{r_k^i} = \left((n(n+2))^{-1} \sum_{t=1}^{n-k-1} (n-t) [r_{k-t}^i + r_{k+t}^i - 2r_k^i r_t^i]^2 \right)^{1/2}$$

Equation (7) is the test statistic of autocorrelation at lag k . The hypothesis test is of individual autocorrelation at significance level α under the assumption of normal autocorrelation values. The null hypothesis of this individual test is rejected at significance level α if $|\tau_i(k)| > z^{1-\frac{\alpha}{2}}$. Rejecting the null hypothesis for at least one individual test means that autocorrelation violates the assumption of the residuals' independence, but the bootstrapping method can still be valid even the residuals are not independent. Formally, the auto-dependency of residuals $\hat{\varepsilon}_{it}$ can be modeled using p^{th} autoregressive model; AR(p).

$$\begin{aligned} \hat{\varepsilon}_{it} &= \sum_{j=1}^p \varphi_{ij} \hat{\varepsilon}_{i(t-j)} + e_{it} \\ e_{it} &\sim N(0, \sigma_{e_{it}}^2) \end{aligned} \tag{8}$$

where φ_{ij} is the j^{th} coefficient of autoregressive model of variable i , and e_{it} is a white noise with zero mean and variance $\sigma_{e_{it}}^2$. The underlying white noise e_{it} should be tested subject to normality and autocorrelation to insure their independence. After estimating the autoregressive process using equation (8), the bootstrapping process can be applied by resampling the white noise (e_{it}) parametrically or non-parametrically to obtain e_{it}^{*b} . Based on the autoregressive model, both steps 5 and 6 of the residual bootstrapping procedure given above can be revised as follows:

Step 5: Use equation (8) to estimate the autoregressive residual terms using $\hat{\varepsilon}_{it} = \sum_{j=1}^p \varphi_{ij} \hat{\varepsilon}_{i(t-j)}$, and compute the estimated 'white noise' terms using $e_{it} = \hat{\varepsilon}_{it} - \hat{\varepsilon}_{it}$. Resample those terms e_{it} using a parametric or nonparametric technique to obtain new residual values (e_{it}^{*b}), where e_{it}^{*b} represents the new residual value of the i^{th} variable at time t in bootstrap iteration b .

Step 6: Generate new data sets (D_{it}^{*b}) by adding the re-sampled error terms to the model output. The new generated data are given by equation (9).

$$D_{it}^{*b} = Y_{it}^b + \hat{\varepsilon}_{it} + e_{it}^{*b} \tag{9}$$

2.5 Confidence Intervals Estimation using the Bootstrap Method

The bootstrap method is particularly useful as an approach to estimating standard errors, and obtaining confidence intervals for location parameters based on these robust estimators. The trouble with standard intervals is that they are based on an asymptotic approximation that, in practice, can be quite inaccurate. There are different methods available for estimating bootstrap confidence intervals for estimated parameters, of which this paper utilizes two: bootstrap percentile confidence interval and bias corrected bootstrap percentile interval. (For complete discussion and detailed algorithms of both methods, see Davison 1997; Efron and Tibshirani 1994; DiCiccio and Efron 1996).

First, the bootstrap percentile confidence interval method is based on the quantile of the bootstrap distribution of the parameters estimate. Formally, the bootstrap percentile confidence interval at the $100\alpha\%$ significance level is given as:

$$[\hat{\theta}^{*(\alpha/2)}, \hat{\theta}^{*(1-\alpha/2)}] \tag{10}$$

where $\hat{\theta}^{*(\frac{\alpha}{2})}$ and $\hat{\theta}^{*(1-\frac{\alpha}{2})}$ represent the actual quantiles of the bootstrap distribution of estimates. To figure out the confidence intervals in equation (13), first sort the bootstrap estimates in ascending order. The lower limit of $\hat{\theta}^{*(\alpha/2)}$ is obtained by locating the $(\alpha/2)100^{\text{th}}$ position in the ordered list, while $\hat{\theta}^{*(1-\alpha/2)}$ is located by the $(1 - \alpha/2)100^{\text{th}}$ position in the ordered list.

Second, the bias correction bootstrap confidence interval method basically tries to take estimated bias into account. According to Efron and Tibshirani (1994) the bias correction constant is defined as:

$$\hat{z}_0 = \Phi^{-1}(B^{-1} \sum_{b=1}^B I(\hat{\theta}^{*b} < \hat{\theta})) \tag{11}$$

where $I(\cdot)$ is the indicator function and $\Phi^{-1}(\cdot)$ is the inverse cumulative standard normal distribution. So the bias correction bootstrap percentile confidence interval is given by:

$$[\hat{\theta}^{*\alpha_1}, \hat{\theta}^{*\alpha_2}] \tag{12}$$

where α_1 and α_2 are modified quantities of the location of the confidence interval's endpoints. The confidence interval endpoints at significant level $100\alpha\%$ are defined as:

$$\alpha_1 = \Phi(2\hat{z}_0 + Z^{\alpha/2}), \text{ and} \tag{13}$$

$$\alpha_2 = \Phi(2\hat{z}_0 + Z^{(1-\alpha/2)}) \tag{14}$$

where Φ is the cumulative standard normal distribution. (Further comments on the sample size are provided by Polansky 2000 and by Chernick 2011).

3 DIRECT RESIDUALS BOOTSTRAPPING PROCEDURE (DRBS)

It is worth pointing out a problem associated with estimating statistical inferences using residual bootstrap method in large-size SD models is the estimation process might potentially require larger bootstraps. Hence the computational process can become lengthy and expensive. To overcome this drawback, this study proposes applying a direct bootstrap method on the empirical distribution estimated using the residual bootstrapping method. The idea here is simply resampling from the empirical distribution estimated from RBS. Roughly speaking, the residual bootstrapping method aforementioned in section 2 is computationally expensive due to the auto calibration processes that require simulation optimization procedures to estimate the best fit parameters. Instead using large numbers of residual bootstraps to estimate unknown parameters, we suggest using much smaller number of residual bootstraps and applying a direct bootstrap method with a large number of bootstraps. Direct bootstrapping is computationally inexpensive, as it simply involves resampling from the estimated empirical distribution.

Formally, suppose that the $\hat{\theta}^* = \{\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*B}\}$ are residual bootstrap estimates of parameter theta, a sample that has been obtained using the residual bootstrap procedure. The empirical distribution function F of $\hat{\theta}^*$ is given by equation (5). B independent drawn samples are taken with replacement from F to obtain a sample $\hat{\theta}^{**}(\mathcal{M}) = \{\hat{\theta}^{**1}, \hat{\theta}^{**2}, \dots, \hat{\theta}^{**B}\}$. This sample can be used to compute $\bar{\theta}^{**}$ using the same formula as for $\bar{\theta}^*$. Repeat the sampling with replacement B cases from F and computing $\bar{\theta}^{**m}$ results in $\{\bar{\theta}^{**1}, \bar{\theta}^{**2}, \dots, \bar{\theta}^{**\mathcal{M}}\}$, where m is the direct bootstrap sample and \mathcal{M} is the total number

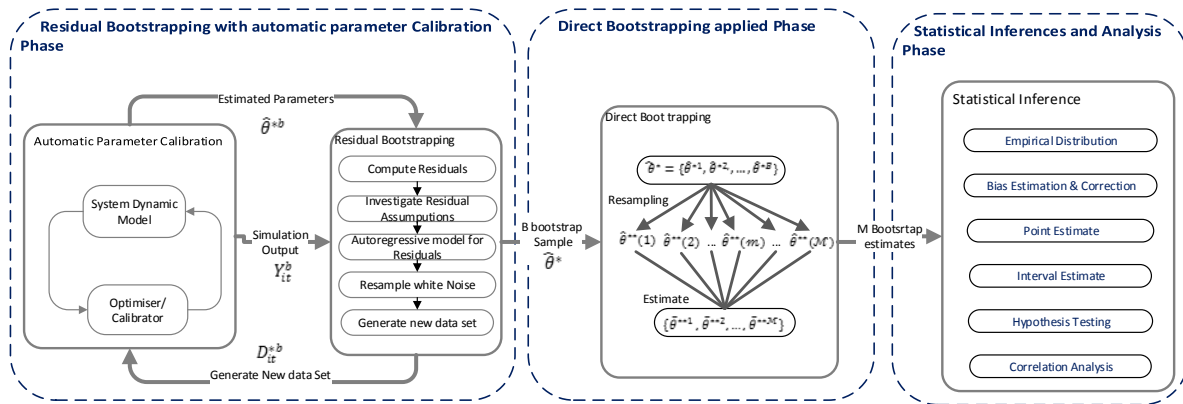


Figure 1: The proposed framework for estimating SD model's parameters statistical inference.

of direct bootstrap samples drawn given $\mathcal{M} \gg B$. The proposed direct bootstrap integration with residual bootstrap for statistical inference of SD model's parameters framework is represented in Figure 1.

4 EMPIRICAL RESULTS

The empirical results were obtained by applying the RBS and DRBS to the SD model of delayed elderly patients discharge from Irish public healthcare (model developed by Rashwan et al. 2013). The historical data is used in order to fit parameters of delayed discharge patients in Irish public hospitals in 2012, which represents (D_t). Three parameters (θ) of the model are estimated: (1) **'Rehabilitation to Long Term Care Ratio'** represents the ratio of patients referred to long care units directly from rehabilitation services; (2) **'Average time in Rehabilitation'** represents the average the patients spend in rehabilitation services; and (3) **'Maximum Waiting Time to Access Rehabilitation Services'** represents the time taken by patients to access rehabilitation services. Short in data sources and inconsistency in available data were the main challenges that faced the developer of SD model. This issue inspires the idea of applying RBS and DRBS in the SD model to estimate statistical inferences for the three aforementioned parameters.

For simplicity, the three parameters are denoted as θ_1, θ_2 and θ_3 respectively. Two types of experiments were run, both using RBS and DBS but with different assumptions about the parameters. The first (represented as $\hat{\theta}$) used prior information (i.e., guess, judgments, or estimations based on past data) - in this case values computed based on Irish delayed discharge data for 2011. The second experiment addressed the problem of the unavailability of prior information about the model parameters (i.e., unknown parameter values). This experiment aimed to investigate the usefulness of bootstrapping method to construct an empirical distribution for unknown model parameters. This would be practicable when the parameters are completely unknown, then the first estimates of the parameters ($\hat{\theta}$) are used. In the experimentation, RBS(B) refers to the residual bootstrapping method (as presented in section 2.), with number of bootstrap replications B, while DRBS(B, M) refers to the DRBS, method (as proposed in section 3) with residual bootstrap replications B and direct bootstrap replications \mathcal{M} .

The Elderly Delayed Discharge SD model in the experiments was developed using Microsoft Excel spreadsheets, and the visual basic for application (VBA) programming language was used to develop the RBS and DRBS methods. The 'Solver add-ins' integrated in Microsoft Excel were used as the optimizer/estimator for each calibration that minimizes the RMSEs required. The RBS requires a minimum of 1000 residual re-sampling iterations to get good results (Efron and Tibshirani 1994; DiCiccio and Efron 1996). Hence, the SD model used in this case was calibrated for 1000 times. With a relatively large SD model such as the elderly care model, one RBS run demands huge computational time (run execution usually takes approx. 5 hours), so the RBS and DRBS experiments presented in this section used 1000 iterations per run.

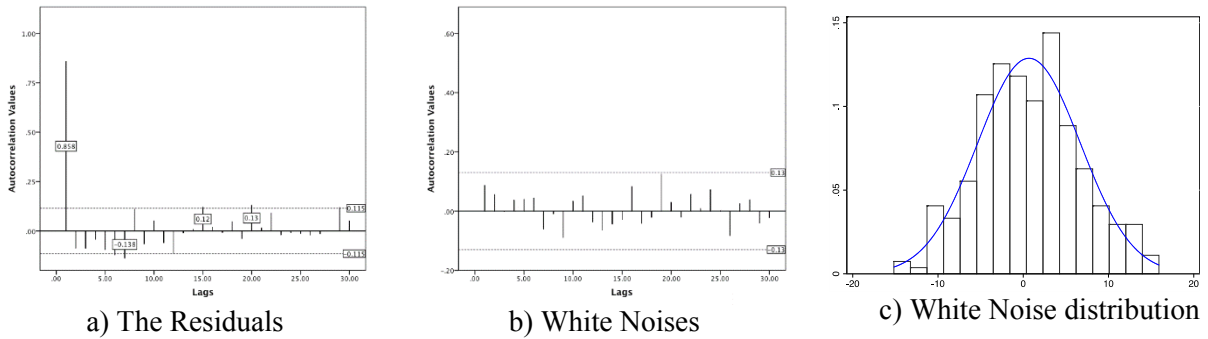


Figure 2: The Autocorrelation Values (Lags 1 to 30) for the residuals & underlying ‘white noise.’

4.1 Availability of Prior Information

The residuals were computed using equation (3) and step 3a (subsection 2.2), and tested for autocorrelation according to the method described in subsection 2.5. The autocorrelation function values were calculated and checked for the first 30 lags (see Figure 2a). Although there are significant autocorrelations at lags $k = 1, 6, 7, 15$ and 20 , the first was the most significant lag ($r_1 = 0.858$) compared to the others. Hence a first-order autoregressive process AR(1) provided by equation (8) was used to model the auto-dependency of the residuals.

Autocorrelation tests of the underlying white noises revealed that they were independent (Figure 2b), and the normality test showed that they were normally distributed (Figure 2c). The residuals were re-sampled with parametrical re-sampling of the underlying white noises using normal distribution, and these samples were used to estimate percentile confidence intervals (PCI) and bias-corrected percentile confidence intervals (BCPCI) for each parameter. Table 1 shows the estimated 95% bootstrap CIs based on the RBS and DRBS with sizes of 25, 50, 100, 200 and 500, while Figure 3 illustrates the empirical bootstrap sampling distributions for each of three parameters. All estimated CIs for the parameters suggest that none of the three parameters were significantly different from their initial guesses at the 95% confidence level. The estimated CIs using DRBS with sizes of 100, 200 and 500 are very close to those estimated using the original RBS for the three parameters. On the other hand, the CIs estimated using DRBS with the size of 25 are close to those estimated based on bias-corrected RBS for the three parameters. Overall, the bias correction for the RBS results produces tighter CIs.

4.2 Unavailability of Prior Information

Assume that the pre-mentioned parameters are unknown. In this experiment (i.e., the second) the residuals are calculated using the first estimate given in step 3b (subsection 2.2), and are modeled using AR(1) as given by equation (8). The test of the underplaying white noises showed that they are independent and normally distributed, so normal distribution was employed to parametrically re-sample

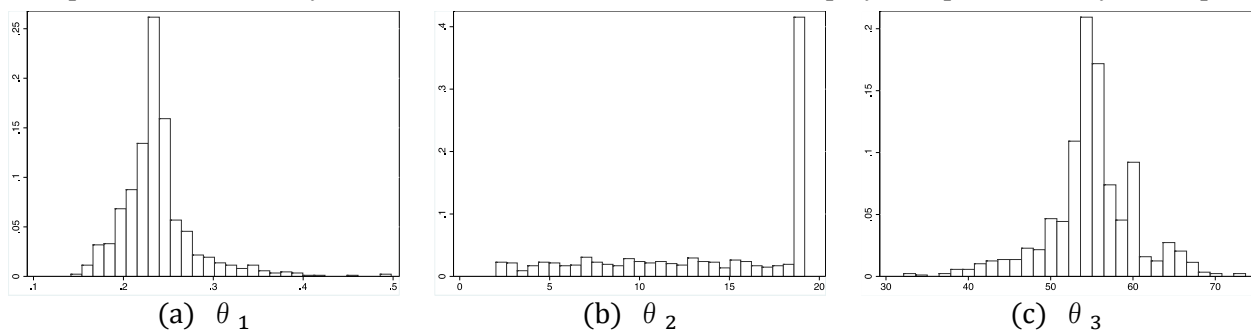


Figure 3: Empirical distributions for the parameters with prior information of the parameters.

Table 1: The estimated CIs for the parameters using RBS and DRBS with prior information of the parameters.

95% CI	CI Method	Lower Limit	Parameter's Guess ($\tilde{\theta}$)	Upper Limit	Width of Interval
θ_1			0.20		
RBS (1000)	PCI	0.17		0.35	0.18
Bias Corrected RBS (1000)	BCPCI	0.14		0.22	0.08
DRBS (25,2000)	PCI	0.17		0.27	0.10
DRBS (50,2000)	PCI	0.17		0.27	0.10
DRBS (100,2000)	PCI	0.16		0.33	0.16
DRBS (200,2000)	PCI	0.17		0.35	0.18
DRBS (500,2000)	PCI	0.17		0.35	0.18
θ_2			19.00		
RBS (1000)	PCI	2.61		19.15	16.54
Bias Corrected RBS (1000)	BCPCI	5.08		19.18	14.09
DRBS (25,2000)	PCI	6.89		19.18	12.29
DRBS (50,2000)	PCI	2.61		19.14	16.54
DRBS (100,2000)	PCI	2.61		19.15	16.54
DRBS (200,2000)	PCI	2.60		19.15	16.54
DRBS (500,2000)	PCI	2.78		19.15	16.37
θ_3			60.00		
RBS (1000)	PCI	41.81		66.14	24.33
Bias Corrected RBS (1000)	BCPCI	55.27		73.67	18.40
DRBS (25,2000)	PCI	50.54		67.00	16.47
DRBS (50,2000)	PCI	49.84		67.00	17.16
DRBS (100,2000)	PCI	43.92		67.87	23.94
DRBS (200,2000)	PCI	38.54		67.29	28.76
DRBS (500,2000)	PCI	41.81		66.47	24.66

them. Table 2 presents the estimates of 95% bootstrap CIs with various bootstrap samples, and Figure 4 gives the empirical bootstrap sampling distribution for the three parameters. Except for the biased corrected CIs, all estimated CIs suggest that the three parameters are not significantly different from their initial guesses at 95% confidence level. Although the bias correction method in this case created CIs that were wider than the RBS, they also suggest that the parameters θ_1 and θ_3 were significantly different from their initial guesses. The results for the DRBS are very close to the regular RBS for all bootstrap sample sizes.

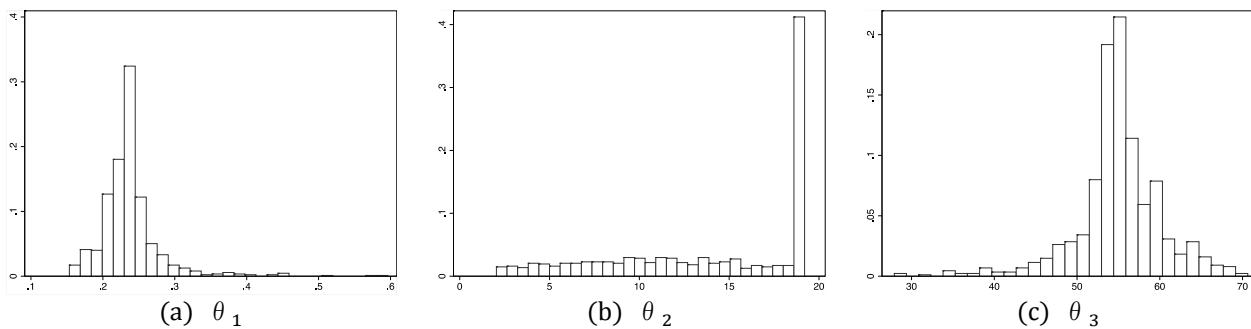


Figure 4: Empirical distributions for the parameters without prior information of the parameters.

Table 2: The estimated CIs for the parameters using RBS and DRBS without prior information of the parameters.

95% CI	CI Method	Lower Limit	$\tilde{\theta}$	$\hat{\theta}$	Upper Limit	Width of Interval
θ_1			0.20	0.27		
RBS (1000)	PCI	0.18			0.36	0.18
Bias Corrected RBS (1000)	BCPCI	0.24			0.60	0.35
DRBS (25,2000)	PCI	0.17			0.33	0.16
DRBS (50,2000)	PCI	0.19			0.39	0.20
DRBS (100,2000)	PCI	0.18			0.39	0.21
DRBS (200,2000)	PCI	0.18			0.37	0.20
DRBS (500,2000)	PCI	0.17			0.34	0.17
θ_2			19.0	19.11		
RBS (1000)	PCI	2.61			19.15	16.23
Bias Corrected RBS (1000)	BCPCI	5.08			19.20	5.50
DRBS (25,2000)	PCI	6.89			19.14	15.55
DRBS (50,2000)	PCI	2.61			19.13	16.29
DRBS (100,2000)	PCI	2.61			19.14	16.31
DRBS (200,2000)	PCI	2.60			19.14	15.92
DRBS (500,2000)	PCI	2.78			19.15	16.23
θ_3			60.0	49.99		
RBS (1000)	PCI	41.36			65.51	24.15
Bias Corrected RBS (1000)	BCPCI	27.92			53.86	25.93
DRBS (25,2000)	PCI	43.55			67.06	23.50
DRBS (50,2000)	PCI	39.01			63.46	24.45
DRBS (100,2000)	PCI	39.01			64.65	25.64
DRBS (200,2000)	PCI	39.01			64.91	25.91
DRBS (500,2000)	PCI	40.14			65.50	25.36

5 DISCUSSION AND CONCLUSION

This paper discusses the theoretical framework of the residual bootstrapping method and its application in system dynamics models. Residual Bootstrapping (RBS) is used to estimate statistical distributions for the model parameters. These distributions can be used to build confidence intervals for the parameters or to perform hypotheses testing. They can also be used for model sensitivity and policy analyses, and to help estimating correlations between model parameters.

This paper presents RBS in an application of a system dynamics model for the Irish elderly care capacity planning, using historical data of delayed discharges from Irish public hospitals in 2012. The literature recommends a large number of iterations (min 1000) for the RBS to replicate good estimation for distribution of the model parameter(s), and this large number imposes a long execution time, which grows rapidly with the increase in model complexity. The idea presented is to integrate RBS with the direct bootstrap method (DRBS) in order to reduce overall execution times. In DRBS, a relatively small number of iterations are used for the RBS, and the direct bootstrap method is then applied to generate a larger sample of the parameter estimates. The proposed framework is generic and can be applied to any simulation model.

The results in Tables 1 and 2 suggest that using regular RBS with few iterations (25, 50 or 100) and applying the direct bootstrapping on estimated boots, can give PCI estimates that are close to those by using RBS with much larger numbers of iterations (e.g. 1000 or more). This suggests that the proposed integration between RBS and direct bootstrapping has an advantage over the perspective computational time. The estimated samples for the parameters could be used for other purposes than CI estimation and hypothesis testing, such as sensitivity analysis. Further statistical analysis is required to test whether the samples estimated using DRBS are different from those estimated using regular RBS or not, such as pair wise hypotheses testing between the two. The null hypothesis (H0) claims that the two samples have equal variances, while the alternative hypothesis (H1) claims the two samples' variances differ from

each other. The results showed that H_0 was rejected for all samples estimated using DRBS with subsets sizes of 25 and 50 for the three parameters: but that the test results did not reject H_0 for the three parameters for sizes 100, 200 and 500, with exceptions of θ_2 at size 100; θ_1 and θ_3 at size 200; and θ_1 at size 500 ($\alpha = 0.05$).

The empirical results of the two experiments (using prior information versus no information for the parameters) demonstrate that the medians of the estimated samples in both experiments did not differ significantly from the initial conjecture of the parameters. However, the bias correction in the second experiment (no information available) appears to contradict this finding. The historical data used in the SD model were for the year 2012, and the parameters values used were estimated based on healthcare data statistics for 2011. Assuming that there were no dramatic changes in demographic characteristics between 2011 and 2012, these parameters values were also assumed to be close to each other over the two years. Empirically, bias correction might result in estimated distribution that could be significantly different from the original parameters' values, and might negatively affect the estimated inferences for the parameters. Nevertheless, empirical experimentation showed that the method proposed in here help to reduce RBS computational times by 90%, and yield promising results with minimum inaccuracy.

REFERENCES

- Alexopoulos, C., D. Goldsman, and P. Tang. 2013. "A Sequential Procedure For Estimating The Steady-State Mean Using Standardized Time Series." In *2013 Winter Simulation Conference (WSC)*, edited by Eds R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 613–622. Piscataway, NJ: IEEE.
- Barlas, Y. 1989. "Multiple Tests for Validation of System Dynamics Type of Simulation Models." *European Journal of Operational Research* 42 (1) (September): 59–87.
- Barlas, Y. 1994. "Model Validation in System Dynamics." In *Proceedings of the 1994 International System Dynamics Conference. Sterling, Scotland*, 1–10.
- Barlas, Y., H. Topaloglu, and S. Yilankaya. 1997. "A Behavior Validity Testing Software (BTS)." In *Proceedings of the 15th International System Dynamics Conference: Systems Approach to Learning and Education into the 21st Century, Istanbul*, 19–22.
- Calvin, J. M., and M. K. Nakayama. 2013. "Confidence Intervals For Quantiles With Standardized Time Series." In *2013 Winter Simulation Conference (WSC)*, edited by Eds R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 601–612. Piscataway, NJ: IEEE.
- Chernick, M. R. 2011. *Bootstrap Methods: A Guide for Practitioners and Researchers*. Vol. 619. John Wiley & Sons.
- Davison, A. C. 1997. *Bootstrap Methods and Their Application*. Vol. 1. Cambridge university press.
- DiCiccio, T. J., and B. Efron. 1996. "Bootstrap Confidence Intervals." *Statistical Science* 11 (3): 189–228.
- Dogan, G. 2004. "Confidence Interval Estimation in System Dynamics Models: Bootstrapping vs. Likelihood Ration Method." In *In 22nd International Conference of the System Dynamics Society. Oxford, UK*.
- Dogan, G. 2007. "Bootstrapping for Confidence Interval Estimation and Hypothesis Testing for Parameters of System Dynamics Models." *System Dynamics Review* 23 (4): 415–436.
- Efron, B., and R. J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Vol. 57. CRC press.
- "Hongyi Li," G S, and Maddala. 1996. "Bootstrapping Time Series Models." *Econometric Reviews* 15 (2): 115–158.
- Jun, Y., and S. H. Ng. 2013. "An Entropy Based Sequential Calibration Approach For Stochastic Computer Models." In *2013 Winter Simulations Conference (WSC)*, edited by Eds R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 589–600. Piscataway, NJ: IEEE.
- Kleijnen, J. P. C., and R. G. Sargent. 2000. "A Methodology for Fitting and Validating Metamodels in Simulation." *European Journal of Operational Research* 120 (1): 14–29.

- Lyneis, J. M., and A. L. Pugh. 1996. "Automated Vs. 'hand' Calibration of System Dynamics Models: An Experiment with a Simple Project Model." In *In Proceedings of the 1996 International System Dynamics Conference*. System Dynamics Society, Cambridge, MA, 317–320.
- Meadows, D. H. 1980. "The Unavoidable A Priori. Elements of the Systems Dynamics Method." *J. Randers*. Cambridge, Massachusetts, Productivity Press: 22–57.
- Miller, J. H. 1998. "Active Nonlinear Tests (ANTs) of Complex Simulation Models." *Management Science* 44 (6): 820–830.
- Moxnes, E. 2005. "Policy Sensitivity Analysis: Simple versus Complex Fishery Models." *System Dynamics Review* 21 (2): 123–145.
- Ng, T. S., C. L. Sy, and L. H. Lee. 2012. "Robust Parameter Design for System Dynamics Models: A Formal Approach Based on Goal-Seeking Behavior." *System Dynamics Review* 28 (3): 230–254.
- Oliva, R. 2003. "Model Calibration as a Testing Strategy for System Dynamics Models." *European Journal of Operational Research* 151 (3): 552–568.
- Polansky, A. M. 2000. "Stabilizing Bootstrap-T Confidence Intervals for Small Samples." *Canadian Journal of Statistics* 28 (3): 501–516.
- Rahmandad, H., S. J. Mohammed, and G. E. Hamed. 2013. "Estimation Of Unknown Parameters In System Dynamics Models Using The Method Of Simulated Moments." In *In 2013 Winter Simulations Conference (WSC)*, edited by Eds R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 1212–1222. Piscataway, NJ: IEEE.
- Rashwan, W., M. Ragab, W. Abo-Hamad, and A. Arisha. 2013. "Evaluating Policy Interventions for Delayed Discharge: A System Dynamics Approach." In *2013 Winter Simulations Conference (WSC)*, edited by Eds R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 2463–2474. Piscataway, NJ: IEEE.
- Reichelt, K. S., J. M. Lyneis, and C. G. Bespolka. 1996. "Calibration Statistics: Selecting a Statistic and Setting a Standard." In *In Proceedings of the 1996 International System Dynamics Conference*. System Dynamics Society, 425–428.
- Senge, P. M. 1977. "Statistical Estimation of Feedback Models." *Simulation* 28 (6): 177–184.
- Sterman, J. D. 2000. "Business Dynamics: Systems Thinking and Modeling for a Complex World." *Irwin/McGraw-Hill*. Boston: 928.

AUTHOR BIOGRAPHIES

MOHAMMED MESABBAH, M.Sc. is a researcher in the 3S Group - a research unit in the College of Business at the Dublin Institute of Technology (DIT) - specialized in complex systems simulation and optimization. He has a B.Sc. in Operations Research and Decision Support from Cairo University and an M.Sc. in System Dynamics from University of Bergen - Norway. His research interests include Decision Support, Modeling and Simulation, Optimization, Computational intelligence and Game Theory. His email address is <mohammedBayoumi.mesabbah@mydit.ie>.

WAEEL RASHWAN, M.Sc. is a researcher in the 3S Group since 2013. He has a B.Sc. and M.Sc. degree in Operations Research from Cairo University, as well as extensive experience in industry. His research interests include: Modeling and Simulation, Optimization, Game Theory and Probability Collectives. His email address is <wael.rashwan@dit.ie>.

AMR ARISHA, Ph.D. is the director of the 3S Group and the head of international business department in the College of Business at the Dublin Institute of Technology, which he joined in 2005. He received his PhD in Industrial Engineering from Dublin City University (DCU). Intel-Ireland sponsored his research from 2000 - 2005. His research interests include analysis and optimization of complex dynamic systems in manufacturing and service industries. He is a member in IIE, IMECH, IEI, ESE, ORS, IEEE and ASME and Chief Examiner at MII. His e-mail address is <amr.arisha@dit.ie>.