PERGAMON Mathematical and Computer Modelling 34 (2001) 1159–1170

MATHEMATICAL AND COMPUTER MODELLING

www.elsevier.com/locate/mcm

Bootstrap Confidence Intervals for the Simultaneous Equations Model under Heavy-Tailed Contamination

R. GATTO
Institut für Mathematische Statistik und Versicherungslehre
University of Bern, Sidlerstrasse 5, CH-3012, Bern, Switzerland

Abstract—A simulation study illustrates the good accuracy of the bootstrap bias-corrected accelerated confidence interval (BC_a) by Efron [1] in the simultaneous equations model when the errors are generated from a normal distribution contaminated by a heavy-tailed distribution. When the sample size is small, the BC_a confidence interval is shown to have a substantially smaller coverage error than the normal asymptotic interval, which is, however, often used in practice.

Keywords—α-stable distribution, Asymptotic approximation, Bias-corrected accelerated confidence interval, Contaminated distribution, Influence function, Likelihood function, M-estimator, Monte Carlo resampling, Small sample size.

1. INTRODUCTION

This article presents a simulation study regarding the accuracy, in terms of coverage rate, of bootstrap confidence intervals in a linear simultaneous equation model. The errors of the model are generated from a contamination of a normal distribution with a Cauchy distribution. The Cauchy distribution belongs to the class of α-stable heavy-tailed distributions. The parameters of the simultaneous equation model are estimated by the "full information maximum likelihood" (FIML) estimator. Comparisons with asymptotic normal intervals are also given. A particularity of this simulation study is that our model is defined in a high-dimensional parameter space, whereas most bootstrap studies concern models with a smaller number of parameters.

The idea of using α-stable distributions for describing the behavior of economic variables was introduced in the economic literature by Mandelbrot [2]. An α-stable random variable can be denoted as $S_\alpha(\sigma, \beta, \mu)$, where the characterizing parameters $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma \in \mathbb{R}_+$, and $\mu \in \mathbb{R}$ are the indexes of stability, skewness, scale, and shift, respectively. When $\beta = 0$, $S_\alpha(\sigma, \beta, \mu)$ is symmetric about μ. Two well-known examples are the normal $S_2(\sigma, 0, \mu)$ and the Cauchy $S_1(\sigma, 0, \mu)$ random variables. With these two exceptions, no simple expressions are known for the density or the distribution of α-stable random variables. An α-stable random variable can be characterized by the fact that it has a domain of attraction; i.e., there exist a sequence $\{Y_n\}$ of i.i.d. random variables and sequences $\{a_n\}$ and $\{b_n\}$ of real positive numbers such that

$$\frac{Y_1 + \cdots + Y_n}{b_n} + a_n \xrightarrow{\mathcal{D}} S_\alpha(\sigma, \beta, \mu).$$

0895-7177/01/\$ - see front matter Typeset by $\mathcal{A}_\mathcal{M}\mathcal{S}$-TEX
PII: S0895-7177(01)00124-8

Note that this would be the standard Central Limit Theorem if $E|Y_1|^2 < \infty$ was assumed. (For $\alpha \in (0,2)$, $E|S_\alpha(\sigma,\beta,\mu)|^p < \infty$ only when $p \in (0,\alpha)$.) More generally, a random vector is α-stable if every linear combination of its components is univariate α-stable, and the multivariate normal and Cauchy admit closed form expressions for their density functions. For a complete presentation, see, e.g., [3] or [4]. α-stable distributions with $\alpha \in (0,2)$ are special cases of heavy-tailed distributions, i.e., distributions with infinite variance. In fact, α is indicative of the weight of the tail(s) of the density since the tail(s) vanish as $|x|^{-(1+\alpha)}$, as $|x| \to \infty$, x representing an abscissa point of the density. Heavy-tailed distributions play a key role in economic modeling, where variables often behave as though their variance were infinite. If one chooses to use a heavy-tailed distribution, the restriction to α-stable distributions is motivated by the domain of attraction. As an example, the price of a stock, supposed as the sum of several i.i.d. small fluctuations, is an α-stable random variable; see, e.g., [5]. In this context, a realistic deviation from the normal model would be in the direction of an α-stable, so that a relevant model could be obtained through the contamination of a normal distribution with a heavy-tailed α-stable distribution, namely,

$$\varepsilon S_\alpha(\sigma,\beta,\mu) + (1-\varepsilon) S_2(\sigma,0,\mu),$$

for $\alpha \in (0,2)$ and $\varepsilon \in (0,1)$. In Section 3, we study the accuracy of some nonparametric bootstrap confidence intervals, as well as the accuracy of the asymptotic normal confidence interval, when the errors of the model are generated from a normal distribution with a Cauchy heavy-tailed contamination. (See Appendix B for a method of simulation.)

2. THE BIAS-CORRECTED ACCELERATED CONFIDENCE INTERVAL

Since the pioneering article by Efron [6], bootstrap techniques have become important tools of statistical inference. Textbooks on the bootstrap are now available; see, e.g., [7,8], etc. For a survey regarding applications of the bootstrap in econometrics, see [9]. This simulation study concerns the application of bootstrap confidence intervals in the simultaneous equations model, widely used in economic modeling, where some economic variables can appear in more than one structural equation. A typical example would be the demand and the supply equations in a market equilibrium model. We compute by simulation the coverage rates of the asymptotic normal confidence interval (N) and of three types of bootstrap confidence intervals: the "percentile" (P), the "bias-corrected" (BC), and the "bias-corrected accelerated" (BC_a) intervals. We consider small and moderate sample sizes ($n = 20$ and 40). The BC_a bootstrap confidence interval, due to Efron [1], is known to lead to second-order accuracy in both coverage (the error with respect to the nominal coverage rate is of the order n^{-1}) and correctness (the error with respect to the "exact" confidence bounds, in the sense of Hall [10], is of the order $n^{-3/2}$), in contrast to the P, BC, and N intervals, which have first-order accuracy only (errors $n^{-1/2}$ and n^{-1} for coverage and correctness, respectively). For technical explanations about these asymptotic orders of accuracy, refer to, e.g., [1,10]. For a recent overview of bootstrap confidence intervals, see also [11]. However, the theoretical comparisons based on asymptotic orders of accuracy represent only one part of the information needed for a complete evaluation of bootstrap confidence intervals, because a simulation study for a specific model interest is also necessary in order to check the numerical performance. This simulation study confirms the importance of using the BC_a interval instead of more popular methods such as the N or even the P intervals, in the presence of a very small sample size and for the type of error distributions considered, namely the normal and the normal with a heavy-tailed contamination.

The implementation of the BC_a confidence interval can be summarized as follows. Suppose we have n (multivariate) independent and identically distributed observations $w_1, \dots, w_n$, and denote by $\hat{F}$ the empirical distribution of these observations (i.e., the distribution function which puts mass n^{-1} over each observation). In what follows, $\hat{t} = \hat{t}(w_1, \dots, w_n) \in \mathbb{R}^p$ is an estimator of an unknown parameter $t_0 \in \mathbb{R}^p$ of the underlying model. We are interested in obtaining confidence intervals for the j^{th} element of t_0, $1 \leq j \leq p$. The main steps for the construction of the (two-sided) BC_a confidence interval are as follows.

STEP 1. Compute the bootstrap distribution function

$$\hat{G}(s) = P\left(\hat{\theta}^* \leq s\right) \tag{1}$$

by randomly drawing $w_1^*, \ldots, w_n^*$ from $\hat{F}$ (i.e., from the original sample with replacement) and by computing the bootstrap replicates $\hat{\theta}^* = \hat{t}_j^*$, where $\hat{t}^* = \hat{t}(w_1^*, \ldots, w_n^*)$, a large number of times (usually 1000 or more).

STEP 2. Compute the "bias-correction constant" z_0 and the "acceleration constant" $\hat{a}$ by

$$z_0 = \Phi^{(-1)}\left\{\hat{G}\left(\hat{\theta}\right)\right\}, \tag{2}$$

and by

$$\hat{a} = \frac{1}{6}\frac{\sum_{i=1}^{n} \widehat{IF}_j^3(w_i)}{\left[\sum_{i=1}^{n} \widehat{IF}_j^2(w_i)\right]^{3/2}}, \tag{3}$$

where $\widehat{IF}(w_i)$ is the empirical influence function of $\hat{t}$ at sample point w_i, see (12), and where $\Phi^{(-1)}(\cdot)$ is the inverse standard normal distribution function.

STEP 3. Compute the $(1-2\alpha)$-level BC_a confidence interval by

$$\left(\hat{G}^{(-1)}\left\{\Phi\left(z_0 + \frac{z_0 + z^{(\alpha)}}{1 - \hat{a}\left(z_0 + z^{(\alpha)}\right)}\right)\right\}, \hat{G}^{(-1)}\left\{\Phi\left(z_0 + \frac{z_0 - z^{(\alpha)}}{1 - \hat{a}\left(z_0 - z^{(\alpha)}\right)}\right)\right\}\right), \tag{4}$$

where $z^{(\alpha)}$ is the α-quantile of the standard normal distribution $\Phi(\cdot)$, and where $\hat{G}^{(-1)}(u) = \inf\{s \mid \hat{G}(s) \geq u\}$ is the quantile function associated to $\hat{G}$.

The BC or the P confidence intervals can be obtained by setting $\hat{a} = 0$ or $\hat{a} = z_0 = 0$, respectively, in the previous steps.

The derivations of the P, BC, and BC_a confidence intervals can be outlined as follows. To simplify, we suppose for the moment that $\hat{\theta}$ is an estimator of an unknown parameter $\theta \in \mathbb{R}$ based on n observations having common distribution depending on θ only. The $(1-2\alpha)$-level P interval is given by

$$\left(\hat{G}^{(-1)}(\alpha), \hat{G}^{(-1)}(1-\alpha)\right), \tag{5}$$

where $\hat{G}(s) = P(\hat{\theta}^* \leq s)$ is the bootstrap distribution of $\hat{\theta}$. The confidence interval (5) is transformation respecting in the sense that if $\phi = g(\theta)$ and $\hat{\phi} = g(\hat{\theta})$, for a given monotone transform $g(\cdot)$, then the P interval for θ is the g-transform of the bounds of the P interval for ϕ. Suppose now that we are in the case where

$$\frac{\hat{\phi} - \phi}{\tau} \sim \mathcal{N}(0, 1), \tag{6}$$

where τ is the known standard deviation of $\hat{\phi}$, supposed fixed. Under (6), the standard normal interval

$$\left(\hat{\phi} - z^{(1-\alpha)}\tau, \hat{\phi} - z^{(\alpha)}\tau\right) \tag{7}$$

holds exactly, and the bootstrap distribution of $\hat{\phi}$ would be given by $\Phi((\cdot - \hat{\phi})/\tau)$. Hence, the P interval would correspond to the exact one given by (7). Moreover, because a P interval is transformation respecting, the $(1-2\alpha)$-level P confidence interval for θ, given by (5), can be re-expressed as

$$\left(g^{(-1)}\left(\hat{\phi}-\tau z^{(1-\alpha)}\right), g^{(-1)}\left(\hat{\phi}-\tau z^{(\alpha)}\right)\right).$$

This means that the P confidence interval would hold exactly if there is a monotone normalizing transform $g(\cdot)$ leading to (6), which needs not be identified. We could hence suppose that the P confidence interval automatically incorporates this normalizing transform, and this explains the advantage with respect to the standard normal confidence interval whose accuracy depends on the scale chosen for the parameter. Assumption (6) allows for another important generalization. In fact, we could assume that there exists a monotone transformation g so that $\hat{\phi}=g(\hat{\theta})$ and $\phi=g(\theta)$ which satisfies

$$\frac{\hat{\phi}-\phi}{\tau_{\phi}} \sim \mathcal{N}(-z_0, 1), \tag{8}$$

where $\tau_{\phi} = 1 + a\phi$, and z_0 is the bias of $\hat{\phi}$. Under assumption (8), the exact $(1-2\alpha)$-level confidence interval is given by $(\phi[\alpha], \phi[1-\alpha])$, where

$$\phi[\alpha] = \hat{\phi} + \tau_{\hat{\phi}} \frac{z_0 + z^{\alpha}}{1 - a(z_0 + z_{\alpha})}.$$

The distribution of $\hat{\phi}$ under (8) is given by $H(s) = \Phi((s-\phi)/\tau_{\phi} + z_0)$, and its bootstrap version would be $\hat{H}(s) = \Phi((s-\hat{\phi})/\tau_{\hat{\phi}} + z_0)$. By defining

$$z[\alpha] = z_0 + \frac{z_0 + z^{\alpha}}{1 - a(z_0 + z_{\alpha})},$$

some standard manipulations would show that

$$\hat{H}^{(-1)}(\Phi(z[\alpha])) = \phi[\alpha],$$

so that the exact confidence interval in the ϕ-scale can be viewed as a P interval based on the corrected tail probabilities $\Phi(z[\alpha])$. The relationship $\hat{H}(g(s)) = \hat{G}(s)$ implies $\hat{H}^{(-1)}(\Phi(z[\alpha])) = g\{\hat{G}^{(-1)}(\Phi(z[\alpha]))\}$, so that $g\{\hat{G}^{(-1)}(\Phi(z[\alpha]))\} = \phi[\alpha]$. This means that the confidence bound $\hat{G}^{(-1)}(\Phi(z[\alpha]))$ transforms to the exact one in the ϕ-scale. From there, we define the BC_a $(1-2\alpha)$-level confidence interval as $(\hat{G}^{(-1)}(z[\alpha]), \hat{G}^{(-1)}(z[1-\alpha]))$, exactly as given by (4). When the acceleration constant a is equal to zero, then the BC_a confidence interval is the BC confidence interval, which is exact under assumption (8) with $a=0$. Also, $a = z_0 = 0$ implies $z[\alpha] = z^{(\alpha)}$ so that the P, the BC, and the BC_a confidence intervals are the same one. Clearly, (8) implies $P(\hat{\phi} < \phi) = \Phi(z_0)$, which is also equal to $P(\hat{\theta} < \theta)$, leading to $z_0 = \Phi^{(-1)}(P[\hat{\theta} < \theta])$ or to the equivalent expression (2). For a detailed interpretation of the acceleration constant a and its approximations, we refer to [1]. The approximation $\hat{a}$ given by (3) is valid only when $\hat{G}(\cdot)$ is obtained by Monte Carlo resampling from the original sample (as considered in our simulations), and holds in the multiparameter setting also. In fact, this description generalizes to multiparameter families, as given by Steps 1 and 2 above; see [1] for details. For a version of the BC_a confidence interval which avoids the Monte Carlo resampling, by means of a saddlepoint approximation, see [12].

The remaining part of this article is divided as follows. Section 3 gives additional explanations concerning the implementation of bootstrap confidence intervals with the FIML estimator. Section 3 shows the results of the numerical simulations. The second-order derivatives of the Lagrangian associated with the likelihood are given in Appendix A, and Appendix B recalls a method for generating multivariate Cauchy vectors.

3. THE FIML ESTIMATOR AS AN M-ESTIMATOR

Consider the model with g linear simultaneous equations

$$\mathrm{B}y_i + \Gamma z_i = u_i, \qquad i = 1, \dots, n, \tag{9}$$

where the dimensions of the matrices appearing in (9) are expressed inside the following brackets:

$$\mathrm{B}\,(g \times g), \qquad \Gamma\,(g \times k), \qquad y_i\,(g \times 1), \qquad z_i\,(k \times 1), \qquad \text{and} \qquad u_i\,(g \times 1).$$

We distinguish between y_i, the variable explained by the model, also called "endogenous" variable, and z_i, the independent variable, also called "exogenous" variable. It is convenient to adopt the compact notation $\Theta^\top w_i = u_i$, $i = 1, \dots, n$, equivalent to (9), where $\Theta^\top = (\mathrm{B}, \Gamma)$ and $w_i = (y_i^\top, z_i^\top)^\top$. By grouping all observations, $W = (w_1, \dots, w_n)$ and $U = (u_1, \dots, u_n)$, the entire system can be simply re-expressed as $\Theta^\top W = U$. We suppose that the conditional distribution of y_i given z_i is the multivariate normal with mean $-\mathrm{B}^{-1}\Gamma z_i$ and covariance matrix $\mathrm{B}^{-1}\Sigma(B^{-1})^\top = \Omega$, where $\Sigma\,(g \times g)$ is the covariance matrix of u_i, and we obtain the likelihood from this conditional distribution. As usual, we also suppose that the model has c linear constraints on the regression parameters, expressed as

$$R\,\mathrm{vec}\,(\Theta) = r, \tag{10}$$

where $R\ (c \times g(g+k))$ is a matrix uniquely defined by the constraints, $r(c \times 1)$ is an appropriate vector of constants, and $\mathrm{vec}(\cdot)$ is the "vectorization" operator which stacks the columns of a matrix by putting the $(i+1)^{\mathrm{th}}$ column under the i^{th} column. In addition to this, we consider d linear constraints on the covariance matrix

$$C\,\mathrm{vec}\,(\Sigma) = 0, \tag{11}$$

where $C(d \times g^2)$ is a matrix fully determined by the constraints (generally symmetry restrictions).

In order to estimate Θ and Σ, we construct the Lagrangian ($\mathcal{L}$) associated with the normal logarithmic likelihood, with the constraints (10), and with the restrictions (11). After some standard multivariate manipulations, we can express it as

$$\mathcal{L} = n \log\det\left(L^\top\Theta\right) - \frac{n}{2}\,\log\det(\Sigma) - \frac{1}{2}\,\mathrm{trace}\left(\Theta^\top W W^\top \Theta \Sigma^{-1}\right) \\ - \lambda^\top (R\,\mathrm{vec}\,\Theta - r) - \mu C\,\mathrm{vec}\,(\Sigma),$$

where λ, μ are the Lagrange's multipliers associated with two sets of constraints (10) and (11), respectively, and where L is such that $\mathrm{B}^\top = L^\top\Theta$. The first-order derivatives of the Lagrangian can be written as

$$\frac{\partial\mathcal{L}}{\partial\,\mathrm{vec}\,\Theta} = \sum_{i=1}^{n} \left\{ \mathrm{vec}\left[L\left(\Theta^\top L\right)^{-1}\right] - \frac{1}{2}\,\left[I \otimes \left(w_i w_i^\top \Theta\right)\right]\,\mathrm{vec}\,\Sigma^{-1} \right. \\ \left. - \frac{1}{2}\,\left[I \otimes \left(w_i w_i^\top \Theta\right)\right] P_{g,g}\,\mathrm{vec}\,\Sigma^{-1} - \frac{1}{n}\,R^\top\lambda \right\} = \sum_{i=1}^{n} \psi_1(\Theta, \Sigma;\, w_i),$$

$$\frac{\partial\mathcal{L}}{\partial\,\mathrm{vec}\,\Sigma} = \sum_{i=1}^{n} \left\{ -\frac{1}{2}\,P_{g,g}\,\mathrm{vec}\,\Sigma^{-1} + \frac{1}{2}\,P_{g,g}\,\mathrm{vec}\,\left(\Sigma^{-1}\Theta^\top w_i w_i^\top \Theta \Sigma^{-1}\right) \right\} = \sum_{i=1}^{n} \psi_2(\Theta,\, \Sigma;\, w_i),$$

$$\frac{\partial\mathcal{L}}{\partial\lambda} = \sum_{i=1}^{n} -\frac{1}{n}\,(R\,\mathrm{vec}\,\Theta - r) = \sum_{i=1}^{n} \psi_3(\Theta),$$

$$\frac{\partial\mathcal{L}}{\partial\mu} = \sum_{i=1}^{n} -\frac{1}{n}\,C\,\mathrm{vec}\,\Sigma = \sum_{i=1}^{n} \psi_4(\Sigma).$$

In the above formulas, $P_{m,n}$ is the "permutated unit matrix", which is defined as an $(mn \times mn)$ matrix composed of (mn) blocks of order $(m \times n)$. The block in position (i,j) has an element of value 1 in its j^{th} row and i^{th} column and zero elements elsewhere. It is now convenient to assemble all the parameters of the model in the single vector

$$t = \left((\operatorname{vec}\Theta)^\top, (\operatorname{vec}\Sigma)^\top, \lambda^\top, \mu\right)^\top.$$

By defining $\psi = (\psi_1^\top, \psi_2^\top, \psi_3^\top, \psi_4^\top)^\top$, the FIML estimator can be expressed as the solution $\hat{t}$ of

$$\sum_{i=1}^{n} \psi\left(w_i, \hat{t}\right) = 0.$$

Thus, the FIML estimator is an M-estimator, and we can easily obtain the empirical influence function required for the computation of the acceleration constant (3), namely

$$\widehat{IF}(w_i) = \left[M^{-1}\left(\psi, \hat{F}\right)\right]^\top \psi\left(w_i, \hat{t}\right), \tag{12}$$

where

$$M\left(\psi, \hat{F}\right) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial t^\top}\psi(w_i, t)\bigg|_{t=\hat{t}};$$

see [13]. The derivatives of ψ can be obtained from Appendix A.

The FIML estimator can be computed by solving the following iterating technique. Suppose that the r^{th} iteration has provided us the updated values B_r, Γ_r, Σ_r, Ω_r. By defining $Y = (y_1, \dots, y_n)$ and $Z = (z_1, \dots, z_n)$, the r^{th} iteration of the algorithm is given by the equations

$$\begin{pmatrix} \Sigma_r^{-1} \otimes \begin{pmatrix} YY^\top - n\Omega_r & YZ^\top \\ ZY^\top & ZZ^\top \end{pmatrix} & R^\top \\ R & 0 \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\Theta_{r+1}) \\ \lambda_{r+1} \end{pmatrix} = \begin{pmatrix} 0 \\ r \end{pmatrix},$$

$$\Sigma_{r+1} = \frac{1}{n}\Theta_{r+1}^\top W \left(\Theta_{r+1}^\top W\right)^\top, \qquad \text{and} \qquad \Omega_{r+1} = \mathrm{B}_{r+1}^{-1}\Sigma_{r+1}\left(\mathrm{B}_{r+1}^{-1}\right)^\top.$$

The initial conditions for the iterations can be given by $\Sigma_0 = I$ (the identity matrix) and by $\Omega_0 = n^{-1}Y(I - Z^\top(ZZ^\top)^{-1}Z)Y^\top$. This is mainly a Newton-Raphson algorithm.

4. NUMERICAL RESULTS

In this section, we compute by Monte Carlo simulation the coverage rates of the asymptotic normal confidence interval, the percentile, the bias-corrected, and the bias-corrected accelerated bootstrap confidence intervals. The asymptotic normal interval is obtained from the asymptotic covariance matrix of the M-estimator $\hat{t}$,

$$\hat{V}\left(\hat{t}\right) = \frac{1}{n} M^{-1}\left(\psi, \hat{F}\right) Q\left(\psi, \hat{F}\right) \left[M^{-1}\left(\psi, \hat{F}\right)\right]^\top,$$

where $\hat{F}$ is the empirical distribution of the observations, and

$$Q\left(\psi, \hat{F}\right) = \frac{1}{n}\sum_{i=1}^{n} \psi\left(w_i, \hat{t}\right)\psi^\top\left(w_i, \hat{t}\right).$$

We consider a model with $g = 2$ simultaneous equations with two endogenous variables and two exogenous variables plus a constant term, giving $k = 3$. The first equation is overidentified and the second equation is just identified (see [14]). The two simultaneous equations are

$$\beta_{11}y_{1i} + \beta_{12}y_{2i} + \gamma_{11}z_{1i} + \gamma_{12}z_{2i} + \gamma_{13} = u_{1i},$$
$$\beta_{21}y_{1i} + \beta_{22}y_{2i} + \gamma_{21}z_{1i} + \gamma_{22}z_{2i} + \gamma_{23} = u_{2i},$$

with $i = 1, \ldots, n$. The constraints of type (10) are given by

$$\beta_{11} = 1, \quad \gamma_{11} = 0, \quad \gamma_{12} = 0, \quad \beta_{21} = -1, \quad \text{and} \quad \gamma_{22} = -1,$$

and concerning (11) we impose symmetry to Σ. The true values of the underlying parameters of this study are given by

$$\beta_{12} = -0.5, \quad \gamma_{13} = -0.25, \quad \beta_{22} = 0.9, \quad \gamma_{21} = -0.3, \quad \gamma_{23} = -0.15, \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

The z_{1i} are uniform pseudorandom numbers in the range of $[0, 1]$, and z_{2i} and v_i are standard normal pseudorandom variables, $i = 1, \ldots, n$. The errors u_i are obtained by transforming the v_i with the covariance matrix Σ. We use Matlab's random number generator, which is based on a congruential formula. In the following tables, we estimate the coverage level from the proportion of simulations having a true value inside a generated confidence interval, and we indicate by "% left", "% center", and "% right" the proportion of cases where the true value fell, respectively, on the left side of the interval, inside it, and on its right side. We compute the coverage rates of the asymptotic normal interval by means of 10,000 Monte Carlo simulations, and the coverages of the bootstrap intervals by 400 Monte Carlo simulations, each one containing 1,000 resamplings (400,000 simulations in total). Also, LB and UB are the medians of the lower bounds and the upper bounds, respectively, of all the simulated intervals.

Table 1. Confidence intervals for β_{12}, normal errors (upper part), and normal with 5% Cauchy contamination errors (lower part), $n = 20$.

Interval	% Left	% Center	% Right	LB	UB	Level
N	0.24	80.60	19.16	−0.642	−0.407	90% normal errors
P	0.25	93.00	6.75	−0.619	−0.312	
BC	2.00	89.25	8.75	−0.617	−0.309	
BC_a	2.25	88.25	9.50	−0.617	−0.315	
N	0.07	84.71	15.22	−0.662	−0.386	95% normal errors
P	0.00	98.00	2.00	−0.635	−0.231	
BC	0.25	96.75	3.00	−0.633	−0.210	
BC_a	0.75	96.25	3.00	−0.632	−0.215	
N	4.25	85.06	10.69	−0.682	−0.296	90% normal-Cauchy errors
P	5.05	90.15	4.80	−0.684	−0.208	
BC	5.30	88.13	6.57	−0.689	−0.203	
BC_a	5.56	88.13	6.31	−0.685	−0.204	
N	3.14	89.41	7.45	−0.718	−0.252	95% normal-Cauchy errors
P	2.53	94.19	3.28	−0.745	−0.113	
BC	3.03	93.18	3.79	−0.745	−0.114	
BC_a	3.03	92.93	4.04	−0.751	−0.121	

% left, % center, and % right: proportion of simulations having the true value on the left side, inside, or on the right side of the simulated intervals, respectively. LB and UB: medians of the simulated lower and upper bounds, respectively, of the intervals.

Table 2. Confidence intervals for β_{22}, normal with 5% Cauchy contamination errors, $n = 20$.

Interval	% Left	% Center	% Right	LB	UB	Level
N	0.39	83.73	15.88	0.654	1.075	90% normal-Cauchy errors
P	1.01	91.41	7.58	0.631	1.293	
BC	1.77	90.15	8.08	0.632	1.301	
BC_a	1.77	89.90	8.33	0.634	1.310	
N	0.13	88.78	11.09	0.604	1.118	95% normal-Cauchy errors
P	0.25	95.45	4.29	0.567	1.536	
BC	0.51	95.20	4.29	0.571	1.552	
BC_a	0.51	94.95	4.55	0.568	1.527	

% left, % center, and % right: proportion of simulations having the true value on the left side, inside, or on the right side of the simulated intervals, respectively. LB and UB: medians of the simulated lower and upper bounds, respectively, of the intervals.

Table 3. Confidence intervals for γ_{13}, normal errors (upper part), and normal with 5% Cauchy contamination errors (lower part), $n = 20$.

Interval	% Left	% Center	% Right	LB	UB	Level
N	10.77	86.93	2.30	−0.585	0.170	90% normal errors
P	5.25	94.00	0.75	−0.743	0.169	
BC	8.25	89.50	2.25	−0.747	0.160	
BC_a	8.25	89.00	2.75	−0.751	0.150	
N	6.85	92.07	1.08	−0.652	0.242	95% normal errors
P	3.00	96.75	0.25	−0.922	0.256	
BC	4.00	95.75	0.25	−0.943	0.247	
BC_a	4.00	95.50	0.50	−0.953	0.247	
N	5.34	90.84	3.82	−0.826	0.260	90% normal-Cauchy errors
P	2.78	95.20	2.02	−1.132	0.529	
BC	3.28	93.18	3.54	−1.126	0.496	
BC_a	3.79	92.17	4.04	−1.157	0.498	
N	2.59	95.75	1.66	−0.943	0.367	95% normal-Cauchy errors
P	1.77	96.72	1.52	−1.387	0.759	
BC	1.77	96.46	1.77	−1.414	0.699	
BC_a	2.02	96.21	1.77	−1.419	0.682	

% left, % center, and % right: proportion of simulations having the true value on the left side, inside, or on the right side of the simulated intervals, respectively. LB and UB: medians of the simulated lower and upper bounds, respectively, of the intervals.

The important results of the simulations are presented in Tables 1–6. These tables give the coverage for all the confidence intervals discussed here: the asymptotic N, and the bootstrap P, BC, and BC_a. More precisely, % left, % center, and % right indicate the proportions of occurrences

Table 4. Confidence intervals for γ_{23}, normal with 5% Cauchy contamination errors, $n = 20$.

Interval	% Left	% Center	% Right	LB	UB	Level
N	8.55	86.83	4.62	−1.202	1.013	90% normal-Cauchy errors
P	2.78	93.94	3.28	−2.076	1.505	
BC	4.55	92.42	3.03	−1.986	1.391	
BC_a	4.80	91.67	3.54	−1.963	1.416	
N	4.87	92.76	2.37	−1.427	1.237	95% normal-Cauchy errors
P	1.52	95.96	2.53	−2.833	1.967	
BC	1.77	95.96	2.27	−2.747	1.939	
BC_a	2.02	95.71	2.27	−2.748	1.953	

% left, % center, and % right: proportion of simulations having the true value on the left side, inside, or on the right side of the simulated intervals, respectively. LB and UB: medians of the simulated lower and upper bounds, respectively, of the intervals.

Table 5. Confidence intervals for β_{12}, normal errors (upper part), and normal with 5% Cauchy contamination errors (lower part), $n = 40$.

Interval	% Left	% Center	% Right	LB	UB	Level
N	0.19	86.21	13.60	−0.607	−0.414	90% normal errors
P	1.77	92.93	5.30	−0.585	−0.354	
BC	3.54	89.14	7.32	−0.582	−0.354	
BC_a	3.28	88.64	8.08	−0.583	−0.354	
N	0.04	89.89	10.07	−0.625	−0.396	95% normal errors
P	0.25	96.97	2.78	−0.597	−0.302	
BC	0.51	96.46	3.03	−0.597	−0.307	
BC_a	0.25	96.46	3.28	−0.597	−0.306	
N	3.75	90.09	6.16	−0.658	−0.303	90% normal-Cauchy errors
P	5.05	91.41	3.54	−0.678	−0.199	
BC	6.06	89.39	4.55	−0.679	−0.217	
BC_a	6.31	88.89	4.80	−0.672	−0.221	
N	2.57	93.65	3.78	−0.692	−0.262	95% normal-Cauchy errors
P	2.78	95.20	2.02	−0.738	−0.109	
BC	2.78	94.95	2.27	−0.738	−0.134	
BC_a	2.78	94.95	2.27	−0.738	−0.120	

% left, % center, and % right: proportion of simulations having the true value on the left side, inside, or on the right side of the simulated intervals, respectively. LB and UB: medians of the simulated lower and upper bounds, respectively, of the intervals.

for which the true value of the of the regression coefficient of interest is in the left side, inside, or on the right side of the simulated intervals, respectively. Also, LB and UB are medians of the simulated lower and upper bounds, respectively, of the confidence intervals. Tables 1–4 show the results for the sample size $n = 20$, whereas Tables 5 and 6 refer to the sample size $n = 40$. All intervals are constructed in order to have coverage rates 90% and 95%. In Table 1, we see that in both cases where the errors of the models are generated from a normal distribution or where they are generated from the same normal with a 5% of Cauchy contamination, the Cauchy chosen with Ω as dispersion matrix, the bootstrap confidence intervals tend to reproduce coverage rates close to the nominal ones. This is especially true when the BC_a confidence interval is used. On the other side, the coverage rates of the asymptotic N intervals are not very close to the nominal ones. Even under the Cauchy heavy-tailed contamination, the BC_a confidence intervals remain accurate. Still with $n = 20$, the same kind of conclusion could be derived from Tables 2–4. Note that the cases of β_{12} and γ_{13} have been studied under both the normal and the contaminated model, as well as for the larger sample size $n = 40$. Not surprisingly, when $n = 40$, we can see the accuracy of the asymptotic N improves, especially for the the case of γ_{13}; see Tables 5 and 6. The BC and BC_a intervals do not differ substantially. Both yield more accurate coverage rates than the N and P intervals. Note that from a theoretical point of view, the BC is only first-order accurate, whereas the BC_a is second-order accurate. Hence, the bias correction appears more important than the variance stabilization, i.e., the consideration of the acceleration constant.

Table 6. Confidence intervals for γ_{13}, normal errors (upper part), and normal with 5% Cauchy contamination errors (lower part), $n = 40$.

Interval	% Left	% Center	% Right	LB	UB	Level
N	8.57	89.66	1.77	−0.514	0.054	90% normal errors
P	1.52	96.46	2.02	−0.686	0.035	
BC	3.03	93.69	3.28	−0.684	0.024	
BC_a	3.28	93.43	3.28	−0.679	0.032	
N	5.08	94.32	0.60	−0.568	0.112	95% normal errors
P	1.52	97.98	0.51	−0.805	0.091	
BC	2.27	96.97	0.76	−0.771	0.080	
BC_a	2.27	96.97	0.76	−0.770	0.083	
N	5.34	90.84	3.82	−0.826	0.260	90% normal-Cauchy errors
P	1.52	95.96	2.53	−1.030	0.365	
BC	3.28	93.94	2.78	−0.971	0.414	
BC_a	3.54	92.42	4.04	−0.985	0.422	
N	2.59	95.75	1.66	−0.943	0.366	95 % normal-Cauchy errors
P	1.01	97.47	1.52	−1.237	0.540	
BC	1.26	97.47	1.26	−1.209	0.581	
BC_a	1.52	96.72	1.77	−1.209	0.580	

% left, % center, and % right: proportion of simulations having the true value on the left side, inside, or on the right side of the simulated intervals, respectively. LB and UB: medians of the simulated lower and upper bounds, respectively, of the intervals.

The general conclusion we could draw from these results is that there is an underlying risk in using the N and the P confidence intervals, since they can lead to coverage rates far away from the nominal rates, especially at high levels. On the other side, both the BC and the BC_a confidence intervals guarantee coverage rates close to the nominal rates. Although the acceleration constant does not seem to have a crucial role, it leads to theoretical second-order accuracy, which is an additional guarantee of accuracy, and it can be easily computed.

APPENDIX A
SECOND-ORDER DERIVATIVES OF THE LAGRANGIAN

As a result, the second-order derivatives of the Lagrangian are given by

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\Theta)^\top \partial \operatorname{vec}\Theta} &= -n\left\{\left[\left(L^\top\Theta\right)^{-1} L^\top\right] \otimes \left[L\left(\Theta^\top L\right)^{-1}\right]\right\} P_{g+k,g} \\
&\quad - \frac{1}{2}\left(\Sigma^{-1}\right)^\top \otimes \left(WW^\top\right) - \frac{1}{2}\Sigma^{-1} \otimes \left(WW^\top\right), \\
\frac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\Theta)^\top \partial \operatorname{vec}\Sigma} &= \frac{1}{2}\left[I \otimes \left(WW^\top\Theta\right)\right]\left(I + P_{g,g}\right)\left[\left(\Sigma^{-1}\right)^\top \otimes \Sigma^{-1}\right], \\
\frac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\Sigma)^\top \partial \operatorname{vec}\Theta} &= \left[\frac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\Theta)^\top \partial \operatorname{vec}\Sigma}\right]^\top, \\
\frac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\Sigma)^\top \partial \operatorname{vec}\Sigma} &= \frac{n}{2} P_{g,g}\left[\left(\Sigma^{-1}\right)^\top \otimes \Sigma^{-1}\right] - \frac{1}{2} P_{g,g}\left[\left(\Sigma^{-1}\right)^\top \otimes \left(\Sigma^{-1}\Theta^\top WW^\top\Theta\Sigma^{-1}\right)\right] \\
&\quad - \frac{1}{2} P_{g,g}\left[\left\{\left(\Sigma^{-1}\right)^\top \Theta^\top WW^\top\Theta\left(\Sigma^{-1}\right)^\top\right\} \otimes \Sigma^{-1}\right].
\end{aligned}$$

With the above formulae, we can obtain the complete matrix of second-order derivatives

$$\frac{\partial^2 \mathcal{L}}{\partial t^\top \partial t} = \begin{pmatrix}
\dfrac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\Theta)^\top \partial \operatorname{vec}\Theta} & \dfrac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\Sigma)^\top \partial \operatorname{vec}\Theta} & -R^\top & 0 \\
\dfrac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\Theta)^\top \partial \operatorname{vec}\Sigma} & \dfrac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\Sigma)^\top \partial \operatorname{vec}\Sigma} & 0 & -C^\top \\
-R & 0 & 0 & 0 \\
0 & -C & 0 & 0
\end{pmatrix}.$$

APPENDIX B
GENERATION OF MULTIVARIATE CAUCHY RANDOM VARIABLES

A multivariate Cauchy random variable in $\mathbb{R}^p$ has the density function with parameters $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p\times p}$ given by

$$\Gamma\left(\frac{p+1}{2}\right)\pi^{-(p+1)/2}\det{}^{-1/2}(\Sigma)\left[1 + (x-\mu)^\top\Sigma^{-1}(x-\mu)\right]^{-(p+1)/2}.$$

An algorithm for generating a Cauchy (μ, Σ) random vector x is the following.

STEP 1. Generate y multinormal $(0, \Sigma)$, and u standard normal.

STEP 2. Compute $x = \mu + y/|u|$.

For general methods for generating α-stable random variables, see, e.g., [15].

REFERENCES

1. B. Efron, Better bootstrap confidence intervals, *Journal of the American Statistical Association* **82**, 171–200, (1987).
2. B. Mandelbrot, The Pareto-Lévy law and the distribution of income, *International Economic Review* **1**, 79–106, (1960).

3. W. Feller, *An Introduction to Probability Theory and Its Applications*, Volume 2, Wiley, (1966).
4. E. Fama and R. Roll, Some properties of symmetric stable distributions, *Journal of the American Statistical Association* **63**, 817–836, (1968).
5. E. Fama, The behaviour of stock market prices, *Journal Business* **38**, 34–105, (1965).
6. B. Efron, Bootstrap methods: Another look to the jackknife, *The Annals of Statistics* **7**, 1–26, (1979).
7. B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, (1993).
8. A.C. Davison and D.V. Hinkley, *Bootstrap Methods and Their Application*, Cambridge University Press, (1997).
9. H.D. Vinod, Bootstrap methods: Applications in econometrics, In *Handbook of Statistics: Econometrics*, Volume 11, (Edited by G.S. Maddala *et al.*), Chapter 23, pp. 629–661, North Holland, (1993).
10. P. Hall, Theoretical comparisons of bootstrap confidence intervals (with discussion), *The Annals of Statistics* **16**, 927–985, (1988).
11. T.J. DiCiccio and B. Efron, Bootstrap confidence intervals, *Statistical Science* **11**, 189–212 (Discussion: 212–228), (1996).
12. R. Gatto, Saddlepoint approximations of marginal densities and confidence intervals for regression *R*-estimators, *Journal of Nonparametric Statistics* **7**, 239–253, (1997).
13. F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, *Robust Statistics. The Approach Based on Influence Functions*, Wiley & Sons, (1986).
14. J. Johnston, *Econometric Methods*, McGraw-Hill, (1972).
15. L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, (1986).