Beatlestrap

Alessandro Palandri*
Warwick Business School and CREATES†

This Version: October 19, 2010

Abstract

The bootstrap of test statistics requires the re-estimation of the model's parameters for each bootstrap sample. When parameter estimates are not available in closed form, this procedure becomes computationally demanding as each replication requires the numerical optimization of an objective function. This paper investigates the feasibility of the Beatlestrap, an optimization-free approach to bootstrap. It is shown that, *ex-post*, M-estimators may be expressed in terms of simple arithmetic averages therefore reducing the bootstrap of Wald statistics to the bootstrap of averages. Similarly, it is shown how the Lagrange Multiplier and the Likelihood Ratio statistics may be bootstrapped bypassing the objective function's multiple optimizations. The proposed approach is extended to simulation based Indirect Estimators. The finite sample properties of Beatlestrap are investigated via Monte Carlo simulations.

Keywords: Bootstrap, Wald test, Lagrange Multiplier test, Likelihood Ratio test, Indirect Estimators, Indirect Inference, Efficient Method of Moments.

JEL classification: C01, C12, C15.

*Mailing address: Warwick Business School, University of Warwick, Coventry, CV4 7AL, United Kingdom, Email: alessandro.palandri@gmail.com

†CREATES, Center for Research in Econometric Analysis of Time Series, is funded by the Danish National Research Foundation.

1 Introduction

Since the seminal article of Efron (1979), bootstrap has been widely employed in empirical works as well as studied and extended along various dimensions in theoretical papers. Nevertheless, the range of practical applications of the bootstrap methodology is constrained by the computational requirements of the underlying estimator. Specifically, B bootstrap replications involve a magnification of the computational burden by a factor B. For estimators that are available in closed form, B estimates may be replicated essentially at no cost. Estimators that are not available in closed form and simulation-based estimators, for which the objective function itself is not available in closed form, often require time consuming optimizations. While the latter may be feasibly carried out once to compute the parameters' point estimates, B calls to the optimizer required to bootstrap the distribution of the statistic of interest are at best impractical[1].

In this paper it is shown that, *ex-post*, M-estimators[2] and the associated Wald, Lagrange Multiplier (LM) and Likelihood Ratio (LR) test statistics may be expressed, without approximations, in terms of simple arithmetic averages. The terms of these averages are unknown *ex-ante* but may be computed *ex-post* the estimation. Hence, while the estimators' correspondence with averages is of no help in attaining the point estimates, it may be effectively exploited to bootstrap the estimators' distribution. This is what Beatlestrap is about: reinterpret the class of M-estimators from a slightly different angle and be able to bypass the B optimizations of the classic bootstrap.

The *ex-post* correspondence between M-estimators and arithmetic averages is obtained as a byproduct of the Mean Value Theorem (MVT). In particular, for a non-linear function evaluated at the point estimates $\widehat{\theta}$, take the MVT expansion around θ and then evaluate it in $\theta = \widehat{\theta}$. This results in a tautology and therefore an exact relationship, which will be used throughout the paper to perform optimization-free bootstraps, henceforth Beatlestrap. Since Beatlestrap is essentially bootstrap performed on averages, it is straightforward to extend to existing bootstrap schemes such as the block- and wild-bootstrap.

The proposed Beatlestrap is neither a competitor nor an alternative to bootstrap.

[1] A first attempt to address this issue may be found in Andrews(2002) who advocates the use of a k-step estimator to minimize the computational burden of bootstrap. However, it must be noted that the first steps of an optimizer are generally the most delicate and unless good initial conditions are provided, the algorithm is likely to diverge.

[2] Here the term M-estimators is used in the broad sense, including the Generalized Method of Moments and Indirect Estimators such as Indirect Inference and the Efficient Method of Moments.

It is a feasible finite sample alternative to the use of asymptotic approximations when the high computational demands of the estimation procedure render bootstrap particularly inconvenient. Beatlestrap, on the other hand, is always a convenient alternative: it relies on the same elements needed to compute the asymptotic variance-covariance matrix of the estimator in conjunction with their bootstrap, which reads as the bootstrap of one or two arithmetic averages.

The paper is organized as follows. The intuition behind Beatlestrap is presented in Section 2 which describes the classic bootstrap of the mean estimator and its *reverse engineering* into Beatlestrap. Beatlestrap of the Wald, Likelihood Ratio and Lagrange Multiplier statistics is described in Sections 3, 4 and 5 respectively. Beatlestrap of Wald statistics for simulation based Indirect Estimators is derived in Section 6. A proposed refinement is presented in Section 7, while the applicability of Beatlestrap is discussed in Section 8. Setup and results of the Monte Carlo simulations investigating the finite sample properties of the proposed approach are shown in Section 9. Section 10 concludes.

2 From a Different Angle

Consider estimating the mean of N observables y_i and bootstrapping the estimator's distribution. Assume that the data generating process is:

$$y_i = \mu + \epsilon_i$$

with ϵ_i i.i.d. random variables with mean zero and variance σ^2. Let $\overline{y}$ be the sample mean and $\widehat{\epsilon}_i \equiv y_i - \overline{y}$ be the residuals. In the *Residual-Resampling* scheme, at the b-th bootstrap replication, the synthetic response variables $\widehat{y}_{b,i}$ are generated according to:

$$\widehat{y}_{b,i} = \overline{y} + \widehat{\epsilon}_{u(b,i)}$$

where for each pair (b, i) the function $u(\bullet)$ randomly selects, with equal probability, a residual $\{\widehat{\epsilon}_i\}_{i=1}^N$. The $b-$th replication estimate is then computed as:

$$\overline{y}_b = \frac{1}{N}\sum_{i=1}^N \widehat{y}_{b,i}$$

and the empirical distribution of $\overline{y}$ characterized by $\{\overline{y}_b\}_{b=1}^B$.

The starting point of Beatlestrap is the characterization of the estimator $\overline{y}$ by a tautological expression:

$$\overline{y} = \overline{y} + \frac{1}{N}\sum_{i=1}^N \widehat{\epsilon}_i \tag{1}$$

which reads: "$\overline{y}$ is equal to the constant $\overline{y}$ and the arithmetic average of the random variables $\widehat{\epsilon}_i$, which have an in-sample mean of zero." *Ex-post*, this allows for the characterization of the generally abstract quantity $\overline{y}_i$:

$$\overline{y}_i = \overline{y} + \widehat{\epsilon}_i \tag{2}$$

Interpreting $\overline{y}_i$ as the *ex-post* observation of the value of the mean for individual i, immediately allows to define the bootstrap-synthetic response variable as:

$$\overline{y}_{b,i} = \overline{y} + \widehat{\epsilon}_{u(b,i)}$$

and the b-th replication estimate as:

$$\overline{y}_b = \overline{y} + \frac{1}{N}\sum_{i=1}^{N} \widehat{\epsilon}_{u(b,i)}$$

The idea behind this trivial example is relatively simple but applied *mutatis mutandis* to the class of M-estimators allows to Beatlestrap the distribution of the Wald, LM and LR test statistics free of optimizations.

3 Wald Test

Consider the M-estimator $\widehat{\theta}$ of the vector of parameters θ:

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} \frac{1}{N}\sum_{i=1}^{N} \psi_i\left(\theta\right)$$

where, with abuse of notation, $\psi_i(\theta) \equiv \psi(y_i, X_i; \theta)$ is function of the endogenous and exogenous variables corresponding to individual i. From the first order conditions (FOC) it follows that:

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial\theta}\psi_i(\widehat{\theta}) = 0$$

and from the Mean Value Theorem:

$$0 = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial\theta}\psi_i(\theta) + \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'}\psi_i(\theta^*) \cdot (\widehat{\theta} - \theta)$$

for any $\theta \in \Theta$ and with $\theta^* \in (\theta, \widehat{\theta})$. Rearranging terms yields:

$$\widehat{\theta} = \theta + \left[-\frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'}\psi_i(\theta^*)\right]^{-1} \cdot \frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial\theta}\psi_i(\theta) \tag{3}$$

Equation (3) is standard in deriving the asymptotic approximation to the distribution of the estimator $\widehat{\theta}$. In Beatlestrap, it is employed to obtain an alternative and exact representation of the estimator itself. In particular, evaluating the MVT expansion in $\theta = \widehat{\theta}$ leads to the following tautology:

$$\begin{aligned} \widehat{\theta} &= \widehat{\theta} + \left[-\frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'} \psi_i(\widehat{\theta}) \right]^{-1} \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial\theta} \psi_i(\widehat{\theta}) \\ &= \widehat{\theta} + H(\widehat{\theta})^{-1} \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial\theta} \psi_i(\widehat{\theta}) \end{aligned} \quad (4)$$

It must be stressed that equation (4) is not an approximation exactly in the same way that equation (1) is not an approximation. Equation (4) reads: "$\widehat{\theta}$ is equal to the constant $\widehat{\theta}$ and the arithmetic average of the random variables $\partial\psi_i(\widehat{\theta})/\partial\theta$, which have an in-sample mean of zero, weighted by the matrix $H(\widehat{\theta})^{-1}$." Following the same reasoning that brought about equation (1), it is possible to define the abstract quantity $\widehat{\theta}_i$:

$$\widehat{\theta}_i = \widehat{\theta} + H(\widehat{\theta})^{-1} \cdot \frac{\partial}{\partial\theta} \psi_i(\widehat{\theta}) \quad (5)$$

which, in turn, may be used to rewrite the estimator $\widehat{\theta}$ as:

$$\widehat{\theta} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\theta}_i \quad (6)$$

Equation (6) redefines the M-estimator $\widehat{\theta}$ as a simple arithmetic average of the quantities $\widehat{\theta}_i$. *Ex-ante*, equation (6) is not particularly useful as the elements $\widehat{\theta}_i$ are a function of the unknown $\widehat{\theta}$. *Ex-post*, having estimated $\widehat{\theta}$ it is possible to compute $\{\widehat{\theta}_i\}_{i=1}^{N}$ rendering the redundant-looking equation (6) computable in closed form. As a consequence, *ex-post*, Beatlestrap replicates are straightforward to compute as simple averages:

$$\begin{aligned} \widehat{\theta}_b &= \frac{1}{N} \sum_{i=1}^{N} \widehat{\theta}_{u(b,i)} \\ &= \widehat{\theta} + \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{\theta}_{u(b,i)} - \widehat{\theta} \right) \\ &= \widehat{\theta} + H(\widehat{\theta})^{-1} \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial\theta} \psi_{u(b,i)}(\widehat{\theta}) \end{aligned}$$

The empirical distribution of the M-estimator $\widehat{\theta}$ may be characterized, free of optimizations[3], by $\{\widehat{\theta}_b\}_{b=1}^{B}$. Retaining the quantiles of interest[4] allows to promptly Beatlestrap the Wald statistics.

4 Likelihood Ratio Test

Standard bootstrap of the LR statistic requires $2 \times B$ optimizations: B for the constrained- and B for the unconstrained-model. Beatlestrap, instead, may be performed free of optimizations. The starting point of the Beatlestrap approach to LR is the characterization of a LR tautology. Taking the MVT expansion of the unconstrained log-likelihood function $L(\theta)$ around the vector of unconstrained estimates $\widehat{\theta}$ gives:

$$L(\theta) = L(\widehat{\theta}) + \sum_{i=1}^{N} \frac{\partial}{\partial\theta'} l_i(\widehat{\theta}) \cdot (\theta - \widehat{\theta}) + \frac{1}{2}(\theta - \widehat{\theta})' \sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'} l_i(\mathring{\theta}) \cdot (\theta - \widehat{\theta}) \tag{7}$$

with $L(\theta) = \sum_{i=1}^{N} l_i(\theta)$ and $\mathring{\theta} \in (\theta, \widehat{\theta})$. For compactness of notation, define the quantities:

$$\begin{aligned} S(\theta) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial}{\partial\theta} l_i(\theta) \\ H(\theta) &= -\frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'} l_i(\theta) \end{aligned}$$

and substitute them in equation (7):

$$\begin{aligned} L(\theta) &= L(\widehat{\theta}) + \sqrt{N} S(\widehat{\theta})'(\theta - \widehat{\theta}) - \frac{N}{2}(\theta - \widehat{\theta})' H(\mathring{\theta})(\theta - \widehat{\theta}) \\ &= L(\widehat{\theta}) - \frac{N}{2}(\widehat{\theta} - \theta)' H(\mathring{\theta})(\widehat{\theta} - \theta) \end{aligned} \tag{8}$$

For $\psi = l$, $\theta^* \in (\theta, \widehat{\theta})$ and the definitions of $S(\theta)$ and $H(\theta)$, equation (3) may be rewritten as:

$$\sqrt{N}(\widehat{\theta} - \theta) = H(\theta^*)^{-1} S(\theta) \tag{9}$$

Substituting (9) in (8) gives:

$$L(\theta) = L(\widehat{\theta}) - \frac{1}{2} S(\theta)' H(\theta^*)^{-1} H(\mathring{\theta}) H(\theta^*)^{-1} S(\theta) \tag{10}$$

[3] Gonçalves and White (2004) and Corradi and Iglesias (2008) bootstrap the elements of the likelihood function but for each artificial sample re-optimize the vector of parameters.

[4] This includes the quantiles of the parameters as well as the quantiles of functions of the parameters.

Following the same steps, *mutatis mutandis*, for the constrained vector of parameters α, yields:

$$L(\alpha) = L(\widehat{\alpha}) - \frac{1}{2} S_\alpha(\alpha)' H_\alpha(\alpha^*)^{-1} H_\alpha(\mathring{\alpha}) H_\alpha(\alpha^*)^{-1} S_\alpha(\alpha) \tag{11}$$

where $S_\alpha = N^{-1/2} \sum_{i=1}^{N} \partial l_i / \partial \alpha$ and $H_\alpha = -N^{-1} \sum_{i=1}^{N} \partial^2 l_i / \partial \alpha \partial \alpha'$. Taking twice the difference between equations (10) and (11):

$$\begin{aligned} 2\left[L(\theta) - L(\alpha)\right] &= 2\left[L(\widehat{\theta}) - L(\widehat{\alpha})\right] + S_\alpha(\alpha)' H_\alpha(\alpha^*)^{-1} H_\alpha(\mathring{\alpha}) H_\alpha(\alpha^*)^{-1} S_\alpha(\alpha) \\ &- S(\theta)' H(\theta^*)^{-1} H(\mathring{\theta}) H(\theta^*)^{-1} S(\theta) \end{aligned}$$

and rearranging terms:

$$\begin{aligned} \widehat{LRT} &= 2\left[L(\theta) - L(\alpha)\right] + S(\theta)' H(\theta^*)^{-1} H(\mathring{\theta}) H(\theta^*)^{-1} S(\theta) \\ &- S_\alpha(\alpha)' H_\alpha(\alpha^*)^{-1} H_\alpha(\mathring{\alpha}) H_\alpha(\alpha^*)^{-1} S_\alpha(\alpha) \end{aligned} \tag{12}$$

Under the null hypothesis that the constraints are true $\theta = r(\alpha)$ and therefore:

$$S_\alpha(\alpha) = R' S(r(\alpha)) = R' S(\theta) \qquad \text{with } R' = \frac{\partial \theta'}{\partial \alpha} \tag{13}$$

Substituting equation (13) in (12) gives:

$$\begin{aligned} \widehat{LRT} &= 2\left[L(\theta) - L(\alpha)\right] + S(\theta)' H(\theta^*)^{-1} H(\mathring{\theta}) H(\theta^*)^{-1} S(\theta) \\ &- S(\theta)' R H_\alpha(\alpha^*)^{-1} H_\alpha(\mathring{\alpha}) H_\alpha(\alpha^*)^{-1} R' S(\theta) \end{aligned}$$

Evaluating at $\theta = \widehat{\theta}$ and $\alpha = \widehat{\alpha}$ produces the following tautology:

$$\begin{aligned} \widehat{LRT} &= 2\left[L(\widehat{\theta}) - L(\widehat{\alpha})\right] + S(\widehat{\theta})' H(\widehat{\theta})^{-1} S(\widehat{\theta}) - S(\widehat{\theta})' R H_\alpha(\widehat{\alpha})^{-1} R' S(\widehat{\theta}) \\ &= 2\left[L(\widehat{\theta}) - L(\widehat{\alpha})\right] + S(\widehat{\theta})' \left[H(\widehat{\theta})^{-1} - R H_\alpha(\widehat{\alpha})^{-1} R'\right] S(\widehat{\theta}) \end{aligned} \tag{14}$$

Finally, substituting $H_\alpha = R' H R$ in equation (14) yields:

$$\widehat{LRT} = 2\left[L(\widehat{\theta}) - L(\widehat{\alpha})\right] + S(\widehat{\theta})' \left[H(\widehat{\theta})^{-1} - R(R' H(\widehat{\theta}) R)^{-1} R'\right] S(\widehat{\theta}) \tag{15}$$

Since the unconstrained Score in the above equation is evaluated at the unconstrained parameters' estimates, the second term on the RHS is identically zero. To better interpret equation (15) rewrite it as:

$$\widehat{LRT} = \left(S(\widehat{\theta}) + \kappa\right)' \left[H(\widehat{\theta})^{-1} - R(R' H(\widehat{\theta}) R)^{-1} R'\right] \left(S(\widehat{\theta}) + \kappa\right)$$

where κ is such that:

$$\kappa' \left[H(\widehat{\theta})^{-1} - R(R' H(\widehat{\theta}) R)^{-1} R'\right] \kappa = \widehat{LRT}$$

Differently from the Wald statistic, the LR statistic is not a straight average but a quadratic from of the arithmetic average $\left(S(\widehat{\theta})+\kappa\right)$. Beatlestrap of $\widehat{LRT}$ is achieved keeping the weighting matrix fixed and bootstrapping $\left(S(\widehat{\theta})+\kappa\right)$ under the null hypothesis. This means re-centering at $S(\widehat{\theta})$ to obtain a zero mean vector. The b-th replication is therefore computed as:

$$\widehat{LRT}_b = S_b(\widehat{\theta})' \left[H^{-1} - R(R'HR)^{-1}R'\right] S_b(\widehat{\theta}) \tag{16}$$

where:

$$S_b(\widehat{\theta}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial}{\partial\theta} l_{u(b,i)}(\widehat{\theta})$$

The empirical distribution of the LR test statistic is given by $\{\widehat{LRT}_b\}_{b=1}^{B}$.
To get a better understanding of the role of the weighting matrix, factor out H^{-1}:

$$\widehat{LRT}_b = [H^{-1/2}S_b(\widehat{\theta})]' \left[I - H^{1/2}R(R'HR)^{-1}R'H^{1/2}\right] [H^{-1/2}S_b(\widehat{\theta})]$$

The matrix in square brackets is idempotent with rank equal to the number of constraints $dim(\theta) - dim(\alpha)$. This determines the degrees of freedom of the LR statistic. Asymptotically, the outer terms of the above equation are zero mean, unit variance Gaussian random vectors. Hence, $\widehat{LRT}_b$ will be a draw from a χ^2 distribution with degrees of freedom equal to the number of restrictions. However, if the likelihood function is not correctly specified, the test statistic does not have a standard asymptotic distribution since the Hessian H is the variance-covariance matrix of the Score only under correct specification.

5 Lagrange Multiplier Test

Let θ and α be, respectively, the unconstrained and the constrained vectors of parameters. Under the null hypothesis that the restrictions $\theta = r(\alpha)$ are true, the expected value of the unconstrained Score (or orthogonality conditions) evaluated at the constrained estimates is zero:

$$\mathbb{E}\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial}{\partial\theta}\psi_i(\widehat{\alpha})\right] = 0$$

Denoting with $\widehat{V}$ some estimator of the variance-covariance matrix of the FOC, the LM statistic is computed as:

$$\widehat{LM} = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial}{\partial\theta}\psi_i(\widehat{\alpha})\right)' \widehat{V}^{-1} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial}{\partial\theta}\psi_i(\widehat{\alpha})\right)$$

The Beatlestrap of the LM-test would seem straightforward: simply bootstrap the elements of the unconstrained FOC evaluated at the constrained vector of parameters. However, there is a subtle but crucial difference between the Score vector appearing in the LM-test and the Score vectors in the Wald- and LRT-tests. Specifically, the unconstrained FOC are evaluated at the unconstrained parameters' estimates in the latter cases while at the constrained estimates in the former case. Direct bootstrap of the unconstrained Score would ignore the restrictions defined by the constrained Score. In other words, bootstrapping the unconstrained Score generates draws with $dim(\theta)$ degrees of freedom while the target distribution only has $dim(\theta) - \dim(\alpha)$ degrees of freedom. Specifically, bootstrapping the elements $\partial\psi_i/\partial\theta$ of:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\psi_i(\widehat{\alpha}) \tag{17}$$

replicates the variability of the Score conditional on the estimates neglecting the fact that:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial}{\partial\alpha}\psi_i(\widehat{\alpha}) = 0 \tag{18}$$

imposes $dim(\alpha)$ constraints on the $dim(\theta)$-dimensional elements of the orthogonality conditions of equation (17).

Beatlestrap of the LM-test requires imposing the conditions in (18) to the Score in (17). An expansion of the Score evaluated at the parameters' estimates allows to correctly identify the elements to bootstrap free of optimizations. Let $S_i = \partial\psi_i/\partial\theta$, $\overset{\circ}{S}_i = \partial\psi_i/\partial\alpha$ and $\widetilde{\theta} = r(\widehat{\alpha})$ and take the MVT expansion of the constrained Score around α:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overset{\circ}{S}_i(\widehat{\alpha}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overset{\circ}{S}_i(\alpha) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial}{\partial\alpha'}\overset{\circ}{S}_i(\alpha^*)\cdot(\widehat{\alpha}-\alpha)$$

where $\alpha^* \in (\alpha, \widehat{\alpha})$ and from which it follows that:

$$(\widehat{\alpha}-\alpha) = \left[-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial}{\partial\alpha'}\overset{\circ}{S}_i(\alpha^*)\right]^{-1}\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overset{\circ}{S}_i(\widehat{\alpha}) \tag{19}$$

Applying the MVT to the unconstrained FOC evaluated at the constrained estimates yields:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}S_i(\widetilde{\theta}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}S_i(\theta) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial}{\partial\alpha'}S_i(\alpha^+)\cdot(\widehat{\alpha}-\alpha) \tag{20}$$

with $\alpha^+ \in (\alpha, \widehat{\alpha})$. Putting together equations (19) and (20) gives:

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} S_i(\widehat{\alpha}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} S_i(\alpha) - \left[-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\alpha'}S_i(\alpha^+)\right] \\ &\cdot \left[-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\alpha'}\overset{\circ}{S}_i(\alpha^*)\right]^{-1}\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overset{\circ}{S}_i(\alpha)\end{aligned}$$

where, with abuse of notation, $S_i(\alpha) \equiv S_i(r(\theta))$. Evaluating at $\alpha = \widehat{\alpha}$ yields:

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} S_i(\widehat{\alpha}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} S_i(\widehat{\alpha}) - \left[-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\alpha'}S_i(\widehat{\alpha})\right] \\ &\cdot \left[-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\alpha'}\overset{\circ}{S}_i(\widehat{\alpha})\right]^{-1}\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overset{\circ}{S}_i(\widehat{\alpha}) \qquad (21)\end{aligned}$$

Equation (21) reproduces a tautology apparently similar to that derived for the Wald Test in equation (4). The term on the LHS and the first term on the RHS are identical while the second term on the RHS is identically zero due to the presence of the constrained Score evaluated at the constrained parameters' estimates. However, the LM- and Wald-tautologies are intrinsically different and the correct interpretation of equation (21) is crucial for the optimization-free Beatlestrap. Here, correct bootstrap of the term on the LHS requires the bootstrap of the first term on the RHS (which reproduces the variability in the unconstrained Score conditional on the constrained estimates) and the last element of the second term on the RHS (which enforces the restrictions of the constrained Score). Differently put, the first term on the RHS of equation (4) is a realization from the distribution described by the second term on the RHS, while in equation (21) this is not the case and both terms on the RHS describe different aspects of the overall distribution of the term on the LHS. Using the following equalities:

$$\begin{aligned}R' &= \partial\theta'/\partial\alpha \\ \overset{\circ}{S}_i &= \frac{\partial\theta'}{\partial\alpha}\cdot\frac{\partial\psi_i}{\partial\theta} = R'S_i \\ \frac{\partial S_i}{\partial\alpha'} &= \frac{\partial S_i}{\partial\theta'}\cdot\frac{\partial\theta}{\partial\alpha'} = \frac{\partial S_i}{\partial\theta'}\cdot R \\ \frac{\partial \overset{\circ}{S}_i}{\partial\alpha'} &= R'\frac{\partial S_i}{\partial\alpha'} = R'\frac{\partial}{\partial\theta'}\cdot R\end{aligned}$$

and letting $H = -N^{-1}\sum_{i=1}^{N} \partial S_i/\partial\theta'$, equation (21) may be rewritten as:

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} S_i(\widehat{\alpha}) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N} S_i(\widehat{\alpha}) - [H(\widehat{\alpha})R]\cdot[R'H(\widehat{\alpha})R]^{-1}\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N} S_i(\widehat{\alpha}) \\ &= \left[I - H(\widehat{\alpha})R(R'H(\widehat{\alpha})R)^{-1}R'\right]\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N} S_i(\widehat{\alpha}) \qquad (22)\end{aligned}$$

Since under the null hypothesis that the restrictions are true, the expected value of the unconstrained Score evaluated at the constrained estimates is zero, Beatlestrap of the LM-statistic under the null requires the re-centering of the Score vector:

$$\widehat{LM}_b = \widehat{lm}_b'\widehat{V}^{-1}\widehat{lm}_b$$

with

$$\widehat{lm}_b = \left[I - H(\widehat{\alpha})R(R'H(\widehat{\alpha})R)^{-1}R'\right]\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \overset{\bullet}{S}_{u(b,i)}(\widehat{\alpha})$$

where

$$\overset{\bullet}{S}_i(\widehat{\alpha}) = S_i(\widehat{\alpha}) - \frac{1}{N}\sum_{i=1}^{N} S_i(\widehat{\alpha})$$

The matrix in square brackets in equation (22) is idempotent with rank equal to the number of constraints $dim(\theta) - dim(\alpha)$. This determines the correct degrees of freedom of the Beatlestrap replicates. The second term on the RHS of equation(22) is a random vector of dimensions $dim(\theta)$ and its product with the idempotent matrix of rank $dim(\theta) - dim(\alpha)$ produces the random vector $\widehat{lm}_b$ where the number of orthogonal elements is equal to the number of restrictions being tested. Asymptotically, $\widehat{LM}_b$ is distributed as a χ^2 with degrees of freedom equal to the number of constraints.

6 Wald Test for Indirect Estimators

The class of Indirect Estimators (IE) is defined by those estimators for which the objective function itself is not available in closed form and which make use of Monte Carlo simulations to compute the conditional value of such function. Estimators in this class are generally computationally intensive: partly due to the complexity of the model itself partly due to the optimization process involving observed and simulated data. Within this class, the model of interest is generally referred to as the *maintained model* and θ is its vector of parameters. While direct estimation of the maintained model is not feasible it is assumed to be relatively easy to simulate from it. The IE of θ matches moment conditions present in the data with moment conditions present in the simulations produced by the maintained model. Such moment conditions are typically generated by an *auxiliary model* which is easier to handle than the maintained model. Different measures of discrepancy between data- and simulated-moments lead to different IE.

The optimization of the objective function B times, required to generate classic bootstrap replicates, becomes practically infeasible for estimators that do not have a closed form of the objective function such as the IE. On the other hand, Beatlestrap of the Wald statistic of Indirect Estimators is always feasible. As for the class of M-estimators it is possible to derive an *ex-post* correspondence between the simulation-based Indirect Estimators and simple arithmetic averages, allowing to bypass the B simulation-based optimizations of the classic bootstrap altogether.

6.1 Efficient Method of Moments

The Efficient Method of Moments (EMM) estimator of Gallant and Tauchen (1996) first estimates the parameters $\lambda_{(q\times 1)}$ of the auxiliary model on the data. Such estimates will satisfy the associated moment conditions:

$$M_N \equiv M(y, X; \widehat{\lambda}) \equiv \frac{1}{N}\sum_{i=1}^{N} m(y_i, X_i; \widehat{\lambda}) = 0$$

Simulated values of the endogenous variables $y_i^s(\theta)$ are generated from the maintained model given $\theta_{(p\times 1)}$. If $p = q$, then the EMM estimate is that $\widehat{\theta}$ such that:

$$M_S \equiv M(y^s(\widehat{\theta}), X^s; \widehat{\lambda}) \equiv \frac{1}{S\cdot N}\sum_{s=1}^{S\cdot N} m(y_s(\widehat{\theta}), X_s; \widehat{\lambda}) = 0$$

In the general case of $q > p$, it will not be possible to solve q moment conditions with p parameters. Hence, the EMM estimator is defined as:

$$\widehat{\theta} = \arg\min_{\theta} M\left(y^s(\theta), X^s; \widehat{\lambda}\right)' \cdot \Omega \cdot M\left(y^s(\theta), X^s; \widehat{\lambda}\right)$$

where Ω is a positive definite symmetric matrix of weights. Let $\phi' = (\lambda' \quad \theta')$ and $\Psi(\phi)$ be the collection of all moment conditions associated with the EMM:

$$\Psi(\widehat{\phi}) = \begin{pmatrix} \Psi_1(\widehat{\phi}) \\ \Psi_2(\widehat{\phi}) \end{pmatrix} = \begin{cases} M(y, X; \widehat{\lambda}) \\ \frac{\partial}{\partial\theta} M\left(y^s(\widehat{\theta}), X^s; \widehat{\lambda}\right)' \cdot \Omega \cdot M\left(y^s(\widehat{\theta}), X^s; \widehat{\lambda}\right) \end{cases}$$

Taking a MVT expansion around ϕ, noticing that $\Psi(\widehat{\phi}) = 0$ and rearranging terms:

$$\widehat{\phi} = \phi - \left[\left. \frac{\partial \Psi}{\partial \phi'} \right|_{\phi^*} \right]^{-1} \cdot \Psi(\phi) \tag{23}$$

with $\phi^* \in (\widehat{\phi}, \phi)$. The derivative in square brackets and its inverse are equal to:

$$\left[\begin{array}{cc} \frac{\partial \Psi_1}{\partial \lambda'} & 0 \\ \frac{\partial \Psi_2}{\partial \lambda'} & \frac{\partial \Psi_2}{\partial \theta'} \end{array} \right]^{-1} = \left[\begin{array}{cc} \left(\frac{\partial \Psi_1}{\partial \lambda'}\right)^{-1} & 0 \\ -\left(\frac{\partial \Psi_2}{\partial \theta'}\right)^{-1} \frac{\partial \Psi_2}{\partial \lambda'} \left(\frac{\partial \Psi_1}{\partial \lambda'}\right)^{-1} & \left(\frac{\partial \Psi_2}{\partial \theta'}\right)^{-1} \end{array} \right]$$

from which it follows that:

$$\widehat{\theta} \;\; = \;\; \theta + \left(\frac{\partial \Psi_2}{\partial \theta'} \right)^{-1} \frac{\partial \Psi_2}{\partial \lambda'} \left(\frac{\partial \Psi_1}{\partial \lambda'} \right)^{-1} \Psi_1 - \left(\frac{\partial \Psi_2}{\partial \theta'} \right)^{-1} \Psi_2$$

where, Ψ_1 and Ψ_2 are evaluated at ϕ and their derivatives at ϕ^*. In deriving the asymptotic distribution of the estimator $\widehat{\theta}$, three constraints, which hold asymptotically for a correctly specified maintained model, are imposed. Only one of them will be imposed in this version of Beatlestrap for EMM. In particular:

$$\frac{\partial \Psi_2}{\partial \theta'} = \frac{\partial M_S'}{\partial \theta} \Omega \frac{\partial M_S}{\partial \theta'} + Q \cdot M_S = \frac{\partial M_S'}{\partial \theta} \Omega \frac{\partial M_S}{\partial \theta'}$$

since, for a correctly specified model, $M_S \to 0$. Substituting this restriction in the expression for $\widehat{\theta}$ gives:

$$\widehat{\theta} = \theta + \left(\frac{\partial M_S'}{\partial \theta} \Omega \frac{\partial M_S}{\partial \theta'} \right)^{-1} \frac{\partial M_S'}{\partial \theta} \Omega \cdot \left\{ \frac{\partial M_S}{\partial \lambda'} \left(\frac{\partial M_N}{\partial \lambda'} \right)^{-1} M_N - M_S \right\} \tag{24}$$

Evaluating equation (24) at the true parameters' value $\theta = \theta_0$, imposing two additional restrictions and noticing that $\mathbb{V}(M_S) = S^{-1} \cdot \mathbb{V}(M_N)$, implies an asymptotic variance-covariance matrix for the EMM estimator as in Gallant and Tauchen (1996). However, if evaluated at $\theta = \widehat{\theta}$, it gives rise to the following tautology:

$$\widehat{\theta} = \widehat{\theta} + \left(\frac{\partial M_S'}{\partial \theta} \Omega \frac{\partial M_S}{\partial \theta'} \right)^{-1} \frac{\partial M_S'}{\partial \theta} \Omega \cdot \left\{ \frac{\partial M_S}{\partial \lambda'} \left(\frac{\partial M_N}{\partial \lambda'} \right)^{-1} M_N - M_S \right\}$$

where M_N, M_S and all their derivatives are also evaluated at $\widehat{\theta}$. Beatlestrap replicates, in the form of arithmetic averages, are again straightforward to compute:

$$\begin{aligned} \widehat{\theta}_b \;\; = \;\; & \widehat{\theta} + \left(\frac{\partial M_S'}{\partial \theta} \Omega \frac{\partial M_S}{\partial \theta'} \right)^{-1} \frac{\partial M_S'}{\partial \theta} \Omega \cdot \left\{ \frac{\partial M_S}{\partial \lambda'} \left(\frac{\partial M_N}{\partial \lambda'} \right)^{-1} \right. \\ & \left. \cdot \;\; \frac{1}{N} \sum_{i=1}^{N} m(y_{u(b,i)}, X_{u(b,i)}; \widehat{\lambda}) - \frac{1}{S \cdot N} \sum_{s=1}^{S \cdot N} m(y_{v(b,s)}, X_{v(b,s)}; \widehat{\lambda}) \right\} \end{aligned} \tag{25}$$

where the function $v(\bullet)$ does what $u(\bullet)$ does, but independently.

6.2 Indirect Inference

The Indirect Inference (II) estimator of Gouriéroux et al. (1993) estimates the parameters of the auxiliary model $\lambda_{(q\times 1)}$ on the data and on the simulations from the maintained model given $\theta_{(p\times 1)}$. If $p = q$, then the II estimate is that $\widehat{\theta}$ such that:

$$\widehat{\lambda}_S(\widehat{\theta}) \equiv \lambda(y^S(\widehat{\theta}), X) = \lambda(y, X) \equiv \widehat{\lambda}_N$$

Equating $\widehat{\lambda}_S(\theta)$ to $\widehat{\lambda}_N$ is not possible when $q > p$. Therefore, in the general case, the II estimator is defined as:

$$\widehat{\theta} = \arg\min_{\theta} \left(\widehat{\lambda}_S(\theta) - \widehat{\lambda}_N\right)' \Omega \left(\widehat{\lambda}_S(\theta) - \widehat{\lambda}_N\right)$$

where Ω is a positive definite symmetric matrix of weights. Let $\phi' = (\lambda_N' \quad \lambda_S' \quad \theta')$ and $\Psi(\phi)$ be the collection of all moment conditions associated with the II:

$$\Psi(\widehat{\phi}) = \begin{pmatrix} \Psi_1(\widehat{\phi}) \\ \Psi_2(\widehat{\phi}) \\ \Psi_3(\widehat{\phi}) \end{pmatrix} = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} m\left(y_i, X_i; \widehat{\lambda}_N\right) \\ \frac{1}{S\cdot N}\sum_{s=1}^{S\cdot N} m\left(y_s(\widehat{\theta}), X_s; \widehat{\lambda}_S(\widehat{\theta})\right) \\ \frac{\partial \widehat{\lambda}_S(\widehat{\theta})'}{\partial\theta}\Omega\left(\widehat{\lambda}_S(\theta) - \widehat{\lambda}_N\right) \end{cases}$$

A MVT expansion of Ψ around ϕ yields an equation identical to (23) only with different dimensions. For II the derivative in square brackets and its inverse in equation (23) are equal to:

$$\begin{bmatrix} \frac{\partial\Psi_1}{\partial\lambda_N'} & 0 & 0 \\ 0 & \frac{\partial\Psi_2}{\partial\lambda_S'} & 0 \\ 0 & 0 & \frac{\partial\Psi_3}{\partial\theta'} \end{bmatrix}^{-1} = \begin{bmatrix} \left(\frac{\partial\Psi_1}{\partial\lambda_N'}\right)^{-1} & 0 & 0 \\ 0 & \left(\frac{\partial\Psi_2}{\partial\lambda_S'}\right)^{-1} & 0 \\ 0 & 0 & \left(\frac{\partial\Psi_3}{\partial\theta'}\right)^{-1} \end{bmatrix}$$

The diagonal structure of the matrix follows from the fact that: Ψ_1 is only a function of λ_N; Ψ_2 is not a function of λ_N; $\partial\Psi_2/\partial\theta = 0$ since as θ varies, λ_S varies to keep $\Psi_2 = 0$; λ_N and λ_S are pre-determined in Ψ_3. Evaluating equation (23) at $\phi = \widehat{\phi}$, it follows that:

$$\widehat{\lambda}_N = \widehat{\lambda}_N - \left(\frac{\partial\Psi_1}{\partial\lambda_N'}\right)^{-1}\Psi_1 \tag{26}$$

$$\widehat{\lambda}_S = \widehat{\lambda}_S - \left(\frac{\partial\Psi_2}{\partial\lambda_S'}\right)^{-1}\Psi_2 \tag{27}$$

$$\widehat{\theta} = \widehat{\theta} - \left(\frac{\partial\Psi_3}{\partial\theta'}\right)^{-1}\Psi_3 \tag{28}$$

Expanding Ψ_3, equation (28) may be rewritten as:

$$\widehat{\theta} = \widehat{\theta} - \left(\frac{\partial \Psi_3}{\partial \theta'}\right)^{-1} \frac{\partial \lambda'_S}{\partial \theta} \Omega \left(\widehat{\lambda}_S - \widehat{\lambda}_N\right)$$

and substituting in equations (26) and (27) gives:

$$\widehat{\theta} = \widehat{\theta} - \left(\frac{\partial \Psi_3}{\partial \theta'}\right)^{-1} \frac{\partial \lambda'_S}{\partial \theta} \Omega \left(\widehat{\lambda}_S - \left(\frac{\partial \Psi_2}{\partial \lambda'_S}\right)^{-1} \cdot \Psi_2 - \widehat{\lambda}_N + \left(\frac{\partial \Psi_1}{\partial \lambda'_N}\right)^{-1} \cdot \Psi_1\right)$$

From the moment conditions in Ψ_3 it may be simplified to:

$$\widehat{\theta} = \widehat{\theta} + \left(\frac{\partial \Psi_3}{\partial \theta'}\right)^{-1} \frac{\partial \lambda'_S}{\partial \theta} \Omega \left[\left(\frac{\partial \Psi_2}{\partial \lambda'_S}\right)^{-1} \cdot \Psi_2 - \left(\frac{\partial \Psi_1}{\partial \lambda'_N}\right)^{-1} \cdot \Psi_1\right]$$

Asymptotically valid restrictions are imposed on the derivatives when deriving the asymptotic variance-covariance matrix. Only one will be imposed in Beatlestrap, specifically:

$$\frac{\partial \Psi_3}{\partial \theta'} \rightarrow \frac{\partial \lambda'_S}{\partial \theta} \Omega \frac{\partial \lambda_S}{\partial \theta'}$$

which yields :

$$\widehat{\theta} = \widehat{\theta} + \left(\frac{\partial \lambda'_S}{\partial \theta} \Omega \frac{\partial \lambda_S}{\partial \theta'}\right)^{-1} \frac{\partial \lambda'_S}{\partial \theta} \Omega \left\{\left(\frac{\partial \Psi_2}{\partial \lambda'_S}\right)^{-1} \Psi_2 - \left(\frac{\partial \Psi_1}{\partial \lambda'_N}\right)^{-1} \Psi_1\right\}$$

Beatlestrap replicates, in the form of simple arithmetic averages, may be computed from:

$$\begin{aligned}
\widehat{\theta}_b &= \widehat{\theta} + \left(\frac{\partial \lambda'_S}{\partial \theta} \Omega \frac{\partial \lambda_S}{\partial \theta'}\right)^{-1} \frac{\partial \lambda'_S}{\partial \theta} \Omega \cdot \\
&\cdot \left\{\left(\frac{\partial \Psi_2}{\partial \lambda'_S}\right)^{-1} \frac{1}{S \cdot N} \sum_{s=1}^{S \cdot N} m(y_{v(b,s)}, X_{v(b,s)}; \widehat{\lambda}_S)\right. \\
&- \left.\left(\frac{\partial \Psi_1}{\partial \lambda'_N}\right)^{-1} \frac{1}{N} \sum_{i=1}^{N} m(y_{u(b,i)}, X_{u(b,i)}; \widehat{\lambda}_N)\right\} \qquad (29)
\end{aligned}$$

6.3 Alternative Calibration Procedures

Gouriéroux and Monfort (1996) present *symmetrical* calibration procedures to Indirect Inference and Efficient Method of Moments. These are:

$$\widehat{\theta} = \arg\max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \psi\left(y_i, X_i; \widehat{\lambda}_S(\theta)\right) \qquad (30)$$

and

$$\widehat{\theta} = \arg\min_{\theta} \left[\frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \lambda} \psi(y_i, X_i; \widehat{\lambda}_S(\theta) \right]' \Omega \left[\frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \lambda} \psi(y_i, X_i; \widehat{\lambda}_S(\theta) \right] \tag{31}$$

First proposed by Smith (1993), the estimator in equation (31) is closely related to GMM. Here, the $q \geq p$ moment conditions are generated by the auxiliary model and the $(p \times 1)$ vector of estimates minimizes their distance from the origin. The estimator in equation (30), instead, is closely related to a classical M-estimator except that the optimization is performed indirectly. Specifically, the objective function of the auxiliary model is not maximized with respect to the $(q \times 1)$ vector of parameters λ (which would yield parameters' estimates of the auxiliary model) but with respect to the $(p \times 1)$ vector θ of parameters of the maintained model. This is achieved by projecting θ over the λ space via Monte Carlo simulations. The use of this particular estimator has been advocated by Gallant and McCulloch (2010) in situations of sparse data. In such circumstances, an auxiliary model that provides a good statistical description of the data may not be feasibly estimated due to its parameters' dimensionality q. In turn, this will render infeasible both EMM and II estimators which rely on the estimation of λ_N. Instead, in the estimator of equation (30) simulations from the maintained model will result in a $(q \times 1)$ vector $\lambda_S(\theta)$ embedding $(q-p)$ constraints and p degrees of freedom.

Let $\phi' = (\lambda' \quad \theta')$ and $\Psi(\phi)$ be the collection of the moment conditions associated with the estimator of equation (30):

$$\Psi(\widehat{\phi}) = \begin{pmatrix} \Psi_1(\widehat{\phi}) \\ \Psi_2(\widehat{\phi}) \end{pmatrix} = \begin{cases} \frac{1}{S \cdot N} \sum_{s=1}^{S \cdot N} \frac{\partial}{\partial \lambda} \psi \left(y_s(\widehat{\theta}), X_s; \widehat{\lambda}_S \right) \\ \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \theta} \psi \left(y_i, X_i; \widehat{\lambda}_S \right) \end{cases}$$

The corresponding derivative in square brackets and its inverse in equation (23) are equal to:

$$\begin{bmatrix} \frac{\partial \Psi_1}{\partial \lambda'} & 0 \\ \frac{\partial \Psi_2}{\partial \lambda'} & \frac{\partial \Psi_2}{\partial \theta'} \end{bmatrix}^{-1} = \begin{bmatrix} \left(\frac{\partial \Psi_1}{\partial \lambda'} \right)^{-1} & 0 \\ -\left(\frac{\partial \Psi_2}{\partial \theta'} \right)^{-1} \frac{\partial \Psi_2}{\partial \lambda'} \left(\frac{\partial \Psi_1}{\partial \lambda'} \right)^{-1} & \left(\frac{\partial \Psi_2}{\partial \theta'} \right)^{-1} \end{bmatrix}$$

from which it follows that:

$$\widehat{\theta} = \theta + \left(\frac{\partial \Psi_2}{\partial \theta'} \right)^{-1} \cdot \left[\frac{\partial \Psi_2}{\partial \lambda'} \left(\frac{\partial \Psi_1}{\partial \lambda'} \right)^{-1} \Psi_1 - \Psi_2 \right] \tag{32}$$

There are constraints, which hold asymptotically for a correctly specified model, that are imposed in the derivation of the asymptotic variance-covariance matrix of the estimator and that will also be imposed in this version of the Beatlestrap. In particular:

$$\frac{\partial \Psi_2}{\partial \lambda'} = \frac{\partial \lambda'}{\partial \theta} \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial \lambda \partial \lambda'} \psi\left(y_i, X_i; \widehat{\lambda}_S\right) \tag{33}$$

$$\rightarrow \frac{\partial \lambda'}{\partial \theta} \cdot \frac{\partial \Psi_1}{\partial \lambda'} \tag{34}$$

which, substituted in equation(32), gives:

$$\widehat{\theta} = \theta + \left(\frac{\partial \Psi_2}{\partial \theta'}\right)^{-1} \cdot \left[\frac{\partial \lambda'}{\partial \theta} \Psi_1 - \Psi_2\right] \tag{35}$$

Further, notice that:

$$\begin{aligned}
\frac{\partial \Psi_2}{\partial \theta'} &= \frac{\partial}{\partial \theta'} \left[\frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \theta} \psi(y_i, X_i; \widehat{\lambda}_S)\right] \\
&= \frac{\partial}{\partial \theta'} \left[\frac{\partial \lambda'}{\partial \theta} \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \lambda} \psi(y_i, X_i; \widehat{\lambda}_S)\right] \\
&= \frac{\partial \lambda'}{\partial \theta} \left[\frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial \lambda \partial \lambda'} \psi(y_i, X_i; \widehat{\lambda}_S)\right] \frac{\partial \lambda}{\partial \theta'} \\
&= \frac{\partial \lambda'}{\partial \theta} H \frac{\partial \lambda}{\partial \theta'} \qquad (36)
\end{aligned}$$

where H is the Hessian of the auxiliary objective function. Substituting (36) in (35) and evaluating at $\theta = \widehat{\theta}$ gives:

$$\widehat{\theta} = \widehat{\theta} + \left(\frac{\partial \lambda'}{\partial \theta} H \frac{\partial \lambda}{\partial \theta'}\right)^{-1} \cdot \left[\frac{\partial \lambda'}{\partial \theta} \Psi_1 - \Psi_2\right]$$

Beatlestrap replicates may be computed as simple arithmetic averages from:

$$\begin{aligned}
\widehat{\theta}_b &= \widehat{\theta} + \left(\frac{\partial \lambda'}{\partial \theta} H \frac{\partial \lambda}{\partial \theta'}\right)^{-1} \frac{\partial \lambda'}{\partial \theta} \cdot \qquad (37) \\
&\cdot \left\{\frac{1}{S \cdot N} \sum_{s=1}^{S \cdot N} \frac{\partial}{\partial \lambda} \psi(y_{v(b,s)}, X_{v(b,s)}; \widehat{\lambda}_S) - \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \lambda} \psi(y_{u(b,i)}, X_{u(b,i)}; \widehat{\lambda}_S)\right\}
\end{aligned}$$

7 Beatlestrap*: A Refinement

The presented Beatlestrap implicitly conditions on the assumption that the estimated weighting matrices (WM) are, for all practical purposes, indistinguishable from the corresponding population quantities. However, in finite samples this may not be the case and the WM estimates may behave more like random variables than constants. This introduces a potential source of errors in the Beatlestrap of the test statistics. Here, a simple refinement, henceforth Beatlestrap*, is introduced which is aimed at mitigating distortions arising from possible finite sample correlations between the WM and the Score. The idea is to independently bootstrap the WM and Score to loosen potential finite sample correlations. The Beatlestrap* of the Wald test is:

$$\widehat{\theta}_b = \widehat{\theta} + H_b(\widehat{\theta})^{-1} \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \theta} \psi_{u(b,i)}(\widehat{\theta})$$

with:

$$H_b(\widehat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial \theta \partial \theta'} \psi_{v(b,i)}(\widehat{\theta})$$

and $v(b,i) \perp u(b,i)$. For the Likelihood Ratio test the refinement is:

$$\widehat{LRT}_b = S_b(\widehat{\theta})' \left[H_b^{-1} - R(R'H_bR)^{-1}R' \right] S_b(\widehat{\theta}) \tag{38}$$

where:

$$\begin{aligned} S_b(\widehat{\theta}) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial}{\partial \theta} l_{u(b,i)}(\widehat{\theta}) \\ H_b &= -\frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial \theta \partial \theta'} l_{v(b,i)}(\widehat{\theta}) \end{aligned}$$

and $v(b,i) \perp u(b,i)$. Similarly, the refined Lagrange Multiplier test is:

$$\widehat{LM}_b = \widehat{lm}_b' \widehat{V}^{-1} \widehat{lm}_b$$

with:

$$\begin{aligned} \widehat{lm}_b &= \left[I - H_b(\widehat{\alpha}) R (R' H_b(\widehat{\alpha}) R)^{-1} R' \right] \cdot \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \overset{\bullet}{S}_{u(b,i)}(\widehat{\alpha}) \\ H_b(\widehat{\alpha}) &= -\frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial \theta \partial \theta'} \psi_{v(b,i)}(\widehat{\alpha}) \end{aligned}$$

where:

$$\overset{\bullet}{S}_i(\widehat{\alpha}) = S_i(\widehat{\alpha}) - \frac{1}{N} \sum_{i=1}^{N} S_i(\widehat{\alpha})$$

and $v(b,i) \perp u(b,i)$.

Beatlestrap* of the Wald test for Indirect Estimators follows the same logic of loosening the correlation between the weighting matrices and the Score. With respect to the EMM tautology of equation (25) this implies the independent bootstrap of the matrices Ω and $\partial M_N/\partial\lambda'$ potentially correlated with each other and the Score M_N and that of the matrix $\partial M_S/\partial\theta'$ potentially correlated with M_S:

$$\widehat{\theta}_b = \widehat{\theta} + \left(\frac{\partial M'_{S,b}}{\partial\theta}\Omega_b\frac{\partial M_{S,b}}{\partial\theta'}\right)^{-1}\frac{\partial M'_{S,b}}{\partial\theta}\Omega_b\left\{\frac{\partial M_S}{\partial\lambda'}\left(\frac{\partial M_{N,b}}{\partial\lambda'}\right)^{-1}M_{N,b} - M_{S,b}\right\}$$

where:

$$\begin{aligned}
M_{N,b} &= \frac{1}{N}\sum_{i=1}^{N} m(y_{u(b,i)}, X_{u(b,i)}; \widehat{\lambda}) \\
M_{S,b} &= \frac{1}{S\cdot N}\sum_{s=1}^{S\cdot N} m(y_{v(b,s)}, X_{v(b,s)}; \widehat{\lambda}) \\
\frac{\partial M_{N,b}}{\partial\lambda'} &= \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\lambda'}m(y_{\eta(b,i)}, X_{\eta(b,i)}; \widehat{\lambda}) \\
\frac{\partial M_{S,b}}{\partial\theta'} &= \frac{1}{S\cdot N}\sum_{s=1}^{S\cdot N}\frac{\partial}{\partial\theta'}m(y_{\nu(b,s)}, X_{\nu(b,s)}; \widehat{\lambda})
\end{aligned}$$

and $u(\bullet) \perp v(\bullet) \perp \eta(\bullet) \perp \nu(\bullet)$. The matrix Ω_b is computed by bootstrapping the elements of the chosen estimator of the optimal weighting matrix using the orthogonal random number generator $z(\bullet)$.

Beatlestrap* of the Indirect Inference tautology of equation (29) is refined by independently bootstrapping the weighting matrices $\partial\Psi_1/\partial\lambda'_N$ and Ω possibly correlated with each other and Ψ_1 and $\partial\Psi_2/\partial\lambda'_S$ possibly correlated with each other and Ψ_2:

$$\widehat{\theta}_b = \widehat{\theta} + \left(\frac{\partial\lambda'_S}{\partial\theta}\Omega_b\frac{\partial\lambda_S}{\partial\theta'}\right)^{-1}\frac{\partial\lambda'_S}{\partial\theta}\Omega_b\left\{\left(\frac{\partial\Psi_{2,b}}{\partial\lambda'_S}\right)^{-1}\Psi_{2,b} - \left(\frac{\partial\Psi_{1,b}}{\partial\lambda'_N}\right)^{-1}\Psi_{1,b}\right\}$$

with:

$$\begin{aligned}
\Psi_{1,b} &= \frac{1}{N}\sum_{i=1}^{N} m(y_{u(b,i)}, X_{u(b,i)}; \widehat{\lambda}) \\
\Psi_{2,b} &= \frac{1}{S\cdot N}\sum_{s=1}^{S\cdot N} m(y_{v(b,s)}, X_{v(b,s)}; \widehat{\lambda}) \\
\frac{\partial\Psi_{1,b}}{\partial\lambda'_N} &= \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\lambda'}m(y_{\eta(b,i)}, X_{\eta(b,i)}; \widehat{\lambda}) \\
\frac{\partial\Psi_{2,b}}{\partial\lambda'_S} &= \frac{1}{S\cdot N}\sum_{i=1}^{S\cdot N}\frac{\partial}{\partial\lambda'}m(y_{\nu(b,s)}, X_{\nu(b,s)}; \widehat{\lambda})
\end{aligned}$$

and $u(\bullet) \perp v(\bullet) \perp \eta(\bullet) \perp \nu(\bullet) \perp z(\bullet)$, where $z(\bullet)$ is the random number generator used to bootstrap the elements of the optimal weighting matrix.

Refinement of the Alternative Calibration tautology of equation (37) is achieved by bootstrapping the weighting matrix H probably correlated with Ψ_2:

$$\widehat{\theta}_b = \widehat{\theta} + \left(\frac{\partial \lambda'}{\partial \theta} H_b \frac{\partial \lambda}{\partial \theta'} \right)^{-1} \frac{\partial \lambda'}{\partial \theta} \{ \Psi_{2,b} - \Psi_{1,b} \}$$

where:

$$\begin{aligned} \Psi_{1,b} &= \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \lambda} \psi(y_{u(b,i)}, X_{u(b,i)}; \widehat{\lambda}_S) \\ \Psi_{2,b} &= \frac{1}{S \cdot N} \sum_{s=1}^{S \cdot N} \frac{\partial}{\partial \lambda} \psi(y_{v(b,s)}, X_{v(b,s)}; \widehat{\lambda}_S) \\ H_b &= \frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial \lambda \partial \lambda'} \psi(y_{\eta(b,i)}, X_{\eta(b,i)}; \widehat{\lambda}_S) \end{aligned}$$

with $u(\bullet) \perp v(\bullet) \perp \eta(\bullet)$.

8 Applicability

Let the regularity conditions for the consistency and the asymptotic normality of the estimator be satisfied. Then, Beatlestrap is a valid approach which simply replaces: the direct estimation of the Score's variance with an estimator based on the bootstrap of the Score and the Gaussian approximation of the distribution of the Score with an approximation based on the bootstrap of the Score. Proof of these statements is straightforward as the Scores are averages of Score-contributions. In the case of misspecified models, resulting in non i.i.d. Score-contributions, the use of HAC estimators for the Score's variance-covariance matrix in the asymptotic case may be mirrored by the appropriate bootstrap scheme such as the block-bootstrap in the Beatlestrap approach.

The Beatlestrap* refinement differs from the plain Beatlestrap and asymptotic approximation in that the direct estimation of the weighting matrices of the Score, generally the Hessian, is replaced by a bootstrap estimator. Such refinement has been proposed to loosen the finite sample correlation between Score and weighting matrix. However, the refinement is asymptotically irrelevant as the rate of convergence of the bootstrapped weighting matrices is $N^{1/2}$ faster than the bootstrapped Score.

Furthermore, Beatlestrap and Beatlestrap* are computed exclusively from the same elements needed to compute the asymptotic variance-covariance matrix of the estimator. The sole difference being that the averages (Score and Hessian) are bootstrapped rather than kept constant.

9 Monte Carlo Simulations

Simulations have been conducted to evaluate the finite sample performance of the proposed Beatlestrap and the refined Beatlestrap* for a range of representative models. Since the computational time needed to generate $B = 2000$ Bootstrap replicates for $MC = 100000$ Monte Carlo replications is rather demanding, only asymptotic and Beatlestrap statistics have been computed and reported[5].

9.1 GARCH(1,1)

The observables y_t are equal to the product $h_t^{1/2} \cdot z_t$. The shocks z_t are independent and identically distributed with mean zero and unit variance. h_t follows a GARCH(1,1) process:

$$h_t = (1 - \alpha - \beta)\sigma^2 + \beta h_t + \alpha y_{t-1}^2$$

with $\sigma^2 = 1$, $\beta = 0.8$ and $\alpha = 0.1$. In the Monte Carlo simulations, the distribution of z_t is either a standardized Gaussian or a standardized Student-t with 5 degrees of freedom. The sample size varies from $T = 500$ to $T = 2000$ while the number of Beatlestrap replicates is kept fixed at $B = 2000$ and the number of simulations is $MC = 100000$. Estimation is carried out by Gaussian Quasi-Maximum-Likelihood.

Asymptotic and Beatlestrap test statistics require the evaluation of the Hessian of the log-likelihood function. However, the matrix of second derivatives of the log-likelihood delivers quite poor estimates of the Hessian, particularly in terms of positive definiteness and estimation error. In turn, this reflects negatively on the test statistics. An estimator of the Hessian with better finite sample properties is the expected information:

$$H(\widehat{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{2h_t^2} \cdot \frac{\partial h_t}{\partial \theta} \cdot \frac{\partial h_t}{\partial \theta'}\right]$$

[5] Comparison of the Beatlestrap approach with standard bootstrap is superfluous given the good finite sample performance of the proposed method in terms of size of the various test statistics.

This estimator has been the preferred choice for the simulations. The asymptotic variance-covariance matrix of the parameters is based on the sandwich form in which the variance of the Score is sandwiched between the inverse Hessians.

Table 1 reports the summary statistics of the simulations for the β parameter. Regardless of the simulations' setup, the estimates exhibit a significant negative bias.

Table 2 reports the empirical sizes of the t-statistics for the test of the null hypothesis $\widehat{\beta} = 0.8$. For $T = 500$, independently of the distribution of the shocks z_t, both the asymptotic approximation and Beatlestrap of Wald over-reject the null. This is due to the fact that the estimates $\widehat{\beta}$ are downward biased (Table 1) and therefore the distribution of the estimates is not centered around 0.8.

The size of the asymptotic approximation is spot-on for the likelihood ratio test when the shocks are normally distributed. However, when the shocks have a Student-t distribution with 5 degrees of freedom, it clearly over-rejects. This arises from the misspecification of the log-likelihood function: the Hessian is no longer equal to the variance-covariance matrix of the Score. As it can be seen such effect neither vanishes nor diminishes as the sample size increases. To the contrary, the larger sample size $T = 2000$ highlights it even more strongly. When z_t is normally distributed, Beatlestrap of LR-test slightly over-rejects while Beatlestrap* slightly under-rejects for $T = 500$. However, they are both nearly spot-on for $T = 2000$. On the other hand, when z_t is Student-t distributed, Beatlestrap clearly over-rejects even though less severely than asymptotic. Furthermore, such rejections appear to converge to the nominal size as the sample increases from $T = 500$ to $T = 2000$. The source of these effects is to be found in the correlation between the Hessian and the Score. Asymptotically, this correlation converges to zero but in finite samples, even as large as $T = 2000$, it may be enough to significantly bias the empirical distribution of the LR statistic. Beatlestrap* has been proposed to loosen the finite sample correlation between Hessian and Score. The Monte Carlo results are very favorable: for $T = 500$ the empirical sizes are 11.76%, 6.19% and 1.41%, exhibiting a slight over-rejection but a substantial improvement with respect to the asymptotic approximation for which the empirical sizes are 27.38%, 19.49% and 9.30% respectively. When $T = 2000$ the empirical sizes of LR-Beatlestrap* are 10.86%, 5.59% and 1.19%, nearly spot-on and unquestionably better than the asymptotic 31.93%, 23.76% and 12.56%.

For Gaussian z_t, the asymptotic approximation of the distribution of the LM statistic exhibits minimal over-rejections for $T = 2000$ and gradually more pronounced rejections for $T = 500$. Beatlestrap is uniformly better than asymptotic and Beatlestrap* is uniformly better than Beatlestrap for $T = 500$ and $T = 2000$. When z_t is Student-t distributed, the asymptotic approximation displays evident signs of over-rejection: 18.30%, 11.68% and 4.49% when $T = 500$ and 14.89%, 8.74% and 2.89% when $T = 2000$. Beatlestrap, on the other hand, is very accurate in both configurations: 11.76%, 5.60% and 0.92% when $T = 500$ and 10.55%, 5.02% and 0.90% when $T = 2000$. The empirical sizes of Beatlestrap*, with respect to Beatlestrap, show improvements at the 10% nominal, equivalence at the 5% nominal and slight under-rejection at the 1% nominal. Specifically, at the nominal 1%, the empirical size of Beatlestrap* is 0.76% when $T = 500$ and 0.84% when $T = 2000$.

9.2 TOBIT

The left-censored observables y_i are generated by a univariate TOBIT model:

$$y_i = \max\left(\beta x_i + \epsilon_i, 0\right)$$

where $\beta = 1$ and ϵ_i are i.i.d. Gaussian random variables. The distribution of the regressor x_i is either Gaussian or Log-normal $LN(0, 1)$. This particular setup might seem unusual given that it looks at the impact of different distributional properties of the regressor rather than the error term. Nevertheless, there is a good reason for doing so: in the tautologies derived in the paper it is the regressor that will most likely wreak havoc the asymptotic as well as the Beatlestrap results. This is due to the fact that while the error term enters the Score but may or may not enter the Hessian, the regressor always enters in both. The parameter β is estimated using the three Indirect Estimators of Sections 6.1, 6.2 and 6.3. The auxiliary model is a simple linear regression:

$$y_i = \lambda_0 + \lambda_1 x_i + u_i$$

In these Monte Carlo simulations the distribution of the regressor x_i is known. Therefore, in the simulation step of the three Indirect Estimators, draws of both x_s and ϵ_s are made in order to simulate the endogenous y_s. The Efficient Method of Moments estimates are computed using the optimal weighting matrix $\widehat{\Omega}$. The Indirect Inference estimates, on the other hand, are first-round estimates with the optimal weighting matrix set equal to the identity matrix I. Computing the II estimator's optimal Ω would require a two-stage estimation and additional independent simulations for its evaluation. In a simulation experiment designed to evaluate the finite sample properties of the estimator's distribution, focusing on the optimal II estimator does not seem crucial. Hence, the attention will be on the first-round estimator (consistent but not efficient) which allows to investigate the performance of the Beatlestrap approach to Indirect Inference at reasonable costs in terms of simulations and optimizations.

The study considers sample sizes $N = \{200, 400, 1000, 2000\}$, the number of Beatlestrap replications is fixed at $B = 2000$ and the number of Monte Carlo simulations is $MC = 100000$. The number of simulations used in the Indirect Estimators is $S = \{1, 10\}$ where $S = 1$ corresponds to a simulated sample of the same size as the sample N and $S = 10$ corresponds to a simulated sample which is ten times longer than the sample size.

Tables (3) and (4) report summary statistics of the simulations for Gaussian and Log-normal regressors respectively. It should be noted the presence of a significant bias, predominant when the regressor is Log-normally distributed. While correcting for bias is beyond the scope of this paper, its presence will affect the rejection rates of the Wald test statistics. Thus, to better analyze the size of the test statistics two sets of Tables will be presented: one containing standard rejection rates, the other containing rejection rates of re-centered statistics. In the latter, the null hypothesis $\beta = 1$ is replaced with the null β equals the Monte Carlo means of the simulations, as they appear in Tables (3) and (4).

Tables (5) and (6) contain size results of the Wald test for the three estimators when the regressor is Gaussian. Beatlestrap results are found in the second column while those of the Beatlestrap* are reported in the third column. Overall, the size of the asymptotic approximation is reasonable for all three estimators with a slight tendency to over-reject for samples of size less than $N = 2000$ observations. The refined- provides better coverage than the plain-Beatlestrap in roughly 90% of the cases for both the raw and re-centered statistics. For all sample sizes N and $S = 1$, the Beatlestrap* has better rejection rates than the asymptotic approximation in 97% of the cases. It is still superior in 90% of the cases for $S = 10$ and $N < 2000$. When the sample size is $N = 2000$ and $S = 10$ the asymptotic approximation becomes relatively good therefore leaving not much space for improvement to the Beatlestrap.

Tables (7) and (8) report the empirical sizes of the Wald statistics for Log-normal regressors. The asymptotic approximation presents a pronounced over-rejection for all three Indirect Estimators. The bias reported in Table (4) is a factor that contributes to over-reject the null hypothesis. Therefore, to evaluate the behavior of the asymptotic and Beatlestrap approximations, net of the bias, the attention should be focused on Table (8) which reports the empirical size of the re-centered statistics. For the Efficient Method of Moments and the Alternative Calibration the empirical size is consistently sensibly better than that of the asymptotic approximation. The same is true for the Indirect Inference estimator with the exception of the nominal 1% coverage: in all cases but one, Beatlestrap of II has a slightly higher over-rejection than the asymptotic approximation. The Beatlestrap* performs significantly better than the asymptotic approximation for all estimators and every sample size. The average rejection rates, taken *cum grano salis*, summarize relatively well the different magnitudes. To the nominal 10%, 5% and 1% rejection rates correspond average asymptotic empirical rates of 21.0%, 14.6% and 7.4% and Beatlestrap* rates of 12.7%, 6.6% and 1.5% respectively. The difference is obviously substantial. Furthermore, it emerges that the empirical size is quite close to nominal for EMM when $S = 1$: 9.52%, 4.75% and 1.13% on average. Thus, slight under-rejection at the 10% and 5% and slight over-rejection at 1%. When $S = 1$, the proposed Beatlestrap* over-rejects for both the Indirect Inference and Alternative Calibration estimators, still the improvements with respect to the asymptotic approximation are obvious. For $S = 10$, the Beatlestrap* continues to perform significantly better than the asymptotic approximation. However, it should be noted that it now exhibits a slightly higher degree of over-rejection when bench-marked against the nominal sizes. Perhaps this is due to the fact that the proposed refinement does not fully capture the behavior of the empirical distribution as a function of the parameter S and that further refinements might be needed. Nevertheless, looking at the average empirical rejection rates shows that to the nominal 10%, 5% and 1% correspond 21.6%, 14.9% and 7.5% for the asymptotic approximation and 13.5%, 7.1% and 1.6% for Beatlestrap*. Again, a substantial difference in favor of the proposed approach.

10 Conclusions

The proposed Beatlestrap approach allows to bootstrap Wald-, LM- and LR-test free of optimizations. Monte Carlo results highlight the effectiveness of Beatlestrap for the LM-test for which the empirical size is very close to nominal. Simulations also show good size properties for the Beatlestrap of the LR-test which could be improved by further refinements. On the other hand, Beatlestrap of the Wald-test, like the asymptotic approximation, is negatively affected by the estimator's finite sample bias. While the effects of the bias may be taken into account to re-center the Wald-statistic, the Monte Carlo simulations suggest the use of a different strategy. Usually the choice among Wald, LM and LR is dictated by convenience: whether the model is easier to estimate under the null (LM), the alternative (Wald), or both estimates are readily available (LR). However, when the asymptotic approximation is not considered to be reliable and standard bootstrap requires a non-negligible number of optimizations, the Beatlestrap-LM alternative becomes very attractive[6].

The optimization-free bias correction of the Wald-test, the definition of the Beatlestrap for the LM-test based on auxiliary regressions and the application of different bootstrap schemes, particularly with respect to the Indirect Estimators for which the auxiliary score is misspecified by construction, are left as areas for future research.

Acknowledgements

I would like to thank Valentina Corradi and the participants to the WBS Finance Group Workshops for very useful comments and suggestions.

References

Andrews, D. W. K. (2002), "Equivalence of the Higher Order Asymptotic Efficiency of K-Step and Extremum Statistics", *Econometric Theory*, **18** 1040-1085.

Efron, B. (1979), "Bootstrap methods: Another look at the jackknife", *Annals of Statistics*, **7** 1-26.

Corradi V. and E. M. Iglesias (2008), "Bootstrap Refinements for QML Estimators of the GARCH(1,1) Parameters", *Journal of Econometrics*, **144** 500-510.

Gallant, A. R. and R. E. McCulloch (2010), "On the Determination of General Scientific Models", *Journal of the American Statistical Association*, (forthcoming).

Gallant, A. R. and G. Tauchen (1996), "Which Moments to Match?", *Econometric Theory*, vol. 12, n. 4, 657-681.

Gonçalves, S. and H. White (2004), "Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models", *Journal of Econometrics*, **119** 199-219.

Gouriéroux, C. and A. Monfort (1996), Simulation-Based Econometric Methods, New York, New York, Oxford University Press. 61-91.

[6]Consider the case in which k parameters will be tested individually. Bootstrap of the Wald-tests will require B estimations while the Beatlestrap of the LM-tests requires only k estimations and the computation of kB arithmetic averages.

Gouriéroux, C., A. Monfort and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics*, **8** 85-118.

Smith, A. (1993), "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions", *Journal of Applied Econometrics*, **8** 63-84.

Table 1: GARCH: Summary Statistics of Simulations for the β parameter.

	mean	bias	var	mse
$T = 500$ $\epsilon_t \sim N(0,1)$	0.7488	-0.0512^{***}	0.0221	0.0247
$T = 500$ $\epsilon_t \sim \sqrt{\frac{3}{5}} \cdot t_5$	0.7280	-0.0720^{***}	0.0337	0.0389
$T = 2000$ $\epsilon_t \sim N(0,1)$	0.7912	-0.0088^{***}	0.0023	0.0024
$T = 2000$ $\epsilon_t \sim \sqrt{\frac{3}{5}} \cdot t_5$	0.7843	-0.0157^{***}	0.0049	0.0051

Table 2: GARCH: Size of Wald, LM and LR statistics. The null hypothesis is $\widehat{\beta} = \beta$.

		Asymptotic			Beatlestrap			Beatlestrap*		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$T = 500$	Wald	12.59	8.47	4.30	12.56	8.42	4.26	12.17	8.04	3.91
$\epsilon_t \sim N(0,1)$	LM	12.37	6.79	1.72	11.25	5.76	1.19	10.91	5.53	1.10
	LR	10.17	5.07	0.97	11.97	6.50	1.62	9.38	4.73	0.94
$T = 500$	Wald	16.48	11.93	6.97	17.04	12.53	7.46	16.38	11.75	6.62
$\epsilon_t \sim \sqrt{\frac{3}{5}} \cdot t_5$	LM	18.30	11.68	4.49	11.76	5.60	0.92	10.66	4.95	0.76
	LR	27.38	19.49	9.30	15.53	9.47	3.23	11.76	6.19	1.41
$T = 2000$	Wald	10.89	6.22	2.18	10.88	6.24	2.23	10.80	6.15	2.19
$\epsilon_t \sim N(0,1)$	LM	10.66	5.49	1.18	10.35	5.20	1.03	10.27	5.12	1.02
	LR	10.11	5.06	1.00	10.58	5.43	1.19	9.68	4.91	1.03
$T = 2000$	Wald	13.41	8.30	3.50	14.33	9.28	4.38	14.12	9,09	4.22
$\epsilon_t \sim \sqrt{\frac{3}{5}} \cdot t_5$	LM	14.89	8.74	2.89	10.55	5.02	0.90	10.31	4.91	0.84
	LR	31.93	23.76	12.56	12.78	7.00	1.78	10.86	5.59	1.19

Table 3: TOBIT: Summary Statistics of Simulations.

		mean	bias	var	mse
$N = 200$	EMM	1.0065	0.0065^{***}	0.0248	0.0248
$x_i \sim N(0,1)$	II	1.0022	0.0022^{***}	0.0250	0.0250
$S = 1$	ALT	1.0007	0.0007	0.0249	0.0249
$N = 400$	EMM	1.0030	0.0030^{***}	0.0122	0.0122
$x_i \sim N(0,1)$	II	1.0008	0.0008^{**}	0.0125	0.0125
$S = 1$	ALT	0.9999	-0.0000	0.0124	0.0124
$N = 1000$	EMM	1.0013	0.0013^{***}	0.0048	0.0048
$x_i \sim N(0,1)$	II	1.0005	0.0005^{**}	0.0049	0.0049
$S = 1$	ALT	1.0002	0.0002	0.0049	0.0049
$N = 2000$	EMM	1.0007	0.0007^{***}	0.0024	0.0024
$x_i \sim N(0,1)$	II	1.0002	0.0002	0.0025	0.0025
$S = 1$	ALT	1.0000	0.0000	0.0025	0.0025
$N = 200$	EMM	0.9999	-0.0001	0.0132	0.0132
$x_i \sim N(0,1)$	II	0.9989	-0.0011^{***}	0.0136	0.0136
$S = 10$	ALT	0.9973	-0.0026^{***}	0.0136	0.0136
$N = 400$	EMM	0.9998	-0.0002	0.0066	0.0066
$x_i \sim N(0,1)$	II	0.9994	-0.0006^{**}	0.0068	0.0068
$S = 10$	ALT	0.9986	-0.0014^{***}	0.0068	0.0068
$N = 1000$	EMM	0.9998	0.0002	0.0026	0.0026
$x_i \sim N(0,1)$	II	0.9997	-0.0003^{*}	0.0027	0.0027
$S = 10$	ALT	0.9994	-0.0006^{***}	0.0027	0.0027
$N = 2000$	EMM	0.9999	-0.0001	0.0013	0.0013
$x_i \sim N(0,1)$	II	0.9997	-0.0001	0.0013	0.0013
$S = 10$	ALT	0.9997	-0.0003	0.0013	0.0013

Table 4: TOBIT: Summary Statistics of Simulations.

		mean	bias	var	mse
$N = 200$	EMM	0.9887	-0.0113^{***}	0.0111	0.0112
$x_i \sim LN(0,1)$	II	1.0024	0.0024^{***}	0.0112	0.0112
$S = 1$	ALT	1.0043	0.0043^{***}	0.0122	0.0122
$N = 400$	EMM	0.9892	-0.0109^{***}	0.0057	0.0058
$x_i \sim LN(0,1)$	II	1.0011	0.0011^{***}	0.0057	0.0057
$S = 1$	ALT	1.0024	0.0024^{***}	0.0065	0.0065
$N = 1000$	EMM	0.9913	-0.0087^{***}	0.0025	0.0026
$x_i \sim LN(0,1)$	II	1.0004	0.0004^{**}	0.0025	0.0025
$S = 1$	ALT	1.0010	0.0010^{***}	0.0030	0.0030
$N = 2000$	EMM	0.9933	-0.0067^{***}	0.0013	0.0014
$x_i \sim LN(0,1)$	II	1.0002	0.0002	0.0014	0.0014
$S = 1$	ALT	1.0007	0.0007^{***}	0.0017	0.0017
$N = 200$	EMM	0.9653	-0.0347^{***}	0.0062	0.0074
$x_i \sim LN(0,1)$	II	0.9766	-0.0234^{***}	0.0059	0.0064
$S = 10$	ALT	0.9743	-0.0257^{***}	0.0064	0.0070
$N = 400$	EMM	0.9764	-0.0236^{***}	0.0033	0.0039
$x_i \sim LN(0,1)$	II	0.9858	-0.0142^{***}	0.0031	0.0033
$S = 10$	ALT	0.9842	-0.0158^{***}	0.0036	0.0038
$N = 1000$	EMM	0.9863	-0.0137^{***}	0.0014	0.0016
$x_i \sim LN(0,1)$	II	0.9929	-0.0071^{***}	0.0014	0.0015
$S = 10$	ALT	0.9921	-0.0079^{***}	0.0017	0.0018
$N = 2000$	EMM	0.9910	-0.0090^{***}	0.0007	0.0008
$x_i \sim LN(0,1)$	II	0.9958	-0.0042^{***}	0.0008	0.0008
$S = 10$	ALT	0.9953	-0.0047^{***}	0.0010	0.0010

Table 5: TOBIT: Size of Wald statistics for the Efficient Method of Moments, Indirect Inference and Alternative Calibration Procedure. The null hypothesis is $\widehat{\beta} = \beta$.

		Asymptotic			Beatlestrap			Beatlestrap*		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$N = 200$	EMM	12.51	7.24	2.35	11.52	6.22	1.71	10.14	5.06	1.15
$x_i \sim N(0,1)$	II	11.51	6.18	1.61	11.16	5.88	1.46	10.63	5.42	1.19
$S = 1$	ALT	11.56	6.25	1.63	11.24	5.90	1.47	10.81	5.49	1.20
$N = 400$	EMM	11.30	6.14	1.70	10.78	5.67	1.40	10.08	5.08	1.11
$x_i \sim N(0,1)$	II	10.80	5.66	1.31	10.60	5.59	1.26	10.37	5.33	1.13
$S = 1$	ALT	10.79	5.67	1.33	10.72	5.57	1.27	10.48	5.38	1.16
$N = 1000$	EMM	10.53	5.49	1.29	10.32	5.34	1.19	10.03	5.10	1.08
$x_i \sim N(0,1)$	II	10.34	5.32	1.15	10.29	5.27	1.19	10.21	5.14	1.12
$S = 1$	ALT	10.36	5.33	1.16	10.36	5.30	1.16	10.26	5.22	1.11
$N = 2000$	EMM	10.25	5.11	1.09	10.15	5.04	1.07	9.98	4.94	1.06
$x_i \sim N(0,1)$	II	10.08	5.08	1.05	10.07	5.05	1.10	10.05	5.00	1.04
$S = 1$	ALT	10.10	5.09	1.05	10.07	5.04	1.08	9.99	5.02	1.06
$N = 200$	EMM	12.28	7.10	2.27	12.09	6.86	2.06	10.59	5.52	1.32
$x_i \sim N(0,1)$	II	11.32	6.16	1.55	11.24	5.98	1.46	11.87	5.61	1.26
$S = 10$	ALT	11.33	6.16	1.57	11.28	6.17	1.66	10.80	5.71	1.39
$N = 400$	EMM	11.18	6.08	1.61	11.09	5.98	1.53	10.36	5.32	1.19
$x_i \sim N(0,1)$	II	10.59	5.50	1.30	10.56	5.44	1.27	10.38	5.28	1.18
$S = 10$	ALT	10.59	5.53	1.31	10.60	5.59	1.33	10.39	5.36	1.22
$N = 1000$	EMM	10.48	5.42	1.26	10.43	5.42	1.21	10.16	5.17	1.10
$x_i \sim N(0,1)$	II	10.27	5.22	1.12	10.28	5.24	1.11	10.21	5.16	1.06
$S = 10$	ALT	10.27	5.22	1.14	10.35	5.27	1.18	10.27	5.19	1.14
$N = 2000$	EMM	10.12	5.12	1.06	10.11	5.17	1.06	9.97	5.01	1.02
$x_i \sim N(0,1)$	II	9.96	5.03	1.00	10.01	5.08	1.08	9.97	5.05	1.06
$S = 10$	ALT	9.97	5.04	1.01	10.01	5.04	1.05	9.96	5.01	1.02

Table 6: TOBIT: Size of re-centered Wald statistics for the Efficient Method of Moments, Indirect Inference and Alternative Calibration Procedure. The null hypothesis is $\widehat{\beta} = \beta$.

		Asymptotic			Beatlestrap			Beatlestrap*		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$N = 200$	EMM	12.67	7.42	2.46	11.62	6.44	1.80	10.06	5.05	1.12
$x_i \sim N(0,1)$	II	11.53	6.20	1.62	11.19	5.90	1.48	10.64	5.44	1.22
$S = 1$	ALT	11.57	6.26	1.63	11.24	5.91	1.47	10.91	5.50	1.20
$N = 400$	EMM	11.35	6.21	1.75	10.85	5.73	1.41	10.05	5.05	1.12
$x_i \sim N(0,1)$	II	10.82	5.68	1.31	10.62	5.59	1.26	10.39	5.34	1.15
$S = 1$	ALT	10.79	5.66	1.33	10.72	5.57	1.27	10.48	5.38	1.16
$N = 1000$	EMM	10.54	5.52	1.31	10.35	5.38	1.21	9.99	5.10	1.08
$x_i \sim N(0,1)$	II	10.36	5.33	1.16	10.34	5.27	1.19	10.21	5.14	1.12
$S = 1$	ALT	10.36	5.33	1.16	10.35	5.29	1.16	10.28	5.22	1.11
$N = 2000$	EMM	10.27	5.14	1.10	10.15	5.05	1.08	9.99	4.95	1.04
$x_i \sim N(0,1)$	II	10.10	5.09	1.05	10.07	5.06	1.10	10.06	5.00	1.04
$S = 1$	ALT	10.10	5.09	1.05	10.06	5.04	1.08	9.99	5.02	1.06
$N = 200$	EMM	12.28	7.09	2.26	12.08	6.86	2.06	10.59	5.52	1.32
$x_i \sim N(0,1)$	II	11.31	6.12	1.53	11.24	5.99	1.46	10.84	5.62	1.25
$S = 10$	ALT	11.29	6.10	1.54	11.21	6.10	1.62	10.77	5.68	1.36
$N = 400$	EMM	11.18	6.08	1.61	11.08	5.98	1.53	10.37	5.32	1.19
$x_i \sim N(0,1)$	II	10.59	5.47	1.28	10.55	5.44	1.28	10.38	5.27	1.18
$S = 10$	ALT	10.58	5.50	1.31	10.57	5.54	1.32	10.37	5.33	1.20
$N = 1000$	EMM	10.46	5.41	1.26	10.42	5.41	1.20	10.16	5.16	1.09
$x_i \sim N(0,1)$	II	10.26	5.22	1.12	10.29	5.22	1.11	10.23	5.16	1.05
$S = 10$	ALT	10.27	5.22	1.13	10.31	5.25	1.17	10.23	5.17	1.13
$N = 2000$	EMM	10.11	5.11	1.06	10.10	5.17	1.06	9.97	5.01	1.02
$x_i \sim N(0,1)$	II	9.97	5.02	1.01	10.01	5.07	1.08	9.99	5.05	1.06
$S = 10$	ALT	9.97	5.02	1.01	10.02	5.04	1.05	9.96	5.01	1.01

Table 7: TOBIT: Size of Wald statistics for the Efficient Method of Moments, Indirect Inference and Alternative Calibration Procedure. The null hypothesis is $\widehat{\beta} = \beta$.

		Asymptotic			Beatlestrap			Beatlestrap*		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$N = 200$	EMM	26.06	21.38	15.90	23.36	18.29	12.52	8.39	3.97	0.79
$x_i \sim LN(0,1)$	II	21.31	14.51	6.42	19.29	13.14	6.54	13.30	6.81	1.40
$S = 1$	ALT	25.13	17.93	8.94	19.91	13.72	7.05	13.74	7.09	1.67
$N = 400$	EMM	23.77	18.98	13.41	21.09	15.87	10.10	8.51	4.02	0.80
$x_i \sim LN(0,1)$	II	19.62	12.74	5.18	17.77	11.80	5.58	13.79	7.43	1.77
$S = 1$	ALT	22.54	15.38	6.91	17.14	11.06	4.89	12.91	6.64	1.55
$N = 1000$	EMM	21.18	16.21	10.61	18.36	13.08	7.42	8.79	4.01	0.79
$x_i \sim LN(0,1)$	II	17.66	11.18	4.07	16.19	10.24	4.40	14.40	7.95	1.99
$S = 1$	ALT	19.71	12.90	5.29	14.30	8.33	2.89	11.96	5.89	1.43
$N = 2000$	EMM	19.18	14.13	8.62	16.22	10.99	5.54	8.76	4.08	0.78
$x_i \sim LN(0,1)$	II	16.14	10.00	3.46	14.78	9.02	3.56	13.90	7.95	2.05
$S = 1$	ALT	17.65	11.38	4.50	12.46	6.72	1.93	11.02	5.36	1.16
$N = 200$	EMM	33.89	28.88	22.21	30.35	24.75	17.05	14.55	7.68	1.63
$x_i \sim LN(0,1)$	II	23.01	15.80	6.99	19.38	12.40	4.92	15.36	8.37	2.04
$S = 10$	ALT	27.78	20.22	10.26	21.61	14.04	5.26	14.91	7.28	1.36
$N = 400$	EMM	29.69	24.41	17.55	26.33	20.43	12.99	14.53	7.88	1.87
$x_i \sim LN(0,1)$	II	21.35	14.12	5.92	18.55	11.62	4.62	15.43	8.52	2.17
$S = 10$	ALT	25.07	17.60	8.36	19.81	12.34	4.31	15.25	7.92	1.71
$N = 1000$	EMM	24.84	19.17	12.33	21.84	15.71	8.55	13.57	7.57	1.91
$x_i \sim LN(0,1)$	II	18.83	12.05	4.62	17.04	10.71	4.06	14.96	8.58	2.24
$S = 10$	ALT	21.49	14.51	6.33	17.27	10.50	3.35	14.87	8.17	2.10
$N = 2000$	EMM	21.62	16.08	9.36	19.12	13.21	6.34	12.94	7.11	1.87
$x_i \sim LN(0,1)$	II	17.12	10.67	3.78	15.86	9.65	3.48	14.47	8.26	2.21
$S = 10$	ALT	19.35	12.67	5.13	15.70	9.33	2.75	14.35	7.94	2.14

Table 8: TOBIT: Size of re-centered Wald statistics for the Efficient Method of Moments, Indirect Inference and Alternative Calibration Procedure. The null hypothesis is $\widehat{\beta} = \beta$.

		Asymptotic			Beatlestrap			Beatlestrap*		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$N = 200$	EMM	25.07	20.19	14.49	22.35	17.02	11.24	9.36	4.64	1.03
$x_i \sim LN(0,1)$	II	21.33	14.49	6.36	19.16	13.01	6.38	13.36	6.87	1.42
$S = 1$	ALT	25.26	18.01	8.91	19.64	13.37	6.70	13.52	6.84	1.57
$N = 400$	EMM	22.49	17.30	11.80	19.84	14.37	8.69	9.60	4.86	1.17
$x_i \sim LN(0,1)$	II	19.63	12.76	5.15	17.73	11.72	5.50	13.84	7.45	1.79
$S = 1$	ALT	22.62	15.43	6.94	16.93	10.79	4.68	12.72	6.48	1.48
$N = 1000$	EMM	19.85	14.62	8.98	17.05	11.61	6.12	9.64	4.79	1.18
$x_i \sim LN(0,1)$	II	17.70	11.19	4.06	16.16	10.19	4.35	14.40	7.95	1.99
$S = 1$	ALT	19.80	12.95	5.35	14.16	8.21	2.80	11.84	5.78	1.38
$N = 2000$	EMM	18.02	12.72	7.30	15.23	9.88	4.49	9.48	4.71	1.12
$x_i \sim LN(0,1)$	II	16.14	10.01	3.48	14.76	9.01	3.54	13.90	7.95	2.06
$S = 1$	ALT	17.69	11.46	4.58	12.38	6.60	1.87	10.95	5.29	1.12
$N = 200$	EMM	29.78	23.06	15.57	26.81	19.95	11.96	14.17	7.49	1.58
$x_i \sim LN(0,1)$	II	22.41	15.38	7.05	19.98	13.62	6.55	14.55	7.73	1.69
$S = 10$	ALT	26.16	18.91	9.82	20.12	13.08	5.44	12.94	5.98	0.89
$N = 400$	EMM	25.42	19.15	12.09	22.99	16.51	8.90	13.80	7.42	1.71
$x_i \sim LN(0,1)$	II	20.50	13.49	5.64	18.70	12.46	5.68	14.75	8.01	1.90
$S = 10$	ALT	23.48	16.15	7.52	18.13	11.06	3.80	13.32	6.31	1.09
$N = 1000$	EMM	21.15	15.04	8.40	18.88	12.66	5.87	12.68	6.87	1.70
$x_i \sim LN(0,1)$	II	17.89	11.34	4.15	16.91	10.88	4.67	14.48	8.16	2.11
$S = 10$	ALT	20.05	13.11	5.36	15.78	9.03	2.63	13.11	6.57	1.42
$N = 2000$	EMM	18.69	12.64	6.47	16.75	10.61	4.39	11.87	6.26	1.57
$x_i \sim LN(0,1)$	II	16.21	9.77	3.32	15.48	9.61	3.76	13.98	7.78	2.08
$S = 10$	ALT	17.94	11.28	4.22	14.19	7.87	2.14	12.71	6.46	1.53