Quantitative Finance and Economics

http://www.aimspress.com/journal/QFE

QFE, 3(1): 165–186.
DOI:10.3934/QFE.2019.1.165
Received: 03 January 2019
Accepted: 01 April 2019
Published: 08 April 2019

Research article

Wild multiplicative bootstrap for M and GMM estimators in time series

Francesco Audrino[1], **Lorenzo Camponovo**[2,*] **and Constantin Roth**[1]

[1] Department of Economics, University of St.Gallen, St.Gallen, Switzerland

[2] Department of Innovative Technologies, SUPSI, University of Applied Sciences and Arts of Southern Switzerland, Manno, Switzerland

* **Correspondence:** Email: lorenzo.camponovo@supsi.ch; Tel: +41918292020.

Abstract: We introduce a wild multiplicative bootstrap for M and GMM estimators in nonlinear models when autocorrelation structures of moment functions are unknown. The implementation of the bootstrap algorithm does not require any parametric assumptions on the data generating process. After proving its validity, we also investigate the accuracy of our procedure through Monte Carlo simulations. The wild bootstrap algorithm always outperforms inference based on standard first-order asymptotic theory. Moreover, in most cases the accuracy of our procedure is also better and more stable than that of block bootstrap methods. Finally, we apply the wild bootstrap approach to study the forecast ability of variance risk premia to predict future stock returns. We consider US equity from 1990 to 2010. For the period under investigation, our procedure provides significance in favor of predictability. By contrast, the block bootstrap implies ambiguous conclusions that heavily depend on the selection of the block size.

Keywords: M and GMM estimators; time series; wild bootstrap
JEL codes : C12, C13, C15

1. Introduction

Extremum estimators, such as M and generalized method of moments (GMM) estimators, have attained widespread applicability in various statistics and econometrics problems; see, e.g., Huber (1964) and Hansen (1982). The GMM provides a powerful tool for introducing statistical inference in several economic and financial models that are specified by some moment conditions; see, e.g., Hall (2005) for a review of the GMM. Unfortunately, recent research indicates that there are considerable issues with M and GMM estimators, in particular in their finite sample performance. More precisely, the asymptotic theory may provide very poor approximations of the sampling distribution of M and GMM estimators and related test statistics; see, e.g., the special issue of the *Journal of Business and*

Economic Statistics (Volume 14 (3), 1996).

To overcome this problem, a common approach consists of applying bootstrap methods. In time series settings, in the absence of parametric assumptions on the data generating process, the standard approach to bootstrapping is the block bootstrap; see, e.g., Hall (1985), Carlstein (1986), and Künsch (1989). Under strong regularity conditions on the data generating process and the general estimating functions, the block bootstrap may provide asymptotic refinements relative to standard first-order asymptotic theory; see, e.g., Hall and Horowitz (1996), Götze and Künsch (1996), Lahiri (1996), Andrews (2002), and Inoue and Shintani (2006). However, the magnitude of these improvements is not as large as that of the iid bootstrap or the parametric bootstrap. A main issue is that the independence of the blocks does not correctly mimic the structure of the true data generating process. Moreover, from a practical point of view, to ensure accurate approximations, the definition of the block bootstrap also requires an appropriate selection of the block size. The bootstrap literature proposes several ways of selecting this tuning parameter; see, e.g., Hall et al. (1995). Unfortunately, many of these approaches rely on asymptotic arguments, and the practical implementation in finite samples remains unclear.

In this paper, we introduce a wild multiplicative bootstrap for time series settings with unknown structures of the autocorrelation function that does not require the selection of block sizes, but depends on a different lag truncation tuning parameter. Unlike conventional bootstrap procedures proposed in the literature, in our algorithm we do not construct random samples by resampling from the observations. Rather, we propose to perturbate the general estimating functions using correlated innovations. More precisely, to generate the covariance matrix of these innovations, we apply the same kernel function principle adopted for the computation of the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix in the efficient GMM estimation criterion; see, e.g., Newey and West (1987) and Andrews (1991) for seminal works on HAC estimation, and Müller (2014) and Lazarus et al. (2018) for more recent studies on heteroskedasticity and autocorrelation robust (HAR) inference. By introducing this time series dependence, our approach is able to properly capture the autocorrelation of the true moments. Similar multiplicative bootstrap procedures have also been proposed in Minnier et al. (2011), Kline and Santos (2012), and Chernozhukov et al. (2014) in iid settings. Furthermore, dependent wild bootstrap methods for time series are also developed in Politis and Romano (1992), Shao (2010), Zhu and Li (2015) and Bücher and Kojadinovic (2016), among others. In contrast to these studies, instead of generating new random bootstrap observations by introducing correlated error terms, our bootstrap algorithm fixes the original observations and perturbates the (nonlinear) general estimating functions of M and GMM estimators.

In the Monte Carlo analysis, our bootstrap method always outperforms inference based on standard first-order asymptotic theory. Furthermore, the accuracy of our procedure is in general superior to that of block bootstrap methods, and less sensitive to the selection of tuning parameters. Finally, we also consider a real data application. Using the wild multiplicative bootstrap and the block bootstrap, we study the ability of variance risk premia to predict future returns. We consider US equity data from 1990 to 2010 from Shiller (2000) and Bollerslev et al. (2009). For the period under investigation, the wild multiplicative bootstrap provides significance in favor of predictability. By contrast, the block bootstrap implies ambiguous conclusions that heavily depend on the selection of the block size. The reason for these divergent conclusions could be related to the lack of robustness of the block bootstrap in the presence of anomalous observations; see, e.g., Singh (1998), Salibian-Barrera and Zamar (2002) and Camponovo et al. (2012, 2015) for more details on the robustness properties of resampling methods. Indeed, the period under investigation is characterized by several unusual observations, linked to the recent credit crisis, that may easily corrupt inference based on block bootstrap procedures.

The rest of the paper is organized as follows. In Section 2, we introduce M and GMM estimators. In Section 3, we present the wild bootstrap algorithm and prove its validity. In Section 4, we study the accuracy of our approach and block bootstrap procedures through Monte Carlo simulations. In Section 5, we consider the real data application. Finally, Section 6 concludes. A proof and assumptions related to the main theorem about the bootstrap validity discussed in Section 3 are presented in the Appendix.

2. Extremum estimators

In this section, we introduce M and GMM estimators. As noted in Andrews (2002), M estimators can be written as GMM estimators. However, because of the different identification conditions, we prefer to introduce these classes of estimators separately; see, e.g., Andrews (2002).

2.1. M estimators

Let $(X_1, \ldots, X_n)$ be a sample from a process $\mathcal{X} = \{X_t, t \in \mathbb{Z}\}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, where $X_t \in \mathbb{R}^{d_x}$. Furthermore, let $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ be an unknown parameter. We consider M estimators $\hat{\theta}_n$ of θ defined as

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta \subset \mathbb{R}^{d_\theta}} \frac{1}{n} \sum_{t=1}^{n} \rho(X_t, \theta), \tag{1}$$

where $\rho : \mathbb{R}^{d_x} \times \mathbb{R}^{d_\theta} \to \mathbb{R}$ is a known smooth function. Examples of M estimators include maximum likelihood, quasi-maximum likelihood, and least squares estimators, among others; see, e.g., Andrews (2002).

Let θ_0 denote the true value of the unknown parameter θ. Then, under some regularity conditions, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges weakly to a normally distributed random vector with mean 0 and covariance matrix $V_0 = D_0^{-1}\Omega_0 D_0^{-1}$, where $D_0 = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta'}\rho(X_t, \theta_0)\right]$, and $\Omega_0 = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial}{\partial\theta}\rho(X_i, \theta_0)\frac{\partial}{\partial\theta}\rho(X_j, \theta_0)'\right]$. Therefore, the normal distribution provides valid approximations of the sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Unfortunately, the asymptotic distribution may work poorly in finite samples. To overcome this problem, in Section 3 we analyze bootstrap approximations.

2.2. GMM estimators

For simplicity, we adopt the same notation introduced in the previous section. Let $(X_1, \ldots, X_n)$ be a sample from a process $\mathcal{X} = \{X_t, t \in \mathbb{Z}\}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, where $X_t \in \mathbb{R}^{d_x}$. Consider the moment condition $E[g(X_t, \theta_0)] = 0$, where $g(\cdot, \cdot)$ is an $\mathbb{R}^{d_g}$-valued function with $d_g \geq d_\theta$, and θ_0 denotes the true value of the unknown parameter $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$. We focus on GMM estimators $\hat{\theta}_n$ of θ_0 defined as

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta \subset \mathbb{R}^{d_\theta}} \left(\frac{1}{n}\sum_{t=1}^{n} g(X_t, \theta)\right)' W_n \left(\frac{1}{n}\sum_{t=1}^{n} g(X_t, \theta)\right), \tag{2}$$

where W_n is a positive-definite symmetric matrix. Examples of matrix W_n also include the efficient weighting matrix $W_n = \left(\Omega_n(\bar{\theta}_n)\right)^{-1}$, where $\bar{\theta}_n$ is a preliminary estimator of θ_0,

$$\Omega_n(\theta) = \sum_{i=-(n-1)}^{n-1} k(i/h)\Gamma_i(\theta), \tag{3}$$

$$\Gamma_i(\theta) = \frac{1}{n}\sum_{t=1}^{n-i} g(X_t,\theta)g(X_{t+i},\theta)', \tag{4}$$

$k(\cdot)$ is a kernel function, and h is the lag truncation.

Suppose that W_n converges in probability to a non-random positive-definite symmetric matrix W_0. Then, under some further regularity conditions, the GMM statistic $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges weakly to a normally distributed random vector with mean 0 and covariance matrix $V_0 = (D_0'W_0D_0)^{-1}D_0'W_0\Omega_0W_0D_0(D_0'W_0D_0)^{-1}$, where $D_0 = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}g(X_t,\theta_0)\right]$, and $\Omega_0 = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g(X_i,\theta_0)g(X_j,\theta_0)'\right]$. Therefore, in this case as well the normal distribution provides valid approximations of the sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Alternatively, in the next section we analyze bootstrap approximations.

3. Bootstrap approximations

In Section 3.1, we briefly present the block bootstrap approach, while in Section 3.2 we introduce our wild multiplicative bootstrap procedure.

3.1. Block bootstrap

Since in our setting we do not have parametric information on the data generating process, the standard approach to bootstrapping is the block bootstrap; see, e.g., Carlstein (1986). More precisely, given the observation sample $(X_1,\ldots,X_n)$, consider the non-overlapping blocks $(X_{im+1},\ldots,X_{(i+1)m})$, $i = 0,\ldots,n/m-1$, of size m, where for simplicity we assume $n/m = b \in \mathbb{N}$. The non-overlapping block bootstrap constructs random samples $(X_1^\star,\ldots,X_n^\star)$ by selecting b non-overlapping blocks with replacement. Let $\hat{\theta}_n^\star$ be the bootstrap M or GMM estimator solution of (1) or (2), respectively, based on the bootstrap sample $(X_1^\star,\ldots,X_n^\star)$. Then, the non-overlapping block bootstrap approximates the sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ with the conditional distribution of $\sqrt{n}(\hat{\theta}_n^\star - \hat{\theta}_n)$ given the observations $(X_1,\ldots,X_n)$; see also Künsch (1989) for the definition of block bootstrap approximations based on overlapping blocks.

Under strong regularity conditions on the data generating process and on the general estimating functions, the block bootstrap may provide asymptotic refinements relative to standard first-order asymptotic theory; see, e.g., Inoue and Shintani (2006). However, to ensure accurate approximations of the sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$, the definition of the block bootstrap also requires an appropriate selection of the block size. The bootstrap literature proposes several ways of selecting m; see, e.g., Hall et al. (1995). Unfortunately, many of these approaches rely on asymptotic arguments, and the practical implementation in finite samples remains unclear. In the next section, we introduce a wild multiplicative bootstrap approach that does not require the selection of block sizes.

3.2. Wild multiplicative bootstrap

First, we introduce the wild multiplicative bootstrap algorithm. In a second step, we clarify the key rationale of our approach. Finally, we prove the validity of the wild bootstrap approximation.

Algorithm 1. Wild Multiplicative Bootstrap.

(i) Compute either the M or the GMM estimators $\hat{\theta}_n$ defined in (1) and (2), respectively.
(ii) Generate a random sample $(e_1, \ldots, e_n)$ of positive correlated observations with following properties $E[e_t|(X_1, \ldots, X_n)] = 1$ and $Cov(e_t, e_{t+i}|(X_1, \ldots, X_n)) = k(i/h)$, where $k(\cdot)$ is an appropriate kernel function, and h is the lag truncation parameter. For $t = 1, \ldots, n$, let

$$\rho^*(X_t, \theta) = \rho(X_t, \theta)e_t \tag{5}$$

$$g^*(X_t, \theta) = \left(g(X_t, \theta) - \frac{1}{n}\sum_{i=1}^{n} g(X_i, \hat{\theta}_n)\right)e_t. \tag{6}$$

(iii) Compute either the wild multiplicative bootstrap M or GMM estimators $\hat{\theta}_n^*$ defined as, respectively,

$$\hat{\theta}_n^* = \arg\min_{\theta\in\Theta\subset\mathbb{R}^{d_\theta}} \frac{1}{n}\sum_{t=1}^{n} \rho^*(X_t, \theta), \tag{7}$$

$$\hat{\theta}_n^* = \arg\min_{\theta\in\Theta\subset\mathbb{R}^{d_\theta}} \left(\frac{1}{n}\sum_{t=1}^{n} g^*(X_t, \theta)\right)' W_n \left(\frac{1}{n}\sum_{t=1}^{n} g^*(X_t, \theta)\right). \tag{8}$$

(iv) Repeat steps (ii)-(iii) B times, where B is a large number. The empirical distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ approximates the sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

Unlike conventional bootstrap procedures proposed in the literature, in our approach we do not construct random samples by resampling from the observations. Rather, in step (ii) of Algorithm 1, we perturbate the general estimating functions using correlated innovations. By introducing this time series dependence, our bootstrap method is able to properly capture the autocorrelation of the true moments. In equation (8), we compute the wild multiplicative bootstrap GMM estimator. To this end, as in Hall and Horowitz (1996) and Andrews (2002), we recenter the bootstrap moment by subtracting off $\frac{1}{n}\sum_{i=1}^{n} g(X_i, \hat{\theta}_n)$. The recentering ensures that the bootstrap moment $E^*[\frac{1}{n}\sum_{t=1}^{n} g^*(X_t, \theta)] = 0$, when $\theta = \hat{\theta}_n$. In the next theorem, we prove the validity of our bootstrap algorithm.

Theorem 3.1. *Let Assumptions 6.1-6.3 in the Appendix hold. Then,*

(i) For M estimators, the conditional law of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ *converges weakly to a normal distribution with mean* 0 *and covariance matrix* $V_0 = D_0^{-1}\Omega_0 D_0^{-1}$*, as* $n \to \infty$.
(ii) For GMM estimators, the conditional law of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ *converges weakly to a normal distribution with mean* 0 *and covariance matrix* $V_0 = (D_0' W_0 D_0)^{-1} D_0' W_0 \Omega_0 W_0 D_0 (D_0' W_0 D_0)^{-1}$*, as* $n \to \infty$.

Theorem 3.1 shows that both for M and GMM estimators, the wild multiplicative bootstrap algorithm provides a valid method for approximating the sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

Remark 1. To verify the validity of the wild bootstrap approximation, in the proof of Theorem 3.1 first we show that $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ minimizes a particular random process. Then, we compute the limit of this random process. To this end, we consider the conditional probability given the sample $(X_1, \ldots, X_n)$, and compute the limit by successively conditioning on a sequence of samples, as $n \to \infty$. Suppose that $\frac{1}{n}\sum_{t=1}^{n} g(X_t, \hat{\theta}_n) = 0$. Then, note that

$$Var\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g^*(X_t, \hat{\theta}_n)\middle|(X_1, \ldots, X_n)\right) = \sum_{i=-(n-1)}^{n-1} k(i/h)\Gamma_i(\hat{\theta}_n) \tag{9}$$

where $\Gamma_i(\theta) = \frac{1}{n}\sum_{t=1}^{n-i} g(X_t, \theta)g(X_{t+i}, \theta)'$, which converges in probability to Ω_0 under Assumptions 6.1-6.3. Finally, we compute the limit, and apply results in Geyer (1994).

Remark 2. In Algorithm 1, we can observe that the definition of the wild multiplicative bootstrap does not require the selection of block sizes m. However, the multiplicative bootstrap still requires the selection of the lag truncation tuning parameter h. As highlighted in our Monte Carlo analysis in Section 4, the wild multiplicative bootstrap is less sensitive to the selection of the tuning parameter h than is the block bootstrap to the selection of the block size m, yielding more stable results; see, e.g., Shao (2010) for similar empirical findings.

Remark 3. Suppose that in equation (2) we adopt the optimal weighting matrix $W_n = \left(\Omega_n(\bar{\theta}_n)\right)^{-1}$. Then, the natural selection of the weighting matrix in equation (8) in the wild bootstrap algorithm is given by $W_n = \left(\Omega_n^*(\bar{\theta}_n^*)\right)^{-1}$, where

$$\Omega_n^*(\theta) = \sum_{i=-(n-1)}^{n-1} k(i/h)\Gamma_i^*(\theta), \tag{10}$$

$$\Gamma_i^*(\theta) = \frac{1}{n}\sum_{t=1}^{n-i} \bar{g}^*(X_t, \theta)\bar{g}^*(X_{t+i}, \theta)', \tag{11}$$

$$\bar{g}^*(X_t, \theta) = \left(g(X_t, \theta) - \frac{1}{n}\sum_{i=1}^{n} g(X_i, \hat{\theta}_n)\right)(e_t - 1), \tag{12}$$

and $\bar{\theta}_n^*$ is a preliminary bootstrap GMM estimator. Note that since $E[e_t] = 1$, in equation (12) we replace $g^*(X_t, \theta)$ with $\bar{g}^*(X_t, \theta) = \left(g(X_t, \theta) - \frac{1}{n}\sum_{i=1}^{n} g(X_i, \hat{\theta}_n)\right)(e_t - 1)$.

Remark 4. Using similar arguments adopted in the proof of Theorem 3.1, we can show that $\Omega_n^*(\bar{\theta}_n^*)$ converges in conditional probability to Ω_0, as $n \to \infty$. Similarly, we can easily introduce consistent bootstrap estimators of D_0. These results indicate that the wild multiplicative bootstrap may also provide valid approximations of the sampling distribution of asymptotically pivotal statistics such as t-statistics or J-statistics.

Remark 5. As correctly pointed out by a Referee, in step (iii) of Algorithm 1, instead of re-estimating the unknown parameter of interest, we could simply perturb the estimating equations and use directly for the construction of confidence intervals. However, this approach is not investigated in the Monte Carlo analysis, and left for future research.

Remark 6. Our wild multiplicate bootstrap has some analogies with the (multiplier) bootstrap methods proposed in Minnier et al. (2011), Kline and Santos (2012), Chernozhukov et al. (2014), Politis and Romano (1992), Shao (2010), Zhu and Li (2015), Zhu and Ling (2015), Bcher and Kojadinovic (2016), and Zhu (2016, 2019). However, it is important to highlight that our approach is conceptually different from the procedures developed in previous studies, and in particular from the wild dependent bootstrap introduced in Shao (2010). Specifically, Shao (2010) proposes to generate new random bootstrap observations by introducing correlated error terms. On the other hand, in our bootstrap algorithm we fix the original observations, and propose to perturbate the general estimating functions in a multiplicative way using correlated innovations. Therefore, the wild dependent bootstrap proposed in Shao (2010) cannot be applied to the simplified version of an asset pricing model proposed in our Section 4.2. On the other hand, our approach works in this setting as well.

Remark 7. Inference provided by conventional (block) bootstrap procedures may be easily inflated by a small fraction of anomalous observations in the data; see, e.g., Singh (1998), Salibian-Barrera and Zamar (2002), and Camponovo et al. (2012, 2015). Intuitively, this feature is explained by the overly high fraction of anomalous data that is often simulated by conventional block bootstrap procedures compared to the actual fraction of anomalous observations in the original data. On the other hand, since the wild multiplicative bootstrap does not construct random samples by resampling from the observations, our procedure ensures a desirable accuracy and stability even in the presence of contaminated data. Indeed, preliminary Monte Carlo simulations in the predictive regression setting of Section 4.1 with a small fraction of additive outlying observations confirm a better stability of the wild bootstrap with respect to the block bootstrap. We conjecture that using the breakdown point theory developed in Camponovo et al. (2015), it is possible to establish the superior robustness properties of the wild multiplicative bootstrap with respect to the moving block bootstrap. A complete analysis of the robustness properties of the wild bootstrap is left for future research.

4. Monte carlo simulations

In this section, we study through Monte Carlo simulations the accuracy of our wild bootstrap approach. In Section 4.1, we present the results for a predictive regression model with different form of heteroskedasticity. Subsequently, in Section 4.2, we consider the simplified version of an asset pricing model analyzed in Hall and Horowitz (1996). Finally, in Section 4.3, we analyze a regression model with a time series structure as proposed in Inoue and Shintani (2006).

We use the Parzen kernel in order to construct the covariance matrix of the correlated innovations in step (ii) of the wild multiplicative bootstrap algorithm. As in other contexts, the choice of the kernel has only a marginal and negligible impact on the accuracy of the results. The number of bootstrap replications is $B = 999$ and the nominal coverage probability is 90%. Unreported Monte Carlo simulations for other coverage probabilities, e.g. 95%, produced similar results and confirmed the findings illustrated in the next subsections. For simplicity, in Sections 4.2 and 4.3 we focus on GMM estimators with identity matrix as the weighting matrix. Furthermore, we construct confidence intervals for the unknown parameter of interest θ_0 using approximations of the sampling distribution of the non-studentized statistic $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Unreported empirical results with optimal weighting matrix and based on studentized statistics are qualitatively very similar. However, in this case the wild multiplicative bootstrap seems to be slightly more sensitive to the selection of the tuning parameter h.

The source of this instability may be related to the estimation of the optimal weighting matrix; see, e.g., Altonji and Segal (1996) for similar computational issues.

Finally, for brevity, we report results only for our bootstrap approach and the non-overlapping block bootstrap. Monte Carlo investigations with alternative block bootstrap procedures such as the stationary block bootstrap and the stationary block-of-blocks bootstrap based on the resampling of the estimating functions using the block bootstrap produce similar results to those shown for the non-overlapping block bootstrap; robustness checks are available from the authors upon request. These findings are not too surprising since stationary block bootstrap methods cannot address the problem of breaking up the dependence structure either and since the block-of-blocks bootstrap also mitigates the problem only at the break points of the subsamples.

4.1. Predictive regression model

We consider the predictive regression model,

$$Y_t = \alpha + \theta Z_{t-1} + U_t, \tag{13}$$

$$Z_t = \mu + \rho Z_{t-1} + V_t, \tag{14}$$

where, for $t = 1, \ldots, n$, Y_t denotes the dependent variable at time t, and Z_{t-1} is assumed to predict Y_t. The parameters $\alpha \in \mathbb{R}$ and $\mu \in \mathbb{R}$ are the unknown intercepts of the linear regression model and the autoregressive model, respectively, $\theta \in \mathbb{R}$ is the unknown parameter of interest, $\rho \in \mathbb{R}$ is the unknown autoregressive coefficient, and $U_t \in \mathbb{R}$, $V_t \in \mathbb{R}$ are error terms.

In the first exercise, we generate 5000 Monte Carlo samples of size $n = 180$ according to model (13)–(14), with $U_t \sim N(0,1)$, $V_t \sim N(0,1)$, $\alpha_0 = \mu_0 = 0$, $\rho_0 = 0.3, 0.5, 0.7$, and $\theta_0 = 0$. We estimate the unknown parameter of interest through the least squares estimators,

$$(\hat{\alpha}_n, \hat{\theta}_n) = \arg\min_{(\alpha,\theta)} \frac{1}{n} \sum_{t=1}^{n-1} (Y_{t+1} - \alpha - \theta Z_t)^2. \tag{15}$$

We construct 90% confidence intervals for θ_0 using the block bootstrap with block sizes $m = 2, 5, 10, 15, 20$, and the wild multiplicative bootstrap with lag truncation $h = 2, 5, 10, 15, 20$. Table 1 reports the empirical coverages.

In Table 1, we can observe that both bootstrap procedures provide empirical coverages quite close to the nominal coverage probability 90%. However, the wild multiplicative bootstrap seems to be less sensitive to the selection of the tuning parameter h than the block bootstrap is to the selection of the block size m. For instance, when $\rho_0 = 0.3$, the empirical coverages of the wild bootstrap range from 90.6 to 91.4 for $h = 5$ and $h = 20$, respectively. On the other hand, in the same setting the empirical coverages of the block bootstrap range from 91.3 to 87.2 for $m = 5$ and $m = 20$, respectively. In particular, in the lines "Variation Block" and "Variation Wild" we report the maximal difference between empirical coverages implied by the block bootstrap and the wild bootstrap for different values of the block size and the lag truncation tuning parameter, respectively. We can observe that the variation for the block bootstrap is always larger than 4.5%. On the other hand, for the wild bootstrap the difference is below 2.0%.

In the second exercise, we consider the same parameter selection as in the previous study. However, in this case the error terms are heteroskedastic and correlated. More precisely, let $\sigma_t^2 = \frac{1}{t-1} \sum_{i=1}^{t-1} Z_i^2$.

Table 1. Predictive regression model. Empirical coverage probabilities for the predictive regression model analyzed in Section 4.1. We consider first-order asymptotic theory, the block bootstrap with block size $m = 2, 5, 10, 15, 20$, and the wild multiplicative bootstrap with lag truncation $h = 2, 5, 10, 15, 20$. The degree of persistence is $\rho_0 = 0.3, 0.5, 0.7$. The sample size is $n = 180$, and the nominal coverage probability is 90%. The error terms are standard normal distributed. In the lines "Variation Block" and "Variation Wild" we report the maximal difference between empirical coverages implied by the block bootstrap and the wild bootstrap for different values of the block size and the lag truncation tuning parameter, respectively.

ρ_0					0.3	0.5	0.7
	Asymptotic theory				89.1	88.9	88.9
	Block	m	=	2	92.4	92.3	91.9
		m	=	5	91.3	90.5	90.7
		m	=	10	89.3	89.1	89.5
		m	=	15	88.3	88.4	88.4
		m	=	20	87.2	87.5	86.6
	Variation Block				5.2	4.8	5.3
	Wild	h	=	2	90.4	90.3	90.6
		h	=	5	90.6	90.4	90.8
		h	=	10	90.8	90.9	91.2
		h	=	15	91.2	91.8	92.0
		h	=	20	91.4	92.1	92.5
	Variation Wild				1.0	1.8	1.9

Then, for the distribution of the error terms we consider following model,

$$V_t \sim N(0, \sigma_t^2), \tag{16}$$

$$U_t = -0.5V_t + E_t, \tag{17}$$

where $E_t \sim N(0, 1)$. In Table 2, we report the empirical coverages using the block bootstrap and the wild multiplicative bootstrap. Also in this case, we can observe that both bootstrap procedures provide empirical coverages quite close to the nominal coverage probability 90%. However, the wild multiplicative bootstrap is again less sensitive to the selection of the tuning parameter h than the block bootstrap is to the selection of the block size m. Indeed, in the lines "Variation Block" and "Variation Wild" we note that the maximal variation for the block bootstrap is always larger than 5%. On the other hand, for the wild bootstrap the difference is always below 2.0%.

In the last exercise, we study the power properties of the bootstrap procedures. To this end, we generate 5000 Monte Carlo samples of size $n = 180$ according to model (13)–(14), with $U_t \sim N(0, 1)$, $V_t \sim N(0, 1)$, $\alpha_0 = \mu_0 = 0$, $\rho_0 = 0.3, 0.5, 0.7$, and $\theta_0 \in [0, 3/\sqrt{n}]$. Finally, using the block and the wild bootstrap, we test the null hypothesis $\mathcal{H}_0 : \theta_0 = 0$ versus $\mathcal{H}_1 : \theta_0 \neq 0$, for $\theta_0 \in [0, 3/\sqrt{n}]$. Figure 1 reports the power curves for different selections of the block size m and the lag truncation h.

In Figure 1, we can observe that both bootstrap procedures have quite similar power properties. When $\theta_0 = 0$, the empirical rejection frequencies of the null hypothesis are very close to the

Table 2. Predictive regression model. Empirical coverage probabilities for the predictive regression model analyzed in Section 4.1. We consider first-order asymptotic theory, the block bootstrap with block size $m = 2, 5, 10, 15, 20$, and the wild multiplicative bootstrap with lag truncation $h = 2, 5, 10, 15, 20$. The degree of persistence is $\rho_0 = 0.3, 0.5, 0.7$. The sample size is $n = 180$, and the nominal coverage probability is 90%. The error terms are heteroskedastic and correlated. In the lines "Variation Block" and "Variation Wild" we report the maximal difference between empirical coverages implied by the block bootstrap and the wild bootstrap for different values of the block size and the lag truncation tuning parameter, respectively.

ρ_0					0.3	0.5	0.7
	Asymptotic theory				88.4	88.3	88.1
	Block	m	=	2	90.6	90.4	90.5
		m	=	5	88.8	88.6	88.3
		m	=	10	87.9	87.8	87.5
		m	=	15	86.8	86.6	86.5
		m	=	20	85.3	85.2	85.0
	Variation Block				5.3	5.2	5.5
	Wild	h	=	2	89.8	89.8	89.7
		h	=	5	90.4	90.4	90.5
		h	=	10	90.6	90.7	91.0
		h	=	15	91.3	91.4	91.4
		h	=	20	91.5	91.4	91.6
	Variation Wild				1.7	1.6	1.9

significance level 10%. As expected, when $\theta_0 \neq 0$, the empirical rejection frequencies increase. However, in this case as well we can observe that the wild multiplicative bootstrap seems to be less sensitive to the selection of the tuning parameter h than the block bootstrap for the selection of the block size m. Given that power results in the next two settings are perfectly in line with those presented for predictive regressions, for the sake of brevity we do not report them in detail.

4.2. Hall and horowitz (1996)

We consider the example introduced in Hall and Horowitz (1996), who introduce a simplified version of an asset pricing model defined by the moment conditions

$$E[g(X,\theta_0)] = E\left[\begin{pmatrix} 1 \\ X_2 \end{pmatrix}(\exp(\mu - \theta_0(X_1 + X_2) + 3X_2) - 1)\right] = 0, \tag{18}$$

where $X = (X_1, X_2)'$, $\theta_0 = 3$ is the parameter of interest, μ is a known normalization constant, and X_1, X_2 are independent random scalars. In particular, we consider the case where $X_1 \sim N(0, 0.2^2)$ and X_2 follows a strictly stationary AR(1) process with no intercept, first-order serial correlation coefficient ρ, and standard normal innovations.

In Table 3, we report empirical coverage probabilities of 90% confidence intervals for parameter θ_0 based on 5000 Monte Carlo samples of size $n = 48, 96$, and 256. For the first-order serial correlation

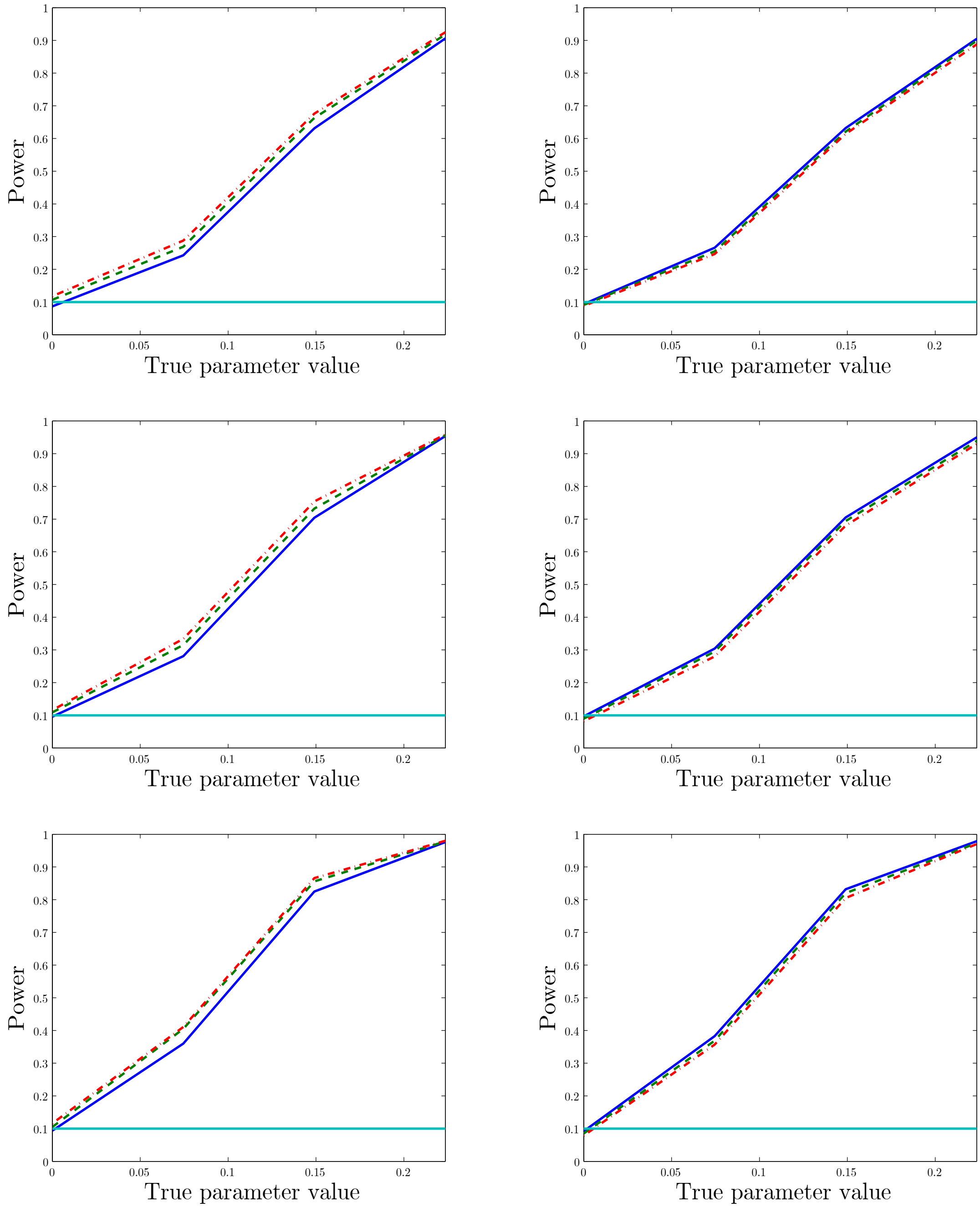

Figure 1. Power curves predictive regression model. We plot the proportion of rejections of the null hypothesis $\mathcal{H}_0 : \theta_0 = 0$ versus $\mathcal{H}_1 : \theta_0 \neq 0$, for $\theta_0 \in [0, 3/\sqrt{n}]$. In the left panels, we consider the block bootstrap with block size $m = 5$ (solid line), $m = 10$ (dashed line) and $m = 15$ (dash-dotted line). In the right panels, we consider the wild multiplicative bootstrap with lag truncation $h = 5$ (solid line), $h = 10$ (dashed line) and $h = 15$ (dash-dotted line). From the top to the bottom, the degree of persistence is $\rho_0 = 0.3, 0.5, 0.7$, respectively. The sample size is $n = 180$.

coefficient in the data generating process, we consider the cases $\rho_0 = 0.3, 0.5, 0.7$. We construct the confidence intervals using first-order asymptotic theory and bootstrap approximations. More precisely, for the wild bootstrap and the block bootstrap, we consider as lag truncation and block sizes $h = m = 2, 4, 6, 8, 10, 12$, $h = m = 4, 8, 12, 16, 20$ and $h = m = 4, 8, 16, 24, 32$, for $n = 48$, $n = 96$, and $n = 256$, respectively. The values we consider are similar to those in Hall and Horowitz (1996) who focused on block sizes $m = 5, 10, 20$ for $n = 50, 100$.

The results for $\rho_0 = 0.3, 0.5, 0.7$ are qualitatively very similar. The first observation we make is that the wild multiplicative bootstrap significantly outperforms inference based on standard first-order asymptotic theory for all values of h we consider. The second observation is that the accuracy of both the wild bootstrap and the block bootstrap depends on the choice of the parameters h and m, respectively. Furthermore, for the same values of h and m, we see that the wild bootstrap is closer to the nominal coverage probability 90% for most of the settings. Finally, when comparing the wild and block bootstrap, we also observe that the wild bootstrap is much less sensitive to the choice of h than the block bootstrap is to the choice of m. Indeed, in the lines "Variation Block" and "Variation Wild" we note that the maximal difference between empirical coverages implied by the block bootstrap is always larger than 5%. On the other hand, the maximal difference for the wild bootstrap is around 1%. As mentioned above, there is no clear method to determine the block size in finite samples, which makes this dependence problematic in practice. The higher stability of the wild bootstrap with respect to the lag truncation h is therefore a major advantage in practice, as the procedure is quite accurate for a wide range of values, unlike the block bootstrap.

4.3. Inoue and Shintani (2006)

In this section, we consider the linear regression model

$$Y_t = \theta Z_t + U_t, \tag{19}$$

where $Y_t \in \mathbb{R}$, $\theta \in \mathbb{R}$, and the disturbance and the regressors are generated according to the following autoregressive processes with common ρ,

$$\begin{aligned} U_t &= \rho U_{t-1} + V_{1t}, & (20)\\ Z_t &= \rho Z_{t-1} + V_{2t}, & (21) \end{aligned}$$

with $V_t = (V_{1t}, V_{2t})' \sim N(0, I_2)$. We generate 5000 samples according to this model with $\theta_0 = 0$, $\rho_0 = 0.3, 0.5, 0.7$, and $n = 48, 96, 256$. Note that in this setting, the unknown parameter of interest satisfies the moment conditions

$$E[g(X_t, \theta_0)] = E\begin{bmatrix} (Y_t - Z_t\theta_0) \\ (Y_t - Z_t\theta_0)Z_t \\ (Y_t - Z_t\theta_0)Z_{t-1} \\ (Y_t - Z_t\theta_0)Z_{t-2} \end{bmatrix} = 0, \tag{22}$$

where $X_t = (Y_t, Z_t, Z_{t-1}, Z_{t-2})'$. Again, we construct 90% confidence intervals using first-order asymptotic theory and bootstrap approximations. More precisely, for the wild bootstrap and the block bootstrap, we consider as lag truncation and block sizes $h = m = 2, 4, 6, 8, 10, 12$,

Table 3. Hall and horowitz (1996). Empirical coverage probabilities of 90% confidence intervals based on 5000 Monte Carlo samples for three sample sizes $n = 48, 96, 256$. Results are reported for the first-order asymptotic theory, the block bootstrap with different values of the block size parameter m, and our wild bootstrap algorithm with different values of the lag truncation h. In the lines "Variation Block" and "Variation Wild" we report the maximal difference between empirical coverages implied by the block bootstrap and the wild bootstrap for different values of the block size and the lag truncation tuning parameter, respectively.

ρ_0					0.3	0.5	0.7
$n = 48$	Asymptotic theory				65.2	67.1	68.2
	Block	m	=	2	90.5	92.3	92.8
		m	=	4	89.7	91.5	91.8
		m	=	6	88.0	89.6	91.0
		m	=	8	85.1	87.8	89.4
		m	=	10	85.1	87.0	87.9
		m	=	12	79.8	82.5	84.5
	Variation Block				10.7	9.8	8.3
	Wild	h	=	2	92.5	92.8	91.8
		h	=	4	92.4	92.6	91.6
		h	=	6	92.4	92.8	92.2
		h	=	8	92.3	92.6	92.7
		h	=	10	92.0	92.5	92.6
		h	=	12	91.8	92.3	92.6
	Variation Wild				0.7	0.5	1.0
$n = 96$	Asymptotic theory				64.2	64.4	67.3
	Block	m	=	4	89.4	90.8	92.3
		m	=	8	87.6	89.6	91.2
		m	=	12	84.8	87.3	89.2
		m	=	16	83.1	84.3	87.8
		m	=	20	84.0	84.6	87.2
	Variation Block				5.4	6.2	5.1
	Wild	h	=	4	90.9	92.2	91.2
		h	=	8	90.9	92.2	92.0
		h	=	12	90.9	92.4	92.4
		h	=	16	91.3	92.2	92.1
		h	=	20	91.3	92.2	92.0
	Variation Wild				0.4	0.2	1.2
$n = 256$	Asymptotic theory				63.9	63.5	64.0
	Block	m	=	4	90.3	90.4	90.9
		m	=	8	89.0	88.3	90.0
		m	=	16	87.8	86.5	88.8
		m	=	24	87.6	85.9	88.5
		m	=	32	84.5	83.1	85.3
	Variation Block				5.8	7.3	5.6
	Wild	h	=	4	91.0	90.6	90.5
		h	=	8	91.0	90.5	90.4
		h	=	16	91.0	89.9	90.6
		h	=	24	90.9	89.9	90.8
		h	=	32	90.8	89.9	90.9
	Variation Wild				0.2	0.7	0.4

$h = m = 4, 8, 12, 16, 20$ and $h = m = 4, 8, 16, 24, 32$, for $n = 48$, $n = 96$, and $n = 256$, respectively. The empirical coverage probabilities are summarized in Table 4.

Table 4. Inoue and shintani (2006). Empirical coverage probabilities of 90% confidence intervals based on 5000 Monte Carlo samples for three sample sizes $n = 48, 96, 256$. Results are reported for the first-order asymptotic theory, the block bootstrap with different values of the block size parameter m, and our wild bootstrap algorithm with different values of the lag truncation h. In the lines "Variation Block" and "Variation Wild" we report the maximal difference between empirical coverages implied by the block bootstrap and the wild bootstrap for different values of the block size and the lag truncation tuning parameter, respectively.

ρ_0					0.3	0.5	0.7
$n = 48$	Asymptotic theory				76.2	67.4	55.7
	Block	m	=	2	86.2	83.6	78.2
		m	=	4	85.6	81.7	76.8
		m	=	6	84.1	81.0	79.2
		m	=	8	82.0	79.8	77.7
		m	=	10	83.0	80.7	79.2
		m	=	12	76.9	76.1	74.4
	Variation Block				9.3	7.5	4.8
	Wild	h	=	2	89.3	85.8	85.9
		h	=	4	89.3	85.8	85.8
		h	=	6	90.0	88.2	85.6
		h	=	8	90.7	89.3	87.8
		h	=	10	91.3	89.9	89.8
		h	=	12	91.7	90.7	90.2
	Variation Wild				2.4	4.9	4.3
$n = 96$	Asymptotic theory				79.4	70.4	62.5
	Block	m	=	4	86.5	82.6	79.2
		m	=	8	84.8	82.1	80.4
		m	=	12	82.8	80.6	79.7
		m	=	16	80.6	78.4	78.2
		m	=	20	81.7	80.3	79.4
	Variation Block				5.9	4.2	2.2
	Wild	h	=	4	88.5	86.9	86.3
		h	=	8	89.8	88.2	87.6
		h	=	12	90.3	89.2	89.6
		h	=	16	91.0	90.0	90.9
		h	=	20	91.2	90.6	91.9
	Variation Wild				2.7	3.7	5.6
$n = 256$	Asymptotic theory				81.8	74.7	67.2
	Block	m	=	4	87.6	86.9	83.7
		m	=	8	86.8	85.9	82.5
		m	=	16	85.5	85.0	82.6
		m	=	24	85.7	85.2	83.1
		m	=	32	82.5	82.4	80.6
	Variation Block				5.1	4.5	3.1
	Wild	h	=	4	88.6	88.7	88.6
		h	=	8	88.9	89.0	88.5
		h	=	16	89.0	89.8	88.5
		h	=	24	89.5	90.5	89.6
		h	=	32	90.0	91.1	90.7
	Variation Wild				1.4	2.4	2.1

In this setting as well, the wild multiplicative bootstrap clearly outperforms inference based on standard first-order asymptotic theory, regardless of the choice of the lag truncation. The higher precision of the wild bootstrap with respect to that of the block bootstrap when using the same parameter values is even more evident than in the previous setting. Moreover, results again show that the accuracy of the block bootstrap is much more sensitive to the block size parameter than is that of the wild bootstrap with respect to the lag truncation parameter, even for quite large samples and low persistence.

5. Real data application

In this section, we study the forecast ability of variance risk premia to predict future stock returns. Recently, a large number of studies have investigated whether stock returns can be predicted by economic variables such as the price-dividend ratio, the interest rate or the variance risk premia; see, e.g., Rozeff (1984), Fama and French (1988), Campbell and Shiller (1988), Nelson and Kim (1993), Campbell and Yogo (2006), Jansson and Moreira (2006), Polk et al. (2006), and Bollerslev et al. (2009).

In this empirical analysis, we consider monthly S&P 500 index data (1871–2010) from Shiller (2000). We define the one-period real total return as

$$R_t = (P_t + d_t)/P_{t-1}, \tag{23}$$

where P_t is the end-of-month real stock price and d_t is the real dividends paid during month t. Finally, we consider the predictive regression model,

$$\frac{1}{k}\ln(R_{t+k,t}) = \alpha + \theta \cdot VRP_t + \epsilon_{t+k,t}, \tag{24}$$

where $\ln(R_{t+k,t}) := \ln(R_{t+1}) + \cdots + \ln(R_{t+k})$ and the variance risk premium $VRP_t := IV_t - RV_t$ is defined by the difference between the S&P 500 index option-implied volatility at time t, for one month maturity options, and the ex-post realized return variation over the period $[t-1, t]$. Bollerslev et al. (2009) show that the variance risk premium is the most significant predictive variable of market returns over a quarterly horizon. Therefore, we test the predictive regression model (24) for $k = 3$.

We estimate the unknown parameter of interest through the least squares estimators

$$(\hat{\alpha}_n, \hat{\theta}_n) = \arg\min_{(\alpha,\theta)} \frac{1}{n}\sum_{t=1}^{n-3}\left(\frac{1}{3}\ln(R_{t+3,t}) - \alpha - \theta \cdot VRP_t\right)^2. \tag{25}$$

We are interested in testing the hypothesis of no predictability $\mathcal{H}_0 : \theta_0 = 0$. To this end, using the block bootstrap and the wild multiplicative bootstrap, we construct 90% confidence intervals for the unknown parameter of interest θ_0. More precisely, we apply the procedures under investigation to the period 1990–2010, consisting of 240 observations. Table 5 reports our empirical results. For the period under investigation, our wild bootstrap procedure always provides significance in favor of predictability. Similarly, inference based on standard first-order asymptotic theory also rejects the null hypothesis. By contrast, the block bootstrap implies larger and less stable confidence intervals that lead to ambiguous conclusions depending on the selection of the block size. For instance, for $m = 5, 10, 15$ the block bootstrap also rejects the hypothesis of no predictability. However, for $m = 20$, the block bootstrap does not reject $\mathcal{H}_0$.

Table 5. Stock return predictability. We report 90% confidence intervals for the parameter θ_0 in model (24). We consider the block bootstrap with block sizes $m = 5, 10, 15, 20$ and the wild multiplicative bootstrap with lag truncation $h = 5, 10, 15, 20$, for the period 1990–2010, consisting of 240 observations.

Block	m	=	5	[0.1014 ; 0.4819]
	m	=	10	[0.0958 ; 0.4866]
	m	=	15	[0.0451 ; 0.5373]
	m	=	20	[-0.0064 ; 0.5888]
Wild	h	=	5	[0.1251 ; 0.4573]
	h	=	10	[0.1035 ; 0.4789]
	h	=	15	[0.0973 ; 0.4851]
	h	=	20	[0.0776 ; 0.5048]

A possible source of the divergent conclusions could be related to the lack of robustness of the block bootstrap in the presence of anomalous observations. Indeed, the year 2008 is characterized by several unusual observations linked to the recent credit crisis. As shown in Camponovo et al. (2015), inference provided by block bootstrap procedures may be easily inflated by a small fraction of anomalous observations in the data. Intuitively, this feature is explained by the excessively high fraction of anomalous data that is often simulated by conventional block bootstrap procedures, when compared to the actual fraction of anomalous observations in the original data. On the other hand, since the wild multiplicative bootstrap does not construct random samples by resampling from the observations, our procedure may preserve a desirable accuracy even in the presence of anomalous observations.

6. Conclusions

In time series models, in absence of parametric assumptions on the data generating process, the standard approach to bootstrapping is the block bootstrap. After splitting the original sample into (non)-overlapping blocks, the block bootstrap constructs random samples by selecting the (non)-overlapping blocks with replacement. Under strong regularity conditions on the data generating process and on the estimating functions, the block bootstrap may provide asymptotic refinements relative to standard first-order asymptotic theory. However, to achieve this objective, the definition of the block bootstrap and the selection of the block size require some care.

In this paper, we introduce a wild multiplicative bootstrap procedure that does not require the selection of block sizes but still depends on a less sensitive lag truncation parameter. Unlike conventional bootstrap procedures proposed in the literature, in our algorithm we do not construct random samples by resampling from the observations. Instead, we propose perturbating the general estimating functions using correlated innovations. By introducing this time series dependence, our bootstrap method is able to properly capture the autocorrelation of the true moments. Moreover, unlike conventional bootstrap methods, the wild bootstrap may preserve a desirable accuracy and stability even in the presence of anomalous observations. We prove the validity of our bootstrap procedure and in a Monte Carlo analysis show that our approach always outperforms inference based on standard first-order asymptotic theory. Furthermore, the wild multiplicative bootstrap we propose also compares favorably with block bootstrap procedures for values of the block size typically suggested in the literature.

Finally, in a real data application related to the large literature on stock return predictability, we show the advantages of the proposed procedure for obtaining clear results that are not influenced by the presence of possible anomalous observations in the data.

Acknowledgments

We thank the editor, the associate editor and four anonymous referees for useful comments.

Conflict of interest

We have no conflict on interest to declare.

References

Allen J, Gregory AW, Shimotsu K (2011) Empirical likelihood block bootstrapping. *J Econom* 161: 110–121.

Altonji JG, Segal LM (1996) Small-sample bias in GMM estimation of covariance structures. *J Bus Econ Stat* 14: 353–366.

Andrews DWK (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59: 817–858.

Andrews DWK (2002) Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica* 70: 119–162.

Bollerslev T, Tauchen G, Zhou H (2009) Expected stock returns and variance risk premia. *Rev Financ Stud* 22: 4463–4492.

Bücher A, Kojadinovic I (2016) A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli* 22: 927–968.

Campbell JY, Shiller RJ (1988) The dividend ratio model and small sample bias: a Monte Carlo study. *Econ Lett* 29: 325–331.

Campbell JY, Yogo M (2006) Efficient tests of stock return predictability. *J Financ Econ* 81: 27–60.

Camponovo L, Scaillet O, Trojani F (2012) Robust subsampling. *J Econom* 167: 197–210.

Camponovo L, Scaillet O, Trojani F (2015) Predictability hidden by anomalous observations. Working paper.

Carlstein E (1986) The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Ann Stat* 14: 1171–1179.

Chernozhukov V, Chetverikov D, Kato K (2014) Central limit theorems and multiplier bootstrap when p is much larger than n. *Ann Stat*, In press.

Davidson J (1994) *Stochastic Limit Theory*, Oxford University Press, Oxford.

De Jong RM, Davidson J (2000) Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica* 68: 407–424.

Fama E, French K (1988) Dividend yields and expected stock returns. *J Financ Econ* 22: 3–25.

Geyer C (1994) On the asymptotics of constrained M-estimation. *Ann Stat* 22: 1993–2010.

Goncalves S, White H (2004) Maximum likelihood and the bootstrap for nonlinear dynamic models. *J Econom* 119: 199–219.

Götze F, Künsch HR (1996) Second-order correctness of the blockwise bootstrap for stationary observations. *Ann Stat* 24: 1914–1933.

Hall P (1985) Resampling a coverage process. *Stoch Process Their Appl* 19: 259–269.

Hall AR (2005) *Generalized Method of Moments*, Oxford University Press, Oxford.

Hall P, Horowitz J (1996) Bootstrap critical values for tests based on Generalized- Method-of-Moment estimators. *Econometrica* 64: 891–916.

Hall P, Horowitz JL, Jing BY (1995) On blocking rules for the bootstrap with dependent data. *Biometrika* 82: 561–574.

Hansen LP (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–1054.

Huber PJ (1964) Robust estimation of a location parameter. *Ann Math Stat* 35: 73–101.

Politis DN, Romano JP (1992) A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Ann Stat* 20: 1985–2007.

Politis DN, Romano JP (1992) The stationary bootstrap. *J Am Stat Assoc* 89: 1303–1313.

Inoue A, Shintani M (2006) Bootstrapping GMM estimators for time series. *J Econom* 133: 531–555.

Kline P, Santos A (2012) A score based approach to wild bootstrap inference. *J Econom Method* 1: 23–41.

Künsch H (1989) The jacknife and the bootstrap for general stationary observations. *Ann Stat* 17: 1217–1241.

Jansson M, Moreira MJ (2006) Optimal inference in regression models with nearly integrated regressors. *Econometrica* 74: 681–714.

Lahiri S (1996) Edgeworth expansion and moving block bootstrap for studentized M-estimators in multiple linear regression models. *J Multivar Anal* 56: 42–59.

Lazarus E, Lewis DJ, Stock JH, et al. (2018) HAR inference: recommendations for practice rejoinder. *J Bus Econ Stat* 36: 541–559.

Minnier J, Tian L, Cai T (2011) A perturbation method for inference on regularized regression estimates. *J Ame Stat Assoc* 106: 1371–1382.

Müller U (2014) HAC corrections for strongly autocorrelated time series. *J Bus Econ Stat* 32: 311–322.

Nelson CR, Kim MJ (1993) Predictable stock returns: the role of small sample bias. *J Financ* 48: 641–661.

Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55: 703–708.

Polk C, Thompson S, Vuolteenaho T (2006) Cross-sectional forecast of the equity premium. *J Financ Econ* 81: 101–141.

Rozeff M (1984) Dividend yields are equity risk premium. *J Portf Manage* 11: 68–75.

Salibian-Barrera M, Zamar R (2002) Boostrapping robust estimates of regression. Dividend yields are equity risk premium. *Ann Stat* 30: 556–582.

Shao X (2010) The dependent wild bootstrap. *J Am Stat Assoc* 105: 218–235.

Shiller RJ (2000) . Princeton *The dependent wild bootstrap*, University Press, Princeton, NJ.

Singh K (1998) Breakdown theory for bootstrap quantiles. *Ann Stat* 26: 1719–1732.

Zhu K (2016) Breakdown Bootstrapping the portmanteau tests in weak autoregressive moving average models. *J Royal Stat Soc* 78: 463–485.

Zhu K (2019) Statistical inference for autoregressive models under heteroskedasticity of unknown form. *Ann Stat*, In press.

Zhu K, Li WK (2015) A bootstrapped spectral test for adequacy in weak ARMA models. *J Econom* 187: 113–130.

Zhu K, Ling S (2015) LADE-based inference for ARMA models with unspecified and heavy-tailed heteroskedastic noises. *J Am Stat Assoc* 110: 784–794.

Appendix: Assumptions and Proofs

Before proving Theorem 3.1, let us introduce a set of assumptions in line with Goncalves and White (2004) and Allen, et al. (2010), for M and GMM estimators, respectively.

Assumption 6.1.

(a) Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. The observed data are a realization of a stochastic process $X_t : \Omega \to \mathbb{R}^{d_x}$, $d_x \in \mathbb{N}$, with $X_t(\omega) = W_t(\ldots, V_{t-1}(\omega), V_t(\omega), V_{t+1}(\omega), \ldots)$, $V_t : \Omega \to \mathbb{R}^v$, $v \in \mathbb{N}$, and $W_t : \prod_{\tau=-\infty}^{\infty} \mathbb{R}^v \to \mathbb{R}^{d_x}$ is such that X_t is measurable for all t.

(b) *Either for M estimators, the function* $\rho : \mathbb{R}^d \times \Theta \to \mathbb{R}$ *is such that* $\rho(\cdot, \theta)$ *is measurable for each* $\theta \in \Theta$*, a compact subset of* $\mathbb{R}^p$*, and* $\rho(X_t, \cdot)$ *is continuous on* Θ *a.s. for all t; or for GMM estimators, the function* $g : \mathbb{R}^{d_x} \times \Theta \to \mathbb{R}^{d_g}$ *is such that* $g(\cdot, \theta)$ *is measurable for each* $\theta \in \Theta$*, a compact subset of* $\mathbb{R}^{d_\theta}$*, and* $g(X_t, \cdot)$ *is continuous on* Θ *a.s. for all t.*

(c) *Either for M estimators: (i)* θ_0 *is the unique minimum of* $E[\frac{1}{n}\sum_{t=1}^n \rho(X_t, \theta)]$ *over* $\theta \in \Theta$*. (ii)* θ_0 *is an interior point of* Θ*; or for GMM estimators: (i)* θ_0 *is the unique solution of* $E[g(X_t, \theta)] = 0$*,* $\theta \in \Theta$*. (ii)* θ_0 *is an interior point of* Θ*.*

(d) *Either for M estimators: (i)* $\rho(X_t, \theta)$ *is Lipschitz continuous on* Θ*, i.e.,* $|\rho(X_t, \theta_1) - \rho(X_t, \theta_2)| \leq L_t|\theta_1 - \theta_2|$ *a.s. for all* $\theta_1, \theta_2 \in \Theta$*, where* $\frac{1}{n}\sum_{t=1}^n E[L_t] = O(1)$*. (ii)* $\frac{\partial^2}{\partial\theta\partial\theta'}\rho(X_t, \theta)$ *is Lipschitz continuous on* Θ*; or for GMM estimators: (i)* $g(X_t, \theta)$ *is Lipschitz continuous on* Θ*, i.e.,* $||g(X_t, \theta_1) - g(X_t, \theta_2)|| \leq L_t||\theta_1 - \theta_2||$ *a.s. for* $\theta_1, \theta_2 \in \Theta$*, where* $\frac{1}{n}\sum_{t=1}^n E[L_t] = O(1)$*. (ii)* $\frac{\partial}{\partial\theta}g(X_t, \theta)$ *is Lipschitz continuous on* Θ*.*

(e) *For some* $r > 2$*, either for M estimators: (i)* $\rho(X_t, \theta)$ *is r-dominated on* Θ *uniformly in t, i.e., there exists* D_t *such that* $|\rho(X_t, \theta)| \leq D_t$ *for all* $\theta \in \Theta$*, and* D_t *is measurable such that* $E[|D_t|^r] < \infty$ *for all t. (ii)* $\frac{\partial}{\partial\theta}\rho(X_t, \theta)$ *is r-dominated on* Θ *uniformly in t. (iii)*$\frac{\partial^2}{\partial\theta\partial\theta'}\rho(X_t, \theta)$ *is r-dominated on* Θ *uniformly in t; or for GMM estimators: (i)* $g(X_t, \theta)$ *is r-dominated on* Θ *uniformly in t, i.e., there exists* D_t *such that* $||g(X_t, \theta)|| \leq D_t$ *for all* $\theta \in \Theta$*, and* D_t *is measurable such that* $E[||D_t||^r] < \infty$ *for all t. (ii)* $\frac{\partial}{\partial\theta}g(X_t, \theta)$ *is r-dominated on* Θ *uniformly in t.*

(f) $\{V_t\}$ *is an* α*-mixing sequence of size* $-2r/(r-2)$*, with* $r > 2$*.*

(g) *Either for M estimators: the elements of (i)* $\rho(X_t, \theta)$ *are near epoch dependent on* $\{V_t\}$ *of size* $-1/2$*. (ii)* $\frac{\partial}{\partial\theta}\rho(X_t, \theta)$ *are near epoch dependent on* $\{V_t\}$ *of size* -1 *uniformly on* (Θ, f)*, where f is any convenient norm on* $\mathbb{R}^p$*. (iii)* $\frac{\partial^2}{\partial\theta\partial\theta'}\rho(X_t, \theta)$ *are near epoch dependent on* $\{V_t\}$ *of size* $-1/2$ *uniformly on* (Θ, f)*; or for GMM estimators: the elements of (i)* $g(X_t, \theta)$ *are near epoch dependent on* $\{V_t\}$ *of size* -1 *uniformly on* (Θ, f)*, where f is any convenient norm on* $\mathbb{R}^{d_g}$*. (ii)* $\frac{\partial}{\partial\theta}g(X_t, \theta)$ *are near epoch dependent on* $\{V_t\}$ *of size* -1 *uniformly on* (Θ, f)*.*

(h) *Either for M estimators: (i)* $||\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n E[\frac{\partial}{\partial\theta}\rho(X_i, \theta_0)\frac{\partial}{\partial\theta}\rho(X_j, \theta_0)'] - \Omega_0|| \to 0$*, for some positive definite matrix* Ω_0*. (ii)* $||\frac{1}{n}\sum_{t=1}^n E[\frac{\partial^2}{\partial\theta\partial\theta'}\rho(X_t, \theta_0)] - D_0|| \to 0$*, where* D_0 *is of full rank; or for GMM estimators: (i)* $||\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n E[g(X_i, \theta_0)g(X_j, \theta_0)'] - \Omega_0||^2 \to 0$*, for some positive definite matrix* Ω_0*. (ii)* $||\frac{1}{n}\sum_{t=1}^n E[\frac{\partial}{\partial\theta}g(X_t, \theta_0)] - D_0||^2 \to 0$*, where* D_0 *is of full rank. (iii)* W_n *converges in probability to a non-random positive-definite symmetric matrix* W_0*.*

(l) *(i) The kernel function* $k(\cdot)$ *is continuous,* $k(0) = 1$*,* $k(x) = k(-x)$*, and* $\int_{-\infty}^{\infty}|k(x)|dx < \infty$*. (ii) Let* $K(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty}k(x = e^{-ix\lambda}dx$*, then* $\int_{-\infty}^{\infty}|K(\lambda)|d\lambda < \infty$*. (iii) The lag truncation h satisfies* $\frac{1}{h} + \frac{h}{\sqrt{n}} \to 0$*, as* $n \to \infty$*.*

Assumption 6.2.

(a) *For some* $r > 2$*, either for M estimators:* $\frac{\partial}{\partial\theta}\rho(X_t, \theta)$ *is* $3r$*-dominated on* Θ *uniformly in t; or for GMM estimators:* $g(X_t, \theta)$ *is* $3r$*-dominated on* Θ *uniformly in t.*

(b) *Either for M estimators: (i) For small* $\delta > 0$ *and some* $r > 2$*, the elements of* $\frac{\partial}{\partial\theta}\rho(X_t, \theta)$ *are* $L_{2+\delta}$ *near epoch dependent on* $\{V_t\}$ *of size* $-2(r-1)/(r-2)$ *uniformly on* (Θ, f)*. (ii)* $\{V_t\}$ *is* α*-mixing of size* $-r(2+\delta)/(r-2)$*; or for GMM estimators: (i) For small* $\delta > 0$ *and some* $r > 2$*, the elements of* $g(X_t, \theta)$ *are* $L_{2+\delta}$ *near epoch dependent on* $\{V_t\}$ *of size* $-2(r-1)/(r-2)$ *uniformly on* (Θ, f)*. (ii)* $\{V_t\}$ *is* α*-mixing of size* $-r(2+\delta)/(r-2)$*.*

Assumption 6.3.

(a) *Let* $(e_1, \ldots, e_n)$ *be a sample from a stationary process of positive correlated observations with* $E[e_t|(X_1, \ldots, X_n)] = 1$, $Cov(e_t, e_{t+i}|(X_1, \ldots, X_n)) = k(i/h)$, $E[e_t^4|(X_1, \ldots, X_n)] < \infty$ *where* $k(\cdot)$ *is an appropriate kernel function, and h is the lag truncation parameter.*

Assumptions 6.1 and 6.2 are mild conditions typically required for the validity of bootstrap approximations that are satisfied in several time series settings. In particular, Assumption 6.1 provides a set of conditions that are typically required for the consistency and asymptotic normality of M and GMM estimators, whereas in Assumption 6.2, in line with Goncalves and White (2004) and Allen, et al. (2010), we add conditions necessary for the consistency of the bootstrap approximation. Finally, in Assumption 6.3, we add conditions for the error terms in the construction of the wild bootstrap approximation. Unfortunately, these assumptions do not apply to unknown parameters defined through non-differentiable estimating functions.

Proof of Theorem 3.1: First, we consider the M estimator case, and prove statement (i). To this end, consider the random process

$$R_n(u) = \sum_{t=1}^{n} \rho^*(X_t, \hat{\theta}_n + u/\sqrt{n}) - \sum_{t=1}^{n} \rho^*(X_t, \hat{\theta}_n). \tag{26}$$

Note that $R_n(u)$ is minimized at $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$. By considering a Taylor expansion of $\rho^*(X_t, \hat{\theta}_n + u/\sqrt{n})$ around $\hat{\theta}_n$ we have

$$\rho^*(X_t, \hat{\theta}_n + u/\sqrt{n}) = \rho^*(X_t, \hat{\theta}_n) + \frac{u'}{\sqrt{n}} \left(\frac{\partial}{\partial \theta} \rho^*(X_t, \hat{\theta}_n) \right) + \frac{1}{2n} u' \left(\frac{\partial^2}{\partial \theta \partial \theta'} \rho^*(X_t, \hat{\theta}_n) \right) u + o_P(1/n). \tag{27}$$

Therefore, we can rewrite the random process $R_n(u)$ as

$$R_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u' \left(\frac{\partial}{\partial \theta} \rho^*(X_t, \hat{\theta}_n) \right) + \frac{1}{2n} \sum_{t=1}^{n} u' \left(\frac{\partial^2}{\partial \theta \partial \theta'} \rho^*(X_t, \hat{\theta}_n) \right) u + o_p(1). \tag{28}$$

First, consider the second factor $\frac{1}{2n} \sum_{t=1}^{n} u' \left(\frac{\partial^2}{\partial \theta \partial \theta'} \rho^*(X_t, \hat{\theta}_n) \right) u$ in the above expansion. By Theorem 20.21 in Davidson (1994), the term $\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta'} \rho^*(X_t, \hat{\theta}_n)$ converges in conditional probability to D_0. Furthermore, consider now the first factor $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u' \left(\frac{\partial}{\partial \theta} \rho^*(X_t, \hat{\theta}_n) \right)$. By De Jong and Davidson (2000), and Corollary 24.7 in Davidson (1994), the conditional law of $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\frac{\partial}{\partial \theta} \rho^*(X_t, \hat{\theta}_n) \right)$ converges weakly to a normal distribution with mean 0 and covariance matrix Ω_0.

Therefore, the limit $R(u)$ of $R_n(u)$ is given by

$$R(u) = u' v_0 + \frac{u' D_0 u}{2}. \tag{29}$$

where $v_0 \sim N(0, \Omega_0)$. Note that the unique minimum of $R(u)$ is $-D_0^{-1} v_0$, which is normally distributed with mean 0 and covariance matrix $D_0^{-1} \Omega_0 D_0^{-1}$. It turns out that by use of the results in Geyer (1994) the conditional law of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ also converges weakly to a normal distribution with mean 0 and the same covariance matrix.

To prove statement (ii), we adopt the same approach adopted in the proof of statement (i). More precisely, consider the random process

$$S_n(u) = \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g^*(X_t, \hat{\theta}_n + u/\sqrt{n}) \right)' W_n \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g^*(X_t, \hat{\theta}_n + u\sqrt{n}) \right) \tag{30}$$

$$- \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g^*(X_t, \hat{\theta}_n) \right)' W_n \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g^*(X_t, \hat{\theta}_n) \right). \quad (31)$$

Note that $S_n(u)$ is minimized at $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$. By considering a Taylor expansion of the term $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g^*(X_t, \hat{\theta}_n + u/\sqrt{n})$ around $\hat{\theta}_n$ we have,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g^*(X_t, \hat{\theta}_n + u/\sqrt{n}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g^*(X_t, \hat{\theta}_n) + \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} g^*(X_t, \hat{\theta}_n) \right) u + o_p(1). \quad (32)$$

It turns out that using (32) we can rewrite $S_n(u)$ as

$$S_n(u) = u' \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} g^*(X_t, \hat{\theta}_n) \right) W_n \left(\frac{2}{\sqrt{n}} \sum_{t=1}^{n} g^*(X_t, \hat{\theta}_n) \right) \quad (33)$$

$$+ u' \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} g^*(X_t, \hat{\theta}_n) \right) W_n \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} g^*(X_t, \hat{\theta}_n) \right) u + o_p(1). \quad (34)$$

Therefore, by Theorem 20.12 in Davidson (1994), De Jong and Davidson (2000), and Corollary 24.7 in Davidson (1994), the limit $S(u)$ of $S_n(u)$ is given by

$$S(u) = 2u' D_0 W_0 v_0 + u' D_0 W_0 D_0 u, \quad (35)$$

where $v_0 \sim N(0, \Omega_0)$. Note that the unique minimum of $S(u)$ is $-(D_0' W_0 D_0)^{-1} D_0' W_0 v_0$, which is normally distributed with mean 0 and covariance matrix $(D_0' W_0 D_0)^{-1} D_0' W_0 \Omega_0 W_0 D_0 (D_0' W_0 D_0)^{-1}$. By the use of the results in Geyer (1994), the conditional law of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ also converges weakly to a normal distribution with mean 0 and the same covariance matrix.