The multivariate linear process bootstrap

Carsten Jentsch[1] and Dimitris N. Politis[2]

[1]*Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany*

[2]*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA*

Abstract

The linear process bootstrap (LPB) for univariate time series has been introduced by McMurry and Politis (2010) and it is called LPB because it generates linear processes in the bootstrap domain. However, it does not assume that the data are themselves a sample from a linear process. They use tapered and banded estimates for the autocovariance matrix of the whole data stretch [cf. Wu and Pourahmadi (2009)] and i.i.d. resampling of appropriately standardized residuals. Under a physical dependence assumption [cf. Wu (2005)], they show validity of the LPB for the sample mean.

In this paper, we generalize their approach to the case of multivariate time series and show its validity for the sample mean under different assumptions (mixing, weak dependence, linearity). To complement the theory of McMurry and Politis (2010), we show that the multivariate LPB works also for spectral density estimation, but that it fails generally for sample autocovariances. Even in the univariate case and under assumed linearity, this is still the case. However, we show consistency of the LPB for univariate, causal and invertible linear processes for higher order statistics as e.g. autocovariances and for autocorrelations in the univariate case under linearity only. Moreover, the validity of the LPB for the multivariate sample mean can be used in order to correct the invalidity of the univariate LPB for higher order statistics as e.g. autocovariances in the general case. For this purpose, an LPB-of-blocks bootstrap scheme is proposed.

Keywords: Banded covariance matrix estimator, bootstrap, linear process, multivariate time series

AMS subject classifications: 62G09; 62M10

1 Introduction

For dependent time series data, resampling methods and bootstrap techniques in particular have been studied intensively over the last decades. To complement the existing theory, McMurry and Politis (2010) proposed the so-called linear process bootstrap (LPB) for univariate stationary processes and showed its validity for the sample mean actually without assuming linearity of the underlying process.

In comparison to other popular bootstrap procedures in time series analysis as e.g. the AR sieve bootstrap or block bootstrap techniques, it is possible to resample processes with an abruptly dying-out covariance structure as MA processes. In principle, it is of course possible to estimate the parameters of an underlying moving average process by using numerical optimization or the innovations algorithm, but these procedures are much more involved than estimating for example AR coefficients, which can be obtained by stable estimation procedures. Here, the benefit of the LPB is that MA processes can be resampled without estimating their coefficients explicitly.

For an overview of existing bootstrap methods compare the monograph of Lahiri (2003) or the very recent review paper by Kreiss and Paparoditis (2011).

The main idea of the LPB is to consider the data sample of length n as one n-dimensional vector and to estimate appropriately the entire covariance structure of this vector. This is executed by using banded or tapered covariance matrix estimators as proposed by Wu and Pourahmadi (2009) and employed by McMurry and Politis (2010). In order, the resulting covariance matrix is used to whiten the data by pre-multiplying the original (centered) data with its inverse Cholesky matrix, where a modification of its eigenvalues ensures positive definiteness if necessary. After suitable centering and standardizing, the resulting vector is treated as i.i.d. with zero mean and unit variance. Finally, i.i.d. resampling from this vector and pre-multiplying the corresponding bootstrap vector of residuals with the Cholesky matrix itself results in a bootstrap sample that has (approximately) the same covariance structure as the original time series.

The limiting variance of the sample mean is known to depend exclusively on the covariance structure of the underlying process and, heuristically, the LPB is valid here thanks to its capability to capture (asymptotically) its entire covariance structure. However, the LPB is generally *not* able to mimic the complete dependence structure of the underlying process which is potentially a lot more involved and may contain features of higher order than the second as well. Therefore, the LPB as introduced in McMurry and Politis (2010) is typically *not* consistent for statistics that depend asymptotically on higher order characteristics of the underlying process or even on its exact distribution. Nevertheless, there are other important statistics than (smooth functions of) the sample mean around that depend on autocovariances of the process only such that the LPB might be valid in these situations as well. For example, this property is well known for kernel spectral density (matrix) estimates in general and for sample autocorrelations under linearity.

In Section 2 we state and describe the multivariate generalization of the LPB and in Section 3 we discuss our established results concerned with its validity for sample mean, kernel spectral density estimates and sample autocovariances and autocorrelations under general assumptions and in some important special cases.

2 The bootstrap algorithm

Step 1. Let $\mathbf{X}$ be the $(d \times n)$ data matrix consisting of $\mathbb{R}^d$-valued time series data $\underline{X}_1, \ldots, \underline{X}_n$ of sample size n. Compute the centered observations $\underline{Y}_t = \underline{X}_t - \overline{\underline{X}}$, where $\overline{\underline{X}} = \frac{1}{n}\sum_{t=1}^{n} \underline{X}_t$, let $\mathbf{Y}$ be the corresponding $(d \times n)$ matrix of centered observations and define $\underline{Y} = vec(\mathbf{Y})$ to be the dn-dimensional vectorized version of $\mathbf{Y}$.

Step 2. Compute $\underline{W} = (\widehat{\mathbf{\Gamma}}^{\epsilon}_{\kappa,l})^{-1/2}\underline{Y}$, where $(\widehat{\mathbf{\Gamma}}^{\epsilon}_{\kappa,l})^{1/2}$ denotes the lower left triangular matrix $\mathbf{L}$ of the Cholesky decomposition $\widehat{\mathbf{\Gamma}}^{\epsilon}_{\kappa,l} = \mathbf{L}\mathbf{L}^T$. Here, $\widehat{\mathbf{\Gamma}}^{\epsilon}_{\kappa,l}$ is a suitable tapered estimator for the $(dn \times dn)$ covariance matrix $\mathbf{\Gamma}_{dn}$ of $\underline{Y}$ that is ensured to be positive definite.

Step 3. Let $\underline{Z}$ be the standardized version of $\underline{W}$, that is, $Z_i = \frac{W_i - \overline{W}}{\widehat{\sigma}_W}$, $i = 1, \ldots, dn$, where $\overline{W} = \frac{1}{dn}\sum_{t=1}^{dn} W_t$ and $\widehat{\sigma}^2_W = \frac{1}{dn}\sum_{t=1}^{dn}(W_t - \overline{W})^2$.

Step 4. Generate $\underline{Z}^* = (Z^*_1, \ldots, Z^*_{dn})^T$ by i.i.d. resampling from $\{Z_1, \ldots, Z_{dn}\}$.

Step 5. Compute $\underline{Y}^* = (\widehat{\mathbf{\Gamma}}^{\epsilon}_{\kappa,l})^{1/2}\underline{Z}^*$ and let $\mathbf{Y}^*$ be the matrix that is obtained from $\underline{Y}^*$ by putting this vector column-wise into an $(d \times n)$ matrix and denote its columns by $\underline{Y}^*_1, \ldots, \underline{Y}^*_n$.

3 Results

Suppose the data generating process $\{\underline{X}_t, t \in \mathbb{Z}\}$ is an $\mathbb{R}^d$-valued stationary time series with mean vector $\underline{\mu}$ and autocovariance matrices $\mathbf{C}(h)$, $h \in \mathbb{Z}$, which are assumed to be componentwise absolutely summable, i.e. $\sum_{h=-\infty}^{\infty} |C_{ij}(h)| < \infty$ for all i, j. Moreover, we assume that there exists a constant $M < \infty$ independent of h such that

$$\| \sum_{t=1}^{n} (\underline{X}_{t+h} - \overline{\underline{X}})(\underline{X}_t - \overline{\underline{X}})^T - n\mathbf{C}(h) \|_2 \leq M\sqrt{n}, \tag{1}$$

where $\|\mathbf{A}\|_2 = \sqrt{E(|\mathbf{A}|_2^2)}$ and $|\cdot|_2$ denotes the Euclidean matrix norm. Additionally, we have to assume eigenvalues of $\mathbf{\Gamma}_{dn}$ uniformly bounded away from zero for sufficiently large n to ensure positive definiteness.

In this paper, by proving validity of the multivariate LPB as introduced in Section 2 for the bootstrap sample mean vector $\overline{\underline{Y}}^* = \frac{1}{n}\sum_{t=1}^{n} \underline{Y}_t^*$ under suitable assumptions, we generalize the result obtained by McMurry and Politis (2010) to the multivariate case. Note that their result for the univariate case is established under a physical dependence condition [cf. Wu (2005)], which is sufficient for (1).

To complement the results of McMurry and Politis (2010), we consider also bootstrap versions of kernel estimates for the spectral density matrix, i.e.

$$\widehat{\mathbf{f}}^*(\omega) = \frac{1}{n} \sum_{j=-\lfloor \frac{n-1}{2} \rfloor}^{\lfloor \frac{n}{2} \rfloor} K_h(\omega - \omega_j) \mathbf{I}_n^*(\omega_j), \tag{2}$$

where $\mathbf{I}_n^*(\omega) = \underline{J}_n^*(\omega)\overline{\underline{J}_n^*(\omega)}^T$ is the periodogram matrix, $\underline{J}_n^*(\omega) = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} \underline{Y}_t^* e^{-it\omega}$ is the multivariate discrete Fourier transform (DFT) based on $\underline{Y}_1^*, \ldots, \underline{Y}_n^*$, K is a kernel with $\int K(x)dx = 2\pi$, h is a bandwidth and $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$. Under suitable assumptions, we prove validity of the MLPB for kernel spectral density estimates.

In the following, we discuss our obtained results for sample autocovariances

$$\widehat{\mathbf{C}}(h) = \frac{1}{n} \sum_{t=max(1,1-h)}^{min(n,n-h)} (\underline{X}_{t+h} - \overline{\underline{X}})(\underline{X}_t - \overline{\underline{X}})^T \tag{3}$$

in detail. As discussed in Kreiss and Paparoditis (2011), the question whether a certain bootstrap procedure is valid for some statistic T_n depends essentially on the capability of the bootstrap to mimic all characteristics of the underlying distribution of $\{\underline{X}_t, t \in \mathbb{Z}\}$ that crop up in the limiting distribution of T_n. In general, the LPB is designed to mimic the entire covariance structure of $\{\underline{X}_t, t \in \mathbb{Z}\}$ asymptotically, but it does not capture any higher order structure of the underlying process. Being concerned with autocovariances, it is well known, that the limiting distribution of sample autocovariances (3) does depend on fourth order joint cumulants of $\{\underline{X}_t, t \in \mathbb{Z}\}$. This causes the LPB to fail for sample autocovariances (and other statistics that depend on more than first and second order structure of the underlying process) in general. Compare for instance the discussion of the multivariate case in Jentsch and Kreiss (2010).

Even in the univariate case $d = 1$ and under the additional assumption of linearity, i.e. it holds $X_t = \sum_{\nu=-\infty}^{\infty} b_\nu e_{t-\nu}$, $t \in \mathbb{Z}$ with absolutely summable sequence $\{b_j, j \in \mathbb{Z}\}$ and i.i.d. white noise $\{e_t, t \in \mathbb{Z}\}$ with $E(e_t^2) = \sigma_e^2 \in (0, \infty)$ and $E(e_t^4) = \eta\sigma_e^4 \in (0, \infty)$, the limiting covariances of sample autocovariances depend on fourth order structure of the process. Precisely, it holds

$$nCov(\widehat{C}(h), \widehat{C}(k)) \underset{n\to\infty}{\longrightarrow} C(h)C(k)(\eta-3) + \sum_{r=-\infty}^{\infty} \left(C(r)C(r+k-h) + C(r+k)C(r-h)\right).$$

The LPB captures the second part on the right-hand side above correctly, but fails to mimic the first term in general. Consequently, the LPB is not valid asymptotically for sample autocovariances even in the univariate case and under assumed linearity.

However, we show that consistency of the LPB holds for *univariate*, *causal* and *invertible* linear processes for higher order statistics as e.g. autocovariances. At first sight, this may seem to be surprising, because contrary to fully nonparametric bootstrap methods as e.g. the block bootstrap, the LPB mimics by construction only the second order structure of the DGP and not its entire dependence structure. But by taking a closer look and particularly in comparison to the prominent autoregressive sieve bootstrap [cf. the discussion in Kreiss, Paparoditis and Politis (2011)], which is strongly related to the LPB approach, this result fits in the existing literature. Heuristically, the LPB remains valid in this case due to a triangular shape of the matrix A that (approximately) transforms the vector of residuals $(e_1, \ldots, e_n)'$ by pre-multiplication to the vector of observations $(X_1, \ldots, X_n)'$. Thanks to the triangular shape of A and under invertibility, this matrix fits properly together asymptotically with the Cholesky matrix that is involved in the bootstrap algorithm. Eventually, this implies consistency for the LPB in this case. Moreover, we obtain validity of the LPB for autocorrelations under linearity only, because their limiting distributions then depend on second order structure exclusively.

Finally, the validity of the LPB for the multivariate sample mean can be used in order to correct the invalidity of the univariate LPB for higher order statistics as e.g. autocovariances in the general case. For this purpose, an LPB-of-blocks bootstrap scheme is proposed that resamples directly vectors of lagged products $Y_{t+h}Y_t = (X_{t+h} - \overline{X})(X_t - \overline{X})$ to capture their covariance structure properly instead of the covariance structure of the process $\{X_t, t \in \mathbb{Z}\}$ itself. For instance, if the bivariate process $\underline{U}_t = (Y_{t+h}Y_t, Y_t^2)$ satisfies the assumptions imposed at the beginning of this section, we show consistency for sample autocovariances in general and without assuming linearity at all by standard application of $g(x, y) = x/y$ and the Δ-method.

References

[1] Jentsch, C. and Kreiss, J.-P. (2010): The multiple Hybrid Bootstrap - Resampling multivariate linear processes. *Journal of Multivariate Analysis*, 101, 2320–2345.

[2] Kreiss, J.-P. and Paparoditis, E. (2011). Bootstrap for dependent data: a review. To appear.

[3] Kreiss, J.-P., Paparoditis, E. and Politis, D.N. (2011). On the range of the validity of the autoregressive sieve bootstrap. To appear.

[4] Lahiri, S.N. (2003): *Resampling Methods for Dependent Data*, Springer, New York.

[5] McMurry, T.L. and Politis, D.N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis* 31, 471–482.

[6] Wu, W.B. (2005). Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences* 102, 14150–54.

[7] Wu, W.B. and Pourahmadi (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica* 19(4), 1755-68.