Discussion on the review by J.-P. Kreiss and E. Paparoditis by Sílvia Gonçalves and Dimitris N. Politis

May 24, 2011

This review of bootstrap methods for time series is most welcome especially coming from two key figures in the development of these methods. We would like to complement their exposition by focusing on some further issues of current interest.

Block bootstrap methods for time series data have been most intensively studied under the assumption of stationarity and mixing. An important example is the stationary bootstrap of Politis and Romano (1994). As detailed in the review, this method is a variation of the standard block bootstrap that manages to create bootstrap series that are strictly stationary. This property not only replicates the strict stationarity of the original time series, but it also makes the method easy to analyze theoretically. However, some rather involved Mean Square Error calculations of Lahiri (1999) seemed to imply that the stationary bootstrap has poor asymptotic efficiency relative to the usual block bootstrap in the sample mean case. As a result, practitioners have been understandably reluctant in endorsing the stationary bootstrap after the publication of Lahiri's (1999) paper.

It was quite a surprise when Nordman (2009) discovered an error in Lahiri's (1999) calculations; as it turns out, the stationary bootstrap has *identical* asymptotic accuracy as the block bootstrap with non-overlapping blocks; the latter is well-known to be only slightly less efficient than the full-overlap block bootstrap. These new findings are expected to rekindle the interest of researchers since the stationarity of bootstrap sample paths is a very powerful tool. For example, *functional data* are of current interest in the 21st century; see the review by McMurry and Politis (2011). Suffices to note that Politis and Romano (1994b) were able to prove the asymptotic validity of the stationary bootstrap for the sample mean of functional data almost twenty years ago. Such a result does not seem to be available for the regular block bootstrap even to date.

The assumptions of strict stationarity and strong mixing can be too strong for many applications in economics and finance. First, for many of these applications, the time span is rather long, making it unlikely that stationarity holds. A more appropriate assumption is to allow for some time series heterogeneity (for instance, in the form of time varying moments). Similarly, mixing is too strong a dependence condition to be broadly applicable. Andrews (1994) gives an example of a simple AR(1) model that fails to be strong mixing. More generally, although measurable functions of mixing processes are themselves mixing when the function depends on a *finite* number of lagged values of the mixing process, this is not necessarily the case when the observed time series depends on the *infinite* history of an underlying mixing process. For instance, popular ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) processes, which are a function of the entire past history of a fundamental driving innovation, are known to be mixing only when we assume these innovations to be i.i.d. (see e.g. Carrasco and Chen, 2002, and Basrak, Davis and Mikosch, 2002).

The need to accommodate a large class of heterogeneous weakly dependent time series motivated the work by Gonçalves and White (2002, 2004, 2005). These papers demonstrated the consistency of two *"first generation"* block bootstrap methods, namely the stationary bootstrap (SB) and the moving blocks bootstrap (MBB), with data that arise as functions of mixing processes, i.e., processes that are near epoch dependent (NED) on an underlying mixing process (Billingsley, 1968; McLeish, 1975; Gallant and White, 1988). An important example of the usefulness of NED in economics concerns ARCH and GARCH processes, which were shown to be NED on a mixing process under very mild regularity conditions (that do not require the i.i.d. assumption on the fundamental innovations). Other examples of NED processes include the bilinear and threshold autoregressive models (see Davidson, 2002).

An alternative approach towards allowing a degree of heterogeneity is given by the notion of local (as opposed to global) stationarity that applies when the stochastic structure of the data series is slowly evolving. Locally stationary processes were formally defined by Dahlhaus (1996, 1997) following earlier work of Priestley (1988). To that effect, Paparoditis and Politis (2002) and Dowla, Paparoditis and Politis (2003) proposed a modification of the block bootstrap called the local block bootstrap (LBB) for locally stationary processes. The main idea of the LBB is to only resample blocks that are 'close' to each other, i.e. a block that starts at time t can only be replaced with blocks whose starting points are 'close' to t. These papers established the consistency of the LBB for inference on μ in the context of the following model,

$$X_{t,n} = \mu + m\left(\frac{t}{n}\right) + \sigma\left(\frac{t}{n}\right)\varepsilon_t, \quad \text{for} \quad t = 1, \ldots, n \text{ and } n = 1, 2, \ldots. \tag{1}$$

Here, ε_t is a mean-zero, strong mixing process satisfying some moment and mixing conditions, and m and σ are fixed functions from $[0,1]$ into $\mathbb{R}$ satisfying some smoothness conditions. In the above, the sample mean $\bar{X}_n = n^{-1}\sum_{t=1}^{n} X_{t,n}$ estimates $\bar{\mu}_n \equiv n^{-1}\sum_{t=1}^{n}\mu_{t,n}$ where $\mu_{t,n} \equiv E\left(X_{t,n}\right) = m\left(\frac{t}{n}\right) + \mu$ is time varying. Nevertheless, under the assumption that $\int_0^1 m\left(x\right)dx = 0$, we have $\bar{\mu}_n \to \mu$ as $n \to \infty$, and therefore the sample mean $\bar{X}_n$ is an estimate of μ.

As it turns out, the property of NED on a mixing process can be established for many locally stationary processes. For example, this is true in the context of model (1) where $\{X_{t,n}\}$ can be shown to be NED on the mixing process ε_t. To see why, note that $v_m \equiv \sup_{t,n}\left\|X_{t,n} - E_{t-m}^{t+m}\left(X_{t,n}\right)\right\|_2 = 0$, where $E_{t-m}^{t+m}\left(\cdot\right) = E\left(\cdot\,|\varepsilon_{t-m}, \ldots, \varepsilon_{t+m}\right)$ and $\left\|\cdot\right\|_2 = \left(E\left|\cdot\right|^2\right)^{1/2}$ is the L_2-norm. Hence, under some additional conditons, the results of Gonçalves and White (2002) could be used to justify the consistency of the standard moving blocks bootstrap and stationary bootstrap for inference on μ. One crucial condition for MBB/SB to be consistent for the mean is Assumption 2.2 of Gonçalves and White (2002) which requires that the mean heterogeneity is not too strong so as to verify

$$\frac{1}{n}\sum_{t=1}^{n}\left(\mu_{t,n}-\bar{\mu}_n\right)^2 = o\left(\frac{1}{\ell_n}\right), \tag{2}$$

where ℓ_n is the (average) block size and is such that $\ell_n = o\left(\sqrt{n}\right)$; for example, condition (2) is trivially satisfied if $m \equiv 0$. In any case, under condition (2), the moving blocks bootstrap and the stationary bootstrap can be used for inference on μ with data from model (1) despite the fact that the underlying stochastic process is heterogeneous.

The robustness of standard block bootstrap methods to departures from stationarity (for instance, in the form of local stationarity) is easy to understand through the lens of the current review paper. As the authors emphasize, the asymptotic validity of any bootstrap scheme depends crucially on how the dependence—and we would add, heterogeneity—properties of the data enter the limiting distribution of the statistic of interest. To the extent that heterogeneity does not impact the limiting distribution of the statistics of interest, the bootstrap does not need to mimic this feature (or lack thereof). This explains why the stationary bootstrap can be asymptotically valid under some heterogeneity, e.g. model (1) with condition (2). It also helps explain a reverse phenomenon: why the moving blocks bootstrap can be consistent under strict stationarity despite the fact that the bootstrap samples it generates are not stationary.

This robustness property is also quite appealing since regular bootstrap methods only require the choice of one tuning parameter (the block size). This is in contrast to the LBB which requires the choice of an additional tuning parameter, the 'local' window size over which blocks are considered to be close to each other and thus exchangeable. On the other hand, the LBB can be more generally valid as it is in principle applicable to all kinds of locally stationary models, not just model (1). Furthermore, in the case of model (1), the LBB does not require restricting the degree of mean heterogeneity, e.g. condition (2). The implication is that while the standard MBB/SB would treat the mean heterogeneity as a nuisance that is asymptotically negligible under condition (2), the LBB would actively adapt to the underlying nonstationarity, and thus hopefully better account for it. It would be interesting to compare the finite sample performance of these two methods under model (1).

Although the current review paper focuses on first order asymptotic validity of the bootstrap, one of the key properties of any bootstrap method is its ability to provide an asymptotic refinement over the first order asymptotic distribution. This requires the use of asymptotically pivotal statistics, which in the time series context involves studentized statistics based on a spectral density estimator at frequency zero—also known in econometrics as Heteroskedasticity and Autocorrelation Consistent (HAC) variance estimator.

HAC variance estimators are typically consistent under the assumption that the bandwidth parameter grows with the sample size but at a slower rate. Under this assumption, the asymptotic distribution of HAC-based t-tests is standard normal, independently of the kernel and bandwidth used. However, the finite sample performance of HAC-based t-tests that rely on the asymptotic normal distribution can be very poor, one of the main reasons being that the latter does not reflect the choice of the kernel and bandwidth parameter. This motivated Kiefer and Vogelsang (2005) to develop a new asymptotic theory for HAC–based tests that takes into account the kernel and bandwidth choices. In particular, Kiefer and Vogelsang (2005) assumed that the bandwidth is a fixed proportion of the sample size when deriving the asymptotic distribution of the robust t-test. Under this fixed-bandwidth (or fixed-b) asymptotics, the limiting distribution of the usual robust-t test is a complicated functional of Brownian motions—that can nevertheless be simulated—and it depends explicitly on the kernel function and on the bandwidth parameter chosen.

The simulation results of Kiefer and Vogelsang (2005) showed that the fixed-b approximation outperformed the conventional asymptotic normal approximation in finite samples. For a simple Gaussian location model, results by Jansson (2004) and Phillips, Sun and Jin (2006) provided a theoretical explanation of these simulation results. More recently, McElroy and Politis (2011) extended the asymptotic calculations of Kiefer and Vogelsang (2005) to the case of long memory data, also allowing for the possibility of using a *'flat-top'* kernel for smoothing such as the trapezoid of Politis and Romano (1995); see Politis (2011) for more details.

Another important finding by Kiefer and Vogelsang (2005) was that a naïve application of block bootstrap methods (even an i.i.d. bootstrap with block size equal to one) to robust-t tests closely mimics rejections that are obtained when using the fixed-b approximation. These finite sample patterns are not predicted by the existing Edgeworth expansions theory; in particular, see Davison and Hall (1993), and Götze and Künsch (1996).

Gonçalves and Vogelsang (2011) developed an alternative theory to explain the finite sample patterns reported by Kiefer and Vogelsang (2005). Within the fixed-b asymptotic framework, they showed that a naive moving blocks bootstrap HAC-based t-statistic (by naive we mean that we simply replace the bootstrap data for the original data in the formulae used to construct the test statistic) has the same fixed-b asymptotic distribution as the original test statistic. This explains why the naive block bootstrap replicates the rejections obtained with fixed-b critical values. This is true even when the block size equals one, i.e. when an i.i.d. bootstrap is used, and the data are weakly dependent. The reason is that the fixed-b asymptotic distribution of a studentized statistic does not reflect the dependence structure of the time series. Hence, the bootstrap does not need to mimic this dependence when applied to the studentized sample mean from time series data; the denominator—although inconsistent by itself—takes care of the dependence.

Interestingly, an i.i.d. bootstrap approximation to the distribution of the fixed-b studentized statistic (i.e. a naive i.i.d. bootstrap) correctly reflects the kernel and bandwidth choices. These choices determine the finite sample properties of HAC-based tests and—as previously mentioned—they are completely absent from the normal approximation. Therefore, we may expect the naive i.i.d. bootstrap applied to studentized statistics to outperform the normal approximation. Gonçalves and Vogelsang (2011) showed that this is the case for a very simple location model using the Bartlett kernel. Specifically, relying on the theory of strong approximations, it was shown that the i.i.d. bootstrap has an error in rejection probability (ERP) that may converge to zero faster than the ERP of the standard normal approximation under some conditions on the HAC bandwidth choice and the number of finite moments that exist in the data.

Since the fixed-b asymptotic distribution is more accurate than the standard normal approximation, it can be used as the new benchmark for comparisons with the bootstrap. The naive i.i.d. bootstrap is consistent for this new asymptotic benchmark, but the Monte Carlo simulation results of Gonçalves and Vogelsang (201) clearly suggest that using a larger block size outperforms the fixed-b asymptotic approximation. An interesting topic of further research is to propose a theoretical explanation for this finding, and to then develop a way to optimally choose the block size in connection with fixed-b studentized statistics from time series data.

One of the so-called *"second generation"* bootstrap methods for time series is the tapered block bootstrap (TBB) introduced by Paparoditis and Politis (2001b, 2002a). The data tapering of the blocks used in the TBB is designed to decrease the bootstrap bias, and has as a result an increased accuracy of estimation of sampling characteristics (standard error, distribution, etc.) for linear and approximately linear statistics. Although the studentized TBB has yet to be studied in detail, the unstudentized TBB is the most accurate block bootstrap procedure available at the moment; the TBB's accuracy is better by an order of magnitude as compared to the untapered (usual) block bootstrap and/or the stationary bootstrap.

The TBB was recently revisited by Shao (2010a) who developed the extended tapered block bootstrap (ETBB), and thereby widened the scope of the applicability of the TBB. The main idea of the ETBB is to apply the tapering to the random weights in the bootstrapped empirical distribution. Its asymptotic validity in estimating asymptotic variance and sampling distribution has been shown in Shao (2010a) for the smooth function model; it is also expected to work for a large class of approximately linear statistics such as the sample median. The ETBB is identical to the TBB in the sample mean case, and it is asymptotically equivalent to TBB in the case of the smooth function model. Their difference lies in the way approximately linear statistics are handled; the TBB explicitly uses linearization in forming bootstrapped statistics whereas the ETBB uses it implicitly. The favorable bias and mean squared error properties of the TBB over the moving block bootstrap (MBB) are fully preserved by ETBB.

By and large, the development of bootstrap methods for time series has been focused on equally spaced time series. However, irregularly spaced processes (in time and/or space) occur quite often in real applications. The irregularity could be due to missing observations in an evenly spaced time series or to the observations being available at randomly sampled time points. Interestingly, in the presence of a small fraction of missing observations, a practitioner could conceivably proceed with implementing a block bootstrap procedure by simply ignoring the fact that there are missing values. On the other hand, the presence of a large number of missing values and/or irregularly spaced data must necessarily be addressed. Fortunately, the use of block bootstrap methods is possible and theoretically justifiable in the irregularly spaced setting. Pertinent references include Politis et al. (1998, 1999), Lahiri and Zhu (2006), etc.; see also Hall (1985) for an early treatment of point processes that foreshadows the now popular block resampling methods.

Recently, Shao (2010b) proposed a new resampling procedure, the dependent wild bootstrap (DWB) for stationary time series. The DWB extends the traditional wild bootstrap of Wu (1986) to the time series setting by allowing the auxiliary variables involved in the wild bootstrap to be dependent; hence, the DWB is capable of mimicking the dependence in the original series nonparametrically. Unlike block bootstrap methods, the DWB does not involve a partition of the data into blocks and there is no change when it is implemented with irregular spaced data or missing values. Shao (2010b) provided a rigorous theoretical study of the asymptotic properties of the DWB including: (a) the consistency of distribution approximation under the framework of the smooth function model; (b) bias and variance expansions of the DWB variance estimator for irregularly spaced time series on a lattice; and (c) consistency of the DWB for variance estimation and distribution approximation in the mean case for irregularly spaced non-lattice time series. On the downside, the DWB is not as widely applicable as the block bootstrap (in its many variations), and lacks the higher order accuracy property of the latter.

Additional References

Andrews, D.W.K. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930-934.

Billingsley, P. (1968). *Convergence of probability measures*. New York: Wiley.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.

Carrasco, M. and X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17-39.

Dahlhaus, R. (1996) On the Kulback-Leibler information divergence of locally stationary processes. *Stochastic Processes and Their Applications* 62, 139-168.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Annals of Statistics* 25, 1-37.

Davison, A.C. and P. Hall (1993 On studentizing and blocking methods for implementing the bootstrap with dependent data. *Australian Journal of Statistics* 35, 215-224.

Davidson, J. (2002) Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics* 106, 243-269.

Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1006.

Gallant, A.R. and H. White (1988) *A unified theory of estimation and inference for nonlinear dynamic models*. New York: Basil Blackwell.

Gonçalves, S. and T. Vogelsang (2011). Block bootstrap puzzles in HAC robust testing: the sophistication of the naive bootstrap. Forthcoming in *Econometric Theory*.

Goncalves, S. and H. White (2002). The Bootstrap of the mean for dependent heterogeneous arrays, *Econometric Theory*, 2002, 18, 1367-1384.

Goncalves, S. and H. White (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics*, 119, 199-219.

Goncalves, S. and H. White (2005). Bootstrap standard error estimates for linear regressions, *Journal of the American Statistical Association* Vol. 100, No. 471, 970-979.

Götze, F. and H.R. Künsch (1996) Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics* 24, 1914-1933.

Hansen, B.E. (1991) GARCH(1,1) processes are near epoch dependent. *Economics Letters* 36, 181-186.

Hall, Peter (1985). Resampling a coverage pattern. *Stochastic Process. Appl.* 20, no. 2, 231—246.

Jansson, M. (2004) The error rejection probability of simple autocorrelation robust tests. *Econometrica*, 72, 937-946.

Kiefer, N.M. and T.J. Vogelsang (2005) A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130-1164.

Lahiri, S. N. and Zhu, Jun (2006). Resampling methods for spatial regression models under a class of stochastic designs. *Ann. Statist.* 34, no. 4, 1774—1813.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods, *Ann. Statist.* 27, 386–404. GIVEN IN MAIN ARTICLE–DO NOT REPEAT

McLeish, D.L. (1975) A maximal inequality and dependent strong laws. *Annals of Probability* 5, 616-621.

McElroy, T. and D.N. Politis (2011), 'Fixed-b asymptotics for the studentized mean from time series with short, long or negative memory ', to appear in *Econometric Theory* in 2011.

McMurry , T. and D. N. Politis (2011). Resampling methods for functional data, *The Oxford Handbook of Functional Data Analysis*, (F. Ferraty and Y. Romain, Eds.), Oxford Univ. Press, pp. 189-209.

Nordman, D.J. (2009). A note on the stationary bootstrap, 37, 359–370. *Ann. Statist.* 27, 386–404. GIVEN IN MAIN ARTICLE–DO NOT REPEAT

Paparoditis, E. and Politis, D. N. (2001b). Tapered block bootstrap. *Biometrika* **88**, 1105–1119. GIVEN IN MAIN ARTICLE–DO NOT REPEAT

Paparoditis, E. and Politis, D. N. (2002a). The tapered block bootstrap for general statistics from stationary sequences. *Econometrics Journal* **5**, 131-148. GIVEN IN MAIN ARTICLE–DO NOT REPEAT

Politis, D.N. (2011) , 'Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices', to appear in *Econometric Theory* in 2011.

Politis, D.N., E. Paparoditis, J.P.Romano (1998). Large sample inference for irregularly spaced dependent observations based on subsampling, *Sankhya, Ser. A*, Vol. 60, No. 2, pp. 274-292.

Politis, D.N., E. Paparoditis, J.P.Romano (1999). Resampling Marked Point Processes, in *Multivariate Analysis, Design of Experiments, and Survey Sampling: a Tribute to J. N. Srivastava,* Subir Ghosh (Ed.), Marcel Dekker, Inc., New York, pp. 163-185.

Politis, D.N., and J.P.Romano (1994). The Stationary Bootstrap, *J. Amer. Statist. Assoc.*, vol. 89, pp. 1303-1313. GIVEN IN MAIN ARTICLE–DO NOT REPEAT

Politis, D.N., and J.P.Romano (1994b). Limit Theorems for Weakly Dependent Hilbert Space Valued Random Variables with Applications to the Stationary Bootstrap, *Statistica Sinica*, vol. 4, pp. 461-476.

Politis, D.N., and J.P.Romano (1995). 'Bias-Corrected Nonparametric Spectral Estimation', *J. Time Ser. Anal.,* vol. 16, No. 1, pp. 67-104.

Priestley, M.B. (1988) Non-linear and non-stationary time series analysis. Academic Press, London.

Sun, Y., P.C.B. Phillips and S. Jin (2008) Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* 76, 175-194.

Shao, X. (2010a) Extended tapered block bootstrap. *Statistica Sinica*, 20, 807-821.

Shao, X. (2010b) The dependent wild bootstrap. *Journal of the American Statistical Association*, 105, 218-235. GIVEN IN MAIN ARTICLE–DO NOT REPEAT (BUT PLEASE CORRECT THE DATE OF THIS ARTICLE'S PUBLICATION WHICH IS LISTED INCORRECTLY IN MAIN ARTICLE)

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics*, **14**, 1261-1350. GIVEN IN MAIN ARTICLE–DO NOT REPEAT