arXiv: arXiv:0000.0000

Model-free Bootstrap for a General Class of Stationary Time Series

Yiren Wang, Dimitris N. Politis

Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112, USA
e-mail: yiw518@ucsd.edu; dpolitis@ucsd.edu

Abstract: A model-free bootstrap procedure for a general class of stationary time series is introduced. The theoretical framework is established, showing asymptotic validity of bootstrap confidence intervals for many statistics of interest. In addition, asymptotic validity of one-step ahead bootstrap prediction intervals is also demonstrated. Finite-sample experiments are conducted to empirically confirm the performance of the new method, and to compare with popular methods such as the block bootstrap and the autoregressive (AR)-sieve bootstrap.

1. Introduction

The bootstrap, since its introduction by Efron (1979), has been an invaluable tool for statistical inference with independent data. Resampling for time series has also been a flourishing topic since the late 1980s. However, there are a plethora of ways to resample a stationary time series. It is always important to validate the correctness of such bootstrap procedures, i.e., to show their asymptotic validity and range of applicability with respect to common statistics. These problems have been well studied for popular methods like the block bootstrap and the autoregressive (AR)-sieve bootstrap. For a summary of the state-of-the-art, see McElroy and Politis (2019) and Kreiss and Paparoditis (2020).

In a dependent setup, the main purpose of bootstrap is two-fold: One is to obtain confidence intervals for a parameter of interest and/or conduct a hypothesis test. Another important aspect of time series analysis is forecasting. A standard setup is the following: Given the time series data $\{Y_t\}_{t=1}^n$, the goal is h-step ahead prediction, i.e., predicting Y_{t+h} for some integer $h \geqslant 1$. An optimal h-step ahead point predictor $\widehat{Y}_{t+h}$ should minimize the expected loss between the true Y_{t+h} and itself, conditioned on the current data $\{Y_1, \cdots, Y_n\}$. The most widely used loss functions are L^1 and L^2. The L^2 loss, $E\left(\left(\widehat{Y}_{t+h} - Y_{t+h}\right)^2 | Y_1, \cdots, Y_n\right)$ is minimized by the conditional expectation $\widehat{Y}_{t+h} = E(Y_{n+h}|Y_1, \cdots Y_n)$. The L^1 loss, $E\left(|\widehat{Y}_{t+h} - Y_{t+h}||Y_1, \cdots, Y_n\right)$ is minimized by the conditional median $med(Y_{t+h}|Y_1, \cdots, Y_n)$ instead.

Besides point predictors, prediction intervals and joint prediction regions are quite useful; since any point predictor will invariably incur an error, it is

important to provide a range of values where the future point Y_{t+h} will be found with high probability. Prediction intervals can be constructed by approximating the distribution of the so-called *predictive root*, i.e. $Y_{t+h} - \widehat{Y}_{t+h}$, and using the respective quantiles to produce upper and lower bounds. Approximating this distribution typically requires one to fit a specific model to the data, which enables a model-based resampling for Y_{t+h} and $\widehat{Y}_{t+h}$ separately; see Pan and Politis (2016a) for a review.

However, model-fitting and prediction are two separate notions with very different objective functions. Cross-validation ideas that are currently popular attempt to link the two notions, in choosing a model that is actually good for predictive purposes. Nevertheless, it is possible for the practitioner to proceed directly to prediction without the intermediate step of model-fitting; this is the essence of the *model-free* prediction principle of Politis (2013), Politis (2015). To describe it, the goal is to find an invertible transformation that transforms the data vector $(Y_1, \ldots, Y_n)'$ to a new data vector $(e_1, \ldots, e_n)'$ whose entries are independent and identically distributed (i.i.d.). One can then employ the i.i.d. bootstrap on the $e_1, \ldots, e_m$ to generate $e_1^*, \ldots, e_m^*$, and use the inverse transform to get bootstrap samples $Y_1^*, \ldots, Y_m^*$ in the domain of the original data. Using $m = n$ is the standard framework for estimation and confidence intervals; interestingly, using $m = n + h$ allows us to equally address the problem of forecasting Y_{t+h} with prediction intervals.

Under regularity conditions, such a transformation always exists but is not unique; see Ch. 2.3.3 of Politis (2015). The challenge for the practitioner is to use the structure of the data at hand in order to devise a transformation that works in the given setting, having features that can be estimated from the data. In a model-based approach, these steps are analogous to choosing a model, and then fitting the model using the data. Indeed, any model driven by i.i.d. errors can be used to define a transformation of the data towards the i.i.d. target; however, the power of the model-free approach is that it can work without restricting oneself to a model equation.

To elaborate, if the data arise as a stretch of a strictly stationary time series $\{Y_t\}$ with (absolutely) continuous distributions, then the Rosenblatt transformation (Rosenblatt (1952)) can be used to transform $Y_1, \ldots, Y_n$ to a set of n i.i.d. Uniform random variables. In general, this application of the Rosenblatt transformation can not be implemented in practice because it involves n unknown conditional distribution functions. However, if additional structure is assumed, e.g., when $\{Y_t\}$ is a stationary Markov sequence, then this approach is feasible; see Pan and Politis (2016b) and Ch. 8 of Politis (2015).

To describe a different approach, recall the Linear Process Bootstrap (LPB) of McMurry and Politis (2010) which essentially transforms the the data vector $(Y_1, \ldots, Y_n)'$ to a data vector $(W_1, \ldots, W_n)'$ that has uncorrelated entries, i.e., $\{W_t\}$ is a 'white noise'. If $\{Y_t\}$ is a linear time series, then $\{W_t\}$ can further be claimed to be i.i.d. (under some conditions). The LPB has parallels with the AR-sieve bootstrap since both are applicable to nonlinear time series as long as the statistic of interest has a large-sample distribution that only depends on the first and second order moment structure of the data; see Kreiss, Paparoditis and

Politis (2011) and Jentsch and Politis (2015).

Nevertheless, in the search of a transformation that renders the data i.i.d., it may be helpful to first devise a transformation into Gaussianity; see e.g., Ch. 2.3.2 of Politis (2015). For example, we can use a version of the Probability Integral Transform (PIT) in order to transform our time series data to Gaussian; the latter can then be transformed to i.i.d. by a decorrelating/whitening operation as in the LPB. This approach was first suggested in Ch. 9 of Politis (2015), and was practically implemented to the setting of a locally stationary time series by Das and Politis (2020a).

In the paper at hand, we focus on stationary time series data, with the goal of establishing the realm of applicability of the above mentioned procedure which, for lack of a better word, we will call the *model-free bootstrap* (MFB). We will show asymptotic validity of the MFB for a general class of stationary processes, and for many types of statistics of interest. We will also establish MFB's validity for the construction of one-step-ahead prediction intervals, i.e., to fix ideas we will focus on the case $h = 1$ in the above.

The remaining of the paper is organized as follows. Section 2 restates the MFB algorithm carefully. Section 3 introduces some necessary tools and assumptions to be used, and summarizes some useful preliminary results for our proofs. Section 4 proves MFB's asymptotic validity for various estimation problems, while Section 5 shows its validity for prediction intervals. Numerical experiments that back up our asymptotic results are presented in Section 6. Technical proofs are given in the Appendix.

2. Model-free bootstrap algorithm

2.1. The MFB algorithm

Here we describe the model-free bootstrap (MFB) algorithm for inference and prediction as proposed in Chaper 9 of Politis (2015). Given a time series $\{Y_t\}_{t\in\mathbb{Z}}$ that is strictly stationary, let F_Y be the cumulative distribution function (CDF) of Y_0. The PIT defined by

$$U_t = F_Y(Y_t)$$

implies that U_t is uniformly distributed on $[0, 1]$, assuming F_Y is continuous. See Angus (1994). Let Φ be the CDF of standard normal distribution and $\Phi^{-1}(p) = \inf\{x \in \mathbb{R} : \Phi(x) \geqslant p\}$ be the quantile function; then, $Z_t = \Phi^{-1}(U_t)$ is $\mathcal{N}(0, 1)$ distributed. Also, stationarity is preserved for $\{U_t\}$ and $\{Z_t\}$.

Let Σ_n denote the covariance matrix of $\underline{Z}_n = (Z_1, \cdots, Z_n)$, and denote by $\Sigma_n^{-\frac{1}{2}}$ the lower triangular matrix from the Cholesky decomposition of Σ_n^{-1}. Then, $\underline{\xi}_n = \Sigma_n^{-\frac{1}{2}}\underline{Z}_n$ is a vector of i.i.d. $\mathcal{N}(0, 1)$ entries, provided $Z_1, \cdots, Z_n$ are *jointly* normal.

Suppose we use a resampling scheme to create the i.i.d. bootstrap sample $\xi_1^*, \xi_2^*, \cdots, \xi_n^*$. Then, letting $\underline{Z}_n^* = \Sigma_n^{\frac{1}{2}}\underline{\xi}_n^*$ where $\underline{\xi}_n^* = (\xi_1^*, \xi_2^*, \cdots, \xi_n^*)'$, and $Y_t^* = F_Y^{-1}(\Phi(Z_t^*))$, then $\{Y_t^*\}$ is our bootstrapped sample.

Moreover, ξ^*_{n+1} can also be generated through i.i.d. sampling, and Z^*_{n+1} can be generated through the relation $(\underline{Z}_n, Z^*_{n+1}) = \Sigma^{1/2}_{n+1}(\underline{\xi}_n, \xi_{n+1})$. Using the inverse of the previously mentioned transforms, the next bootstrap value can be generated by $Y^*_{n+1} = F_Y^{-1}(\Phi(Z^*_{n+1}))$. It can be shown that by using these theoretical transforms, $Y^*_{n+1}|\underline{Y}_n$ has the same distribution as $Y_{n+1}|\underline{Y}_n$.

Nevertheless, to use the above steps for practical purposes, each transform must also be estimated in a consistent manner from the data at hand. Furthermore, the validity of the bootstrap procedure has to be investigated, both for estimation and prediction. Thus, several questions arise:

- Under what circumstance are the entries of $\underline{Z}_n$ jointly normal?
- What estimators for F_Y and Σ_n should we use so that the above steps lead to validity of the bootstrap?
- How should we create the i.i.d. bootstrap values $\{\xi^*_t\}$?

The first two points will be addressed in the following paragraphs. For the third point, Politis (2015) has proposed two ways to do it. One way is sampling with replacement from $\{\widehat{\xi}_t\}_{t=1}^n$, with $\widehat{\xi}_t$ calculated from Y_t using estimated transform functions. A second way is to generate ξ^*_t as i.i.d. $\mathcal{N}(0,1)$, which is presumably the limiting distribution of $\widehat{\xi}_t$. The first method is called **model-free** (MF), and the second is referred to as **limit model-free** (LMF) since the limit distribution is used.

Frequently used notations include the following. Let $\widehat{F}$ and $\widehat{\Sigma}_n$ denote general estimators for F_Y and Σ_n respectively. The subscript Y is dropped from $\widehat{F}$ for simplicity. Φ is the CDF of a standard normal distribution with Φ^{-1} its quantile function. Let $\tilde{\Phi}$ be the CDF of a thresholded standard normal distribution: suppose $X \sim \mathcal{N}(0,1)$, $X_c = X$ for $|X| \leqslant c$ and $X_c = sgn(X)c$ for $|X| > c$, where $sgn(\cdot)$ is the sign function. Then $\tilde{\Phi}$ denotes the CDF of X_c and its inverse $\tilde{\Phi}^{-1}$ the quantile function. We omit c in the notation for simplification. Asymptotically we also require $c \to \infty$ such that $\tilde{\Phi}^{-1}$ converges to Φ^{-1}. The reason of this augmentation is provided in Section 2.4 and asymptotic details are explained in Section 3.

By using these practical transforms, we can calculate $\widehat{U}_t = \widehat{F}(Y_t)$, and $\tilde{\widehat{Z}}_t = \tilde{\Phi}^{-1}(\widehat{U}_t)$, which are the estimations for the latent series $\{U_t\}$ and $\{Z_t\}$ respectively. Since $\{Z_t\}$ is latent, $\widehat{\Sigma}_n$ can not be directly calculated. Instead, we use $\tilde{\widehat{\Sigma}}_n$ which is the same estimator calculated based on $\{\tilde{\widehat{Z}}_t\}$.

Let $\sigma_Z(k) = EZ_0Z_k$ be the lag-k autocovariance of Z_t, $\widehat{\sigma}_Z(k)$ be its estimator and $\tilde{\widehat{\sigma}}_Z(k)$ be the estimator calculated from $\tilde{\widehat{Z}}_t$. Let $\|\cdot\|_p = E(|\cdot|^p)^{\frac{1}{p}}$ denote the $p-norm$ of a random variable; $\|\cdot\|_{op}$ denotes the operator norm of a matrix, i.e., $\|M\|_{op} = \sup_{x\in\mathbb{R}^n, \|x\|_2=1} \|Mx\|_2$ where M is a $n \times n$ square matrix. Relative quantities in the bootstrap world will be denoted by a superscript $*$.

Given the above introduction, we can now describe the model-free bootstrap algorithm.

2.2. MFB for confidence intervals

Let θ_0 be a population parameter of interest, $\widehat{\theta}_n$ an estimator of θ_0 from data $\{Y_t\}_{t=1}^n$, and $\widehat{\theta}_n^*$ the same estimator from bootstrapped data $\{Y_t^*\}_{t=1}^n$. Define the real-world *root* $r = \theta_0 - \widehat{\theta}_n$; let R denote its CDF. Then, a $(1-\alpha)100\%$ equal-tailed confidence interval (CI) for θ_0 is

$$(\widehat{\theta}_n + R^{-1}(\alpha/2), \widehat{\theta}_n + R^{-1}(1-\alpha/2)),$$

where $R^{-1}(x) = \inf\{r \in \mathbb{R} : R(r) \geqslant x\}$ denotes the quantile function of R. The distribution R could be approximated through bootstrap simulations.

Algorithm 1 (Model-free bootstrap for parameter inference)

1. *Given data* $\{Y_t\}_{t=1}^n$, *let* $\widehat{U}_t = \widehat{F}(Y_t)$; $\tilde{\widehat{Z}}_t = \tilde{\Phi}^{-1}(\widehat{U}_t)$; $\underline{\widehat{\xi}}_n = \widehat{\tilde{\Sigma}}_n^{-\frac{1}{2}} \underline{\tilde{\widehat{Z}}}_n$.
2. *(MF) Let* ξ_t^* *be i.i.d. samples from* $\bar{F}_{\widehat{\xi}}$, *where* $\bar{F}_{\widehat{\xi}}$ *is the empirical CDF of* $\{\widehat{\xi}_t\}_{t=1}^n$, *and* $\underline{Z}_n^* = \widehat{\tilde{\Sigma}}_n^{\frac{1}{2}} \underline{\xi}_n^*$. *Let* $Y_t^* = \widehat{F}^{-1}(\Phi(Z_t^*))$. *Calculate the bootstrap root* $r^* = \widehat{\theta}_n - \widehat{\theta}_n^*$.
3. *Do the above step* B *times to form an empirical CDF* $\bar{R}$ *based on the* B *replicates of* r^*. $\bar{R}$ *is used to approximate* R; *hence, an approximate* $(1-\alpha)100\%$ *CI for* θ_0 *is*

$$(\widehat{\theta}_n + \bar{R}^{-1}(\alpha/2), \widehat{\theta}_n + \bar{R}^{-1}(1-\alpha/2)).$$

For the limit model-free (LMF) bootstrap, replace step 2 with following:

2. (LMF) Let ξ_t^* *be i.i.d. samples from* $\mathcal{N}(0,1)$, *and* $\underline{Z}_n^* = \widehat{\tilde{\Sigma}}_n^{\frac{1}{2}} \underline{\xi}_n^*$. *Let* $Y_t^* = \widehat{F}^{-1}(\Phi(Z_t^*))$. *Calculate the bootstrap root* $r^* = \widehat{\theta}_n - \widehat{\theta}_n^*$. *This is equivalent to sample* $\underline{Z}_n^*$ *from* $\mathcal{N}(\mathbf{0}, \widehat{\tilde{\Sigma}}_n)$ *distribution.*

Below are some simple examples of statistics of interest.

- The mean: $\theta_0 = E(Y_0)$; $\widehat{\theta}_n = \frac{\sum_{t=1}^n Y_t}{n}$.
- Autocovariance: $\theta_0 = \gamma_Y(k)$; $\widehat{\theta}_n = \widehat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(Y_t - \bar{Y}_n)(Y_{t+k} - \bar{Y}_n)$.
- Autocorrelation: $\theta_0 = \rho(k) = \frac{\gamma(k)}{\gamma(0)}$; $\widehat{\rho}(k) = \frac{\widehat{\gamma}(k)}{\widehat{\gamma}(0)}$.

And additional examples will be addressed in Section 4.

2.3. MFB for prediction intervals

For prediction problems, we want to use the bootstrap to simulate the distribution of Y_{n+1} **conditional** on past values $\{Y_t\}_{t=1}^n$. For this purpose, we use bootstrap to approximate the conditional distribution of the predictive root $Y_{n+1} - \widehat{Y}_{n+1}$, where $\widehat{Y}_{n+1}$ is a predictor chosen by the practitioner. Let G denote the conditional distribution of the predictive root defined above. Then a $(1-\alpha)100\%$ equal tailed prediction interval for Y_{n+1} is :

$$(\widehat{Y}_{n+1} + G^{-1}(\alpha/2), \widehat{Y}_{n+1} + G^{-1}(1-\alpha/2))$$

Model-free bootstrap algorithm for one-step ahead prediction is the following:

Algorithm 2 (Model-free bootstrap for one-step ahead prediction)

1. *Given data* $\{Y_t\}_{t=1}^n$, *let* $\widehat{U}_t = \widehat{F}(Y_t)$, $\hat{\tilde{Z}}_t = \tilde{\Phi}^{-1}(\widehat{U}_t)$.
2. *Denote* $\hat{\hat{\Sigma}}_{11} = \hat{\hat{\Sigma}}_n$, $\hat{\hat{\Sigma}}_{12} = \begin{bmatrix} \hat{\hat{\sigma}}(n) \\ \vdots \\ \hat{\hat{\sigma}}(1) \end{bmatrix}$, *and* $\hat{\hat{\Sigma}}_{22} = \hat{\hat{\sigma}}(0)$. *Let* $\hat{\hat{\Sigma}}_{n+1} = \begin{bmatrix} \hat{\hat{\Sigma}}_{11} & \hat{\hat{\Sigma}}_{12} \\ \hat{\hat{\Sigma}}_{12}^T & \hat{\hat{\Sigma}}_{22} \end{bmatrix}$.

   Let $\{\xi_t^*\}_{t=1}^{n+1}$ *be drawn randomly with replacement from* $\{\widehat{\xi}_t\}_{t=1}^n$, *and* $\underline{Z}_n^* = \hat{\hat{\Sigma}}_n^{\frac{1}{2}} \underline{\xi}_n^*$. *Let* Z_{n+1}^* *be the* $(n+1)^{th}$ *element of the vector* $\hat{\hat{\Sigma}}_{n+1}^{\frac{1}{2}}(\widehat{\xi}_n, \xi_{n+1}^*)$. *Denote the distribution of* Z_{n+1}^* *as* $\widehat{F}_Z^{(n+1)}$. *This is also the estimated conditional distribution of* $Z_{n+1}|\underline{Y}_n$. *The form of this distribution is conditional on our data* $\underline{Y}_n = (Y_1, \cdots, Y_n)$. *Let* $Y_{n+1}^* = \widehat{F}^{-1}(\Phi(Z_{n+1}^*))$.
3. *Choose a predictor* $\widehat{Y}_{n+1}$ *for* Y_{n+1} *based on* $\underline{Y}_n$. *For example, the* L^2 *optimal predictor as mentioned in Section 1 is the expectation of* Z_{n+1} *conditioning on* $\underline{Y}_n$ *that can be approximated by*

$$\widehat{Y}_{n+1} = \int \widehat{F}^{-1}(\Phi(z)) d\widehat{F}_Z^{(n+1)}(z).$$

   The above integral can be evaluated through Monte-Carlo simulation. The chosen predictor will be used as the center of our prediction interval, and the bootstrap procedure will be used to capture the distribution of the predictive root in the next steps.
4. *Re-estimate all the transforms, matrices and the distribution* $\widehat{F}_Z^{(n+1)}$ *used in the above calculation, with bootstrapped data* $\underline{Z}_n^* = (Z_1^*, \cdots, Z_n^*)$ *and* $Y_t^* = \widehat{F}^{-1}(\Phi(Z_t^*))$. *Let* $(\widehat{F}_Z^{(n+1)})^*$ *denote the re-estimated distribution function for* Z_{n+1}^* *with bootstrap data* $\underline{Y}_n^* = (Y_1^*, \cdots, Y_n^*)$. *Let* $(Z_{n+1}|Y_n)^*$ *denote the random variable with estimated conditional distribution* $(\widehat{F}_Z^{(n+1)})^*$.

   Let $\widehat{Y}_{n+1}^*$ *denote the one-step ahead predictor with re-estimated transforms based on the bootstrap pseudo data. In the* L^2*-optimal setting,* $\widehat{Y}_{n+1}^* = \int (\widehat{F}^*)^{-1}(\Phi(z)) d(\widehat{F}_Z^{(n+1)})^*(z)$.
5. *The bootstrapped* L^2- *optimal predictive root is:*

$$Y_{n+1}^* - \widehat{Y}_{n+1}^*$$

6. *Denote the empirical CDF of bootstrapped predictive roots as* $\bar{G}$. *The approximate* $(1-\alpha)$ *prediction interval for* Y_{n+1} *is*

$$\left(\widehat{Y}_{n+1} + \bar{G}^{-1}(\alpha/2), \widehat{Y}_{n+1} + \bar{G}^{-1}(1-\alpha/2)\right).$$

Algorithm 3 (Limit model-free bootstrap for 1 step ahead prediction)

1. *Given data* $\{Y_t\}_{t=1}^n$, *let* $\widehat{U}_t = \widehat{F}(Y_t)$, $\hat{\tilde{Z}}_t = \tilde{\Phi}^{-1}(\widehat{U}_t)$.
2. *(LMF) Denote* $\hat{\hat{\Sigma}}_{11} = \hat{\hat{\Sigma}}_n$, $\hat{\hat{\Sigma}}_{12} = \begin{bmatrix} \hat{\hat{\sigma}}(n) \\ \vdots \\ \hat{\hat{\sigma}}(1) \end{bmatrix}$. *Let* $\underline{Z}_n^* \sim \mathcal{N}(0, \hat{\hat{\Sigma}}_n)$; $Z_{n+1}^* \sim \mathcal{N}(\hat{\hat{\Sigma}}_{21}\hat{\hat{\Sigma}}_{11}^{-1}\hat{\tilde{Z}}_n, \hat{\hat{\Sigma}}_{22} - \hat{\hat{\Sigma}}_{21}\hat{\hat{\Sigma}}_{11}^{-1}\hat{\hat{\Sigma}}_{12})$; $Y_{n+1}^* = \widehat{F}^{-1}(\Phi(Z_{n+1}^*))$
3. *Choose a predictor for* Y_{n+1} *based on* $\underline{Y}_n$. *For example, the* L^2 *optimal predictor is*

$$\widehat{Y}_{n+1} = E(\widehat{F}^{-1}(\Phi(Z_{n+1}))|\underline{Y}_n).$$

   Where $Z_{n+1}|\underline{Y}_n \sim \mathcal{N}(\hat{\hat{\Sigma}}_{21}\hat{\hat{\Sigma}}_{11}^{-1}\hat{\tilde{Z}}_n, \hat{\hat{\Sigma}}_{22} - \hat{\hat{\Sigma}}_{21}\hat{\hat{\Sigma}}_{11}^{-1}\hat{\hat{\Sigma}}_{12})$.

4. *Re-estimate all the transforms and matrices used in the above calculation, with* $\underline{Z}_n^*$ *and* $Y_t^* = \widehat{F}^{-1}(\Phi(Z_t^*))$. *The one-step ahead predictor in the bootstrap world is* $\hat{Y}_{n+1}^* = E^*((\widehat{F}^*)^{-1}(\Phi(Z_{n+1}))|\underline{Y}_n)$, *where the expectation in the bootstrap world is calculated through the distribution of* $(Z_{n+1}|\underline{Y}_n)^*$ *that is* $\mathcal{N}(\hat{\widehat{\Sigma}}_{21}^*(\hat{\widehat{\Sigma}}_{11}^*)^{-1}\hat{\tilde{\underline{Z}}}_n, \hat{\widehat{\Sigma}}_{22}^* - \hat{\widehat{\Sigma}}_{21}^*(\hat{\widehat{\Sigma}}_{11}^*)^{-1}\hat{\widehat{\Sigma}}_{12}^*)$ *distributed.*
5. *The bootstrapped* L^2- *optimal predictive root is:*

$$Y_{n+1}^* - \hat{Y}_{n+1}^*$$

6. *Denote the empirical CDF of bootstraped predictive root as* $\bar{G}$. *The approximate* $(1-\alpha)$ *prediction interval for* Y_{n+1} *is*

$$\left(\hat{Y}_{n+1} + \bar{G}^{-1}(\alpha/2), \hat{Y}_{n+1} + \bar{G}^{-1}(1-\alpha/2)\right).$$

Here we provide an explanation to step *5* above. The bootstrap is supposed to capture the distribution of the predictive root $Y_{n+1} - E(\widehat{F}^{-1}(\Phi(Z_{n+1}))|\underline{Y}_n)$, where for $Y_{n+1} = F_Y^{-1}(\Phi(Z_{n+1}))$, Z_{n+1} has the conditional distribution $\mathcal{N}(\Sigma_{21}\Sigma_{11}^{-1}\underline{Z}_n, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$; and in the expectation $Z_{n+1}|\underline{Y}_n \sim \mathcal{N}(\hat{\widehat{\Sigma}}_{21}\hat{\widehat{\Sigma}}_{11}^{-1}\hat{\tilde{\underline{Z}}}_n, \hat{\widehat{\Sigma}}_{22} - \hat{\widehat{\Sigma}}_{21}\hat{\widehat{\Sigma}}_{11}^{-1}\hat{\widehat{\Sigma}}_{12})$ is estimated from data. Clearly, the randomness in the distribution of the predictive root not only comes from the randomness of the series, but also randomness in the estimation. Thus in the bootstrap world, we should replace theoretical transforms with their data-dependent analogue, and also account for all the errors arising from estimation, i.e. replace F_Y with $\widehat{F}$, $\widehat{F}$ with $\widehat{F}^*$, and $\hat{\widehat{\Sigma}}_n$ with $\hat{\widehat{\Sigma}}_n^*$, resulting in the formula in step *5*. In this way, the bootstrap procedure can capture all the randomness in the distribution of predictive root, which helps relieve potential *undercoverage* issue for finite data. See also Ch. 9 of Politis (2015).

The above algorithms require the practitioner to select the parameter c in the threshold normal distribution in order to calculate the inverse $\tilde{\Phi}^{-1}$. Section 2.4 gives a detailed reasoning for this procedure. In the simulation section, we provide an easy to implement rule for c based on the number of data and choice of CDF estimator $\widehat{F}$, which is used to generate the numerical results. As we will show in Section 4, to establish asymptotic results, we also need c to diverge at a rate that depends on the sample size n.

Computationally, the MFB has two procedures that require intensive computational resources. First of all, the time complexity of finding the Cholesky decomposition is $O(n^2)$ for Toeplitz matrices. Secondly, because $\widehat{F}^{-1}$ usually does not have a closed-form formula, for example, when $\widehat{F}$ is the inverse kernel CDF estimator, computing the value $\widehat{F}^{-1}(U_t^*)$ is usually performed through numerical interpolation of a range of sample points $\{(\widehat{F}(x_i), x_i)\}$. This procedure is time-costly as sample size grow larger. It is important to mention that MFB for prediction intervals is more time consuming than for confidence intervals by a factor of B, since we need to re-estimate the transforms and autocovariance matrix (and its Cholesky decomposition) for each bootstrap sample.

2.4. Appropriate estimators for $\widehat{F}$, $\tilde{\Phi}^{-1}$ and $\widehat{\widehat{\Sigma}}_n$

Now we discuss what should be the appropriate estimators $\widehat{F}$, $\widehat{\widehat{\Sigma}}_n$ and also why an augmented version of Φ^{-1} might be needed. Firstly, it is necessary that $\widehat{F}$, $\widehat{F}^{-1}$, $\widehat{\widehat{\Sigma}}_n$ should be consistent in certain forms for F_Y, F_Y^{-1} and Σ_n respectively. For $\widehat{F}$, the first idea is to use the empirical CDF $\bar{F}(y) = \frac{1}{n}\sum_{t=1}^{n} I\{Y_t \leqslant y\}$, where $I\{\cdot\}$ is the indicator function, and its inverse $\bar{F}^{-1}(p) = \inf\{y \in \mathbb{R} : \bar{F}(y) \geqslant p\}$. Under moment and short-range dependence assumptions, consistency of $\bar{F}$ and $\bar{F}^{-1}$ can be established by looking into the empirical process and quantile process. Details are in later sections and will play an important role in our proofs.

Another natural candidate is the kernel smoothed CDF estimator $\widehat{F}_h(y) = \frac{1}{n}\sum_{t=1}^{n} K_h(y - Y_t)$, where $K_h(y - Y_t) = K(\frac{y-Y_t}{h})$ and K is a smooth CDF function with additional assumptions. The obvious advantage of $\widehat{F}_h$ is that it is continuous, which is a property $\bar{F}$ is lacking. An additional implication of using $\bar{F}$ is the resulting $\widehat{U}_t = \bar{F}(Y_t)$ only takes value in $\{\frac{1}{n}, \frac{2}{n}, \cdots, 1\}$ and $Y_t^* = \bar{F}^{-1}(U_t^*)$ only takes value in $\{Y_t\}_{t=1}^n$. But by using the kernel estimator $\widehat{F}_h$ and its inverse, Y_t^* can take values that did not appear in the original series. If the data size n is large, the influence of this is minimal; whereas if n is small, $\widehat{F}_h$ is a better estimator because of its ability to interpolate unseen values compared to the coarse behavior of $\bar{F}$. It is also worth mentioning that when n is large, using $\bar{F}$ and its inverse will save computational time comparing to $\widehat{F}_h$.

Nevertheless, using $\bar{F}$ in the first step of the Model-free procedure, i.e., letting $\widehat{U}_t = \bar{F}(Y_t)$ will result in one of the variables $\widehat{U}_1, \ldots, \widehat{U}_n$ taking the value 1 in which case the inverse transform $\Phi^{-1}(1)$ in the second step is not well defined. To address this issue, we can use an augmented version of the normal quantile denoted by $\tilde{\Phi}^{-1}$, which is the inverse CDF of a thresholded standard normal $\mathcal{N}_c(0,1)$ as defined in Section 2.1. By doing this, $\tilde{\Phi}^{-1}$ is bounded on $[0,1]$, which relieves this problem.

The purpose of using $\tilde{\Phi}^{-1}$ goes beyond being an ad-hoc fix. It further plays an important rule in asymptotic analysis of the bootstrap algorithm, in that the thresholding helps control the fast diverging behavior of Φ^{-1} near the endpoints 0 and 1, thereby bounding the extreme values of $\widehat{\tilde{Z}}_t$ by a controllable rate of divergence . In turn, this enables the key step of analyzing convergence of the covariance estimator $\widehat{\widehat{\Sigma}}_n$ derived from $\widehat{\tilde{Z}}_t$, for both $\bar{F}$ and $\widehat{F}_h$ scenarios.

Consistent estimator $\widehat{\Sigma}_n$ of the autocovariance matrix Σ_n has been well studied. Wu and Pourahmadi (2009) established the first result on consistency of a banded matrix estimator. Here we shall use the more general flat-top estimators of McMurry and Politis (2010). Let $\kappa(x)$ be the tapering weight function:

$$\kappa(x) = \begin{cases} 1 & \text{if } |x| \leqslant 1 \\ g(|x|) & \text{if } 1 < |x| \leqslant c_\kappa \\ 0 & \text{if } |x| > c_\kappa \end{cases} \tag{2.1}$$

where $|g(x)| < 1$. The most commonly used flat-top kernel is defined by

$$\kappa(x) = \begin{cases} 1 & \text{if } |x| \leqslant 1 \\ 2 - |x| & \text{if } 1 < |x| \leqslant 2 \\ 0 & \text{if } |x| > 2; \end{cases} \tag{2.2}$$

see Politis and Romano (1995). Let l be the bandwidth of choice and $\kappa_l(x) = \kappa(x/l)$. The tapered estimator $\widehat{\Sigma}_n = \widehat{\Sigma}_n^{(\kappa,l)}$ at entry (i,j) has value:

$$\widehat{\Sigma}_n^{(\kappa,l)}(i,j) = \kappa_l(k) \left(\frac{1}{n} \sum_{t=1}^{n-k} Z_t Z_{t+k} \right), \tag{2.3}$$

where $|i-j| = k$.

Remark 1. It is worth noting that the above estimator is not guaranteed to be positive definite (PD) for finite samples, but can be corrected towards PD by looking into its Cholesky decomposition; for details see McMurry and Politis (2010), McMurry and Politis (2015). Asymptotically, the estimator is PD with probability tending to one, and the corrected estimator enjoys the same rate of convergence as the original estimator. Either way, it is not a problem in asymptotic studies.

While convergence results exist for $\widehat{\Sigma}_n$, the true series $\{Z_t\}$ is latent and can not be used in the calculation of $\widehat{\Sigma}_n$. We can only use the estimator with the estimated $\tilde{\hat{Z}}_t$ defined in Section 2.1 that are calculated from original data Y_t, resulting in an estimator $\widehat{\tilde{\Sigma}}_n$, of which the entries are given by

$$\widehat{\tilde{\Sigma}}_n^{(\kappa,l)}(i,j) = \kappa_l(k) \left(\frac{1}{n} \sum_{t=1}^{n-k} \tilde{\hat{Z}}_t \tilde{\hat{Z}}_{t+k} \right). \tag{2.4}$$

Consistency of $\widehat{\tilde{\Sigma}}_n$ to Σ_n will be shown in the following.

2.5. Connection with previous work on MFB for time series

The idea of the model-free bootstrap for time series was first proposed in the monograph Politis (2015), as well as in Das and Politis (2020a). The emphasis of Das and Politis (2020a) is on applying the model-free Prediction Principle to provide a novel statistical procedure for prediction inference of locally stationary data. It has more emphasis on methodology, simulation studies, and real data. By contrast, our emphasis is to build a framework for establishing theoretical properties for the MFB under the stationarity assumption, both for inference (i.e., confidence intervals) as well as for prediction purposes (i.e., prediction intervals). This provides grounds and theoretical guarantees for the MFB algorithm, which has not been addressed in previous works. While the assumptions made in this paper do not cover the locally stationary scenario in Das and Politis (2020a), we are actively working on establishing theoretical results for locally stationary time series with the help of recent advances in this field.

3. Assumptions and preliminary results

3.1. Acceptable forms of $\{Y_t\}$

The stationary series $\{Y_t\}$ in our setting will be assumed to have the form:

$$Y_t = f(W_t) \tag{3.1}$$

where f is some continuously differentiable function such that the CDF F_Y of Y_t is strictly increasing and continuously differentiable, and $\{W_t\}$ is a strictly stationary Gaussian process. Without loss of generality, we may assume that $EW_t = 0$ since the mean of W_t can be incorporated into the function f.9

Equation (3.1) is a common form of extension from Gaussian series to the non-Gaussian case. It has been used in the study of long range dependence, as well as in analyzing time series with heavy tails; see Samorodnitsky and Taqqu (1994). This assumption also figures in a completely different setting, namely that of Bayesian machine learning; see Snelson, Ghahramani and Rasmussen (2004).

By the Wold decomposition (see Brockwell and Davis (1991)) coupled with Gaussianity, W_t admits the expansion

$$W_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j} + V_t \tag{3.2}$$

where ϵ_t are i.i.d. $\mathcal{N}(0,1)$ distributed, and V_t is a deterministic process independent from $\{\epsilon_t\}$. If we assume W_t is purely nondeterministic, then V_t vanishes, and it is clear that Y_t is of the form $Y_t = h(\cdots, \epsilon_{t-1}, \epsilon_t)$ for some function h. Such a representation naturally appears in many time series and dynamical system models, and is also a common form used for developing short range dependence conditions; see Wu (2005a) for details.

The importance and necessity of equation 3.1 and the conditions on f and F_Y can be explained by the following lemma:

Lemma 3.1. *Let $Z_t = \Phi^{-1} \circ F_Y(Y_t)$. Then $Z_1, \cdots, Z_n$ are jointly normal if and only if Y_t admits the representation in equation (3.1).*

Proof. If (3.1) holds, $Z_t = \Phi^{-1} \circ F_Y \circ f(W_t)$. Since both W_t and Z_t are normally distributed, $\Phi^{-1} \circ F_Y \circ f$ is a normality preserving, continuously differentiable transform. Therefore the transform is linear by Corollary 2 of Mase (1977). As a result, each Z_t is linearly transformed from W_t, $t \in \{1, 2, \cdots, n\}$. Collecting the n instaneous linear transforms, we obtain a transform of the vector $(W_1, \cdots, W_n)$ to $(Z_1, \cdots, Z_n)$ that is linear. Since $(W_1, \cdots, W_n)$ are jointly normal and a linear transform preserves normality, $(Z_1, \cdots, Z_n)$ is multivariate normal; see Lemma 3.1 of Das and Politis (2020a).

Conversely, given $Z_t = \Phi^{-1} \circ F_Y(Y_t)$ is jointly normal and strictly stationary, we have $Y_t = F_Y^{-1} \circ \Phi(Z_t)$, which is of the form (3.1). □

Remark 2. The above lemma essentially clarifies the class of time series for which the model-free algorithm is applicable. Under condition (3.1), we can recover the standardized underlying Gaussian process by the transform $Z_t = \Phi^{-1} \circ F_Y(Y_t)$ that results in $Z_t = \frac{W_t - \mu_W}{\sigma_W}$, where μ_W and σ_W are the mean and standard deviation of the hidden Gaussian process W_t.

Under some assumptions, it is indeed possible to estimate the monotone transfer function f, see e.g. Zhang et al. (2019). We could then estimate the unobserved process W_t, in an analogous way that regression residuals are estimates of the underlying (and unobserved) errors. From the estimates of W_t, we can construct estimates of its statistical features can be estimated, e.g., its covariance structure. However, consistency of these estimates would not automatically imply the bootstrap validity we need. Since knowledge of f and of the estimates of W_t is not really needed in performing the MFB procedures, we have chosen to work with the transformations detailed here, and focus on proving bootstrap validity instead.

Since $F_Y^{-1} \circ \Phi$ is monotone, it is reasonable to believe that the function f is monotone. In fact, $f(\cdot)$ is equivalent to $F_Y^{-1} \circ \Phi(\cdot)$ modulo affine differences in their arguments. Thus we have the following lemma:

Lemma 3.2. *Under the setup in equation (3.1), given F_Y is strictly increasing, then f is a strictly monotone function.*

One large subclass of strictly monotone, continuously differentiable functions is functions with derivatives bounded away from 0. As it turns out, this subclass can simplify our short range dependence assumption to be introduced next.

3.2. Short range dependence (SRD) assumptions on $\{Y_t\}$

The SRD assumption is a necessary tool typically assumed to establish statistical results under dependent settings. In the context of our bootstrap scheme, the following types of results are necessary for carrying out proof of bootstrap validity:

1. Consistency of estimated transforms can be established in proper mathematical forms.
2. Certain central limit theorems for the statistic of interest can be established.
3. Consistency of banded/tapered sample autocovariance matrix and CLT variance in the bootstrap world, as further explained in Section 4.

One of the most widely used SRD assumptions for strictly stationary time series is the strong mixing condition introduced by Rosenblatt (1956) and extensively studied since. Many useful results are available for series under certain mixing rates. However such conditions are often hard to verify for general time series models; moreover, the mixing conditions were not sufficient for the proof of banded/tapered autocovariance matrix estimator. In Wu (2005a), a new type of dependence measure named physical dependence measure was introduced which

is as follows. Assume $Y_t = h(\cdots, \epsilon_{t-1}, \epsilon_t) \in L^p$ with $\epsilon_t \overset{i.i.d.}{\sim} F_\epsilon$, then the physical dependence measure is defined by

$$\delta_p(j) = \left\| Y_j - Y_j' \right\|_p,$$

where $Y_j' = h(\cdots, \epsilon_{-1}, \epsilon_0', \cdots, \epsilon_j)$ with $\epsilon_0' \sim F_\epsilon$ independent of $\{\epsilon_t\}$. The series Y_t is called strongly $p-$ stable if

$$(C0) \quad \Delta_p = \sum_{j=0}^{\infty} \delta_p(j) < \infty.$$

The p-stable assumption was able to provide the first results on the convergence of banded and/or tapered autocovariance matrix estimators by Wu and Pourahmadi (2009) and McMurry and Politis (2010).

Alas, functional central limit theorem results for the empirical process–which relates to uniform consistency of CDF estimators cannot be readily established with the physical dependence measure, and require further conditions. To simplify assumptions, we hope for a condition that offers more flexibility than what is mentioned above. One short range dependence measure that is gaining interest is the $m-approximation$ assumption developed in a series of papers in Berkes, Hörmann and Schauer (2009), Berkes, Hörmann and Schauer (2011) and Hörmann and Kokoszka (2010).

The following related conditions are used for proving different properties and are very important in our proofs.

Definition 3.1. The following different $m-approximation$ conditions are proposed. Let $\{Y_t\}$ be a strictly stationary time series, there exists an $m-$ dependent series $\{Y_t^{(m)}\}$ that

(C1) $\forall t \in \mathbf{Z}$, $P(|Y_t - Y_t^{(m)}| > \gamma_m) < \delta_m$, *for some sequence* $\gamma_m \to 0$ *and* $\delta_m \to 0$.

(C2) both Y_t *and* $Y_t^{(m)}$ *are in* L^p; $\exists \delta(m) : \mathbb{N} \to \mathbb{R}_+$, *satisfying* $\delta(m) \ll m^{-A}$ *for some* $A > 0$, *and such that* $\left\| Y_t - Y_t^{(m)} \right\|_p \leqslant \delta(m)$.

(C3) both Y_t *and* $Y_t^{(m)}$ *are in* L^p; $\sum_{m=0}^{\infty} \left\| Y_t - Y_t^{(m)} \right\|_p < \infty$.

Here, the notation "$a_n \ll b_n$" means $\limsup \left| \frac{a_n}{b_n} \right| \to 0$.

Assumption (C1) was introduced in Berkes, Hörmann and Schauer (2009) to establish asymptotic behavior of empirical process. (C2) was introduced in Berkes, Hörmann and Schauer (2011) to establish an invariance principle for partial sums. Both are relevant in our setting. (C3) appeared in Hörmann and Kokoszka (2010) in the context of functional time series. While it is not directly related to the setup here, there is an important relation between (C3) and the physical dependence measure (C0), under which consistency of tapered autocovariance matrix estimator can be established. A detailed layout of the relative theorems mentioned above that are necessary for our proofs can be found in the supplementary material.

There are various methods to construct the m-dependent sequence $Y_t^{(m)}$. Typically one can use truncation, substitution and coupling methods on the representation $Y_t = h(\cdots, \epsilon_{t-1}, \epsilon_t)$; for details refer to Berkes, Hörmann and Schauer (2009). The coupling construction is the most desirable: we replace ϵ by i.i.d. independent copies of ϵ' for times that are at least m steps away from current time t, i.e. let $Y_t^{(m)} = h(\cdots, \epsilon'_{t-m-1}, \epsilon'_{t-m}, \epsilon_{t-m+1}, \cdots, \epsilon_t)$. The resulting $Y_t^{(m)}$ is $m-$ dependent, and also has the advantage that it has the same distribution and moments as Y_t. This construction along with (C2) were also used in Wu (2005b) where it was called the geometric-moment contracting property, assuming a faster geometric decay rate. We will use the coupling construction in what follows.

The above SRD conditions are related; the following lemma clarifies.

Lemma 3.3. *Suppose that $\{Y_t^{(m)}\}$ is constructed by coupling as defined above, then:*

1. Assume (C2) with $A = C(\frac{1}{p} + \frac{1}{\theta})$ for some $C > 0$, $\theta \in (0, 1)$; then (C1) holds with

$$P(|Y_t - Y_t^{(m)}| > m^{-C/\theta}) < m^{-C}.$$

2. Assume (C2) with $A > 1$; then (C3) holds.

3. Assume (C3); then (C0) holds.

4. (C2) is preserved (with a new rate) under θ-Lipschitz transforms. To elaborate, let g be a $\theta-$Lipschitz function, i.e., for some constant K and $\theta \in (0, 1]$, $|g(x) - g(y)| \leqslant K|x - y|^\theta$. If $\left\|Y_t - Y_t^{(m)}\right\|_p \leqslant \delta(m)$ with $\delta(m) \ll m^{-A}$, then

$$\left\|g(Y_t) - g(Y_t^{(m)})\right\|_p \leqslant K\delta(m)^\theta \ll m^{-\theta A}.$$

Proof. See Appendix. □

Based on the above lemma, it is convenient that we only assume condition (C2) with the appropriate rate, such that the remaining dependence measures and the results they imply can be derived from it. We summarize our first set of assumptions as follows.

Assumption Set 1. ***Model setup and weak dependence assumptions***

(A1) $Y_t = f(W_t)$, where $f : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function, and W_t is a (zero mean) stationary Gaussian process with spectral density bounded and strictly bounded away from 0. Also assume W_t is purely nondeterministic.

(A2) The CDF $F_Y(\cdot)$ is strictly increasing, θ-Lipschitz and continuously differentiable with density function $f_Y(\cdot) > 0$.

(A3) $Y_t \in L^p$ satisfies (C2) with $p > 2$, and A to be specified later. Also, $\exists c > 0$, $|f'| \geqslant c > 0$.
The following assumption is a slight generalization of (A3):

(A4) W_t satisfies (C2) with $p > 2$, and A to be specified later. Also, f preserves the (C2) property.

Remark 3. (a). Our simulations in Section 6 suggest that we may relax (A1) and (A2) where f is a possibly discontinous but strictly monotone function, and F_Y is an absolutely continuous function. It can be shown that Lemma 3.1 still holds under this assumption; see Corollary 3.1 of Das and Politis (2020a). However some extra constraints are necessary so that our asymptotic results are still true. To avoid making additional assumptions for this edge case, we use stronger assumptions (A1) and (A2).

(b). The assumption (A3) or (A4) establishes SRD for both $\{Y_t\}$ and $\{Z_t\}$. (A3) is a convenient condition under which the inverse of transfer function f is Lipschitz, so that SRD of Y_t is inherited by other series, see Lemma 3.4; meanwhile, (A4) is a more general condition for this matter, and it is implied by (A3).

Lemma 3.4. *Assume (A1)–(A3), then $\{W_t\}$ and $\{Z_t\}$ satisfy (C2).*

Proof. Under (A1), (A2) and by Lemma 3.2, f is continuously differentiable and strictly monotone. Therefore it is invertible and $|(f^{-1})'| = |\frac{1}{f'}|$ is bounded, which means f^{-1} is a Lipschitz function. Since (C2) is preserved under Lipschitz transform and $W_t = f^{-1}(Y_t)$, $\{W_t\}$ also satisfies (C2) with same rate $\delta(m)$ as $\{Y_t\}$. By Lemma 3.1, the transform from W_t to Z_t is linear; thus $\{Z_t\}$ also satisfies (C2) with same rate $\delta(m)$.

□

We now proceed to the proof of bootstrap validity for parameter estimation and prediction.

4. Model-free bootstrap validity for estimation

Let P^* denote the probability measure conditioning on the data. In this section, we show that the MFB is asymptotically valid for constructing confidence intervals of a parameter of interest θ_0, where there exists an estimator $\widehat{\theta}_n$ having the form of the so-called functions of linear statistics, i.e.,

$$\widehat{\theta}_n = q(\frac{1}{n-k+1}\sum_{t=1}^{n-k+1} g(Y_t, \cdots, Y_{t+k-1})) \tag{4.1}$$

where $q : \mathbb{R}^d \to \mathbb{R}^{\tilde{d}}$ and $g : \mathbb{R}^k \to \mathbb{R}^d$ are smooth functions. The most important case is the mean where $\theta_0 = E(Y_t)$ and $\widehat{\theta}_n = \frac{1}{n}\sum_{t=1}^n Y_t$. In fact, the class of statistics of the type (4.1) includes a wide range of estimators beyond the sample mean, such as sample autocovariances and autocorrelations, to name a few. Some other literature that investigated (4.1) include Künsch (1989), Politis and Romano (1992), Bühlmann (1997),Lahiri (2003) and Kreiss, Paparoditis and Politis (2011).

4.1. Bootstrap validity for the mean

As a demonstration of our approach, we focus on bootstrap validity for the mean $\theta_0 = E(Y_t)$. We look to prove that the following relation holds:

$$\sup_{x\in\mathbb{R}} \left| P^*(\sqrt{n}(\bar{Y}_n^* - E^*(Y_t^*)) \leqslant x) - P(\sqrt{n}(\bar{Y}_n - E(Y_t)) \leqslant x) \right| \xrightarrow{P} 0. \tag{4.2}$$

Convergence of the type (4.2) is usually proved in a two step procedure. First of all, a central limit theorem holds for both $\sqrt{n}(\bar{Y}_n - E(Y_t))$, as well as its analogue in the bootstrap world $\sqrt{n}(\bar{Y}_n^* - E^*(Y_t^*))$, where the latter distribution is conditional on the data $\{Y_t\}_{t=1}^n$. This relation holds true under a wide range of SRD conditions including our (C2); see e.g., lemma 4.1. Secondly, we need to show that the variance of the limiting normal distribution, namely the long run variance of the stationary time series denoted by σ_∞^2 and $(\sigma_\infty^*)^2$, are equal in probability as $n \to \infty$. In other words, the proposed bootstrap should correctly mimic the second-order moments of the original time series model in the limit.

Following the two step procedure, the following lemma provides a central limit result:

Lemma 4.1. *Let $\{Y_t\}$ satisfy assumptions (A1)-(A4) with $p > 2$, $\eta \in (0,1)$, $A > \frac{p-2}{2\eta}(1-\frac{1+\eta}{p}) \vee 1$ and $(1+\eta)/p - 1/2 < 0$. Then $\gamma_Y(k)$ is absolutely summable. Also,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n (Y_k - EY_0) \xrightarrow{d} \mathcal{N}(0, \sigma_\infty^2) \tag{4.3}$$

with $\sigma_\infty^2 = \sum_{k\in\mathbb{Z}} \gamma_Y(k)$.

Proof. By Theorem 7.1, $\gamma_Y(k)$ is absolutely summable for $k \in \mathbb{Z}$. Then

$$\begin{aligned} \sqrt{n}(\bar{Y}_n - EY_0) &= \frac{1}{\sqrt{n}} \left(W_1(s_n^2) + W_2(t_n^2) + O_{a.s.}(n^{(1+\eta)/p}) \right) \\ &= W_1(s_n^2/n) + W_2(t_n^2/n) + O_{a.s.}(n^{(1+\eta)/p-1/2}) \\ &\xrightarrow{d} \mathcal{N}(0, \sigma_\infty^2). \end{aligned} \tag{4.4}$$

since $n^{\gamma-1} \to 0$ and $n^{(1+\eta)/p-1/2} \to 0$.

□

Proving second moment consistency requires some additional work. The formulas for estimated transforms $\widehat{F}$ and $\widehat{F}^{-1}$ involve the whole range of data and complicate the expression as well as the dependence structure of the estimated series $\hat{\tilde{Z}}_t$, and we do not have an explicit expression for $(\sigma^*_\infty)^2$ that enables us to analyze its convergence. Alternatively, we have to prove that each step of our transform is consistent; furthermore, a uniform rate of convergence for the forward transform is indispensable for the consistency of autocovariance estimation to that of the hidden Gaussian process. See Remark 4 for detailed explanation.

Lemma 4.2. *Under the assumptions of Lemma 4.1, and the additional assumption* $\frac{A}{\frac{1}{p}+\frac{1}{\theta}} > 4$, *we have*

$$P(|Y_t - Y_t^{(m)}| > m^{-C/\theta}) < m^{-C} \tag{4.5}$$

with $C > 4$, *and*

$$\sup_{t\in[n]} |\widehat{U}_t - U_t| = O_p(n^{-1/2}) \tag{4.6}$$

Remark 4. The above lemma argues that the uniform error for the first step transform $\widehat{U}_t = F_Y(Y_t)$ is of $O_p(n^{-1/2})$ rate. This is important, as consistency of $\hat{\tilde{\Sigma}}_n$ to Σ_n depends on the consistency of the estimated $\hat{\tilde{Z}}_t$ to Z_t, which itself depends on the consistency of $\widehat{U}_t$ to U_t.

Lemma 4.2 depends on a crucial rate of approximation of the associated empirical process of Y_t towards its Gaussian process limit. The use of kernel estimators complicates this issue, as there are no known results on the rates of approximation for empirical processes indexed by kernel functions under the $m-$approximation condition. To achieve the same result as in Lemma 4.2 with kernel estimators $\widehat{F}_h$, we need the following high level assumptions:

Assumption Set 2. ***Additional assumptions for using the kernel CDF estimator:***
(A5) F_Y *is* 2^{nd} *order continuously differentiable such that* $F_Y^{(2)}(x) = \frac{d^2F_Y(x)}{dx^2}$ *is bounded and Hölder continuous.*
(A6) The kernel K *is a distribution function with density* k *satisfying:*

1. $k(x)$ *is bounded.*
2. $k(x) = k(-x)$.
3. $\int |x|^p k(x)dx < \infty$.

(A7) The class of functions $\mathcal{F}_K = \{K(x_0 - x) : x_0 \in \mathbb{R}\}$ *is Donsker with respect to* $\{Y_t\}$ *under condition (C1). In other words, there exists a tight Gaussian process* $G_h(x), x \in \mathbb{R}$, *such that*

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left(K(\frac{x - Y_k}{h}) - E\left(K(\frac{x - Y_0}{h}) \right) \right) - G_h(x) \right| = o_p(n^{1/2}). \tag{4.7}$$

In the above, the covariance of $G_h(x)$ is given by

$$\Gamma_{G_h}(x,x') = \sum_{k\in\mathbb{Z}} E\left(K(\frac{x-Y_0}{h}) - E(K(\frac{x-Y_0}{h}))\right)\left(K(\frac{x'-Y_k}{h}) - E(K(\frac{x'-Y_k}{h}))\right).$$

(A8) $h = o(n^{-1/4})$.

Remark 5. The purpose of assumption (A7) is to help establish tightness for the supremum of the empirical process, which helps establish the uniform $O_p(\frac{1}{\sqrt{n}})$ error rate for the first transform. It is not yet clear if the $m-$approximable condition can lead to (A7); however under similar dependence conditions, (A7) is shown to be correct. For example Theorem 2.1 in Arcones and Yu (1994) established such a result for $\beta-$mixing stationary series with $\mathcal{F}_K$ a V-C subgraph class of functions.

Lemma 4.3. *Let $\widehat{U}_t^{(h)} = \widehat{F}_h(Y_t)$. With assumptions (A1)–(A8) holding,*

$$\sup_{t\in[n]} |\widehat{U}_t^{(h)} - U_t| = O_p(\frac{1}{\sqrt{n}}).$$

With assumptions (A1)-(A8) along with Lemma 4.2 and 4.3 holding, we can prove MFB validity results using either the empirical CDF estimator $\bar{F}$ or the kernel CDF estimator $\widehat{F}_h$. Therefore in the following results, we only consider a general CDF estimator $\widehat{F}$ that can represent either $\bar{F}$ or $\widehat{F}_h$.

The next theorem states that by choosing appropriate threshold c and bandwidth parameter l, we can indeed estimate the autocovariance structure of the hidden Gaussian process with the estimated series $\hat{\hat{Z}}_t$.

Theorem 4.4. *Assume the conditions of Lemma 4.2 or 4.3, with appropriate rate for $l(n) \to \infty$ and $c(n) \to \infty$. Then*

$$\sum_{k=0}^{\lfloor c_\kappa l\rfloor} |\hat{\hat{\sigma}}(k) - \sigma(k)| \xrightarrow{P} 0 \tag{4.8}$$

which implies

$$\left\|\hat{\hat{\Sigma}}_n - \Sigma_n\right\|_{op} \xrightarrow{P} 0 \tag{4.9}$$

and

$$\left\|\hat{\hat{\Sigma}}_n^{-1} - \Sigma_n^{-1}\right\|_{op} \xrightarrow{P} 0. \tag{4.10}$$

The rate of divergence for $l(n)$ and $c(n)$ can be found in appendix.

With the above results, we can show that the pseudo data generated by MF and LMF bootstrap procedure converge in distribution to the truth. Before proceeding, we state one more assumption:

Assumption 3. ***Consistency of matrix square root estimator:***

$$\text{(A9)}\quad \left\|\widehat{\widehat{\Sigma}}_n^{1/2} - \Sigma_n^{1/2}\right\|_{op} = o_p(1) \text{ and } \left\|\widehat{\widehat{\Sigma}}_n^{-1/2} - \Sigma_n^{-1/2}\right\|_{op} = o_p(1).$$

Assumption (A9) holds if we have an additional mild convergence rate for the estimator towards the true covariance matrix, which can be achieved by choosing an appropriate tapering parameter $l(n) \to \infty$ and normal thresholding parameter $c(n) \to \infty$. By Theorem 2.1 of Drmac, Omladic and Veselic (1994), if

$$(\log n)^2 \left\|\widehat{\widehat{\Sigma}}_n - \Sigma_n\right\|_{op} \xrightarrow{P} 0 \tag{4.11}$$

then (A9) holds; see also Jentsch and Politis (2015) and the Corrigendum of McMurry and Politis (2010)).

Lemma 4.5. *Assume (A1)-(A8). (1) For LMF bootstrap, $\forall d \in \mathbb{N}$,*

$$(Y_{t_1}^*, Y_{t_2}^*, \cdots, Y_{t_d}^*) \xrightarrow{d^*} (Y_{t_1}, Y_{t_2}, \cdots, Y_{t_d}), \textit{ in probability.} \tag{4.12}$$

(2) Further assume (A9). Then the model-free procedure is asympototically equivalent to the limit model-free procedure. I.e., the infinite sequence

$$(\xi_1^*, \xi_2^*, \cdots) \textit{ converges in distribution to } (\zeta_1, \zeta_2, \cdots) \textit{ in probability} \tag{4.13}$$

where $(\zeta_1, \zeta_2, \cdots)$ is an infinite sequence with entries being i.i.d. $\mathcal{N}(0,1)$. Furthermore, equation (4.12) holds for the model-free bootstrap.

Let $\gamma_Y^*(k)$ and Γ_n^* denote the bootstrap analogues of $\gamma_Y(k)$ and Γ_n. The long-run variance in the bootstrap world is then $(\sigma_\infty^*)^2 = \sum_{k\in\mathbb{Z}} \gamma_Y^*(k)$. We now establish our main theorem as in (4.2) :

Theorem 4.6. *Under (A1)-(A9). For both MF bootstrap and LMF bootstrap*

$$(\sigma_\infty^*)^2 \to \sigma_\infty^2, \textit{in probability.} \tag{4.14}$$

and

$$\sup_{x\in\mathbb{R}} \left| P^*(\sqrt{n}(\bar{Y}_n^* - E^*(Y_t^*)) \leqslant x) - P(\sqrt{n}(\bar{Y}_n - E(Y_0)) \leqslant x) \right| \xrightarrow{P} 0. \tag{4.15}$$

Proof. See Appendix. □

4.2. Bootstrap validity for smooth functions of linear statistics

Now that the model-free bootstrap validity for the mean has been shown, we can extend the result for statistics of the form (4.1). For convenience, denote $X_t = (Y_t, \cdots, Y_{t+k-1})$, $\widehat{\theta}_n = \frac{1}{n-k+1}\sum_{t=1}^{n-k+1} g(X_t)$ and $\theta_0 = E(g(X_t))$ with bootstrap analogue $\widehat{\theta}_n^* = \frac{1}{n-k+1}\sum_{t=1}^{n-k+1} g(X_t^*)$ and $\theta_0^* = E^*(g(X_t^*))$, respectively. In order to extend previous results using same proof strategy, we need assumptions such that the key points in the proof are checked:

Assumption Set 3. ***Properties for functions of linear statistics***

(A10) $q(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^{\tilde{d}}$ *has continuous partial derivatives in a neighborhood of* θ_0*, and* $\sum_{i=1}^{d} (\partial q/\partial x_i)|_{x=\theta_0} x_i$ *does not vanish.*

(A11) $g = (g_1, \cdots, g_d)$ *is a continuous function and* $\forall i \leqslant d$*,* $g_i(X_t)$ *is* $L^p - m-$*approximable in the sense of (C2) with* $p > 2$ *and* $g_i(X_t^{(m)})$ *its* $m-$*approximation, and appropriate constant* A *such that Lemma 4.1 holds for all* $g_i(X_t)$*.*

Theorem 4.7. *With (A1) - (A11) hold, for MF bootstrap and LMF bootstrap with empirical CDF* $\bar{F}$ *and kernel smoothed CDF* $\widehat{F}_h$*,*

$$\sup_{x \in \mathbb{R}^{\tilde{d}}} |P^*(\sqrt{n}(q(\widehat{\theta}_n^*) - q(\theta_0^*)) \leqslant x) - P(\sqrt{n}(q(\widehat{\theta}_n) - q(\theta_0)) \leqslant x)| \xrightarrow{P} 0. \quad (4.16)$$

The proof is sketched as follows. First of all, by checking the key points in the proof of Theorem 4.6, we can show that every linear combination of $\{\widehat{\theta}_{n,i}\}_{i=1}^{d}$ and its bootstrap analogue satisfy a CLT, and their limiting variances equal in probability. By the Cramér-Wold device, central limit theorem holds for the vector form $\widehat{\theta}_n$. Then by the $\delta-$method, we can prove CLT for the smooth function transform $q(\widehat{\theta}_n)$.

The mean is now a special case of Theorem 4.7. Moreover, it can be applied to more general statistics such as the autocovariance and/or autocorrelation. We will address the autocovariance case in the following.

4.2.1. Autocovariances

For simplicity, let $X_t = (Y_t, Y_{t+1}, \cdots, Y_{t+k}) \in \mathbb{R}^{k+1}$, $g(X_t) = (Y_t Y_{t+1}, \cdots, Y_t Y_{t+k})$ and q is the identity map. Then $\frac{1}{n-k} \sum_{t=1}^{n-k} g(X_t)$ is a version of an estimator that estimates autocovariances up to lag k , assuming $EY_0 = 0$. Note that we can also design g and q more carefully such that we get the usual autocovariance estimators. Then the following corollary holds:

Corollary 4.7.1. *Assume* $Y_t \in L^p$*,* $p > 4$ *such that (C2) is satisfied with rate A. Then* $g(X_t)$ *as defined above also satisfies (C2) with* $p' = \frac{p}{2} > 2$*. Then by Theorem 4.7 the model-free bootstrap procedure is asymptotically correct for bootstrapping a vector of autocovariances.*

Proof. Let $Y_t^{(m)}$ be the approximation series generated by coupling. Then

$$\begin{aligned} \left\| Y_t Y_{t+k} - Y_t^{(m)} Y_{t+k}^{(m)} \right\|_{p'} &= \left\| Y_t Y_{t+k} - Y_t^{(m)} Y_{t+k} + Y_t^{(m)} Y_{t+k} - Y_t^{(m)} Y_{t+k}^{(m)} \right\|_{p'} \\ &\leqslant \|Y_{t+k}\|_p \left\| Y_t - Y_t^{(m)} \right\|_p + \left\| Y_t^{(m)} \right\|_p \left\| Y_{t+k} - Y_{t+k}^{(m)} \right\|_p \\ &= 2 \|Y_t\|_p \left\| Y_t - Y_t^{(m)} \right\|_p \\ &= o(m^{-A}) \end{aligned} \quad (4.17)$$

To complete proof, we apply Theorem 4.7. □

Remark 6. (a). The usual assumption of finite sum of 4^{th} order cumulants of Y_t that ensures finiteness of asymptotic variance can be dropped since it can be deduced by checking assumption (A11) along with Lemma 4.1.

(b). Notably, Kreiss, Paparoditis and Politis (2011) showed that the autoregressive(AR) sieve bootstrap procedure does not work in general for bootstrapping autocovariances of strictly stationary series (even for linear series that are not Gaussian) because of inconsistency of the limiting variance associated the companion AR process. In essence, the AR-sieve bootstrap (and the previously mentioned Linear Process bootstrap) mimic correctly the first and second order moment structure of Y_t; if the statistic of interest has a large-sample distribution that depends on higher order moments, the AR-sieve bootstrap (and the LPB) may fail. However, bootstrap validity holds for the MFB due to the assumption that $Y_t = f(W_t)$ where W_t is a stationary Gaussian process whose covariance structure can be consistently estimated (up to affine transform). Since a Gaussian process is fully described by its second order moment property, consistency of higher moments can also be obtained by the bootstrap procedure. Therefore the model-free bootstrap also works for higher order statistics.

4.3. Approximately linear statistics and sample quantiles

Now consider a statistic $\widehat{\eta}_n$ that can be expressed as

$$\widehat{\eta}_n = \frac{1}{n}\sum_{t=1}^{n} g(Y_t) + o_p(1/\sqrt{n}). \tag{4.18}$$

Under aforementioned conditions, the model-free bootstrap can be applied to the linear statistic $\widehat{\theta}_n = \frac{1}{n}\sum_{t=1}^{n} g(Y_t)$. Since $\widehat{\theta}_n$ will be $\sqrt{n}$-consistent (under the required conditions), it follows that $\widehat{\eta}_n$ and $\widehat{\theta}_n$ are asymptotically equivalent, i.e., $\sqrt{n}(\widehat{\eta}_n - \theta_0)$ has the same asymptotic distribution with $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ where $\theta_0 = E(g(Y_t))$. Therefore, since the model-free bootstrap works to approximate the distribution of $\sqrt{n}(\widehat{\theta}_n - \theta_0)$, it will also work to approximate the distribution of $\sqrt{n}(\widehat{\eta}_n - \theta_0)$.

Focusing on sample quantiles, we have the following result:

Lemma 4.8. *Assume (C2) for $\{Y_t\}$ with appropriate rate A as well as (A2). Let $u_{n,p} = \bar{F}^{-1}(p)$; $u_p = F_Y^{-1}(p)$. Then for the Bahadur-Kiefer process mentioned in Wu (2005b):*

$$\alpha_n = \left| f_Y(u_p)(u_{n,p} - u_p) - (p - \bar{F}(u_p)) \right|. \tag{4.19}$$

We have $\forall p \in (0,1)$, $\alpha_n = O_p(r_n)$ where $r_n = o(n^{-1/2})$. As a result,

$$u_{n,p} - u_p = \frac{p - \bar{F}(u_p)}{f_Y(u_p)} + o_p(n^{-1/2}). \tag{4.20}$$

Under the additional assumptions of (A5)-(A8), the above equation also holds for $u_{n,p} = \widehat{F}_h^{-1}(p)$*, i.e.*

$$u_{n,p} - u_p = \frac{p - \widehat{F}_h(u_p)}{f_Y(u_p)} + o_p(n^{-1/2}). \tag{4.21}$$

Remark 7. As shown in Wu (2005b), for the empirical CDF estimator $\bar{F}$, a stronger version of equation (4.20) holds provided a faster geometric convergence rate for the $m-$ approximation assumption (C2).

In equation (4.20) and (4.21), $\frac{p-\widehat{F}(u_p)}{f_Y(u_p)}$ is of the linear form in equation (4.18). For equation (4.20),

$$\frac{p - \bar{F}(u_p)}{f_Y(u_p)} = \frac{1}{n}\sum_{k=1}^{n} \frac{F_Y(u_p) - I\{Y_k \leqslant u_p\}}{f_Y(u_p)};$$

the same holds for equation (4.21) by the relation

$$\frac{p - \widehat{F}_h(u_p)}{f_Y(u_p)} = \frac{1}{n}\sum_{k=1}^{n} \frac{F_Y(u_p) - K(\frac{u_p - Y_k}{h})}{f_Y(u_p)}.$$

By the previous analysis, we have the following theorem.

Theorem 4.9. *Under assumptions of Theorem 4.7 with respect to the linear part of* $u_{n,p} - u_p$ *and with Lemma 4.8 holding, both MF bootstrap and LMF bootstrap are asymptotically valid for sample quantiles.*

A special case of interest is to let $p = \frac{1}{2}$. Then, the above discussion shows validity of the model-free bootstrap for the sample median.

4.4. Bootstrap validity for kernel smoothed spectral density

Another important parameter of interest is the spectral density evaluated at some frequency ω. For this subsection, let $f_{sp.d}(\omega)$ denote the spectral density of $\{Y_t\}$ which is defined as $f_{sp.d}(\omega) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \gamma_Y(k)e^{-ik\omega}$. The kernel smoothed estimator is a widely used estimator of $f_{sp.d}(\omega)$ in this setting. Let $I_n(\omega_j) = \frac{1}{2\pi n}\left|\sum Y_k e^{-ik\omega_j}\right|^2$ denote the periodogram of $\{Y_t\}$ where $\omega_j \in \{\frac{2\pi j}{n}, j \in [n]\}$ are the Fourier frequencies. The kernel smoothed spectral density estimator is defined as

$$\widehat{f}_{sp.d}(\omega) = \sum_{j\in[n]} \tilde{\kappa}_h(\omega - \omega_j) I_n(\omega_j), \tag{4.22}$$

where $\tilde{\kappa}(\omega)$ is a kernel function under certain assumptions and $\tilde{\kappa}_h(\cdot) = h^{-1}\tilde{\kappa}(\cdot/h)$. Let $\widehat{f}^*_{sp.d}(\omega)$ be the kernel smoothed estimator based on the bootstrap samples $\{Y_t^*\}_{t=1}^n$ generated through model-free bootstrap. We would like to show validity for this procedure. Besides key assumptions for previous theorems to hold, we also need the following:

Assumption Set 4. ***Additional assumptions for bootstrapping spectral density***

(A12) $\{Y_t\}$ *satisfies (C2) with* $p = 4$*;* $\inf_{\omega\in[-\pi,\pi]} f_{sp.d}(\omega) > 0$.

(A13) The kernel $\tilde{\kappa}(\cdot)$ *is a symmetric and bounded square integrable function. Also,* $\int \tilde{\kappa}(u)du = 1$ *and* $\int u^2\tilde{\kappa}(u)du < \infty$.

(A14) $h \to 0$ *and* $nh \to \infty$.

Then the following holds.

Theorem 4.10. *Under (A1)-(A9) and (A12)-(A14), for any fixed* $\omega \in [-\pi, \pi]$

$$\sqrt{nh}\left(\widehat{f}_{sp.d}(\omega) - E\widehat{f}_{sp.d}(\omega)\right) \xrightarrow{d} \mathcal{N}(0, \sigma_\omega^2). \tag{4.23}$$

In the above, $\sigma_\omega^2 = f_{sp.d}^2(\omega)\int \tilde{\kappa}^2(u)du$ *for* $\omega/\pi \notin \mathbb{Z}$ *and* $\sigma_\omega^2 = 2f_{sp.d}^2(\omega)\int \tilde{\kappa}^2(u)du$ *for* $\omega/\pi \in \mathbb{Z}$*. Also, for both model-free and limit model-free bootstrap*

$$\sup_{x\in\mathbb{R}}\left|P^*(\sqrt{nh}\left(\widehat{f}^*_{sp.d}(\omega) - E^*\widehat{f}^*_{sp.d}(\omega)\right) \leqslant x) - P(\sqrt{nh}\left(\widehat{f}_{sp.d}(\omega) - E\widehat{f}_{sp.d}(\omega)\right) \leqslant x)\right| \xrightarrow{P} 0. \tag{4.24}$$

Proof. See Appendix. □

As is well-known, the bandwidth h governs the trade-off between the bias and variance of $\widehat{f}_{sp.d}(\omega)$. If we choose h in such a way that *undersmoothing* occurs, i.e., when the bias of $\widehat{f}_{sp.d}(\omega)$ is of a smaller order of magnitude than its standard deviation, then equation (4.24) holds true with $E\widehat{f}_{sp.d}(\omega)$ replaced by $f_{sp.d}(\omega)$, i.e., we have

$$\sup_{x\in\mathbb{R}}\left|P^*(\sqrt{nh}\left(\widehat{f}^*_{sp.d}(\omega) - E^*\widehat{f}^*_{sp.d}(\omega)\right) \leqslant x) - P(\sqrt{nh}\left(\widehat{f}_{sp.d}(\omega) - f_{sp.d}(\omega)\right) \leqslant x)\right| \xrightarrow{P} 0. \tag{4.25}$$

Equation (4.25) can then be used to construct model-free confidence intervals for $f_{sp.d}(\omega)$ based on the quantiles of the bootstrap distribution of $\sqrt{nh}\left(\widehat{f}^*_{sp.d}(\omega) - E^*\widehat{f}^*_{sp.d}(\omega)\right)$.

5. Model-free bootstrap for prediction

For simplicity, we focus on proving model-free bootstrap validity for one-step ahead prediction; generalizing to the h-step ahead case is also possible. Let Y_{n+1} be the 1- step ahead value of the time series dataset $\{Y_t\}_{t=1}^n$ we wish to predict. We call the conditional distribution D_{n+1} of $Y_{n+1}|\underline{Y}_n$ the predictive distribution. Naturally, the quantiles of D_{n+1} can help us construct prediction interval of Y_{n+1}.

Oftentimes, we also want the prediction interval to be centered at a certain meaningful point predictor, which brings the notion of **predictive root**. Let

$\widehat{Y}_{n+1}$ be a one-step ahead predictor of Y_{n+1} based on the original series $\{Y_t\}_{t=1}^n$, i.e., $\widehat{Y}_{n+1}$ depends entirely on the data $\{Y_t\}_{t=1}^n$. The predictive root, analogous to the case of confidence intervals, is defined as the residual between the data and predictor

$$Y_{n+1} - \widehat{Y}_{n+1}. \tag{5.1}$$

Let $L_{\alpha/2}$ and $R_{\alpha/2}$ be the lower $\alpha/2$ quantile and upper $\alpha/2$ quantile of the (conditional on $\underline{Y}_n$) distribution of the predictive root. Then, an exact two-sided $1-\alpha$ prediction interval for Y_{n+1} is:

$$(\widehat{Y}_{n+1} + L_{\alpha/2}, \widehat{Y}_{n+1} + R_{\alpha/2}); \tag{5.2}$$

in other words,

$$P(\widehat{Y}_{n+1} + L_{\alpha/2} \leqslant Y_{n+1} \leqslant \widehat{Y}_{n+1} + R_{\alpha/2} | \underline{Y}_n) = 1 - \alpha. \tag{5.3}$$

We wish to approximate the quantiles of the predictive root by the bootstrap procedure. Let Y_{n+1}^* be the one-step ahead value generated through the bootstrap; also let $\widehat{Y}_{n+1}^*$ be the one-step ahead predictor of Y_{n+1} conditioning on $\underline{Y}_n$,where its formula is estimated by the bootstrap samples $\{Y_t^*\}_{t=1}^n$. We say that boostrap validity for prediction intervals holds if for any $\alpha \in (0,1)$:

$$L_{\alpha/2}^* \to L_{\alpha/2} \quad \text{and} \quad R_{\alpha/2}^* \to R_{\alpha/2} \tag{5.4}$$

in probability. Here $L_{\alpha/2}^*$ and $R_{\alpha/2}^*$ denote the relative $\alpha/2$ quantiles with respect to the conditional distribution of the bootstrap predictive root $Y_{n+1}^* - \widehat{Y}_{n+1}^*$.

In other words, the two-sided approximate $1-\alpha$ bootstrap prediction interval for Y_{n+1} is

$$(\widehat{Y}_{n+1} + L_{\alpha/2}^*, \widehat{Y}_{n+1} + R_{\alpha/2}^*), \tag{5.5}$$

while equation (5.4) would imply

$$P(\widehat{Y}_{n+1} + L_{\alpha/2}^* \leqslant Y_{n+1} \leqslant \widehat{Y}_{n+1} + R_{\alpha/2}^* | \underline{Y}_n) \to 1 - \alpha. \tag{5.6}$$

The following theorem states that bootstrap validity holds for one-step ahead prediction intervals.

Theorem 5.1. *(Model-free bootstrap validity for one-step ahead prediction) Assume that the conditional distribution of the predictive root $Y_{n+1} - \widehat{Y}_{n+1}$ is continuous. With assumptions (A1) - (A9), **Algorithms** 2 and 3 are asymptotically valid in the sense of equation (5.4).*

Unlike the bootstrap for parameter inference, the conditional distribution of the predictive root is not asymptotically degenerate. Hence, no central limit theorems are required for parameter estimates since their associated variabilty vanishes asymptotically.

Asymptotic validity of bootstrap prediction intervals is of less importance than that of confidence intervals. A good finite-sample prediction interval should

incorporate the variance of all estimated features, or it will result in *undercoverage*. However, as stated above, the property of asymptotic validity has nothing to do with capturing finite-sample variability in estimation; see also Ch. 2.4.1 of Politis (2015). In this respect, the performance of prediction intervals must be quantified in finite-sample numerical experiments as in the following section.

6. Numerical results

We look into three data generating models and evaluate the coverage performance of our proposed MFB methods with respect to both confidence intervals and prediction intervals. All three models satisfy equation (3.1) with

$$f(x) = \begin{cases} -\sqrt{-x} & x < 0 \\ \frac{(x+1)^2}{10} & x \geqslant 0 \end{cases} \tag{6.1}$$

and W_t generated by an autoregressive moving average (ARMA) model, namely

$$W_t - \phi_1 W_{t-1} - \cdots - \phi_p W_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

with $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. The three ARMA models considered are as follows: in one of the following ways:

1. MA model of order one, with $\theta_1 = -0.5$ and all other parameters zero.
2. AR model of order one, with $\phi_1 = 0.5$ and all other parameters zero.
3. MA model of order 30, with $\theta_1 = 2$, $\theta_2 = 1$, $\theta_k = \frac{10}{k^2}$ when $3 \leqslant k \leqslant 30$, and all other parameters zero.

We analyze the performance of model-free and limit model-free bootstrap with both the empirical CDF estimator and the kernel CDF estimator.

6.1. Performance for bootstrap confidence intervals

For confidence intervals, the parameter of interests are: the mean for all three models; lag 1 autocovariance for model 1 and 2; lag 2 autocovariance for model 3. As a comparison, we benchmark our methods with the two most popular alternative methods, i.e. block bootstrap and AR-sieve bootstrap.

The mean is, of course, a linear statistic. The lag-k autocovariance can be written in the form (4.1) with

$$g(Y_t, \cdots, Y_{t+k}) = \begin{bmatrix} Y_t Y_{t+k} \\ Y_t \end{bmatrix}$$

and

$$q(\frac{1}{n-k}\sum_{t=1}^{n-k} g(Y_t, \cdots, Y_{t+k})) = \frac{1}{n-k}\sum_{t=1}^{n-k} Y_t Y_{t+k} - [\frac{1}{n-k}\sum_{t=1}^{n-k} Y_t]^2.$$

Although we can do block bootstrap on the Y_t and recompute $\widehat{\gamma}_k$ on the bootstrap data to construct confidence intervals, this procedure will suffer from end effects resulting into an estimator that is biased towards 0. As a remedy, let $X_t = (Y_t, \cdots, Y_{t+k})$ as we did in Section 4.2, then a block bootstrap on the X_t data will relieve this problem and it is equivalent to the so-called blocks-of-blocks bootstrap on the Y_t data; see Politis and Romano (1992), as well as Paradigm 12.8.11 of the book by McElroy and Politis (2019). For our purposes, we used first level of blocking with $k = 5$ that can be used to capture autocovariances up to lag 4.

For the block bootstrap on the X_t data, we can choose the block size b as $b = const * n^{1/3}$ where *const* is selected according to Patton, Politis and White (2009). As for the AR-sieve bootstrap, the order p of the fitted AR model is selected through minimizing the AIC; see Bühlmann (1997).

For simulation purpose, in the model-free bootstrap algorithm 1 we need to choose the threshold parameter c . Under a finite sample setting, the main goal is to resolve the issue of calculating $\Phi^{-1}(1)$. The following simple empirical rule is used: for $\bar{F}$ scenario, we adjust the value 1 in $\{\widehat{U}_t\}_{t=1}^n$ to $\frac{n-1}{n}$ to avoid the problem of $\Phi^{-1}(1)$, which is equivalent to setting $c = \Phi^{-1}(\frac{n-1}{n})$. In the case of the kernel CDF $\widehat{F}_h$, besides applying the above rule, we can also avoid having 1 in $\{\widehat{U}_t\}_{t=1}^n$ by choosing an appropriate kernel function(e.g., the Gaussian CDF) and the bandwidth, in which case choosing c is no longer necessary.

The metric we use for comparison is the empirical coverage rate(empirical CVR) for the bootstrap intervals. To elaborate, let $N = 1000$ be the number of replicated experiments, $n \in \{100, 200, 500, 1000\}$ be the length of samples, $\alpha = 0.05$ and $B = 250$ the number of bootstrap replications. So for experiment $i \in \{1, \ldots, N\}$, we generate a time series sample of length n, and use different methods to construct a $1 - \alpha$ bootstrap confidence interval (L_i, R_i) based on the B bootstrap replicates. Let θ_0 denote the true parameter of interest; then the empirical coverage rate for the whole experiment is:

$$CVR_{\theta_0} = \frac{1}{1000} \sum_{i=1}^{1000} I_{\theta_0 \in (L_i, R_i)} \tag{6.2}$$

The purpose of looking at CVR is two fold: first, comparing CVR values in a finite-sample size setting will tell which method has better performance. Second, as the sample size n gets large, asymptotic validity can be observed by checking whether the empirical CVR converges to the nominal $100(1-\alpha)\%$ percent, which will provide numerical grounds to our previous conclusions.

The following tables contain CVR results for different settings all with nominal $\alpha = 0.05$ (95% confidence interval). The acronyms in the "Method" column of Table 1 have the following meaning. "MF" for model-free bootstrap; "LMF" for limit model-free bootstrap; "emp" for empirical CDF estimator used in the bootstrap procedure; "ker" for kernel CDF estimator; "BB" for block bootstrap; "AR-sieve" for the autoregressive sieve bootstrap.

For the mean parameter (see Table 1), all the methods being compared are theoretically valid for all three models. We can observe that for each method, as

Method	n=100	200	500	1000	2000
model 1					
MF-ker	92.1	92.5	93.2	93.6	94.8
LMF-ker	92.3	92.5	93.0	94.7	93.9
MF-emp	91.6	92.4	93.1	94.0	94.4
LMF-emp	92.0	92.9	93.5	93.8	94.0
BB	90.4	93.1	94.1	95.0	95.0
AR-sieve	94.0	93.6	94.0	95.2	95.6
model 2					
MF-ker	91.4	92.4	93.2	93.2	93.7
LMF-ker	88.5	92.1	93.3	93.3	94.4
MF-emp	88.7	90.9	92.8	94.1	92.9
LMF-emp	85.3	88.4	92.4	93.0	92.9
BB	85.2	89.6	91.3	92.3	93.0
AR-sieve	90.6	92.9	92.8	93.6	94.2
model 3					
MF-ker	87.9	89.3	90.3	91.7	93.3
LMF-ker	86.9	88.1	90.5	92.1	93.1
MF-emp	82.7	85.5	89.6	91.4	92.7
LMF-emp	80.9	85.8	89.5	91.4	92.8
BB	82.2	85.4	87.3	89.8	91.4
AR-sieve	83.6	88.1	90.0	91.9	92.2

TABLE 1
Empirical CVR for the mean parameter across 3 models

n increases the coverage rate approaches 95% albeit the speed of convergence can be different. The AR sieve holds an obvious advantage over other methods for model 1 with coverage close to 95% even at $n = 100$, whereas the proposed model-free bootstrap works on par with the block bootstrap. However for the other two models, MF and LMF bootstrap with kernel CDF estimator has a noticeable advantage over the AR-sieve and block bootstrap, especially for model 3, which has a more complex data generating process. Interestingly, the behavior of the model-free bootstrap using the empirical CDF is almost the same as that of the block bootstrap. Fortunately enough, the coverage advantage of MF based on kernel CDF over empirical CDF goes away for large n. As a result, we can use the empirical CDF approach under a large sample scenario, to avoid calculating the quantile inverse of a kernel CDF, which becomes computationally expensive as sample size grows.

Tables 2 and 3 present the empirical CVR for the autocovariance with different models. Both lag 1 and lag 2 autocovariances were considered; for conciseness, we present lag 1 results from models 1 and 2 (see Table 2), and lag 2 results from model 3 (see Table 3). Recall that the AR-sieve bootstrap is not asymptotically valid in general for the autocovariance. Comparing Tables 2 and 3, we see that the AR-sieve bootstrap works well for the autocovariance in models 1 and 2 but not with data from model 3; see Table 3 where the AR-sieve CVR appears to converge to around 86% instead of the 95% nominal level.

We can also observe they asymptotic validity of model-free bootstrap methods manifested in Table 2 and 3. Furthermore, model-free methods appear to enjoy a

Method	n=100	200	500	1000	2000
model 1					
MF-ker	82.0	88.2	90.5	92.2	91.1
LMF-ker	89.6	94.3	94.0	94.8	94.1
MF-emp	80.0	88.3	90.1	91.6	92.0
LMF-emp	89.2	93.5	94.0	95.0	93.3
BB	86.0	88.8	91.5	91.1	93.0
AR-sieve	94.3	94.8	94.6	96.3	95.0
model 2					
MF-ker	75.0	86.5	94.2	94.6	94.6
LMF-ker	92.2	94.2	95.4	94.0	93.4
MF-emp	68.6	82.9	90.2	91.9	94.2
LMF-emp	76.9	82.5	89.0	91.6	94.4
BB	82.3	86.2	89.0	90.5	91.5
AR-sieve	88.3	91.8	95.1	95.5	95.5

TABLE 2
Empirical CVR for lag 1 autocovariance for the first 2 models

faster convergence towards nominal compared to block bootstrap. Interestingly, limit model-free with kernel CDF estimator works significantly better than other methods across all 3 models.

Method	n=100	200	500	1000	2000
MF-ker	76.9	87.8	91.7	93.0	94.1
LMF-ker	77.7	86.0	91.3	92.4	93.8
MF-emp	68.2	80.9	86.8	91.6	93.5
LMF-emp	65.1	77.5	85.8	89.5	93.5
BB	64.0	73.9	83.2	87.3	89.4
AR-sieve	69.4	76.4	84.9	86.1	86.1

TABLE 3
Empirical CVR for lag 2 autocovariance for model 3

6.2. Performance of bootstrap prediction intervals

We now move on to the prediction performance for the three models. The empirical coverage rate is defined similarly as

$$CVR_{Y_{n+1}} = \frac{1}{1000}\sum_{i=1}^{1000} I_{Y_{n+1}^{(i)} \in (\hat{Y}_n^{(i)}+L_i, \hat{Y}_n^{(i)}+R_i)} \tag{6.3}$$

where (L_i, R_i) are sample quantiles of the bootstrap predictive root generated by bootstrap and $Y_{n+1}^{(i)}$ is the $n+1$ value of the time series sample from the i^{th} experiment.

Since the block bootstrap is not a viable method for generating one-step ahead prediction value, we benchmark predictive performance of the model-free methods by comparing them with the AR-sieve bootstrap. The bootstrap samples are generated in a forward bootstrap manner as described in Algorithm

3.1 of Pan and Politis (2016a); see also Alonso, Peña and Romo (2002). Table 4 provides the empirical CVR with $n \in \{100, 200, 300, 500\}$; the samples sizes are smaller compared to our previous simulations because of the increasing computational cost for prediction. However, it is reassuring that all methods considered, i.e., the 4 model-free variations as well as the AR-sieve bootstrap, produce prediction intervals with CVR close to the nominal 95% even with n as low as 200. It is difficult to do a finer comparison of these 5 bootstrap methods since, for computational reasons, we had to choose a small number of bootstrap replications ($B = 250$). Ongoing work includes devising an analog of the 'Warp-Speed' method of Giacomini, Politis and White (2013) that will speed up Monte Carlo experiments involving bootstrap in the case of prediction intervals.

Remark 8. An important question is the performance difference of the MFB algorithm proposed here and that proposed in Das and Politis (2020a). The difference is two-fold: First of all, the main thrust of prediction based on locally stationary data is that only local estimates (of the CDF of Y_t and autocovariance, etc.) can be used. If the local window is of size $b \ll n$, the estimates of Das and Politis (2020a) would be less accurate than ours that are based on sample size n. As a consequence, our algorithm will perform better than Das and Politis (2020a) in the case of globally stationary data. On the other hand, if global stationarity fails, our procedure will no longer be consistent, whereas that of Das and Politis (2020a) would still perform well.

In addition, in order to handle boundary effects, Das and Politis (2020b) developed a local linear distribution estimator which was then employed in Das and Politis (2020a). The local linear distribution estimator is not needed in the case of global stationarity which is the focus of the paper at hand.

Acknowledgments. Many thanks are due to Jens-Peter Kreiss, Stathis Paparoditis, and the participants of the August 2015 Workshop on *Recent Developments in Statistics for Complex Dependent Data*, Loccum (Germany), where the seeds of this work were first presented. We are also grateful to three anonymous reviewers for their constructive comments. This research was partially supported by NSF grants DMS 16-13026 and DMS 19-14556.

References

Adler, R. J. (1990). *An introduction to continuity, extrema, and related topics for general Gaussian processes. Lecture Notes–Monograph Series* **Volume 12**. Institute of Mathematical Statistics, Hayward, CA.

Alonso, A. M., Peña, D. and Romo, J. (2002). Forecasting time series with sieve bootstrap. *Journal of Statistical Planning and Inference* **100** 1 - 11.

Angus, J. E. (1994). The Probability Integral Transform and Related Results. *SIAM Review* **36** 652–654.

Method	n=100	200	300	500
model 1				
MF-ker	95.0	94.6	91.6	91.8
LMF-ker	94.1	96.6	93.9	93.0
MF-emp	94.0	94.2	92.2	94.2
LMF-emp	92.2	93.9	94.0	93.8
AR-sieve	94.4	94.6	94.0	93.2
model 2				
MF-ker	93.6	94.0	92.2	93.2
LMF-ker	90.2	95.3	96.4	92.8
MF-emp	94.1	93.4	95.0	94.2
LMF-emp	89.2	90.7	95.0	93.6
AR-sieve	94.0	93.2	94.0	94.4
model 3				
MF-ker	93.8	94.4	95.4	94.8
LMF-ker	91.8	96.7	95.6	93.6
MF-emp	92.2	96.2	96.4	95.2
LMF-emp	92.8	94.8	96.4	93.2
AR-sieve	91.4	94.8	93.2	93.0

TABLE 4
Empirical CVR for prediction intervals across the three models

ARCONES, M. A. and YU, B. (1994). Central limit theorems for empirical and U-processes of stationary mixing sequences. *Journal of Theoretical Probability* **7** 47–71.

BERKES, I., HÖRMANN, S. and SCHAUER, J. (2009). Asymptotic results for the empirical process of stationary sequences. *Stochastic Processes and their Applications* **119** 1298 - 1324.

BERKES, I., HÖRMANN, S. and SCHAUER, J. (2011). Split invariance principles for stationary processes. *Ann. Probab.* **39** 2441–2473.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*. Springer New York, New York, NY.

BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3** 123–148.

DAS, S. and POLITIS, D. N. (2020a). Predictive Inference for Locally Stationary Time Series With an Application to Climate Data. *Journal of the American Statistical Association* **0** 1-16.

DAS, S. and POLITIS, D. N. (2020b). Nonparametric Estimation of the Conditional Distribution at Regression Boundary Points. *The American Statistician* **74** 233-242.

DRMAC, Z., OMLADIC, M. and VESELIC, K. (1994). On the Perturbation of the Cholesky Factorization. *SIAM J. Matrix Anal. Appl.* **15** 1319–1332.

EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.* **7** 1–26.

GIACOMINI, R., POLITIS, D. N. and WHITE, H. (2013). A Warp-speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators. *Econometric Theory* **29** 567–589.

HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884.

Jentsch, C. and Politis, D. N. (2015). Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *Ann. Statist.* **43** 1117–1140.

Kallenberg, O. (1997). *Foundations of Modern Probability*. Springer, New York.

Kreiss, J.-P., Paparoditis, E. and Politis, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.* **39** 2103–2130.

Kreiss, J.-P. and Paparoditis, E. (2020). *Bootstrap for Time Series: Theory and Applications*. Springer New York.

Künsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *Ann. Statist.* **17** 1217–1241.

Lahiri, S. (2003). *Resampling Methods for Dependent Data*. Springer, New York.

Li, Q. and Racine, J. S. (2006). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Liu, W. and Wu, W. B. (2010). Asymptoticis of Spectral Density Estimates. *Econometric Theory* **26** 1218–1245.

Mase, S. (1977). Some Theorems on Normality-Preserving Transformations. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* **39** 186–190.

McElroy, T. S. and Politis, D. N. (2019). *Time Series: A First Course with Bootstrap Starter*. Chapman and Hall/CRC Press, Boca Raton.

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.

McMurry, T. L. and Politis, D. N. (2015). High-dimensional autocovariance matrices and optimal linear prediction. *Electron. J. Statist.* **9** 753–788.

Pan, L. and Politis, D. N. (2016a). Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *Journal of Statistical Planning and Inference* **177** 1 - 27.

Pan, L. and Politis, D. N. (2016b). Bootstrap prediction intervals for Markov processes. *Computational Statistics and Data Analysis* **100** 467 - 494.

Patton, A., Politis, D. N. and White, H. (2009). Correction to Automatic Block-Length Selection for the Dependent Bootstrap by D. Politis and H. White. *Econometric Reviews* **28** 372-375.

Politis, D. N. (2013). Model-free model-fitting and predictive distributions. *Test* **22** 183-221.

Politis, D. N. (2015). *Model-Free Prediction and Regression: A Transformation-Based Approach to Inference*. Springer, New York.

Politis, D. N. and Romano, J. P. (1992). A General Resampling Scheme for Triangular Arrays of α-Mixing Random Variables with Application to the Problem of Spectral Density Estimation. *Ann. Statist.* **20** 1985–2007.

Politis, D. N. and Romano, J. P. (1995). Bias-corrected Nonparametric Spectral Estimation. *Journal of Time Series Analysis* **16** 67-103.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. Springer New York, New York, NY.

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Ann.*

Math. Statist. **23** 470–472.
Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition. *Proceedings of the National Academy of Sciences* **42** 43–47.
Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes*. Routledge, New York.
Snelson, E., Ghahramani, Z. and Rasmussen, C. E. (2004). Warped Gaussian Processes. In *Advances in Neural Information Processing Systems 16* 337–344. MIT Press.
Van Der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
Wu, W. B. (2005a). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* **102** 14150–14154.
Wu, W. B. (2005b). On the Bahadur representation of sample quantiles for dependent sequences. *Ann. Statist.* **33** 1934–1963.
Wu, W. B. and Pourahmadi, M. (2009). Banding Sample Autocovariance Matrices of Stationary Processes. *Statistica Sinica* **19** 1755–1768.
Zhang, Y., Liu, J., Pan, Z. and Politis, D. N. (2019). Estimating transformation function. *Electron. J. Statist.* **13** 3095–3119.

7. Appendix

A. Preliminary results under m-approximation assumptions

Theorem 7.1. *(Berkes, Hörmann, Schauer (2009))*
Define $R(s,t) = \sum_{1\leqslant k\leqslant t}(I\{Y_k \leqslant s\} - F_Y(s))$. *Assume* F_Y *is* θ*-Lipschitz continuous, and (C1) holds with* $\gamma_m = m^{-C/\theta}$, $\delta_m = m^{-C}$ *and some* $C > 4$. *Then there exists a two parameter Gaussian process* $K(s,t)$ *with* $E(K(s,t)K(s',t')) = (t \wedge t')\Gamma(s,s')$, *such that*

$$\sup_{s,t} |R(s,t) - K(s,t)| = o(n^{1/2}(\log n)^{-\alpha}), \ a.s. \tag{7.1}$$

for some $\alpha > 0$. *In addition,*

$$\Gamma(s,s') = \sum_{k\in\mathbb{Z}} E(I(Y_0 \leqslant s) - F_Y(s))(I(Y_k \leqslant s') - F_Y(s')) \tag{7.2}$$

is absolutely convergent for all choices of s, s'.

We are more interested in the case $t = n$ such that the above theorem reduces to a 1-dimensional centered Gaussian process K indexed by $s \in \mathbb{R}$. It turns out that tightness of K plays an important role in the accuracy of the estimated $\widehat{U}_t$. Although Theorem 7.1 does not guarantee the limiting Gaussian process to be tight, it is reasonable to believe so and also it can be shown that for series of the form (3.1) tightness can be shown through continuity and boundedness of the covariance structure.

Theorem 7.2. *(Berkes, Hörmann, Schauer (2011))*
Assume (C2) with $p > 2$, $\eta \in (0,1)$, $A > \frac{p-2}{2\eta}(1-\frac{1+\eta}{p}) \vee 1$. *Then* $\sigma_\infty^2 = \sum_{k\in\mathbb{Z}} \gamma_Y(k)$ *is absolutely convergent. Also,*

$$\sum_{k=1}^{n} (Y_k - EY_0) = W_1(s_n^2) + W_2(t_n^2) + O(n^{(1+\eta)/p}), \ a.s. \tag{7.3}$$

where W_1 *and* W_2 *are two Brownian motions and* $s_n^2 \sim \sigma_\infty^2 n$, $t_n^2 \sim cn^\gamma$ *with* $\gamma \in (0,1)$. *Here* $a_n \sim b_n$ *means* $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$; $a \vee b = \max\{a,b\}$.

Theorem 7.3. *(McMurry & Politis(2010))*
For the tapered estimator defined in equation (2.3) with tapering function (2.1), assume (C0) with $2 < p \leqslant 4$ *for* $\{Z_t\}$ *and* $\sum_{k\in\mathbb{Z}} \sigma(k) < \infty$, *with* $l = o(n^{\frac{p-2}{p}})$ *then,*

$$\sum_{k=0}^{\lfloor c_\kappa l \rfloor} |\widehat{\sigma}(k) - \sigma(k)| \underset{L^{p/2}}{\to} 0, \tag{7.4}$$

and

$$\left\| \widehat{\Sigma}_n^{(\kappa,l)} - \Sigma_n \right\|_{op} \underset{L^{p/2}}{\to} 0. \tag{7.5}$$

Furthermore, assuming Z_t *has spectral density* f_Z *satisfying* $0 < c_1 \leqslant f_Z(w) \leqslant c_2 < \infty, \forall w \in [0, 2\pi]$, $\widehat{\Sigma}_n^{(\kappa,l)}$ *is positive definite with probability tending to 1, and*

$$\left\| (\widehat{\Sigma}_n^{(\kappa,l)})^{-1} - \Sigma_n^{-1} \right\|_{op} = o_p(1) \tag{7.6}$$

B. Proof of relative lemmas and theorems

Proof of Lemma 3.3. 1. Let $\gamma_m = m^{-C/\theta}$ and $\delta_m = m^{-C}$. By Markov's inequality,

$$\begin{aligned} P(|Y_t - Y_t^{(m)}| > m^{-C/\theta}) &= P(|Y_t - Y_t^{(m)}|^p > m^{-Cp/\theta}) \\ &\leqslant \frac{E(|Y_t - Y_t^{(m)}|^p)}{m^{-Cp/\theta}} \\ &\ll \frac{m^{-pA}}{m^{-Cp/\theta}} = m^{-C} \end{aligned} \tag{7.7}$$

2.

$$\sum_{m=0}^{\infty} \left\| Y_t - Y_t^{(m)} \right\|_p \leqslant \sum_{m=0}^{\infty} \delta(m).$$

Since $\delta(m) \ll m^{-A}$ with $A > 1$, $\delta(m)$ is summabe. Thus (C3) holds.

3. Let ϵ_0' be the new independent sample to be used in $\delta_p(t)$. Let $\epsilon_0'', \epsilon_{-1}'', \cdots$ be an infinite sequence of i.i.d. samples from F_ϵ. By triangle inequality we have:

$$\begin{aligned}\delta_p(m) &= \left\|h(\cdots,\epsilon_{-1},\epsilon_0,\cdots,\epsilon_m) - h(\cdots,\epsilon_{-1},\epsilon_0',\cdots,\epsilon_m)\right\|_p \\ &\leqslant \left\|h(\cdots,\epsilon_{-1},\epsilon_0,\cdots,\epsilon_m) - h(\cdots,\epsilon_{-1}'',\epsilon_0'',\epsilon_1,\cdots,\epsilon_m)\right\|_p \\ &+ \left\|h(\cdots,\epsilon_{-1}'',\epsilon_0'',\epsilon_1,\cdots,\epsilon_m) - h(\cdots,\epsilon_{-1},\epsilon_0',\cdots,\epsilon_m)\right\|_p \\ &= 2\left\|Y_m - Y_m^{(m)}\right\|_p\end{aligned} \tag{7.8}$$

Thus (C3) implies (C0).

4. See page 2443 in Berkes, Hörmann and Schauer (2011). □

Proof of Lemma 4.2. Equation (4.5) is deduced by result in Lemma 3.3. Then by Theorem 7.1

$$\begin{aligned}\sup_{t\in[n]}|\widehat{U}_t - U_t| &\leqslant \frac{1}{n}\sup_{s\in\mathbb{R}}\left|\sum_{k=1}^{n}(I(Y_k\leqslant s) - F_Y(s))\right| \\ &= \frac{1}{n}\sup_s|R(s,n)| \\ &\leqslant \frac{1}{\sqrt{n}}\sup_s|K(s,1)| + o(n^{-1/2}(\log n)^{-\alpha}), a.s.\end{aligned} \tag{7.9}$$

where $K(s,1)$ is a centered Gaussian process with covariances

$$\Gamma(s,s') = \sum_{k\in\mathbb{Z}} E(I(Y_0\leqslant s) - F_Y(s))(I(Y_k\leqslant s') - F_Y(s')) \tag{7.10}$$

absolutely convergent $\forall s, s' \in \mathbb{R}$. We also need the limiting Gaussian process $K(s,1)$ to be tight such that $\sup_s|K(s,1)| = O_p(1)$, which usually is true but not mentioned in Theorem 7.1. Another way to show this is to prove certain continuity and boundedness condition for the covariance $\Gamma(s,s')$.

Note that $K(s,1)$ can be reparametrized as $K'(u), u = F_Y(s)$ such that the centered gaussian process is now living in a bounded domain $T = [0,1]$(reparametrization of the empirical process). Also, since F_Y is absolutely continuous and strictly increasing, $\sup_{s\in\mathbb{R}}|K(s,1)| = \sup_{t\in[0,1]}|K'(t)|$. Let $\Gamma'(t,t')$ denote the covariance kernel of $K'(u)$. Obviously $\Gamma'(t,t') = \Gamma(F_Y^{(-1)}(t), F_Y^{(-1)}(t'))$, for which we have $\Gamma'(0,0) = \Gamma'(1,1) = 0$. Thus the Gaussian process $K'(u)$ is pinned to 0 a.s. at enpoints 0 and 1.
We claim that

- (1) $\Gamma'(t,t')$ is uniformly bounded for $t,t' \in [0,1]^2$.
- (2) $\Gamma'(t,t')$ is continuous in t and t'.
- (3) K' satisfies the Kolmogorov-Čentsov condition: $\exists\alpha,\beta,C>0$, such that

$$E|K'(t) - K'(s)|^\alpha \leqslant C|t-s|^{1+\beta}$$

(2) is a consequence of (1). To see this, note that $\forall t, \forall t_1 < t_2$, $|\Gamma'(t_1,t)-\Gamma'(t_2,t)| = |\sum_{k\in\mathbb{Z}} E(I(U_0 \leqslant t_1)-t_1)(I(U_k \leqslant t)-t) - E(I(U_0 \leqslant t_2)-t_2)(I(U_k \leqslant t)-t)| = E\left[I(t_1 \leqslant U_0 \leqslant t_2)-(t_2-t_1)\right]\left[\sum_{k\in\mathbb{Z}}(I(U_k \leqslant t)-t)\right]$. By previous theorem, the expectation exists $\forall t$ and $t_1 < t_2$, and uniformly bounded by $2\sup_{t\in[0,1]}|\Gamma'(t,t)|$. As $|t_2-t_1| \to 0$, the random variable inside the expectation converges to 0 a.s. Thus by dominated convergence theorem we have $|\Gamma'(t_1,t)-\Gamma'(t_2,t)| \to 0$.

For (1), by positive definiteness of covariance kernel we have that

$$\begin{aligned}\sup_{(t,t')\in[0,1]^2}|\Gamma'(t,t')| &= \sup_{t\in[0,1]}|\Gamma'(t,t)| = \sup_{t\in[0,1]}|\sum_{k\in\mathbb{Z}} E(I(U_0 \leqslant t)-t)(I(U_k \leqslant t)-t)| \\ &= \sup_{t\in[0,1]}|\sum_k E\left[I(U_0 \leqslant t)I(U_k \leqslant t)-t^2\right]| \\ &= \sup_{t\in[0,1]}|\sum_k \int_{(-\infty,g^{-1}(t)]^2} f_W^{(k)}(x,y)-f_W(x)f_W(y)dxdy| \\ &\leqslant \int_{\mathbb{R}^2}\sum_k |f_W^{(k)}(x,y)-f_W(x)f_W(y)|dxdy \qquad \cdots(*)\end{aligned} \tag{7.11}$$

Where $g = F_Y \circ f$ is a monotone (increasing) function; $f_W^{(k)}$ is the bivariate normal density of (W_0, W_k); f_W is the normal density of W_t. We show that $(*)$ is bounded.
Rewrite $f_W^{(k)}(x,y)$ as $f_{xy}(a_k,b_k) = \frac{1}{2\pi}a_k^{-1/2}\exp\{-\frac{1}{2}a_k^{-1}b_k\}$,with $a_k = \sigma_W^2(0)-\sigma_W^2(k)$ and $b_k = \sigma_W(0)(x^2+y^2)-2\sigma_W(k)xy$, where $\sigma_W(k)$ is the lag-k autocovaraince for W_t.

$$\nabla f = \begin{bmatrix}\frac{\partial f}{\partial a_k}\\ \frac{\partial f}{\partial b_k}\end{bmatrix} = \begin{bmatrix}\frac{1}{2}a_k^{-\frac{3}{2}}(a_k^{-1}b_k-1)e^{-\frac{1}{2}a_k^{-1}b_k}\\ -\frac{1}{2}a_k^{-\frac{3}{2}}e^{-\frac{1}{2}a_k^{-1}b_k}\end{bmatrix} \tag{7.12}$$

Then

$$\begin{aligned}&\int_{\mathbb{R}^2}|f_{xy}(a_k,b_k)-f_{xy}(\sigma_W^2(0),\sigma_W(0)(x^2+y^2))|dxdy \\ &= \int_{\mathbb{R}^2}|\nabla f^T(\sigma_W^2(0),\sigma_W(0)(x^2+y^2))\begin{bmatrix}-\sigma_W^2(k)\\ -2\sigma_W(k)xy\end{bmatrix}|dxdy + o(\sigma_W(k)+\sigma_W^2(k))\end{aligned} \tag{7.13}$$

and

$$\begin{aligned}(*) &= \int_{\mathbb{R}^2}\sum_k |f_{xy}(a_k,b_k)-f_{xy}(\sigma_W^2(0),\sigma_W(0)(x^2+y^2))|dxdy \\ &= O\left(\int_{\mathbb{R}^2}\sum_k|\nabla f^T(\sigma_W^2(0),\sigma_W(0)(x^2+y^2))\begin{bmatrix}-\sigma_W^2(k)\\ -2\sigma_W(k)xy\end{bmatrix}|dxdy\right) \\ &= O\left(\sum_k\sigma_W(k)+\sum_k\sigma_W^2(k)\right)\end{aligned} \tag{7.14}$$

With $\{W_t\}$ satisfying (C2), $\sum_k \sigma_W(k)$ is absolutely convergent, then so is $\sum_k \sigma_W^2(k)$, the above quantity is bounded.
For (3), Since K' is a centered gaussian process, We have for $s < t$

$$E|K'(t) - K'(s)|^\alpha = \sqrt{\frac{2}{\pi}} \int_{x\in(0,\infty)} x^\alpha \frac{1}{\sigma_{s,t}} e^{-\frac{x^2}{2\sigma_{s,t}^2}} dx = C_3 \sigma_{s,t}^{\alpha-1} \tag{7.15}$$

Where $\sigma_{s,t}^2 = \sum_{k\in\mathbb{Z}} E(I(U_0 \in (s,t]) - (t-s))(I(U_k \in (s,t]) - (t-s)) \quad \cdots (**)$.
It suffice to show that $\exists \alpha > 1, \beta, C > 0$ s.t. $\sigma_{s,t}^2 \leqslant C(t-s)^{\frac{2(1+\beta)}{\alpha-1}}$. Note that

$$(**) \leqslant \int_{(x,y)\in A_{s,t}^2} \sum_k |f_W^{(k)}(x,y) - f_W(x) f_W(y)| dxdy \tag{7.16}$$

Where $A_{s,t} \subseteq \mathbb{R}$ is such that $\int_{A_{s,t}} f_W(x)dx = t - s$. With same Taylor expansion arguments, there exists positive constants C_i such that the right hand side of equation (7.16) is bounded by $C_1 E(W^2 I(W \in A_{s,t})) + C_2 E(|W| I(W \in A_{s,t})) + C_3 E(I(W \in A_{s,t}))$, where W is a normal random variable with density f_W. Since in the calculation of the expectations the dominating term is the exponential decay, $\forall \epsilon > 0$, $E(W^2 I(W \in A_{s,t})) = o((t-s)^{1-\epsilon})$, and so is $E(|W| I(W \in A_{s,t}))$. Thus $\exists \epsilon_0 \in (0,1)$, $(**) \leqslant C(t-s)^{\epsilon_0}$. The Kolmogorov continuity condition is satisfied. By Borell's inequality(Chapter 2.1, Adler (1990)), the tail probability of the supremum is bounded by the tail probability of the Gaussian distribution with variance $\sup_{t\in[0,1]} \Gamma'(t,t) < \infty$. Thus we have proved the desired result. □

Proof of Lemma 4.3. By calculations in Theorem 1.2 in Li and Racine (2006),

$$E(\widehat{F}_h(x)) - F(x) = \frac{c_2}{2} h^2 F^{(2)}(x) + o(h^2). \tag{7.17}$$

where $c_2 = \int x^2 k(x) dx < \infty$. Thus with $F^{(2)}(x)$ bounded and $h = o(n^{-1/4})$,

$$\sup_{x\in\mathbb{R}} |E(\widehat{F}_h(x)) - F(x)| = o(n^{-1/2}). \tag{7.18}$$

Also by (A7) we have

$$\sqrt{n}(\widehat{F}_h(x) - E(\widehat{F}_h(x))) = G_h(x) + o_p(1) \tag{7.19}$$

and $G_h(x)$ is a tight centered Gaussian process with covariance $\Gamma_{G_h}(x,x') = \sum_{k\in\mathbb{Z}} E\left(K(\frac{x-Y_0}{h}) - E(K(\frac{x-Y_0}{h}))\right)\left(K(\frac{x-Y_k}{h}) - E(K(\frac{x-Y_k}{h}))\right) \to \Gamma(x,x')$ as $h(n) \to 0$, $\forall x, x' \in \mathbb{R}$, where Γ is given in equation (7.2). As shown in Lemma 4.2, the limit Gaussian process is tight. This is sufficient for

$$\sup_{x\in\mathbb{R}} \left|\widehat{F}_h(x) - E(\widehat{F}_h(x))\right| = O_p(\frac{1}{\sqrt{n}}). \tag{7.20}$$

Together with equation (7.18) we have $\sup_{x\in\mathbb{R}}|\widehat{F}_h(x)-F(x)|=O_p(\frac{1}{\sqrt{n}})$. Therefore $\sup_{t\in[n]}|\widehat{U}_t^{(h)}-U_t|=O_p(\frac{1}{\sqrt{n}})$.

□

Proof of Theorem 4.4. Note that

$$\sum_{k=0}^{\lfloor c_\kappa l\rfloor}|\widehat{\tilde{\sigma}}(k)-\sigma(k)|\leqslant\underbrace{\sum_{k=0}^{\lfloor c_\kappa l\rfloor}|\widehat{\tilde{\sigma}}(k)-\widehat{\sigma}(k)|}_{(1)}+\underbrace{\sum_{k=0}^{\lfloor c_\kappa l\rfloor}|\widehat{\sigma}(k)-\sigma(k)|}_{(2)} \tag{7.21}$$

First of all, for (2), by Lemma 3.4, $\{Z_t\}$ is (C2) with $A>1$ and therefore satisfies (C0). Then by Theorem7.3, with $l=o(n^{\frac{p-2}{p}})$, (2) $\xrightarrow{P}0$.

For (1), let $\tilde{Z}_t=\tilde{\Phi}^{-1}(U_t)$ and $\hat{\tilde{Z}}_t=\tilde{\Phi}^{-1}(\widehat{U}_t)$, where $\tilde{\Phi}$ is defined in Section 2. Then $\tilde{\Phi}^{-1}$ is a function bounded by c and $-c$.

Now consider:

$$\begin{aligned}
\sum_{k=0}^{\lfloor c_\kappa l\rfloor}|\widehat{\sigma}_Z(k)-\widehat{\sigma}_Z(k)| &= \sum_{k=0}^{\lfloor c_\kappa l\rfloor}\frac{1}{n}|\sum_{t=1}^{n-k}\left(Z_tZ_{t+k}-\widehat{\tilde{Z}}_t\hat{\tilde{Z}}_{t+k}\right)| \\
&= \sum_{k=0}^{\lfloor c_\kappa l\rfloor}\frac{1}{n}|\sum_{t=1}^{n-k}\left(Z_tZ_{t+k}-\tilde{Z}_t\tilde{Z}_{t+k}+\tilde{Z}_t\tilde{Z}_{t+k}-\widehat{\tilde{Z}}_t\hat{\tilde{Z}}_{t+k}\right)| \\
&\leqslant \sum_{k=0}^{\lfloor c_\kappa l\rfloor}|\frac{1}{n}\sum_{t=1}^{n-k}\left(Z_tZ_{t+k}-\tilde{Z}_t\tilde{Z}_{t+k}\right)| \quad (T_1) \\
&+ \sum_{k=0}^{\lfloor c_\kappa l\rfloor}|\frac{1}{n}\sum_{t=1}^{n-k}\left(\tilde{Z}_t\tilde{Z}_{t+k}-\widehat{\tilde{Z}}_t\hat{\tilde{Z}}_{t+k}\right)| \quad (T_2)
\end{aligned} \tag{7.22}$$

Let $X\lesssim Y$ denote $\exists c>0, X\leqslant cY$.

For T_2,

$$\begin{aligned}
T_2 &= \sum_{k=0}^{\lfloor c_\kappa l\rfloor}|\frac{1}{n}\sum_{t=1}^{n-k}\tilde{Z}_t\tilde{Z}_{t+k}-\tilde{Z}_t\hat{\tilde{Z}}_{t+k}+\tilde{Z}_t\hat{\tilde{Z}}_{t+k}-\widehat{\tilde{Z}}_t\hat{\tilde{Z}}_{t+k}| \\
&= \sum_{k=0}^{\lfloor c_\kappa l\rfloor}|\frac{1}{n}\sum_{t=1}^{n-k}\tilde{Z}_t\left(\tilde{Z}_{t+k}-\hat{\tilde{Z}}_{t+k}\right)+\hat{\tilde{Z}}_{t+k}\left(\tilde{Z}_t-\widehat{\tilde{Z}}_t\right)| \\
&\lesssim \sum_{k=0}^{\lfloor c_\kappa l\rfloor}\frac{c}{\phi(c)}\sup_{t\in[n]}|U_t-\widehat{U}_t| \\
&= O_p(lce^{\frac{c^2}{2}}n^{-1/2}).
\end{aligned} \tag{7.23}$$

Choose $c(n)\to\infty$ such that $lce^{\frac{c^2}{2}}=o(\sqrt{n})$ makes $T_2\to 0$ in probability.

Meanwhile,

$$\begin{aligned} ET_1 &\leqslant \sum_{k=0}^{\lfloor c_\kappa l \rfloor} \frac{1}{n} E|\sum_{t=1}^{n-k} Z_t Z_{t+k} - \tilde{Z}_t \tilde{Z}_{t+k}| \\ &= \sum_{k=0}^{\lfloor c_\kappa l \rfloor} \frac{n-k}{n} E|Z_t Z_{t+k} - Z_t \tilde{Z}_{t+k} + Z_t \tilde{Z}_{t+k} - \tilde{Z}_t \tilde{Z}_{t+k}| \\ &\lesssim lE|Z_t(Z_{t+k} - \tilde{Z}_{t+k})| \\ &\leqslant l \left\|Z_t\right\|_2 \left\|Z_{t+k} - \tilde{Z}_{t+k}\right\|_2 \\ &= O(l(ce^{-c^2/2})^{1/2}) \end{aligned} \tag{7.24}$$

with specific calculations for $E(Z_{t+k} - \tilde{Z}_{t+k})^2$ using $Z_{t+k} \sim \mathcal{N}(0,1)$. Thus by choosing $l, c \to \infty$ such that $lce^{c^2/2} = o(\sqrt{n})$ and $lc^{1/2}e^{-c^2/4} = o(1)$ we have both $T_1 \xrightarrow{P} 0$ and $T_2 \xrightarrow{P} 0$.

Equation (4.9) is a direct consequence of equation (4.8). Note that both $\hat{\hat{\Sigma}}_n$ and Σ_n are Toeplitz, then:

$$\begin{aligned} \left\|\hat{\hat{\Sigma}}_n - \Sigma_n\right\|_{op} &\leqslant \sqrt{\left\|\hat{\hat{\Sigma}}_n - \Sigma_n\right\|_1 \left\|\hat{\hat{\Sigma}}_n - \Sigma_n\right\|_\infty} = \sum_{k\in\mathbb{Z}} |\hat{\hat{\sigma}}(k) - \sigma(k)| \\ &= \sum_{|k|\leqslant\lfloor c_\kappa l\rfloor} |\hat{\hat{\sigma}}(k) - \sigma(k)| + \sum_{|k|>\lfloor c_\kappa l\rfloor} |\sigma(k)| \end{aligned} \tag{7.25}$$

Where the 1-norm and ∞-norm of a $n \times n$ matrix A is defined by $\|A\|_1 = \max_{1\leqslant j\leqslant n} \sum_{i=1}^n |a_{i,j}|$ and $\|A\|_\infty = \max_{1\leqslant i\leqslant n} \sum_{j=1}^n |a_{i,j}|$ respectively. If A is Toeplitz, then $\|A\|_1 = \|A\|_\infty$ by definition, thus the first equality holds.

The First term of above is shown to converge to 0 in probability. For second term, note that $\sigma(k)$ is absolutely summable by (C2) condition on Z_t, then with $l(n) \to \infty$, second term converge to 0. Since the spectral density of W_t is both bounded and bounded away from 0, with Z_t a linear transform from W_t, same holds for Z_t. Then by Theorem 2 of McMurry and Politis (2010), $\left\|\hat{\hat{\Sigma}}_n^{-1} - \Sigma_n^{-1}\right\|_{op} \xrightarrow{P} 0$. □

Proof of Lemma 4.5. Before proving the assertion, some preliminary lemmas are required and listed below:

Lemma 7.4. *Given that $\widehat{F}$ is uniformly consistent for F_Y and F_Y is continous and strictly increasing, then $\forall p \in (0,1)$, $\widehat{F}^{-1}(p) \xrightarrow{P} F_Y^{-1}(p)$.*

Proof. See Lemma 1.2.1 in Politis, Romano and Wolf (1999) □

Lemma 7.5. *Let $U \sim Unif(0,1)$. Then $\widehat{Y} = \widehat{F}^{-1}(U) \sim \widehat{F}$ and $Y = F_Y^{-1}(U) \sim F_Y$, With $d_\infty(\widehat{F}, F_Y) = \sup_x |\widehat{F}(x) - F_Y(x)| \xrightarrow{P} 0$, $\widehat{Y} \xrightarrow{d} Y$ in probability.*

Proof. For the first part, see Angus (1994). The second part is straightforward. □

By Lemma 7.5 and Theorem 4.4, also by continuous mapping theorem, (1) in Lemma 4.5 is straightforward. We focus on proving (2).

Note that for both ecdf and kernel cdf with appropriate bandwidth choice, $\sup_x |\widehat{F}(x) - F_Y(x)| \xrightarrow{P} 0$ with relative assumptions. Then by continuous mapping theorem, $\forall k \in \mathbb{N}$, $\tilde{\Phi}^{-1}(\widehat{F}_Y(Y_k))$ converges in probability to $Z_k = \Phi^{-1}(F(Y_k))$. Specifically, $\forall d \leqslant n$, $\hat{\underline{Z}}_d \xrightarrow{P} \underline{Z}_d$. Then with assumption (A9)

$$\hat{\underline{\xi}}_d \xrightarrow{P} \underline{\xi}_d \sim \mathcal{N}(0, I_d) \tag{7.26}$$

Thus by Lemma 3.4 of Kallenberg (1997), sequence $\hat{\xi} = (\hat{\xi}_1, \cdots)$ converges in probability to $\xi = (\xi_1, \cdots)$ with repect to the metric $\rho(\hat{\xi}, \xi) = \sum_{k=1}^{\infty} 2^{-k} |\hat{\xi}_k - \xi_k|$, where $\xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Let $\bar{F}_{\hat{\xi}}$, $\bar{F}_{\xi}$ be the empirical CDF of $\hat{\xi}$ and ξ respectively, i.e. $\bar{F}_{\xi}(x) = \frac{\sum_{k=1}^{n} I(\xi_k \leqslant x)}{n}$ and $L_{n,x}(\xi) = \bar{F}_{\xi}(x)$, $L_x(\xi) = \lim_{n\to\infty} L_{n,x}(\xi)$. By using continous mapping theorem(Theorem 18.11, Van Der Vaart (1998)) on $L_{n,x}(\cdot)$ and $L_x(\cdot)$, we have for x a.e., $\bar{F}_{\hat{\xi}}(x) - \bar{F}_{\xi}(x) \xrightarrow{P} 0$. Since $\sup_x |\bar{F}_{\xi}(x) - \Phi(x)| \to 0, a.s.$, along with continuity of Φ and by Pólya's theorem, $\sup_x |\bar{F}_{\hat{\xi}}(x) - \Phi(x)| \xrightarrow{P} 0$. Since for any finite dimensional vector $(\xi^*_{t_1}, \cdots, \xi^*_{t_d})$, each of the elements are i.i.d. sampled from $\bar{F}_{\hat{\xi}}$, then $(\xi^*_{t_1}, \cdots, \xi^*_{t_d})$ converge in distribution(in bootstrap world) to a d-vector with i.i.d. standard normals in probability. By Theorem 3.29 of Kallenberg (1997), assertion in (4.13) holds.The remaining proof goes back to the setting in (1). □

Proof of Theorem 4.6. The following several key components constitutes the proof:

1. Finite dimensional convergence in distribution shown in Lemma 4.5.
2. The bootstrap sample Y^*_t are $\lfloor c_\kappa l \rfloor$-dependent with respect to P^* due to *i.i.d.* sampling of the ξ^*_ts and bandedness of $\hat{\tilde{\Sigma}}_n$. Also $\exists \delta > 0, \forall t, E^*(Y^*_t)^{2+\delta} = O_p(1)$ from $L^p - m$ approximable assumption of Y_t(with $p > 2$) which implies uniform integrability of $(Y^*_t)^2$ (with respect to n). Also (C2) holds for Y^*_t in probability.
3. Central limit theorem holds by Lemma 4.1 for $\bar{Y}^*_n$ in probability.

The consequences of 1 & 2 above are two fold. One, the long run variance of Y^*_t can be written as $(\sigma^*_\infty)^2 = \sum_{k=-\lfloor c_\kappa l \rfloor}^{\lfloor c_\kappa l \rfloor} \gamma^*_Y(k)$ which is absolutely convergent in probability as $n \to \infty$ by Lemma 4.1. Second, by similar arguments to the proof of Theorem 3.3 in Bühlmann (1997), we have $\forall M > 0$, $\sum_{k=-M}^{M} |\gamma^*_Y(k) - \gamma_Y(k)| = o_p(1)$. Note that $\sum_{k\in\mathbb{Z}} \gamma_Y(k)$ is also absolutely convergent. All together, they imply $\sum_{k=-\lfloor c_\kappa l \rfloor}^{\lfloor c_\kappa l \rfloor} |\gamma^*_Y(k) - \gamma_Y(k)| = o_p(1)$, and as a consequence, $(\sigma^*_\infty)^2 \xrightarrow{P} \sigma^2_\infty$.

Since CLT holds for both $\bar{Y}_n$ and $\bar{Y}_n^*$(in probability), and the limiting variance of $\bar{Y}_n^*$ is consistent to that of $\bar{Y}_n$, we finished proving Theorem 4.6.

□

Proof of Theorem 4.10. The kernel smoothed spectral density estimator

$$\widehat{f}_{sp.d}(\omega) = \sum_{j\leqslant n} \tilde{\kappa}_h(\omega-\omega_j) I_n(\omega_j) \tag{7.27}$$

can be rewritten as the lag-window spectral density estimator with asymptotically negligible error (even after $\sqrt{nh}$- scaling), i.e.:

$$\widehat{f}_{sp.d}(\omega) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} K_h(k)\widehat{\gamma}(k)e^{-i\omega k} + O\left(\frac{1}{n}\right). \tag{7.28}$$

with $\widehat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n}(Y_t-\bar{Y})(Y_{t+k}-\bar{Y})$ and $K_h(k) = \sum_{j=1}^{n}\tilde{\kappa}_h(\omega-\omega_j)e^{i(\omega_j-\omega)k}$. See Politis and Romano (1992). Under assumption (A12) and Lemma 3.3, also with previous results holding, $\{Y_t\}$ and $\{Y_t^*\}$ satisfy (C0) with $p=4$; with additional assumptions (A13) and (A14) holding, conditions in Theorem 2, Liu and Wu (2010) are satisfied. Therefore asymptotic normality in the sense of equation (4.23) holds for both $\widehat{f}_{sp.d}(\omega)$ and $\widehat{f}^*_{sp.d}(\omega)$ with limiting variance $\sigma_\omega^2 = (\eta_\omega+1)f^2(\omega)\int\tilde{\kappa}^2(u)du$ and $(\sigma_\omega^*)^2 = (\eta_\omega+1)(f^*(\omega))^2\int\tilde{\kappa}^2(u)du$ respectively. Since $|f_{sp.d}(\omega)-f^*_{sp.d}(\omega)| \leqslant \frac{1}{2\pi}\sum_{k\in\mathbb{Z}}|(\gamma_Y(k)-\gamma_Y^*(k))e^{i\omega k}| \leqslant \sum_{k\in\mathbb{Z}}|\gamma_Y(k)-\gamma_Y^*(k)| \xrightarrow{P} 0$ and both $f_{sp.d}(\omega)$ and $f^*_{sp.d}(\omega)$ are bounded and bounded a.s. , $|f^2_{sp.d}(\omega)-(f^*_{sp.d}(\omega))^2| \xrightarrow{P} 0$, which implies $(\sigma_\omega^*)^2 \xrightarrow{P} \sigma_\omega^2$. □

Proof of Theorem 5.1. We prove it for $\widehat{F} = \bar{F}$. The case for the kernel estimator $\widehat{F}_h$ is similar with relative assumptions assumed. We mainly show that the distribution of bootstrap predictive root converges in probability to the true distribution in a conditional sense. i.e., $Y^*_{n+1}-\widehat{Y}^*_{n+1} \xrightarrow{d^*} Y_{n+1}-\widehat{Y}_{n+1}$ in probability, conditioning on past values $\underline{Y}_n$.

Firstly, conditioning on $\underline{Y}_n$, $Y^*_{n+1} \xrightarrow{d^*} Y_{n+1}$ in probability:

$$Y^*_{n+1} = \bar{F}_Y^{-1}(\Phi(Z^*_{n+1})) = \left(\bar{F}_Y^{-1}(\Phi(Z^*_{n+1})) - F_Y^{-1}(\Phi(Z^*_{n+1}))\right) + F_Y^{-1}(\Phi(Z^*_{n+1})) \tag{7.29}$$

We show that first term converges to 0 in probability and the distribution of second term converges to that of Y_{n+1}, in probability. Let $U = \Phi(Z^*_{n+1})$. Then $\bar{F}_Y^{-1}(U) = Y_{(i)}$, where $\frac{i-1}{n} < U \leqslant \frac{i}{n}$. It is equivalent to show $Y_{(i)} - F_Y^{-1}(U) \underset{P}{\to} 0$. Consider $F_Y(Y_{(i)}) - U = \left(F_Y(Y_{(i)}) - \bar{F}_Y(Y_{(i)})\right) + \left(\bar{F}_Y(Y_{(i)}) - U\right)$. The first term is $O_p(\frac{1}{\sqrt{n}})$ by Lemma 4.2. For the second term, $\bar{F}_Y(Y_{(i)}) - U = \frac{i}{n} - U$, where $U \in (\frac{i-1}{n}, \frac{i}{n}]$. Hence the second term goes to 0. Therefore $F_Y(Y_{(i)}) - U \to 0$ in probability. Since F_Y^{-1} is continuous, $Y_{(i)} - F_Y^{-1}(U) \xrightarrow{P} 0$.

To show $F_Y^{-1}(\Phi(Z_{n+1}^*)) \xrightarrow{d^*} Y_{n+1} = F_Y^{-1}(\Phi(Z_{n+1}))$, we need to show conditioning on $\underline{Y}_n$, $Z_{n+1}^* \xrightarrow{d^*} Z_{n+1}$ in probability. For MF bootstrap, let $L_n(\cdot,\cdot) : (\underline{Y}_n, \xi_{n+1}^*) \to Z_{n+1}^*$ with its theoretical analogue $L(\cdot,\cdot) : (\underline{Y}_n, \xi_{n+1}) \to Z_{n+1}$. The formula of $L_n(\cdot,\cdot)$ is estimated from $\underline{Y}_n$ and therefore depends on the past values $\underline{Y}_n$; While $L(\cdot,\cdot)$ is the theoretical data-generating mechanism that links the past values $\underline{Y}_n$ and innovation ξ_{n+1} so it does not depend on $\underline{Y}_n$. By results in Lemma 4.5 and assumption (A9), $L_n(x,y)$ converges to $L(x,y)$ in probability; Also the distribution of ξ_{n+1}^* converges to Φ for MF bootstrap. As a result of continuous mapping theorem, $L_n(\underline{Y}_n, \xi_{n+1}^*) \xrightarrow{d^*} L(\underline{Y}_n, \xi_{n+1})$ in probability.

For LMF bootstrap, we only need to show the normal parameters in step 2 of **Algorithm 3** converge to their theoretical analogue, in probability. As an example, we show $\hat{\tilde{\Sigma}}_{21}\hat{\tilde{\Sigma}}_{11}^{-1}\tilde{\underline{Z}}_n \to \Sigma_{21}\Sigma_{11}^{-1}\underline{Z}_n$ in probability. Let $c_n = \Sigma_{21}\Sigma_{11}^{-1}$& $\hat{\tilde{c}}_n = \hat{\tilde{\Sigma}}_{21}\hat{\tilde{\Sigma}}_{11}^{-1}$. we have

$$\begin{aligned} |\hat{\tilde{c}}_n\tilde{\underline{Z}}_n - c_n\underline{Z}_n| &\leqslant \sum_{i=1}^{n} |\hat{\tilde{c}}_{n,i}\tilde{\tilde{Z}}_i - c_{n,i}Z_i| \\ &\leqslant \sum_{i=1}^{n} |\hat{\tilde{c}}_{n,i}||\tilde{\tilde{Z}}_i - Z_i| + |\sum_{i=1}^{n}\left(c_{n,i} - \hat{\tilde{c}}_{n,i}\right)Z_i| \end{aligned} \tag{7.30}$$

The first term is bounded by $\left(\sum_{i=1}^n |\hat{\tilde{c}}_{n,i}|\right)\sup_i |\tilde{\tilde{Z}}_i - Z_i|$, which is $o_p(1)$ by continuous mapping theorem and $\sum_{i=1}^n |\hat{\tilde{c}}_{n,i}|$ is bounded in probability for large n. The second term has $\mathcal{N}(0, (c_n - \hat{\tilde{c}}_n)\Sigma_n(c_n - \hat{\tilde{c}}_n)^T)$ distribution, where the variance is bounded by $\lambda_{\Sigma_n}^{max}\left\|c_n - \hat{\tilde{c}}_n\right\|_2^2$. Thus if $\left\|c_n - \hat{\tilde{c}}_n\right\|_2 \underset{P}{\to} 0$, then second term converge to 0 in probability.

$$\left\|c_n - \hat{\tilde{c}}_n\right\|_2 \leqslant \left\|\hat{\tilde{\Sigma}}_n^{-1} - \Sigma_n^{-1}\right\|_{op}\left\|\hat{\tilde{\Sigma}}_{21}\right\|_2 + \left\|\hat{\tilde{\Sigma}}_{21} - \Sigma_{21}\right\|_2\left\|\Sigma_n^{-1}\right\|_{op} \tag{7.31}$$

Since $\left\|\hat{\tilde{\Sigma}}_n^{-1} - \Sigma_n^{-1}\right\|_{op} \underset{P}{\to} 0$, by previous results, $\left\|\hat{\tilde{\Sigma}}_{21} - \Sigma_{21}\right\|_2 \underset{P}{\to} 0$. Then $\left\|c_n - \hat{\tilde{c}}_n\right\|_2 \underset{P}{\to} 0$. The proof for consistency of the second parameter follows a similar fashion.

What remains to be shown is conditioning on $\underline{Y}_n$,

$$\hat{Y}_{n+1}^* \xrightarrow{P^*} \hat{Y}_{n+1}. \tag{7.32}$$

(In probability), where $\hat{Y}_{n+1}$ is the self-chosen predictor. In L^2 optimal sense, $\hat{Y}_{n+1} = E(Y_{n+1}|\underline{Y}_n) = E(F_Y^{-1}(\Phi(Z_{n+1}))|\underline{Y}_n)$. Therefore it is sufficient to show that:

$$E^*((\bar{F}^*)^{-1}(\Phi(Z_{n+1}))|\underline{Y}_n) \xrightarrow{P^*} E(F_Y^{-1}(\Phi(Z_{n+1}))|\underline{Y}_n) \tag{7.33}$$

Where the expectation is taken with respect to $(Z_{n+1}|\underline{Y}_n)^*$ and $(Z_{n+1}|\underline{Y}_n)$ as defined in **Algorithm** 2 and 3. Since directly calculating the expectation in our setup is impossible, we use the same approach as previously used in the proof of Theorem 4.6 by showing:

1. $(Z_{n+1}|\underline{Y}_n)^* \xrightarrow{d^*} (Z_{n+1}|\underline{Y}_n)$.
2. Consistency of $(\bar{F}^*)^{-1}$ to F_Y^{-1}.
3. Uniform integrability of $(\bar{F}^*)^{-1}(\Phi(Z_{n+1})$ and $F_Y^{-1}(\Phi(Z_{n+1}))$ conditioning on $\underline{Y}_n$.

First of all, we need to show that $(Z_{n+1}|\underline{Y}_n)^* \xrightarrow{d^*} Z_{n+1}|\underline{Y}_n$ in probability. In LMF bootstrap setup as in **Algorithm** 3, this can be checked by showing convergence of the bootstraped autocovariance matrix $\hat{\hat{\Sigma}}_n^*$ to Σ_n in probability. Since $\{Y_t^*\}_{t=1}^n$ are m-dependent and have finite $p^{th}, (p > 2)$ moment in probability, Theorem 7.3 applies to $\{Y_t^*\}_{t=1}^n$ and thus

$$\left\|\hat{\hat{\Sigma}}_n^* - \Sigma_n^*\right\|_{op} \xrightarrow{P^*} 0 \tag{7.34}$$

where Σ_n^* is the autocovariance matrix of the bootstrap samples $\{Y_t^*\}$. Also by Lemma 4.5 and uniform integrability of $(Y_t^*)^2$, We have $\sum_{k\in\mathbb{Z}} |\sigma^*(k) - \sigma(k)| \xrightarrow{P^*} 0$. This implies $\|\Sigma_n^* - \Sigma_n\|_{op} \xrightarrow{P} 0$. Therefore by triangle inequality,

$$\left\|\hat{\hat{\Sigma}}_n^* - \Sigma_n\right\|_{op} \xrightarrow{P^*} 0. \tag{7.35}$$

In MF bootstrap setup as in **Algorithm** 2, we need to show $\forall z \in \mathbb{R}$, $(\hat{F}_Z^{(n+1)})^*(z) \xrightarrow{P^*} \hat{F}_Z^{(n+1)}(z)$. This can also be shown through equation (4.12) in Lemma 4.5 and a slightly stronger result than equation (7.35) above, i.e.: we assume $(\log n)^2 \left\|\hat{\hat{\Sigma}}_n^* - \Sigma_n\right\|_{op} \xrightarrow{P^*} 0$.

Since $\bar{F}_Y^*$ is the empirical CDF of $\{Y_t^*\}$ for which the CDF F_Y^* satisfies $\sup_{y\in\mathbb{R}} |F_Y^*(y) - F_Y(y)| \xrightarrow{P} 0$ by Lemma 4.5, also along with $m-$dependence of $\{Y_t^*\}$, $\sup_{y\in\mathbb{R}} |\bar{F}_Y^*(y) - F_Y^*(y)| \to 0$ almost surely with respect to P^*. Combining these two we have $\sup_{y\in\mathbb{R}} |\bar{F}_Y^*(y) - F_Y(y)| \to 0$ in probability. By Lemma 7.4 and continous mapping theorem used in Lemma 4.5, $(\bar{F}_Y^*)^{-1}(\Phi(Z_{n+1}))$ with Z_{n+1} having the distribution of $(Z_{n+1}|\underline{Y}_n)^*$ converges in distribution to $F_Y^{-1}(\Phi(Z_{n+1}))$ with Z_{n+1} following the conditional distribution of $Z_{n+1}|\underline{Y}_n$, in probability. Also by finite p^{th} moment of Y_t we have uniform integrability of $(\bar{F}_Y^*)^{-1}(\Phi(Z_{n+1}))$(in probability) and $F_Y^{-1}(\Phi(Z_{n+1}))$. Thus equation (7.33) holds. For the L^1 optimal predictor mentioned in Section 1, proving $\hat{Y}_{n+1}^* \xrightarrow{P^*} \hat{Y}_{n+1}$ follows in a similar fashion but may require additional assumptions; for example, continuity and strict monotonicity of the conditional distribution $Y_{n+1}|\underline{Y}_n$ is necessary to ensure consistency of $\hat{Y}_{n+1}$, which is the conditional sample median.

Summing up the previous results and by Slutsky's theorem, the bootstrap predictive root converges in distribution to the true predictive root(in probability). Since the distribution of the true predictive root is continuous, we have consistency of the bootstrap quantiles. □