2017
arXiv: arXiv:0000.0000

GAUSSIAN MULTIPLIER BOOTSTRAP FOR NON-STATIONARY TIME SERIES

By Sayar Karmakar, Wei Biao Wu

University of Chicago, November 30, 2017

Using a novel logarithmic decomposition technique, Karmakar and Wu (2017) obtained an optimal bound for Gaussian approximation of a large class of vector-valued random processes. This paper proposes a multiplier bootstrap technique to simulate the approximating Gaussian process and thus makes the result of [] useful to several applications in statistical inference. We propose an estimate of the block covariances of the empirical process which is then multiplied with Gaussian innovation to compute the Gaussian analogue. Three applications among many are chosen to demonstrate how the Gaussian approximation and the bootstrap technique together can improve many existing results in the literature.

Key Words and Phrases: Gaussian Approximation, Local stationarity, Non-stationary multiple time series, Weak dependence, Trend Estimation, Unit Root testing

1. Introduction. Based on multiplier central limit theorem by Vaart and Wellner (1996, [28]), Gaussian multiplier bootstrap has remained as one of the important tools for the past two decades in the regime of obtaining size and power of statistical tests involving empirical processes. This has been applied in several areas of statistics such as studying local quantile treatment effect (See Hsu, Kan and Lai (2015)) or constructing critical value of stochastic dominance relations (See Barrett and Donald (2003, [1]), Donald, Hsu and Barrett (2012, [11]), and Donald and Hsu (2014, [12]). Generalized regression monotonicity by Hsu, Liu and Shi (2016, [19]) and specification test for moment inequality models by Bugni, Canay and Shi (2017, [5]) are only some of the many applications of this very effective yet intuitive bootstrap procedure.

The multiplier bootstrap essentially depends on some invariance principle where a Gaussian approximation is established for some empirical process. This approximation can be obtained in different sense. Very recently, Chernozukov et al (2014, [6], 2016, [6]) discussed a multiplier bootstrap using a Gaussian approximation of the maximum co-ordinate of an empirical process in an independent set-up. Zhang and Wu (2017) generalized their result allowing weak dependence within a stationary process. For such an invariance principle, in this paper we choose the Gaussian approximation for non-stationary, multiple time series obtained in Karmakar and Wu (2017). We start with a brief overview of the history of strong invariance principle and its applications.

For an empirical process $(X_i) \in \mathbb{R}^d$, let S_i denote the partial sum $S_i = \sum_{j=1}^{i} X_j$. If $d = 1$ and X_i are i.i.d. and have finite pth moment for some $p > 2$, Komlós et al. (1974/75,

[22],[23]) proved that in a probably richer probability space it is possible to construct S_i' and a mean zero Gaussian process G_i such that $(S_i)_{i\leq n}' \stackrel{D}{=} (S_i)_{i\leq n}$ and

$$|S_i' - G_i| = o_{a.s}(n^{1/p}). \tag{1.1}$$

It is also well-known that under the finite pth moment condition this $n^{1/p}$ bound is not improvable. Since a large class of sufficient statistics are based on partial sums of empirical process, there are several areas of statistics and econometrics where result of type (1.1) can be used. This motivation opened the area of research where extensions in multiple directions were sought. See Wu and Zhou (2011, [32]), Berkes et al. ([2]), Karmakar and Wu ([]) and the references therein for a more detailed overview of the history of how the results were generalized to different set-ups.

For $d > 1$, Liu and Lin (2009, [25]) obtained the KMT-type bound for a stationary weak dependent sequence having finite pth moment where $2 < p < 4$. [32] also restricted their research to the case $2 < p < 4$ but allowed non-stationarity and obtained bounds close to $o_p(n^{1/p})$. [] obtained the $\tau_n = n^{1/p}$ bound for a non-stationary weak dependent time series for all $p > 2$ and thus solved a long-standing open problem. This very deep result has several applications in analysis of multiple time-series. We carefully choose three such examples and demonstrate a bootstrap procedure to obtain/simulate the approximating Gaussian process. These should give the reader a general idea of how to use our result in other scenarios. The bootstrap method is kept specifically focused in the context of our first application for presentational clarity.

1.1. *Applications.* Consider the following model

$$y_{i,n} = \mu(i/n) + e_i, \tag{1.2}$$

where μ is a continuous function from $[0, 1] \to \mathbb{R}^d$ and e_i is a dependent non-stationary error process. In an univariate set-up, Wu and Zhao (2007, [31]) obtained simultaneous confidence band for the unknown trend function for the model in (1.2). They assumed stationarity of the error process and double-differentiability of the μ function. In section 3, we treat similar models and extend it in three directions. We allow vector-valued process, use trend function μ that is not Lipschitz and allow the error process to vary in a locally stationary way. This discussion can be viewed as a premier to address a more general non-stationary error process. We also highlight that if the function μ is only Hölder- $1/2$ continuous, then we will need Gaussian approximation with bound sharper than $o_P(n^{1/4})$.

The theoretical simultaneous confidence band still suffers from a few drawbacks. It is well-known in the theory of convergence of extreme value of Gaussian process, that the convergence to Gumbel distribution is logarithmic. Thus it requires a large sample size to have approximately accurate coverage probabilities. Also, one needs to assume some

more structure on how the physical mechanism H_i in (2.1) changes. Lastly, one needs some knowledge about the unobserved ϵ_i variables to construct such a confidence band. To this end, we will show how our bootstrap technique can solve this issue in its most generality and allow us to obtain asymptotically correct simultaneous confidence band.

The Gaussian approximation obtained in [] is not amenable for practical use as the covariance structure of the approximating Gaussian process remains complicated. To address this issue, we first propose an estimate of the covariance structure of the approximating Gaussian process. Next the covariance estimate is used to propose a Gaussian multiplier bootstrap technique. We choose two more applications to exhibit how the bootstrap method can appear as a remedy where obtaining null distribution is difficult under dependence and non-stationarity of error process. We seek the null distribution of the unknown coefficients in the functional regression model (See (6.1)). The deviation from the true value of the coefficients can be approximated by a Gaussian analogue using the invariance principle and thus a Bootstrap method can be proposed to perform inference.

We show a third application in the regime of unit-root testing for panel data. where we put very mild restriction on contemporaneous correlation of the observed data. The inference using the usual estimator of ρ, the AR(1) co-efficient for univariate models of the type $y_i = \rho y_{i-1} + e_i$ has been well-studied for the nearly non-stationary (ρ close to 1) or non-stationary ($\rho = 1$) case. See Dickey and Fuller (1979, [10]), Evans and Savin (1984, [14]) and Nankeris and Savin (1985, [26]) among others. In this paper, we allow panel data as our observation y_i and the error innovations can be a very general non-stationary process. Whereas the usual estimator of ρ easily extends for the panel data, the non-stationarity in the error process makes it difficult to perform statistical inference. We show how our Gaussian bootstrap method can overcome this difficulty. Although we restricted ourselves to the dimension $d = 2$ for this particular example, it can be easily extended to any fixed, higher dimension.

1.2. *Organization of the article.* The rest of the article is organized as follows. In Section 2, we introduce the functional dependence measure and use it to state the Gaussian approximation result. Section 3 discusses the inference on the trend regression model (1.2). We revisit this example in Section 5 using the bootstrap technique in Section 4. We use Section 6 and 7 to discuss two more possible applications of the sharp Gaussian approximation result in the paradigm of functional regression analysis and unit root testing respectively. We defer all the proofs to Section 9.

1.3. *Notation.* For a matrix $A = (a_{ij})$, we define its Frobenius norm as $|A| = (\sum a_{ij}^2)^{1/2}$. For a random vector Y, write $Y \in \mathcal{L}_p, p > 0$, if $\|Y\|_p := [E(|Y|^p)]^{1/p} < \infty$. For $\mathcal{L}_2$ norm write $\|\cdot\| = \|\cdot\|_2$. Define the projection operator P_i by

$$P_i Y = E(Y|\mathcal{F}_i) - E(Y|\mathcal{F}_{i-1}), \qquad Y \in \mathcal{L}_1. \tag{1.3}$$

Throughout the text, $\lfloor x \rfloor$ refers to the greatest integer less than or equal to x. We use C_p to refer to a constant that depends only on p but could take different values on different occurrences. The p-variate normal distribution with mean μ and covariance matrix Σ is denoted by $N_p(\mu, \Sigma)$. For a positive semi-definite matrix A, $A^{1/2}$ refers to the usual Grammian square root of A. If $A = QDQ^T$ is the spectral decomposition of the matrix A then $A^{1/2} = QD^{1/2}A^T$. Say the diagonal entries of D are $\lambda_1 \geq \ldots \lambda_d$. Then $\rho^*(A) = \lambda_1$ and $\rho_*(A) = \lambda_d$. If two quantities A and B satisfy $A \leq cB$ for some $c < \infty$, then we write $A \lesssim B$ or $B \gtrsim A$. If both $A \lesssim B$ and $B \lesssim A$ then we write $A \asymp B$. If $A_n/B_n \to 0$, then we say $A_n \ll B_n$. If $A_n/B_n \to \infty$, then we say $A_n \gg B_n$.

2. Functional dependence and Gaussian Approximation.

2.1. *The functional dependence measure.* To state the structure of dependence we allow throughout the paper, we introduce an uniform functional dependence measure on the underlying process using the idea of coupling as done in Wu (2005, [29]). Assume that X_i has mean 0, $X_i \in \mathcal{L}_p, p > 0$ and it admit the causal representation

$$X_i = H_i(\epsilon_i, \epsilon_{i-1}, \ldots) \tag{2.1}$$

When H_i can potentially take different functional value for different i, it points to the non-stationarity of the process. Suppose $(\epsilon_i')_{i\in\mathbb{Z}}$ is an independent copy of $(\epsilon_i)_{i\in\mathbb{Z}}$. For $j \geq 0$, $0 < r \leq p$, define the functional dependence measure

$$\delta_{j,r} = \sup_{i\geq 1} \|X_i - X_{i,(i-j)}\|_r = \sup_{i\geq 1} \|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,(i-j)})\|_r, \tag{2.2}$$

where $\mathcal{F}_{i,(k)}$ is the coupled version of $\mathcal{F}_i$ with ϵ_k in $\mathcal{F}_i$ replaced by an i. i. d. copy ϵ_k',

$$\mathcal{F}_{i,(k)} = (\epsilon_i, \epsilon_{i-1}, \cdots, \epsilon_k', \epsilon_{k-1}, \cdots) \tag{2.3}$$

and $X_{i,(i-j)} = H_i(\mathcal{F}_{i,(i-j)})$. Also, $\mathcal{F}_{i,k} = \mathcal{F}_i$ if $k > i$. $\|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,(i-j)})\|_r$ measures the dependence of X_i on ϵ_{i-j}. The functional dependence measure $\delta_{j,r}$ quantifies the uniform j-lag dependence in terms of the r^{th} moment. In general, the tail cumulative measure $\Theta_{j,r}$ for $j \geq 0$ and $2 \leq r \leq p$ is defined as

$$\Theta_{j,r} = \sum_{i=j}^{\infty} \delta_{i,r}.$$

One can think of the functional dependence measure as a measurement of an input-output mechanism.

2.2. *General assumptions for Gaussian Approximation:.* For completeness, we rewrite the representation (2.1) of the (X_i) process.

(2.A) X_i is a $d-$dimensional mean 0 vector process with the representation

$$X_i = H_i(\mathcal{F}_i) = (X_{i1}, \ldots, X_{id})^T, \text{ where } \mathcal{F}_i = (..., \epsilon_{i-1}, \epsilon_i),$$

where ϵ_i are i. i. d., and H_i is a measurable function such that X_i is well-defined, and ^{T} denotes matrix transpose.

(2.B) The sequence $(|X_i|^p)_{i\geq 1}$ is uniformly integrable. Namely

$$\sup_{i\geq 1} E(|X_i|^p |\mathbf{1}_{|X_i|\geq u}) \to 0 \text{ as } u \to \infty. \tag{2.4}$$

(2.C) We assume short range dependency,

$$\Theta_{0,p} = \sum_{i=0}^{\infty} \delta_{i,p} < \infty. \tag{2.5}$$

This condition implies the cumulative dependence of $(X_j)_{j\geq k}$ on ϵ_k is finite.

(2.D) There exists a constant $\lambda_* > 0$ such that, for all $t, l \geq 1$ and any unit vector v

$$\inf_t \rho_*(Var(\sum_{i=t+1}^{t+l} X_i)) \geq \lambda_* l. \tag{2.6}$$

Comment 2.1. *(The lower bound on variances) The condition (2.D) is intuitive since if this condition is violated, it will pose the problem of singularity. Similar conditions have been imposed in Gaussian approximation literature. See ChernoZhukov et. al. (2014, []), Wu and Zhou (2011, [32]) for instance. This condition can be relaxed if the normalized incremental process corresponding to only a non-zero fraction of co-ordinates have their minimum eigenvalue bounded away from 0.*

We assume the following upper bound for the cumulative dependence measure $\Theta_{i,p}$.

$$\Theta_{i,p} = O(i^{-\chi}), \quad \chi > 0. \tag{2.7}$$

Result 2.2. *In addition to the assumptions (2.A) to (2.D), we also assume the formulation (2.7) of the cumulative dependence measure $\Theta_{i,p}$. If,*

$$\chi > \chi_0 = \frac{p^2 - 4 + (p-2)\sqrt{p^2 + 20p + 4}}{8p} \tag{2.8}$$

then there exists a probability space (Ω_c, A_c, P_c) *on which we can define random vectors* X_i^c *with the partial sum process* $S_n^c = \sum_{i=1}^n X_i^c$ *and a Gaussian process* G_i^c *with independent increments such that* $S_i^c \stackrel{D}{=} (S_i)_{i \le n}$ *and*

$$\max_{i \le n} |S_i^c - G_i^c| = o_P(n^{1/p}), \qquad \text{in } (\Omega_c, A_c, P_c), \tag{2.9}$$

where $G_i^c = \sum_{t=1}^i Y_t^c$ *with* Y_t^c *being independent Gaussian random variables with mean 0.*

3. Application: Inference on multivariate time-varying models with a non-Lipschtiz trend function. Estimation of trends for an *a posteriori* dataset is a well-studied problem in statistics and related fields. We will consider a non-parametric regression model of the form

$$y_i = \mu\left(\frac{i}{n}\right) + e_i, \tag{3.1}$$

It is important to know if the unknown intrinsic trend function for a given dataset has a particular parametric form and thus simultaneous confidence bands (SCB) are obtained to perform such inference problems. In an univariate set-up see Johnston (1982, [21]), Hall and Titterington (1988, [17]) and Cummins, Filloon and Nychka (2001, [8]) among others for constructing SCB for unknown μ with independent errors. The seminal invariance principle result by Komlos et al. (1975, [23])for independent scalar random variables was used in Eubunk and Speckman (1993, [13]) to construct the SCB. However, it is well-known in the literature that dependence within the stochastic process may appear as a difficult obstacle for constructing the simultaneous bands. Wu and Zhao (2007,[31]) tested for structural breaks in the μ function using the invariance principle by Wu (2007, [30]) to extend the results to a set-up with weakly dependent errors. The Gaussian approximation result in [30] was sub-optimal compared to that of [23] to adjust for the dependence. Recently, Berkes et al. (2014, [2]) achieved the optimal $n^{1/p}$ rate for weakly dependent stationary scalar time-series under the assumption of finite pth moment of the error process.

An invariance principle result for all $p > 2$ similar to that of [2] was not available for a multiple time-series until recently. This was recently solved in [] and thus gives us a motivation to extend the set-up in (3.1) to a multivariate one. Moreover, since the Gaussian approximation available can be sharper than $n^{1/4}$ under the mild assumption of pth moment, it is possible to relax some smoothness criterion on μ. Although historically, many different specifications were put in for μ and tested after observing the data, we will stick to a simple model as in (3.1) as our focus is on loosening the smoothness assumption of the trend function μ. The error process e_i specified in the model of [31] is univariate, dependent but stationary. We will obtain the simultaneous confidence band for μ allowing the following extensions from therein:

- We allow y_i, $\mu(\frac{i}{n})$ and e_i to be vector valued. This is an important extension as this allows us to test for more than one trend function simultaneously.
- μ is not Lipschitz continuous, but only Hölder -α continuous, i.e, μ is a function from $[0,1] \to \mathbb{R}^d$ which satisfy, for all $0, x, y < 1$, and some $\alpha \leq 1$,

$$|\mu(x) - \mu(y)| \leq C|x-y|^{\alpha}. \tag{3.2}$$

  This very weak restriction provides a strong generalization of similar models considered in literature. A function with two smooth derivatives admits a Taylor expansion and that is widely exploited in the literature. On the contrary, little has been researched for the case when the unknown function is not Lipschitz continuous. We know the path of a Brownian motion is not Lipschitz continuous in a stochastic sense. We emphasize that our unknown function is not random but this assumption allows us to incorporate those with similar abrupt changes. For known covariance structure of the data X_i or the error process e_i, we will be able to achieve the class of functions with $\alpha = 1/2$, thus accommodating functions that behave like a Brownian motion. Our result can be substantially useful in many applications of finance and econometrics where one can fit a random walk model. We will also propose how to estimate the covariance matrix leading to a Gaussian multiplier bootstrap.
- The process e_i is allowed to be non-stationary. Naturally, this requires some formulation of the non-stationarity or some regular conditions on it. For the time being, we assume that,

$$e_i = H_i(\epsilon_i, \epsilon_{i-1}, \cdots) = H_i(\mathcal{F}_i) = H(\frac{i}{n}, \mathcal{F}_i)), \tag{3.3}$$

  where ϵ_i are identically distributed independent random variables. We assume local-stationarity for the e_i process. i.e. $H(\cdot, \mathcal{F}_i)$ is stochastic Lipschitz continuous in the sense that there exists $C < \infty$ such that

$$\sup_{0 \leq t < s \leq 1} \frac{\|H(t, \mathcal{F}_i) - H(s, \mathcal{F}_i)\|}{|t-s|} \leq C. \tag{3.4}$$

  The causal representation (3.3) allows us to use the functional dependence framework by Wu (2005,[29]) allows us to handle the dependent process e_i. We will later relax this local-stationarity assumption.

3.1. *Assumptions.* Before using the Gaussian approximation from Theorem 2.2, we provide a list of assumptions for the error process e_i. For completeness, we also add the causal representation, local-stationarity and the short-range dependency assumptions:

A1.1 The error process e_i is a $d-$dimensional mean 0, non-stationary random sequence that assumes the following representation

$$e_i = H_i\left(\frac{i}{n}, \mathcal{F}_i\right) = (e_{i1}, e_{i2}, ...e_{id})^T,$$

where $\mathcal{F}_i = (..., \epsilon_{i-1}, \epsilon_i)$, H_i is a measurable function such that e_i is a well-defined random vector, and ^{T} denotes matrix transpose. The ϵ_i random variables are i. i. d. We also assume that H_i satisfies (3.4).

A1.2 The process e_i is uniformly integrable in $\mathcal{L}_p$.

$$\sup_{i\geq 1} E(|e_i|^p \mathbf{1}_{|e_i|\geq u}) \to 0 \text{ as } u \to \infty. \tag{3.5}$$

Moreover, this implies $e_i \in \lessdot^p$. We also assume short range dependency of the error process

$$\Theta_p(\mathbf{e}, 0) = \sum_{i=0}^{\infty} \delta_p(\mathbf{e}, i) < \infty. \tag{3.6}$$

A1.3 There exists a positive real number m_1 such that for any $d-$ dimensional unit norm vector v and for any integer r,

$$\rho_*(Var(\sum_{i=r+1}^{r+l} e_i)) \geq \lambda_* l. \tag{3.7}$$

A1.4 The cumulative dependence measure $\Theta_p(\mathbf{e}, i)$ satisfies the following

$$\Theta_p(\mathbf{e}, i) = O(i^{-\chi}), \tag{3.8}$$

for some $\chi > \chi_0$ where χ_0 is as described in (2.8).

LEMMA 3.1. *Define*

$$F_n(t) = \sum_{i=1}^{n} V_i K\left(\frac{t_i - t}{b_n}\right),$$

where V_i are i. i. d. $N(0, Id_d)$. If b_n satisfies (3.9), then,

$$\lim_{n\to\infty} (P[\frac{1}{\sqrt{\phi_0 n b_n}} \sup_{t\in\tau} |F_n(t)| - B(r) \leq \frac{u}{\sqrt{2\log(r)}}]) = \exp\{-2\exp(-u)\}.$$

PROOF. See Lemma 1 from Zhou and Wu (2010, [34]). □

Recall P_i from (1.3). Define $\Sigma_e(t) = E(K_i(t)K_i(t)^T)$ where $K_i(t) = \sum_{j=i}^{\infty} P_i(H(t, \mathbb{F}_j))$. We have the following lemma about convergence of $\Sigma_e(t)^{-1/2} D_e(t)$

LEMMA 3.2. *Let $D_e(t) = (nb_n)^{-1} \sum_{i=1}^{n} e_i K((t_i - t)/b_n)$. Suppose K is of bounded variation and $\Sigma_e(t)$ is Lipschitz continuous in $[0, 1]$. We have the following result for constructing the simultaneous confidence band.*

$$\lim_{n\to\infty} (P[\sqrt{nb_n} \sup_{t\in\tau} \{|\Sigma_e(t)^{-1/2} D_e(t)|\} - B(r) \leq \frac{u}{\sqrt{2\log(r)}}]) = \exp(-2\exp(-u)).$$

THEOREM 3.3. *Additional to the assumptions A.1- A.7, assume the kernel function K is of bounded variation and b_n satisfy*

$$\sqrt{\log(1/b_n)}\left(\frac{n^{1/p}}{\sqrt{nb_n}}+b_n\log n+n^{1/2}b_n^{\alpha+\frac{1}{2}}+b_n^{-1/2}n^{-1/2}\right)\to 0. \tag{3.9}$$

Then we have

$$P(\frac{\sqrt{nb_n}}{\phi_0}\sup_{t\in\tau}|\Sigma_e(t)^{-1/2}\{\mu_{b_n}(t)-\mu(t)\}|-B(r)\le\frac{u}{\sqrt{2\log r}})\to\exp\{-2\exp(-u)\}, \tag{3.10}$$

where $r=1/b_n$, $\tau=[b_n,1-b_n]$, and

$$B(r)=\sqrt{2\log(r)}+\frac{\log(C_K)+(d/2-1/2)\log(\log r)-\log(2)}{\sqrt{2\log(r)}},$$

with

$$C_K=\frac{(\int_{-1}^{1}|K'(u)|^2du)^{1/2}}{\Gamma(d/2)}.$$

COMMENT 3.4. *By Theorem (3.3) , if $\hat{\Sigma}_e(t)$ is the uniformly consistent estimate of $\Sigma_e(t)$, then the simultaneous confidence band (SCB) for $\mu(t)$,*

$$\mu_{b_n}(t)+\sqrt{\frac{\phi_0}{nb_n}}[B_K(r)-\frac{\log(\log((1-\alpha)^{-1/2}))}{\sqrt{2\log(1/b_n)}}]\hat{\Sigma}_e(t)\mathcal{B}_d, \tag{3.11}$$

has $100(1-\alpha)\%$ coverage where $\mathcal{B}_d$ is he d-dimensional unit ball.

COMMENT 3.5. *If we assume $b_n\asymp n^{-\delta}$ then by (3.9), we need*

$$\frac{1}{2}<(\alpha+\frac{1}{2})\delta,\quad \frac{2}{p}<(1-\delta),0<\delta<1.$$

If μ is Hölder-1/2 continuous, i.e. $\alpha=1/2$, then the Gaussian approximation obtained by Wu and Zhou (2011, [32]) will not be useful. We will need $o(n^{1/p})$ bound for some $p>4$ and existence of Gaussian approximation of such optimal bound has been proved in Theorem 2.2.

3.2. *A smooth estimate for the covariance matrix.* Note that, $\Sigma_e(t)$ in Lemma 3.2 is difficult to estimate. Also it is only applicable in case the process has the local stationary property. Here, we try to generalize the results for a general non-stationary process.

A careful check of the proof of the Gaussian approximation Result in 2.2 reveals that it says there exists

$$\max_{i\leq n}|S_i - \sum_{j=1}^{[i/2k_0m]}\Sigma_j^{1/2}Z_j| = o_p(n^{1/p}), \tag{3.12}$$

where $\Sigma_j = Var(B_j)$ is an estimate of the variance of the block $X_{2k_0m(j-1)+1}+\cdots+X_{2k_0mj}$. We introduce a smooth version to facilitate using properties of partial sums of standard Gaussian random variables.

$$\Sigma(t) = \begin{cases} \sum_{2ik_0m/n\leq t}\Sigma_i & \text{if } t = 2jk_0m/n \\ \alpha\sum_{i=1}^{j-1}\Sigma_i + (1-\alpha)\sum_{i=1}^{j}\Sigma_i & \text{if } 2(j-1)k)m/n < t < 2jk_0m/n. \end{cases} \tag{3.13}$$

It can be easily shown that,

$$\sup_{0<t\leq 1}|\Sigma(t) - \sum_{2ik_0m/n\leq t}\Sigma_i| = o_p(n^{2/p}), \tag{3.14}$$

Thus we can use results like Lemma 3.2 to exploit the theory for extreme value Gaussian processes.

4. A block Gaussian multiplier resampling procedure. We revisit the trend-regression model (3.1) discussed in Section 3 and implement a bootstrap method to circumvent some of its restrictive assumptions and limitations.

- General Non-stationary error: Note that in Section 3, we assumed the error process in our model (3.1) to be locally stationary (See (3.4)). In this section we introduce a bootstrap technique to relax that assumption.
- Estimating $\Sigma_e(t)$: Looking at the expression of $\Sigma_e(t)$ one can see that it depends on the unobserved ϵ_i process and hence it is difficult to consistently estimate it.
- Logarithmic convergence: The logarithmic convergence to the extreme value distribution in theorem 3.3 makes it practically infeasible as one needs huge n to have appropriate coverage probability. This is our main motivation behind proposing a bootstrap method.

4.1. *Bootstrapped version of SCB.* Compared to the assumptions of Theorem 3.3, we will only need the assumptions made in Theorem 2.2 for the error process e_i to construct a bootstrap SCB for $\mu(t)$. Note that,

$$\mu_{b_n}(t) - \mu(t) \quad = \quad \frac{1}{nb_n}\sum_{i=1}^{n}K\left(\frac{i/n-t}{b_n}\right)e_i + Bias_n, \tag{4.1}$$

where

$$Bias_n = \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n-t}{b_n}\right)(\mu(i/n)-\mu(t)) + \mu(t)\{\frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n-t}{b_n}\right) - 1\}.$$

Also,

$$\sqrt{nb_n}Bias_n = O(\sqrt{n}b_n^{\alpha+1/2}). \tag{4.2}$$

We will use our Gaussian block multiplier to create the quantiles of

$$(nb_n)^{-1/2}\sum_{i=1}^{n} K\left((i/n-t)/nb_n\right)e_i.$$

Assuming K is of bounded variation, by summation-by-parts formula we have,

$$\max_{1\le i\le n}|\frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n-t}{b_n}\right)e_i - \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n-t}{b_n}\right)g_i| = o_P(n^{1/p})O\left(\frac{1}{nb_n}\right), \tag{4.3}$$

where partial sums of the Gaussian random variables g_i form the process G_i^c as described in Theorem (2.2) Define the Gaussian analogue of the estimated μ as following

$$\mu^Z(t) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n-t}{b_n}\right)g_i.$$

We have the following algorithm for constructing the simultaneous confidence band (SCB).

Algorithm for known covariance If $E(\Delta_f)$ are known for the noise process e_i, then we use the following algorithm.

- For $i = 1, 2, 10000$ obtain

$$\sup_{t\in(b_n,1-b_n)}|\mu^{Z_1}(t)|, \cdots, \sup_{t\in(b_n,1-b_n)}|\mu^{Z_{10000}}(t)|.$$

- Get l and u, the $\alpha/2\%$ and $100(1-\alpha/2)\%$ quantiles of the empirical process $\sup_{t\in(b_n,1-b_n)}|\mu^{Z^i}(t)|$.
- Let the two quantiles be l and u. The SCB for $\mu(t)$ using (4.3) is $\mu_{b_n}(t)-l, \mu_{b_n}(t)+u$.

From (4.2) and (4.3), we create the following condition on b_n

$$\frac{n^{1/p}}{\sqrt{nb_n}} + \sqrt{n}b_n^{\alpha+1/2} = o(1). \tag{4.4}$$

Under the above condition on the bandwidth b_n, we have the following theorem for validating our algorithm.

THEOREM 4.1. *Assume b_n satisfies 4.4. Then,*

$$\sup_t |\sqrt{nb_n}(\mu_{b_n}(t) - \mu(t) - \mu^Z(t))| = o_P(1). \tag{4.5}$$

PROOF. Trivially follows from (4.2) and (4.3). □

COMMENT 4.2. *Note that, from (4.4), one can see that if our unknown function μ is only Hölder-1/2 continuous, then one needs $p > 4$ and $b_n \ll n^{-1/2}$. Since the underlying function is not as smooth as a twice differentiable one, one needs a much smaller bandwidth to capture the small changes within the function.*

5. Estimation of covariance of the approximating process. The approximating Gaussian process in Theorem 2.2 G_i^c has a complicated structure arising from different steps of our proof. We first discuss the key strategies behind the proof of the theorem so that it can provide some motivation behind the estimates of the covariance matrix.

5.1. *Key strategies for the proof of Result 2.2.* In this subsection, we throw some light on the key steps of the proof of Result 2.2 as it will serve as a background of the block covariance estimators proposed. There are mainly three stages of the long and involved proof. We will mainly discuss the first of them in detail to remain pertinent to the context of this paper. The other two stages will be briefly mentioned for sake of completeness.

Truncation, m-dependence and blocking approximation A series of approximations are done in the first stage as a pre-processing. The first of them, a truncation approximation is used to allow Rosenthal-type inequality with moments higher than p. For $b > 0$ and $v = (v_1, \ldots, v_d)^T \in \mathbb{R}^d$, define

$$T_b(v) = (T_b(v_1), \ldots, T_b(v_d))^T, \text{ where } T_b(w) = \min(\max(w, -b), b). \tag{5.1}$$

The truncation operator T_b is Lipschitz continuous with Lipschitz constant 1. Using uniform integrability(cf 2.4) of the process X_i t is possible to choose a t_n such that,

$$\max_{1\le i\le n} |S_i - S_i^{\oplus}| = o_P(n^{1/p}), \text{ where } S_l^{\oplus} = \sum_{i=1}^{l} [T_{t_n n^{1/p}}(X_i) - ET_{t_n n^{1/p}}(X_i)]. \tag{5.2}$$

As exhibited in [2] and Karmakar and Wu (2017,[]), it is essential to use such a high power to obtain the optimal rate.

The next approximation is m-dependent approximation, the most important step of the proof. This m-dependence approximation technique has a long history in the literature of dependent data analysis. See Lin and Liu (2011, [24]) for a brief overview. Using the dependence framework by Wu (2005, []), one can create a m-dependent analogue of the

partial sum process with optimal error rate. For a suitably chosen sequence $m_n \to \infty$, using the decay condition (cf 2.7), we have

$$\max_{1\le i\le n} |S_i - \tilde{S}_i| = o_P(n^{1/p}), \text{ where } \tilde{S}_l = \sum_{i=1}^{l} E(T_{t_n n^{1/p}}(X_j)|\epsilon_j, \ldots, \epsilon_{j-m_n}) - E(T_{t_n n^{1/p}}(X_j))$$

Henceforth we drop the suffix of m_n. m is chosen to vary in an almost polynomial rate, $m \approx n^{L_0}$ in []. This approximation gives a very simple yet effective way to handle the original process X_i in terms of only m many ϵ_i's. As we assume that the dependence of X_i and X_{i+k} dies down as k grows, dividing the m-dependent partial sum process in blocks of sufficiently large length seems a natural choice. This allows to use the existing Gaussian approximation results that are only suitable for independent process. To allow the non-stationarity [] used blocks of size $2k_0m$ compared to the blocks of size $3m$ in [2]. For $q_i = \lfloor i/(2k_0m)\rfloor$, we have,

$$\max_{1\le i\le n} |\tilde{S}_i - S_i^{\diamond}| = o_P(n^{1/p}), \text{ where } S_i^{\diamond} = \sum_{j=1}^{q_i} A_j = \sum_{j=1}^{q_i} (\tilde{S}_{2k_0jm} - \tilde{S}_{2(j-1)k_0m}). \tag{5.3}$$

Conditional and Unconditional Gaussian Approximation: The blocks created in the first stage are not independent because two successive blocks possibly share some ϵ_i's in their shared border. In this second step, we look at the partial sum process conditioned on these borderline ϵ_i's. In the last part of the proof, the Gaussian approximation for the unconditional process was obtained by applying Götze and Zaitsev (2008, [15])'s result one more time. The treatment for this part is similar to [2].
In this section, we first propose a simple approximation of the covariance of the partial sums of the Gaussian process G_i^c in Theorem 2.2. This approximated covariance is simple and easy to estimate using the data itself. The following well-known result for Gaussian process will play a key-role in proving efficiency of our bootstrap procedure.

RESULT 5.1. *Let G_{i1} and G_{i2} be two Gaussian process defined as following*

$$G_{i1} = \sum_{j=1}^{i} \Sigma_{j1}^{1/2} Z_j \text{ and } G_{i2} = \sum_{j=2}^{i} \Sigma_{j2}^{1/2} Z_j,$$

where Z_j are standard d-variate normal random variables. Then,

$$\max_{i\le n} |G_{i1} - G_{i2}| = o_P(\max_{i\le n} |\sum_{j=1}^{i} \{\Sigma_{j1} - \Sigma_{j2}\}|^{1/2} (\log n)^{1/2}). \tag{5.4}$$

The m-dependent approximation suggests the following estimator of block covariances. Define,

$$W_l = \sum_{1 \le i,j \le l} I(|i-j| \le m) X_i X_j^T. \tag{5.5}$$

Choose a sequence M such that $m/M \to 0$. Define $\Delta_F = W_{Mf} - W_{Mf-M}$ for a suitable sequence M.

PROPOSITION 5.2. *Under the conditions of Theorem 2.2, $E(\Delta_f)$ is positive definite.*

The original approximating Gaussian process G_i^{raw} has independent but possibly non-identical increment. However, the covariance matrix for each such increment is not easily estimable from data since it is derived after a series of transformation.

5.2. *Approximation of the covariance structure.* First, we prove a deterministic version of the main theorem of this section.

THEOREM 5.3. *Define the process (G_i^E) as follows*

$$G_i^E = \sum_{f=1}^{[i/M]} (E(\Delta_f))^{1/2} Z_f. \tag{5.6}$$

Then we have,

$$\max_{i \le n} |G_i^c - G_i^E| = o_P(n^{max(1/p, 1/2-(1-2/p)\chi/(1+\chi))}). \tag{5.7}$$

COMMENT 5.4. *From (5.7) one can see that we achieve the optimal $n^{1/p}$ bound if $\chi > \chi_0 \ge 1$ for $p > 4$.*

5.3. *Estimating the unknown covariance of the error process.*

PROPOSITION 5.5. *Under the conditions of Theorem 2.2, it is possible to choose a sequence $M \ll n$ such that Δ_f is positive definite with probability going to 1.*

In practice, the covariance will not be known. For the oracle case where e_i is known, we use the natural estimate Δ_f as an estimate of $E(\Delta_f)$ with X_i therein replaced by e_i. Define,

$$G_i^{\Delta} = \sum_{f=1}^{[i/M]} \Delta_f^{1/2} Z_f,$$

where $\Delta_f = W_{Mf} - W_{Mf-M}$ and Z_f are i. i. d. d-dimensional standard normal variables independent of (X_i). Note that, here Δ_f is also a random variable whereas in our approximating process the covariance matrices were deterministic.

PROPOSITION 5.6. *For $p > 4$,*

$$\max_{i \le n} | \sum_{f=1}^{[i/M]} \Delta_f - \sum_{j=1}^{[i/2k_0m]} Var(A_j)| = o_P(n^{2/r}/(\log n)^2), \tag{5.8}$$

Using Result (5.1) and Proposition (5.6) we have the following theorem for the validity of the bootstrap procedure.

THEOREM 5.7.

$$\max_{i \le n} |G_i^{\Delta} - G_i^c| = o_{P^*}((nm)^{1/4}(\log n)^{1/2}), \tag{5.9}$$

where P^* refers to the conditional probability given the data $\{X_i\}$.

COMMENT 5.8. *It is striking that our original Gaussian approximation achieves the $n^{1/p}$ bound where as the bootstrapped version can only obtain roughly $n^{1/4}$ bound. Even in univariate i. i. d. set-up one cannot estimate the long-run variance better than $\sqrt{n}$ rate which translates to $n^{1/4}$ rate for the bootstrapped process with the standard deviation multiplier.*

COMMENT 5.9. *From expression (4.4), it is clear that one needs to have an approximation sharper that $n^{1/4}$ in order to achieve the correct coverage rate for the simultaneous confidence band. On the other hand, note that from the comment (5.8), it is impossible to achieve such a bound. If we can achieve the $(nm)^{1/4}(\log n)^{1/2}$ rate as described in Theorem (5.7), we can still do inference on unknown μ functions that are Hölder-α continuous with $1/2 < \alpha < 1$. To achieve $\alpha = 1/2$ one needs to know the unknown covariance structure of the error process. In practice, for real life data, even if $\alpha = 1/2$, the estimate using Δ_f works fine as a way to simulate the surrogate process.*

Moreover, the e_i's are unobserved. We propose an estimate of G_i^* with X_i therein replaced by $\hat{e}_i = y_i - \hat{\mu}(i/n)$. To show the validity of such a replacement, we need to bound the covariance estimate arising from $E(e_i e_j^T)$ and that from $\hat{e}_i \hat{e}_j^T$.

THEOREM 5.10.

$$\max_{l \le n} | \sum_{1 \le i,j \le l} I(|i-j| \le m)(\hat{e}_i \hat{e}_j^T - E(e_i e_j^T))| = o_P(n^{\delta}), \tag{5.10}$$

for some small δ. Thus the two Gaussian process defined using covariances of the type $\hat{e}_i \hat{e_j}^T$ and $E(e_i e_j^T)$ are close.

6. Application: Functional linear regression model. Consider the functional linear regression model

$$y_i = x_i^T \beta + e_i, \tag{6.1}$$

where y_i and e_i are the response and the noise/error process respectively. x_i stands for the known d-dimensional vector of the covariate for the i^{th} data point and β is the corresponding unknown vector of regression coefficients. The usual ordinary least square estimator of β is

$$\hat{\beta} = (\sum_{i=1}^{n} (x_i x_i^T))^{-1} \sum_{i=1}^{n} x_i y_i. \tag{6.2}$$

Under the assumption of independence and normality of the error process e_i, the distribution of $\hat{\beta} - \beta$ can be easily obtained and hence it can be used for inference problems concerning β. We use the Gaussian multiplier bootstrap to obtain the distribution of $\hat{\beta} - \beta$ and thus can allow a very general dependent, non-linear and non-stationary error process e_i. Note that,

$$\hat{\beta} - \beta = (\sum_{i=1}^{n} x_i x_i^T)^{-1} \sum_{i=1}^{n} x_i e_i. \tag{6.3}$$

Define the corresponding Gaussian analogue

$$\hat{\beta}^g - \beta = (\sum_{i=1}^{n} x_i x_i^T)^{-1} \sum_{i=1}^{n} x_i g_i. \tag{6.4}$$

where g_i are obtained from the Gaussian process so that

$$\max_{i \le n} |\sum_{j=1}^{i} e_j - \sum_{j=1}^{i} g_j| = o_P((nm)^{1/4} (\log n)^{1/2}), \tag{6.5}$$

as described in (5.9). The following theorem states that the two statistic $\hat{\beta} - \beta$ and $\hat{\beta}^g - \beta$ are close to each other.

THEOREM 6.1. *Assume e_i satisfy similar conditions as described in Theorem 2.2 and the cumulative dependence measure for the error process e_i satisfy*

$$\Theta_{i,p}^e = O(i^{-\chi}), \quad \chi > \chi_0.$$

where χ_0 is mentioned in (2.8). Let $g_1, \cdots g_n$ be the simulated Gaussian process satisfying (6.5). Then,

$$(\sum_{i=1}^{n} x_i x_i^T)(\hat{\beta}^g - \hat{\beta}) = (|\sum_{i=1}^{n} |x_i - x_{i-1}||) o_P((nm)^{1/4} (\log n)^{1/2}). \tag{6.6}$$

where m is as described in the discussion following Theorem 2.2.

PROOF. Follows directly by summation-by-parts formula and (5.9) on $\sum_{i=1}^{n} x_i(e_i - g_i)$. □

COMMENT 6.2. *Here we assumed x_i to be known. However, Theorem 6.1 has the potential to be easily extended to the case where x_i can also evolve randomly with time-index i. For random x_i's however, note that, under regular conditions on the covariance structure of the random vectors x_i, it is possible to simplify the convergence rate in (6.6);*

$$n^{c_x}(\hat{\beta} - \hat{\beta}^g) = O_P(1), \tag{6.7}$$

for some uniform constant $c_x > 0$.

COMMENT 6.3. *Theorem 6.1 implies that for inference problems for the unknown β, one can reasonably use $\hat{\beta}^g$ as a surrogate.*

7. Application: Unit root testing for multiple time-series. In this section, we study the problem of unit root testing for multiple time series as an application of our sharp Gaussian approximation result. Over the past three decades, unit root testing in panel data has attracted a great deal of attention. Phillips (1987, [27]) was one of the first to discuss some dependence across time. These were further generalized by incorporating cross-sectional/contemporaneous correlation in Im et. al. (2003, [20]), Breitung (2005 , [3]), Breitung and Das (2005, [4]) and Wachter et. al. (2007,[9]). Also see Harris and Tzavalis (1999, [18]), Hadri (2000, [16]) and Choi (2001, [7]) . However, there are strong assumptions on the contemporaneous correlation structure in these works.

In our work, we do not put any restriction on contemporaneous correlation except the error process following condition (2.D) (2.6). This is a very mild restriction and can also be relaxed in the sense that we need only a positive proportion of co-ordinates to obey such a condition. The dependence restriction we put on the time dependence is through the framework by Wu (2005, [29]) and hence it is easily verifiable. The time dependence of the observed data happen through the error process. We allow non-stationary and non-linear errors that admits the representation (2.1) and exhibit short-range dependency as in (2.5). For simplicity, we concentrate on two-dimensional vector process X_i modeled as follows:

$$\begin{pmatrix} X_{i,1} \\ X_{i,2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \rho \begin{pmatrix} X_{i-1,1} \\ X_{i-1,2} \end{pmatrix} + \begin{pmatrix} e_{i,1} \\ e_{i,2} \end{pmatrix}, \tag{7.1}$$

where the $(e_{i,1}, e_{i,2})^T$ is the error process and $(\mu_1, \mu_2)^T$ is the drift process. For the time being we assume $\mu_1 = \mu_2 = 0$. In the sequel, the data and the error process are denoted by X_i and e_i for convenience. After observing the data $X_1, \cdots X_n$, we will test

$$H_0 : \rho = 1 \text{ vs } H_1 : \rho < 1.$$

If $\rho < 1$ the process is stationary, so this is often used to test stationarity of the observed auto-regressive process as well. We propose the usual slope estimator in a regression context

$$\hat{\rho} = \frac{\sum X_{i-1}^T X_i}{\sum X_{i-1}^T X_{i-1}}. \tag{7.2}$$

For a general non-stationary process e_i, it is difficult to obtain the null distribution of $\hat{\rho}$. To circumvent that, we use the following identity

$$\hat{\rho} - \rho = \frac{\sum_{i=1}^n X_{i-1}^T e_i}{\sum_{i=1}^n X_{i-1}^T X_{i-1}}$$

and the Gaussian multiplier bootstrap proposed in Section 4. Under the null $H_0 : \rho = 1$, since X_i reduces to simply the partial sum process of e_i.

$$\hat{\rho} - \rho = \frac{\sum_{i=1}^n e_i^T (\sum_{0<j<i} e_{i-j})}{\sum_{i=1}^n X_{i-1}^T X_{i-1}}.$$

Since our Gaussian approximation regularizes the partial sum by a Gaussian analogue in the light of Theorem 2.2, we can obtain the approximate null distribution of $\hat{\rho} - \rho$ by bootstrapping from it's Gaussian analogue

$$\hat{\rho}^g - \rho = \frac{\sum_{i=1}^n g_i^T (\sum_{0<j<i} g_{i-j})}{\sum_{i=1}^n X_{i-1}^T X_{i-1}},$$

where $g_1, \cdots g_n$ form a Gaussian process such that

$$\max_{i \le n} |\sum_{j=1}^i e_j - \sum_{j=1}^i g_j| = o_P((nm)^{1/4} (\log n)^{1/2}). \tag{7.3}$$

as described in (5.9). We prove that the two statistics $\hat{\rho}$ and $\hat{\rho}^g$ are close to each other in the following theorem.

THEOREM 7.1. *Assume e_i satisfy similar conditions as described in Theorem 2.2) and the cumulative dependence measure for the error process e_i satisfy*

$$\Theta^e_{i,p} = O(i^{-\chi}), \chi \geq \chi_0,$$

Let $g_1, \cdots g_n$ be the simulated Gaussian process satisfying (7.3). Then, we have,

$$n^{5/4} m^{-1/4} (\log n)^{-1/2} (\hat{\rho}^g - \hat{\rho}) = o_P(1).$$

PROOF. Let $S^e_i = \sum_{j=1}^i e_j$ and $S^g_i = \sum_{j=1}^i g_j$. We bound the distance of the numerators of the expression of $\hat{\rho} - \rho$ and $\hat{\rho}^g - \rho$.

(7.4)

$$\begin{aligned} |\sum_{i=1}^n e_i^T (\sum_{0<j<i} e_j) - \sum_{i=1}^n g_i^T (\sum_{0<j<i} g_j)| &= |\sum_{i=1}^n e_i S^e_{i-1} - \sum_{i=1}^n g_i S^g_{i-1}| \\ &\leq |\sum_{i=1}^n e_i S^e_{i-1} - \sum_{i=1}^n g_i S^e_{i-1}| + |\sum_{i=1}^n g_i S^e_{i-1} - \sum_{i=1}^n g_i S^g_{i-1}| \\ &= o(n^{3/4} m^{1/4} (\log n)^{1/2}), \end{aligned}$$

by summation by-parts formula and (5.9). The denominator of the two expression $\hat{\rho} - \rho$ and $\hat{\rho}^g - \rho$ are roughly of the order of n^2 from the assumption that e_i satisfy the condition (2.6). □

To obtain $\hat{\rho}^g$, we need the Gaussian approximation of the partial sums of e_i. But since they are unobserved, we will use a surrogate $\hat{e}_i$ to obtain the Z_i. We summarize our algorithm of the bootstrap technique below.

1. Obtain $\hat{\rho}$ using (7.2) and using $\hat{\rho}$ obtain $\hat{e_i}$.
2. Obtain the covariance matrix estimates using $\hat{e}_i$ in place of X_i as described in Section 5.
3. Using the covariance matrix estimate, obtain the approximating Gaussian distribution for a large number of times. For each iteration, we obtain a set of g_i such that

$$\max_{i \leq n} |\sum_{j=1}^i \hat{e}_i - \sum_{j=1}^i g_i| = o_P((nm)^{1/4} (\log n)^{1/2}).$$

Using g_i obtain

$$\hat{\rho}^g - \rho = \frac{\sum_{i=1}^n g_i^T (\sum_{0<j<i} g_j)}{\sum_{i=1}^n (\sum_{0<j<i} g_j)^T (\sum_{0<j<i} g_j)}.$$

4. Let q be the $100(1-\alpha)\%$ quantiles of $|\hat{\rho}^g - \rho| = |\hat{\rho}^g - 1|$.

We reject $H_0 : \rho = 1$ if $|\hat{\rho} - 1| > q$.

COMMENT 7.2. *We can use ideas similar to what we discussed in Section 4 to construct confidence interval for the unknown ρ in a general case.*

COMMENT 7.3. *It is easy to extend for the case where drift $\mu \neq 0$. Note that, the estimator of ρ and μ in this case turns out to be the following*

$$\hat{\rho}' = \frac{\sum_{i=1}^n X_{i-1}^T X_i - \frac{1}{n}\sum_{i=1}^n X_{i-1}^T (\sum_{i=1}^n X_i)}{\sum_{i=1}^n X_{i-1}^T X_{i-1} - \frac{1}{n}\sum_{i=1}^n X_{i-1}^T (\sum_{i=1}^n X_{i-1})}. \tag{7.5}$$

and

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^n y_i - \hat{\rho}' \frac{1}{n}\sum_{i=1}^n y_{i-1}.$$

Following similar ideas, one can construct the asymptotic null distribution of $\hat{\mu} - \mu$ and $\hat{\rho} - \rho$ in the same fashion.

Another possible extension is to extend ρ to a matrix R in $\mathbb{R}^{d^2}$ and test whether it is identity versus the alternative that the largest eigenvalue of R lies inside the unit circle.

8. Simulation Studies.

8.1. *Trend Estimation.* ssc:simu trend We work with a simple AR(1) model with different coefficient ρ and different values of n. We will use $\mu(x) = x^{1//2}$ and $b_n = n^{-1/2}$.

8.2. *Unit Root.* We use sample size and dependent error as described in Subsection 8.2 and show that our test statistic based on Gaussian analogue has approximately $\alpha\%$ level of significance.

9. Proofs.

PROOF. of Lemma 3.2 Note that,

$$D_e(t) = \sum_{i=1}^n w_n(t,i) e_i,$$

where $w_n(t,i) = (nb_n)^{-1} K\left(\frac{t_i - t}{b_n}\right)$. Let $\Omega_n(t) = |w_n(t,1)| + \sum_{i=2}^n |w_n(t,i) - w_n(t,i-1)|$. As K is of bounded variation, we have

$$\Omega_n = \max_{0 \leq t \leq 1} \{\Omega_n(t)\} = O(\frac{1}{nb_n}).$$

By summation-by-parts formula and Corollary 2 in Wu, Zhou (2011,[32]), for $\chi > (2p-4)/(p+2)$ in (2.7), we have

$$\sup_{t\in\tau}|D_e(t)-\kappa(t)| \le \Omega_n \max_{i\le n}|\sum_{j=1}^{i} e_j - \sum_{j=1}^{i}\Sigma_e(t_i)^{1/2}V_i| = o\left(\frac{n^{1/p}}{nb_n}\right), \tag{9.1}$$

where

$$\kappa(t) = \frac{1}{nb_n}\sum_{i=1}^{n}\Sigma_e(t_i)V_i K\left(\frac{t_i-t}{b_n}\right).$$

Since $\Sigma_e(t)$ is assumed to be Lipschitz continuous, we also have

$$\sup_{t\in\tau}|\kappa(t) - \frac{1}{nb_n}\Sigma_e(t)\sum_{i=1}^{n}V_i K\left(\frac{t_i-t}{b_n}\right)| = O\left(\sqrt{\frac{b_n}{n}}\log n\right). \tag{9.2}$$

By the choice of our b_n in (3.9) and Lemma 3.1 we conclude the proof. □

Proof. of Theorem 3.3 Suppose, μ is Hölder- α continuous. We will write $\mu_{b_n}(t)-\mu(t)$ as sum of the stochastic and the bias part.

$$\mu_{b_n}(t)-\mu(t) = (\mu_{b_n}(t)-E(\mu_{b_n}(t))) + E(\mu_{b_n}(t)) - \mu(t).$$

where $\mu_{b_n}(t)-E(\mu_{b_n}(t))$ is the stochastic part and $E(\mu_{b_n}(t))-\mu(t)$ is the bias part. Note that, in Lemma (3.2), $D_e(t) = \mu_{b_n}(t) - E(\mu_{b_n}(t)$. For the bias part, our choice of b_n in (3.9) makes

$$\sqrt{nb_n}\sup_t|E(\mu_{b_n}(t))-\mu(t)| \to 0.$$

$$\begin{aligned} & E(\mu_{b_n}(t))-\mu(t) \\ = & \frac{1}{nb_n}\sum_{i=1}^{n}K\left(\frac{t_i-t}{b_n}\right)\mu(\frac{i}{n})-\mu(t) \\ = & \frac{1}{nb_n}\sum_{i=1}^{n}K\left(\frac{t_i-t}{b_n}\right)(\mu(\frac{i}{n})-\mu(t)) + \mu(t)(\frac{1}{nb_n}\sum_{i=1}^{n}K\left(\frac{t_i-t}{b_n}\right)-1) \\ = & I_n + O(1)(II_n-1). \end{aligned} \tag{9.3}$$

Before we deal with the terms I_n and II_n separately, we discuss some properties of the kernel K. For $j = 0, \alpha, 1/2, 1, 2$, we have,

$$\sup_{t\in\tau}\{\int_0^n |K_j(\frac{[1+v]-nt}{nb_n}) - K_j\left(\frac{v-nt}{nb_n}\right)|dv\} = O(1), \tag{9.4}$$

with $K_j(v)$ being $K(v)v^j$. We also have,

$$\frac{1}{nb_n}\int_0^n |K_j(\frac{v-nt}{nb_n})|dv = \int_{\mathbb{R}} |K_j(u)|du.$$

Now,

$$\begin{aligned} \sqrt{nb_n}I_n &= b_n^{\alpha-\frac{1}{2}} n^{-1/2} \int_0^n |K_\alpha(\frac{[1+v]-nt}{nb_n})|dv \\ &= O(b_n^{\alpha-\frac{1}{2}} n^{-\frac{1}{2}}) O(nb_n) \\ &= O(b_n^{\alpha+\frac{1}{2}} n^{\frac{1}{2}}), \end{aligned} \tag{9.5}$$

$$\sqrt{nb_n}II_n = \sqrt{\frac{1}{nb_n}}(O(1) + \int_o^n K_0(\frac{[1+v]-nt}{nb_n})dv) = O(\sqrt{\frac{1}{nb_n}}), \tag{9.6}$$

since $\int_{\mathbb{R}} K(x) = 1$. By Lemma 3.2, 9.3, (9.5), (9.6) we conclude the proof. □

Proof. of Proposition 5.2

$$W_{Mf} - W_{Mf-M} = \sum_{Mf-m+1 \le i,j \le Mf} e_i e_j^T + \sum_{i=1}^m e_{Mf-M+i}(e_{Mf+1} + \cdots e_{Mf+i})^T. \tag{9.7}$$

$$\begin{aligned} E(W_{Mf} - W_{Mf-M}) &= \sum_{Mf-m+1 \le i,j \le Mf} E(e_i e_j^T) + O(\sum_{i=1}^m i\rho_{i-j}) \\ E(\Delta_f) &= E((S_{Mf} - S_{Mf-M})^2) + O(M\Theta_{m+1}) + \sum_{i=1}^m \Theta_{i,p} \\ &\ge \lambda_* M + M\Theta_{m+1,p} + o(m^{-\chi+1}(\log m)^{-A}) \\ &\ge \lambda_* M/2, \end{aligned}$$

since $m/M \to 0$ and $\Theta_{m+1,p} \to 0$. □

Proof. of Proposition 5.5
Note that, we can also argue $E(\Delta_f) \lesssim M$ and

$$||\Delta_f - E(\Delta_f)|| \lesssim (Mm)^{1/2}.$$

This in particular allows us to derive that Δ_f is positive definite in an uniform manner.

$$
\begin{aligned}
P(\inf_f \rho_*(\Delta_f) > 0) &\leq \frac{n}{M} \sup_f P(|\Delta_f - E(\Delta_f)| > M) \\
&\leq \frac{n}{M}\frac{Mm}{M^2} \\
&= o(1),
\end{aligned} \tag{9.8}
$$

provided one can choose M large enough to have $nm/M^2 \to 0$. Since, $m \ll n$ it is possible to choose such a sequence $M \ll n$. □

Proof. of Theorem 5.3 Without loss of generality, we do the proof for the univariate case, $d = 1$. We first discuss a few key results needed for the proof. As described in the key strategies following Theorem 2.2, the covariance matrix of the approximating process goes through several step whereas our estimate here is based on the original X_i process.
Define $\hat{X} = T_{n^{1/p}}(X) - E(T_{n^{1/p}}(X))$ where for a real number $b > 0$

$$T_b(w) = \min(\max(w, -b), b).$$

From [2] (cf expression 3.53), we have,

$$|E(X_i X_j) - E(\hat{X}_i \hat{X}_j)| = o(n^{2/p} - 1). \tag{9.9}$$

Define the blocks A_j as following. For $0 \leq j \leq q_n = n/2k_0m$,

$$A_{j+1} = \sum_{i=2jk0m+1}^{2jk0m} E(T_{n^{1/p}}(X_i|\epsilon_i, \cdots \epsilon_{i-m}) - E(T_{n^{1/p}} X_i). \tag{9.10}$$

Let B_j denote the variance of j th block of the approximating Gaussian process. As described in the "rearrangement of variance" identity and "truncated variance" identity in Karmakar and Wu (2017), assumption (2.6) implies, for any j_0 and t_0,

$$|\sum_{j=j_0}^{j=j_0+t_0} (Var(A_j) - Var(B_j))| = o(m). \tag{9.11}$$

We omit the details here. Note that, using arguments similar to (9.7) we have

$$|E(\Delta_f) - E(S_{Mf} - S_{Mf-M})^2| = O(M\Theta_{m+1,p}). \tag{9.12}$$

Let $E(\Delta_f^j)$ denote $E(W_{Mf-M+2jk_0m} - W_{Mf-M+2(j-1)k_0m})$. By elementary manipulation, one can show that,

$$E(\Delta_f) = \sum_j E(\Delta_f^j) + o(m). \tag{9.13}$$

Denote γ_i and $\tilde{\gamma}_i$ to be the average of $E(X_k X_{k+i})$ and $E(\tilde{X}_k \tilde{X}_{k+i})$ for $k \in \mathbb{Z}$. From (9.9), (9.12),(9.13) and (9.13) we have,

$$\begin{aligned}
&\sum_{f=1}^{[n/M]} ((\sum_j Var(B_j))^{1/2} - E(\Delta_f)^{1/2})^2 \qquad (9.14)\\
= \;& \frac{n}{M} \sup_f ((\sum_j VarB_j)^{1/2} - (E(\Delta_f))^{1/2})^2 \\
= \;& \frac{n}{M} \sup_f \frac{(\sum_j VarB_j) - (E(\Delta_f)))^2}{((VarA_j)^{1/2} + (E(\Delta_f))^{1/2})^2} \\
= \;& \frac{n}{M^2}(M^2\Theta_{m,p}^2 + m^2 + \sup_f (M/m)^2 \sup_j |Var(A_j) - E(\Delta_f^j)|^2) \\
= \;& n(\Theta_{m,p}^2 + m^2/M^2 + |\gamma_0 - \tilde{\gamma}_0 + 2\sum_{i=1}^m \gamma_i - 2\sum_{i=1}^m \tilde{\gamma}_i|^2) \\
= \;& O(n(\Theta_{m,p}^2 + m^2/M^2 + \min_{l\geq 0}(\Theta_{l.p} + ln^{2/r-1})^2)).
\end{aligned}$$

where the index j in the summation in $\sum_j Var(B_j)$ ranges from $(f-1)M/2k_0m + 1$ to $fM/2k_0m$. To choose a proper l one can equate $l^{-\chi} = ln^{2/r-1}$ to obtain optimal l as $l^* = n^{(1-2/r)/(1+\chi)}$. This choice of l also allows us to choose $M \ll n$ so that, $m \leq M\Theta_{l^*,p}$. using $m \ll n^{2/r}$. This leads to,

$$\max_{i\leq n} |G_i^c - G_i^E|^2 \;=\; o_P(n^{\max\{2/r, 1-2\chi(1-2/r)/(1+\chi)\}}). \tag{9.15}$$

□

Proof. of Proposition 5.6 One can view

$$\max_{i\leq n} |\sum_{f=1}^{[i/M]} \Delta_f - \sum_{j=1}^{[i/2k_0m]} Var(B_j)|$$

as

$$\max_{j\leq n/M} |\sum_{f=1}^{j} \{\Delta_f - \sum_{r=(f-1)M/2k_0m+1}^{fM/2k_0m} Var(B_r)\}| = \max_{j\leq n/M} |\sum_{f=1}^{j} N_f| + \max_{j\leq n/M} |\sum_{f=1}^{j} U_f|,$$

where $N_f = \Delta_f - E(\Delta_f)$ and

$$U_f = E(\Delta_f) - \sum_{r=(f-1)M/2k_0m+1}^{fM/2k_0m} Var(B_r).$$

Using Proposition 1 from Wu (2007, [30]) and Lemma 8 from Xiao and Wu (2012, [33]) we assume, $n/M = 2^{d_1}$ and proceed as follows,

$$\begin{aligned} \| \max_{j \le n/M} |\sum_{f=1}^{j} N_f| \|_q &= \sum_{t=0}^{d_1} (\sum_{c=1}^{2^{d_1-t}} \| \sum_{f=2^t c-2^t+1}^{2^t c} N_f \|_q^q)^{1/q} \\ &= \sum_{t=0}^{d_1} (\sum_{c=1}^{2^{d-t}} (2^t Mm)^{q/2})^{1/q} \\ &= (mM)^{1/2} 2^{d_1/q} \sum_{t=0}^{d_1} 2^{t\{1/2-1/q\}}. \end{aligned} \tag{9.16}$$

Choosing $q = 2$, we get,

$$\| \max_{j \le n/M} |\sum_{f=1}^{j} N_f| \|_q = (nm)^{1/2} \log n.$$

The second term involving U_f is easy and is treated similar to Theorem 5.3.

□

References.

[1] Barrett, G. F. and Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica* **71** 71–104. MR1956856

[2] Berkes, I., Liu, W. and Wu, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.* **42** 794–817. MR3178474

[3] Breitung, J. (2000). The local power of some unit root tests for panel data. **15** 161–177. MR1834817

[4] Breitung, J. and Das, S. (2005). Panel unit root tests under cross-sectional dependence. *Statist. Neerlandica* **59** 414–433. MR2254821

[5] Bugni, F. A., Canay, I. A. and Shi, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics* **8** 1–38.

[6] Chernozhukov, V., Chetverikov, D. and Kato, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl.* **126** 3632–3651. MR3565470

[7] Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance* **20** 249-272.

[8] Cummins, D. J., Filloon, T. G. and Nychka, D. (2001). Confidence intervals for nonparametric curve estimates: toward more uniform pointwise coverage. *J. Amer. Statist. Assoc.* **96** 233–246. MR1952734

[9] De Wachter, S., Harris, R. D. F. and Tzavalis, E. (2007). Panel data unit roots tests: the role of serial correlation and the time dimension. *J. Statist. Plann. Inference* **137** 230–244. MR2292853

[10] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74** 427–431. MR548036

[11] Donald, S. G., Hsu, Y.-C. and Barrett, G. F. (2012). Incorporating covariates in the measurement of welfare and inequality: methods and applications. *Econom. J.* **15** C1–C30. MR2892805

[12] Donald, S. G. and Hsu, Y.-C. (2014). Estimation and inference for distribution functions and quantile functions in treatment effect models. *J. Econometrics* **178** 383–397. MR3132439

[13] Eubank, R. L. and Speckman, P. L. (1993). Confidence Bands in Nonparametric Regression. *Journal of the American Statistical Association* **88** 1287-1301.

[14] Evans, G. B. A. and Savin, N. E. (1984). Testing for unit roots. II. *Econometrica* **52** 1241–1269. MR763072

[15] Götze, F. and Zaitsev, A. Y. (2008). Bounds for the rate of strong approximation in the multidimensional invariance principle. *Teor. Veroyatn. Primen.* **53** 100–123. MR2760567

[16] Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *Econom. J.* **3** 148–161. MR1820412

[17] Hall, P. and Titterington, D. M. (1988). On CONFIDENCE bands in nonparametric density estimation and regression. *J. Multivariate Anal.* **27** 228–254. MR971184

[18] Harris, R. D. F. and Tzavalis, E. (1999). Inference for unit roots in dynamic panels where the time dimension is fixed. *J. Econometrics* **91** 201–226. MR1703947

[19] Hsu, Y.-C., Liu, C.-A. and Shi, X. (2016). Testing Generalized Regression Monotonicity IEAS Working Paper : academic research No. 16-A009, Institute of Economics, Academia Sinica, Taipei, Taiwan.

[20] Im, K. S., Pesaran, M. H. and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *J. Econometrics* **115** 53–74. MR1985474

[21] Johnston, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *Journal of Multivariate Analysis* **12** 402 - 414.

[22] Komlós, J., Major, P. and Tusnády, G. (1974). An approximation of partial sums of independent RV's and the sample DF. I. In *Strong laws of invariance principle: a collection of seven papers, Paper No. VI* i+45 pp. Carleton Math. Ser., No. 106. Dept. Math., Carleton Univ., Ottawa, Ont. MR0385955

[23] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 111–131. MR0375412

[24] Lin, Z.-Y. and Liu, W. (2012). m-dependence approximation for dependent random variables. In *Probability approximations and beyond. Lect. Notes Stat.* **205** 117–133. Springer, New York. MR3289380

[25] Liu, W. and Lin, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Process. Appl.* **119** 249–280. MR2485027

[26] Nankervis, J. C. and Savin, N. E. (1985). Testing the autoregressive parameter with the t statistic. *J. Econometrics* **27** 143–161. MR786612

[27] Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica* **55** 277–301. MR882096

[28] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes. Springer Series in Statistics*. Springer-Verlag, New York With applications to statistics. MR1385671

[29] Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154 (electronic). MR2172215

[30] Wu, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389

[31] Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 391–410. MR2323759

[32] Wu, W. B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. MR2827528

[33] Xiao, H. and Wu, W. B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.*

40 466–493. MR3014314

[34] Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 513–531. MR2758526