AUTO-REGRESSIVE APPROXIMATIONS TO NON-STATIONARY TIME SERIES, WITH INFERENCE AND APPLICATIONS

BY XIUCAI DING[1] AND ZHOU ZHOU[2,*]

[1]*Department of Statistics, University of California, Davis xcading@ucdavis.edu*

[2]*Department of Statistical Sciences, University of Toronto* [*] *zhou@utstat.utoronto.ca*

Understanding the time-varying structure of complex temporal systems is one of the main challenges of modern time series analysis. In this paper, we show that every uniformly-positive-definite-in-covariance and sufficiently short-range dependent non-stationary and nonlinear time series can be well approximated globally by a white-noise-driven auto-regressive (AR) process of slowly diverging order. To our best knowledge, it is the first time such a structural approximation result is established for general classes of non-stationary time series. A high dimensional $\mathcal{L}^2$ test and an associated multiplier bootstrap procedure are proposed for the inference of the AR approximation coefficients. In particular, an adaptive stability test is proposed to check whether the AR approximation coefficients are time-varying, a frequently-encountered question for practitioners and researchers of time series. As an application, globally-optimal short-term forecasting theory and methodology for a wide class of locally stationary time series are established via the method of sieves.

1. Introduction. The Wiener-Kolmogorov prediction theory [20, 21, 31] is a fundamental result in time series analysis which, among other findings, guarantees that a weakly stationary time series can be represented as a white-noise-driven auto-regressive (AR) process of infinite order under some mild conditions. The latter structural representation result has had profound influence in the development of the classic linear time series theory. Later, [1, 2] studied the truncation error of AR prediction of stationary processes when finite many past values, instead of the infinite history, were used in the prediction. Nowadays, as increasingly longer time series are being collected in the modern information age, it has become more appropriate to model many of those series as non-stationary processes whose data generating mechanisms evolve over time. Consequently, there has been an increasing demand for a systematic structural representation theory for such processes. Nevertheless, it has been a difficult and open problem to establish linear structural representations for general classes of non-stationary time series. The main difficulty lies in the fact that the profound spectral domain techniques which were essential in the investigation of the AR(∞) representation for stationary sequences are difficult to apply to non-stationary processes where the spectral density function is either difficult to define or only defined locally in time.

The first main purpose of the paper is to establish a unified AR approximation theory for a wide class of non-stationary time series. Specifically, we shall establish that every short memory and uniformly-positive-definite-in-covariance (UPDC) non-stationary time series $\{x_{i,n}\}_{i=1}^n$ can be well approximated globally by a non-stationary white-noise-driven AR process of slowly diverging order; see Theorem 2.5 for a more precise statement. Similar to the spirit of the Wiener-Kolmogorov prediction theory, the latter structural approximation

MSC2020 subject classifications: Primary 62M10, 62M20; secondary 60G07.

Keywords and phrases: Non-stationary time series, AR approximation, High dimensional Gaussian approximation, Multiplier bootstrap, Global optimal forecasting.

result connects a wide range of fundamental problems in non-stationary time series analysis such as optimal forecasting, dependence quantification, efficient estimation and adaptive bootstrap inference to those of AR processes and ordinary least squares (OLS) regression with diverging number of dependent predictors. In fact, the very reason for us to consider the AR approximation instead of a moving average approximation or representation (c.f. Wold decomposition [32]) to non-stationary time series is due to its close ties with the OLS regression and hence the ease of practical implementation. Our proof of the structural approximation result resorts to modern operator spectral theory and classical approximation theory [10] which control the decay rates of inverse of banded matrices. Consequently the decay speed of the best linear projection coefficients can be controlled via the Yule-Walker equations; see Theorem 2.4 for more details.

The last two decades have witnessed the rapid development of locally stationary time series analysis in statistics. Locally stationary time series refers to the subclass of non-stationary time series whose data generating mechanisms evolve smoothly or slowly over time. See [8] for a review. For locally stationary processes, we will show that the UPDC condition is equivalent to the uniform time-frequency positiveness of the local spectral density of $\{x_{i,n}\}_{i=1}^n$ (c.f. Proposition 2.9) and the approximating AR process has smoothly time-varying coefficients (c.f. Theorem 2.11).

In practice, one may be interested in testing various hypotheses on the AR approximation such as whether some approximation coefficients are zero or whether the approximation coefficients are invariant with respect to time. The second main purpose of the paper is to propose a high-dimensional $\mathcal{L}^2$ test and an associated multiplier bootstrap procedure for the inference of the AR approximation coefficients of locally stationary time series. For the sake of brevity we concentrate on the test of stability of the approximation coefficients with respect to time for locally stationary time series (c.f. (3.1)). It is easy to see that similar methodologies can be developed for other problems of statistical inference such as tests for parametric assumptions on the approximation coefficients. Our test is shown to be adaptive to the strength of the time series dependence as well as the smoothness of the underlying data generating mechanism; see Propositions 3.6 and 3.7 and Algorithm 1 for more details. The theoretical investigation of the test critically depends on a result on Gaussian approximations to quadratic forms of high-dimensional locally stationary time series developed in the current paper (c.f. Theorem 3.4). In particular, uniform Gaussian approximations over high-dimensional convex sets [5, 17] as well as m-dependent approximations to quadratic forms of non-stationary time series are important techniques used in the proofs.

Interestingly, the test of stability for the AR approximation coefficients is asymptotically equivalent to testing correlation stationarity in the case of locally stationary time series; see Theorem 3.1 for more details. Here correlation stationarity means that the correlation structure of the time series does not change over time. As a result, our stability test can also be viewed as an adaptive test for correlation stationarity. In the statistics literature, there is a recent surge of interest in testing *covariance* stationarity of a time series using techniques from the spectral domain. See, for instance, [11, 15, 22, 23]. But it seems that the tests for correlation stationarity have not been discussed in the literature. Observe that the time-varying marginal variance has to be estimated and removed from the time series in order to apply the aforementioned tests to checking correlation stationarity. However, it is unknown whether the errors introduced in such estimation would influence the finite sample and asymptotic behaviour of the tests. Furthermore, estimating the marginal variance usually involves the difficult choice of a smoothing parameter. One major advantage of our test when used as a test of correlation stationarity is that it is totally free from the marginal variance as the latter quantity is absorbed into the errors of the AR approximation and hence is independent of the AR approximation coefficients.

Historically, the Wiener-Kolmogorov prediction theory was motivated by the optimal forecasting problem of stationary processes. Analogously, the AR approximation theory established in this paper is directly applicable to the problem of optimal short-term linear forecasting of non-stationary time series. For locally stationary time series, thanks to the AR approximation theory, the optimal short-term forecasting problem boils down to that of efficiently estimating the smoothly-varying AR approximation coefficient functions at the right boundary. We propose a nonparametric sieve regression method to estimate the latter coefficient functions and the associated MSE of forecast. Contrary to most non-stationary time series forecasting methods in the literature where only data near the end of the sequence are used to estimate the parameters of the forecast, the nonparametric sieve regression is global in the sense that it utilizes all available time series observations to determine the optimal forecast coefficients and hence is expected to be more efficient. Furthermore, by controlling the number of basis functions used in the regression, we demonstrate that the sieve method is adaptive in the sense that the estimation accuracy achieves global minimax rate for nonparametric function estimation in the sense of [28] under some mild conditions; see Theorem 4.3 for more details. In the statistics literature, there have been some scattered works discussing non-stationary time series prediction from some different angles. See for instance [9, 12, 18, 19, 26], among others. With the aid of the AR approximation, we are able to establish a unified globally-optimal short-term forecasting theory for a wide class of locally stationary time series asymptotically.

The rest of the paper is organized as follows. In Section 2, we introduce the AR approximation results for both general non-stationary time series and locally stationary time series. In Section 3, we test the stability of the AR approximation using $\mathcal{L}^2$ statistics of the estimated AR coefficient functions for locally stationary time series. A multiplier bootstrapping procedure is proposed and theoretically verified for practical implementation. In Section 4, we provide one important application of our AR approximation theory in optimal forecasting of locally stationary time series. In Section 5, we use extensive Monte Carlo simulations to verify the accuracy and power of our proposed methodologies. In Section 6, we conduct analysis on a financial real data set using our proposed methods. Technical proofs are deferred to the supplementary material [14].

Convention. Throughout the paper, we will consistently use the following notations. For a matrix Y or vector $\boldsymbol{y}$, we use Y^* and $\boldsymbol{y}^*$ to stand for their transposes. For a sequence of random variables $\{x_n\}$ and real values $\{a_n\}$, we use the notation $x_n = O_{\mathbb{P}}(a_n)$ to state that x_n/a_n is stochastically bounded. Similarly, we use the notation $x_n = o_{\mathbb{P}}(a_n)$ to say that x_n/a_n converges to 0 in probability. In this paper, unless otherwise specified, for a sequence of random variables $\{x_{i,n}\}$, we use the notation $x_{i,n} = O_{\mathbb{P}}(a_n)$ to state that $x_{i,n}/a_n$ is stochastically bounded uniformly in the index i. We will always use C as a genetic positive constant independent of n whose value may change from line to line.

2. Auto-Regressive Approximations to Non-stationary Time Series .

In this section, we establish a general AR approximation theory for a non-stationary time series $\{x_{i,n}\}$ under mild assumptions related to its covariance structure. Specifically, in Section 2.1, we study the general non-stationary time series. In Section 2.2, we investigate the special case of locally stationary time series where the covariance structure is assumed to be smoothly time-varying. Before proceeding to our main results, we pause to introduce two mild assumptions. Till the end of the paper, unless otherwise specified, we omit the subscript n and simply write $x_i \equiv x_{i,n}$.

First, in order to avoid erratic behavior of the AR approximation, the smallest eigenvalue of the time series covaraince matrix should be bounded away from zero. For stationary time series, this is equivalent to the uniform positiveness of the spectral density function which is

widely used in the literature. Further note that the latter assumption is mild and frequently used in the statistics literature of covariance and precision matrix estimation; see, for instance, [4, 7, 34] and the references therein. In this paper we shall call this *uniformly-positive-definite-in-covariance (UPDC)* condition and formally summarize it as follows.

ASSUMPTION 2.1 (UPDC). *For all $n \in \mathbb{N}$, we assume that there exists a universal constant $\kappa > 0$ such that*

$$\lambda_n(\mathrm{Cov}(x_1, \cdots, x_n)) \geq \kappa, \tag{2.1}$$

where $\lambda_n(\cdot)$ is the smallest eigenvalue of the given matrix and $\mathrm{Cov}(\cdot)$ is the covariance matrix of the given vector.

As discussed earlier, the UPDC is a mild assumption and is widely used in the literature. Moreover, for the locally stationary time series, we will provide a necessary and sufficient condition (c.f. Proposition 2.9) for practical checking.

Second, we impose the following assumption to control the temporal dependence of the non-stationary time series.

ASSUMPTION 2.2. *For $1 \leq k, l \leq n$, we assume that there exists some constant $\tau > 1$ such that*

$$\max_{k,l} |\mathrm{Cov}(x_k, x_l)| \leq C|k-l|^{-\tau}, \tag{2.2}$$

where $C > 0$ is some universal constant independent of n.

Assumption 2.2 states that the covariance structure of $\{x_i\}$ decays polynomially fast and it can be easily satisfied for many non-stationary time series; see Example 2.7 for a demonstration.

REMARK 2.3. In (2.2), we assume a polynomial decay rate. We can easily obtain analogous results to those established in this paper when the covariance decays exponentially fast; i.e.,

$$\sup_{k,l} |\mathrm{Cov}(x_k, x_l)| \leq Ca^{|k-l|}, \ 0 < a < 1. \tag{2.3}$$

For the sake of brevity, we focus on reporting our main results under the polynomial decay Assumption 2.2. From time to time, we will briefly mention the results under the exponential decay assumption (2.3) without providing extra details.

2.1. *AR approximation for general non-stationary time series.* In this subsection, we establish the approximation theory for general non-stationary time series $\{x_i\}$ satisfying Assumptions 2.1 and 2.2. Denote by $b \equiv b(n)$ a generic value which specifies the order of the AR approximation. In what follows, we investigate the accuracy of an AR(b) approximation to $\{x_i\}$ and provide the error rates using such an approximation. Observe that for theoretical and practical purposes b is typically required to be much smaller than n in order to achieve a parsimonious approximating model. For $i > b$, the best linear prediction (in terms of the mean squared prediction error) $\widehat{x}_i$ of x_i which utilizes all its predecessors $x_1, \cdots, x_{i-1}$, is denoted as

$$\widehat{x}_i = \phi_{i0} + \sum_{j=1}^{i-1} \phi_{ij} x_{i-j}, \ i = b+1, \cdots, n,$$

where $\{\phi_{ij}\}$ are the prediction coefficients. Denote $\epsilon_i := x_i - \widehat{x}_i$. It is well-known that $\{\epsilon_i\}_{i=1}^n$ is a time-varying white noise process, i.e.,

$$\mathbb{E}\epsilon_i = 0\ , \mathrm{Cov}(\epsilon_i, \epsilon_j) = \mathbf{1}(i=j)\sigma_i^2.$$

Armed with the above notation, we write

$$x_i = \phi_{i0} + \sum_{j=1}^{i-1} \phi_{ij} x_{i-j} + \epsilon_i,\ i = b+1, \cdots, n. \tag{2.4}$$

To provide an AR approximation of order b, where b may be much smaller than n, we need to examine the theoretical properties of the coefficients ϕ_{ij}. We summarize the results in Theorem 2.4.

THEOREM 2.4. *Suppose Assumptions 2.1 and 2.2 hold for $\{x_i\}$. For τ in (2.2), there exists some constant $C > 0$, when n is sufficiently large, we have that*

$$\max_i |\phi_{ij}| \le C \left(\frac{\log j + 1}{j} \right)^{\tau - 1},\ \textit{for all } j \ge 1. \tag{2.5}$$

Moreover, analogously to (2.4), denote by $\{\phi_{ij}^b\}$ the best linear forecast coefficients of x_i based on $x_{i-1}, \cdots, x_{i-b}$, i.e.,

$$x_i = \phi_{i0}^b + \sum_{j=1}^{b} \phi_{ij}^b x_{i-j} + \epsilon_i^b,\ i > b. \tag{2.6}$$

Then we have that

$$\begin{aligned} \max_{i>b} \max_{1 \le j \le b} |\phi_{ij} - \phi_{ij}^b| &\le C (\log b)^{\tau-1} b^{-(\tau-3)}, \\ \max_{i>b} |\phi_{i0} - \phi_{i0}^b| &\le C (\log b)^{\tau-1} b^{-(\tau-3.5)}. \end{aligned} \tag{2.7}$$

One one hand, Theorem 2.4 is general and only needs mild assumptions on the covariance structure of $\{x_i\}$. On the other hand, all the error bounds in Theorem 2.4 are adaptive to the decay rate of the temporal dependence and the order of the AR approximation. Particularly, we only need $\tau > 1$ (c.f. (2.5)) to assure the existence of a meaningful AR approximation with a diverging order much smaller than i.

Based on Theorem 2.4, we establish the AR approximation theory for the time series $\{x_i\}$ in Theorem 2.5. Particularly, we show that, under the short-range dependence and UPDC assumptions, any non-stationary time series can be efficiently approximated by an AR process of slowly diverging order (c.f. (2.10)). And the order of a consistent AR approximation is adaptive to the temporal decay rate of the time series dependence.

Denote the process $\{x_i^*\}$ by

$$x_i^* = \begin{cases} x_i, & i \le b; \\ \phi_{i0} + \sum_{j=1}^b \phi_{ij} x_{i-j}^* + \epsilon_i, & i > b. \end{cases} \tag{2.8}$$

Since $\{\epsilon_i\}$ is a time-varying white noise process, by construction, we have that $\{x_i^*\}_{i \ge 1}$ is an AR(b) process.

THEOREM 2.5. *Suppose the assumptions of Theorem 2.4 hold. Then we have that*

$$x_i = \phi_{i0} + \sum_{j=1}^{\min\{b, i-1\}} \phi_{ij} x_{i-j} + \epsilon_i + O_{\mathbb{P}}((\log b)^{\tau-1} b^{-(\tau-1.5)}). \tag{2.9}$$

As a result, we have

$$\max_{1\le i\le n}|x_i - x_i^*| = O_{\mathbb{P}}((\log b)^{\tau-1}b^{-(\tau-1.5)}). \tag{2.10}$$

Note that to guarantee a consistent AR approximation, we only require $\tau > 1.5$. This is a very mild assumption.

REMARK 2.6. In this paper, our discussions are carried out under Assumption 2.2. Our results can be easily extended to the case when the temporal dependence is of exponential decay, i.e., (2.3) holds true. In this case, (2.5) can be updated to

$$|\phi_{ij}| \le C\max\{n^{-2}, a^{j/2}\},\ j > 1,\ C > 0 \text{ is some constant},$$

and the magnitude of the error bounds in equations (2.7), (2.9) and (2.10) can all be changed to $\max\{b^{3/2}/n^2, n^{-1}, a^{b/2}\}$.

Before concluding this subsection, we provide an example of a general class of non-stationary time series using its physical representation [33, 36].

EXAMPLE 2.7. *For general non-stationary time series $\{x_i\}$, we assume that it has the following form*

$$x_i = G_i(\mathcal{F}_i),\ i = 1, 2, \cdots, n, \tag{2.11}$$

where $\mathcal{F}_i := (\cdots, \eta_{i-1}, \eta_i)$ and $\eta_i, i \in \mathbb{Z}$ are i.i.d. random variables and the sequence of functions $G_i : \{1, 2, \cdots, n\} \times \mathbb{R}^\infty \to \mathbb{R}$ are measurable such that for all $1 \le i_0 \le n$, $G_{i_0}(\mathcal{F}_i)$ is a properly defined random variable. The above representation is very general since any non-stationary time series can be represented in the form of (2.11) via the Rosenblatt transform [25].

Under the setting (2.11), the temporal dependence can be quantified using the physical dependence measure [33, 35, 37]. For any random variable X and some constant $q > 0$, denote $\|X\|_q = (\mathbb{E}|X|^q)^{1/q}$. Let $\{\eta_i'\}$ be an i.i.d. copy of $\{\eta_i\}$. For $j \ge 0$, we define the physical dependence measure of $\{x_i\}$ by

$$\delta(j, q) := \sup_i ||G_i(\mathcal{F}_0) - G_i(\mathcal{F}_{0,j})||_q, \tag{2.12}$$

where $\mathcal{F}_{0,j} := (\mathcal{F}_{-j-1}, \eta'_{-j}, \eta_{-j+1}, \cdots, \eta_0)$.

There are several advantages of using the device of physical dependence measure. On one hand, the concentration inequalities regarding the summation of $\{x_i\}$ can be fully expressed in terms of the physical dependence measure; see Lemma 6 of [35] or Lemma D.8 of [14] for more details. On the other hand, the covariance structure, i.e., Assumption 2.2, can be represented in terms of $\delta(j, q)$. In fact, it can be seen from Lemma D.7 of [14] that Assumption 2.2 will be satisfied if we assume

$$\delta(j, 2) \le Cj^{-\tau}. \tag{2.13}$$

Finally, we mention that many commonly used time series can be represented using (2.11). For example, the non-stationary linear process can be written as

$$x_i = \sum_{j=0}^{\infty} a_{ij} z_{i-j},$$

where $\{z_i\}$ are i.i.d. random variables with constant variance. In this case, it is easy to see that (2.13) is satisfied if $\sup_i |a_{ij}| \le Cj^{-\tau}$. For more examples in the form of (2.11), we refer the readers to [33] and [13, Section 2.1].

2.2. *AR approximation for locally stationary time series.* From the discussion of Section 2.1, we have seen that every UPDC and short-range dependent non-stationary time series can be well approximated by an AR process with diverging order (c.f. (2.8) and Theorem 2.5). However, from an estimation viewpoint, since we assume that only one time series is observed, the Yule-Walker's equations by which the AR coefficients (c.f. (2.4) or (2.6)) are governed, are clearly underdetermined linear systems. Driven by this challenge, we need to put some constraints for this ill-posed estimation problem. In the current paper, we consider the smoothness assumption and focus our study on an important subclass of non-stationary time series, the locally stationary time series [8, 11, 12, 30, 36]. This class of non-stationary time series is characterized by the fact that the underlying data generating mechanism evolves *smoothly* over time.

In this subsection, we will establish the AR approximation theory for the locally stationary time series under certain smoothness assumption. This will be a base for our later discussion regarding the inference and applications. We emphasize that thanks to the smoothness assumption, we will be able to prove a necessary and sufficient condition (c.f. Proposition 2.9) for checking the UPDC condition in Assumption 2.1.

We start with introducing the locally stationary time series. Compared to the general non-stationary time series discussed in Section 2.1, we need to impose smoothness assumption on the covariance structure.

DEFINITION 2.8 (Locally stationary time series). *The non-stationary time series $\{x_i\}$ is a locally stationary time series (in covariance) if there exists a function $\gamma(t,k) : [0,1] \times \mathbb{N} \to \mathbb{R}$ such that*

$$\mathrm{Cov}(x_i, x_j) = \gamma(t_i, |i-j|) + O\left(\frac{|i-j|+1}{n}\right), \; t_i = \frac{i}{n}. \tag{2.14}$$

Moreover, we assume that γ is Lipschitz continuous in t and for any fixed $t \in [0,1]$, $\gamma(t,\cdot)$ is the autocovariance function of a stationary process.

Our Definition 2.8 only imposes smoothness assumptions on its covariance structure. It covers the locally stationary time series used in the literature [8, 11, 12, 30, 36]. For more detail, we refer the readers to Example 2.13 below.

We first provide a necessary and sufficient condition for checking UPDC condition for locally stationary time series. For stationary time series, Herglotz's theorem asserts that UPDC holds if the spectral density function is bounded from below by a constant; see [3, Section 4.3] for more detail. Our next proposition extends such results to locally stationary time series with short-range dependence.

PROPOSITION 2.9. *If $\{x_i\}$ is locally stationary time series satisfying Assumption 2.2 and Definition 2.8, and there exists some constant $\kappa > 0$ such that $f(t,\omega) \geq \kappa$ for all t and ω, where*

$$f(t,\omega) = \sum_{j=-\infty}^{\infty} \gamma(t,j) e^{-\mathrm{i}j\omega}, \; \mathrm{i} = \sqrt{-1}, \tag{2.15}$$

then $\{x_i\}$ satisfies UPDC in Assumption 2.1. Conversely, if $\{x_i\}$ satisfies Assumptions 2.1 and 2.2 and Definition 2.8, then there exists some constant $\kappa > 0$, such that $f(t,\omega) \geq \kappa$ for all t and ω.

Note that $f(t,w)$ is the local spectral density function. Proposition 2.9 implies that the verification of UPDC reduces to showing that the local spectral density function is uniformly

bounded from below by a constant, which can be easily checked for many non-stationary processes. We refer the readers to Example 2.13 below for a demonstration.

Next, we establish the AR approximation theory for the locally stationary time series. As mentioned earlier, we will impose some smoothness assumptions such that the AR coefficients in (2.6) can be efficiently calculated. Till the end of the paper, unless otherwise specified, we use the following Assumption 2.10, which states that the mean and covariance functions of x_i are d-times continuously differentiable, for some positive integer d.

ASSUMPTION 2.10. *For some given integer $d > 0$, we assume that there exists a smooth function $\mu(\cdot) \in C^d([0,1])$, where $C^d([0,1])$ is the function space on $[0,1]$ of continuous functions that have continuous first d derivatives, such that*

$$\mathbb{E}x_i = \mu(t_i),\ t_i = \frac{i}{n}.$$

Moreover, we assume that $\gamma(t,j) \in C^d([0,1])$ for any $j \geq 0$.

Armed with Assumption 2.10, we proceed to state the AR approximation theory for locally stationary time series (c.f. Theorem 2.11). We first prepare some notations. Denote $\boldsymbol{\phi}(t) = (\phi_1(t), \cdots, \phi_b(t))^* \in \mathbb{R}^b$ such that

$$\boldsymbol{\phi}(t) = \Gamma(t)^{-1}\boldsymbol{\gamma}(t), \tag{2.16}$$

where $\Gamma(t) \in \mathbb{R}^{b\times b}$ and $\boldsymbol{\gamma}(t) \in \mathbb{R}^b$ are defined as

$$\Gamma_{ij}(t) = \gamma(t, |i-j|),\ \boldsymbol{\gamma}_i(t) = \gamma(t,i).$$

As we will see in the proof of Theorem 2.11, it is always valid to calculate $\Gamma(t)^{-1}$ under the UPDC assumption. With the above notations, we further define $\phi_0(t)$ as

$$\phi_0(t) = \mu(t) - \sum_{j=1}^{b} \phi_j(t)\mu(t). \tag{2.17}$$

Analogous to (2.8), denote

$$x_i^{**} = \begin{cases} x_i, & i \leq b; \\ \phi_0(\frac{i}{n}) + \sum_{j=1}^{b} \phi_j(\frac{i}{n}) x_{i-j}^{**} + \epsilon_i, & i > b. \end{cases}$$

THEOREM 2.11. *Consider the locally stationary time series from Definition 2.8. Suppose Assumptions 2.1, 2.2 and 2.10 hold true. Then we have that*

$$\phi_j(t) \in C^d([0,1]),\ 0 \leq j \leq b.$$

Furthermore, there exists some constant $C > 0$, such that

$$\max_{1\leq j\leq b} \max_{i>b} \left|\phi_{ij} - \phi_j(\frac{i}{n})\right| \leq C\left((\log b)^{\tau-1} b^{-(\tau-3)} + \frac{b^2}{n}\right). \tag{2.18}$$

Moreover, we have that

$$\max_{i>b} \left|\phi_{i0} - \phi_0(\frac{i}{n})\right| \leq C\left((\log b)^{\tau-1} b^{-(\tau-3.5)} + \frac{b^{2.5}}{n}\right). \tag{2.19}$$

Finally, we have

$$x_i = \phi_0(\frac{i}{n}) + \sum_{j=1}^{b} \phi_j(\frac{i}{n}) x_{i-j} + \epsilon_i + O_{\mathbb{P}}\left((\log b)^{\tau-1} b^{-(\tau-3.5)} + \frac{b^{2.5}}{n}\right), \tag{2.20}$$

and

$$\max_{1\le i\le n}|x_i - x_i^{**}| = O_{\mathbb{P}}\left((\log b)^{\tau-1}b^{-(\tau-4)} + \frac{b^3}{n}\right). \tag{2.21}$$

Theorem 2.11 states that for locally stationary time series from Definition 2.8 satisfying Assumptions 2.1 (or equivalently the assumptions of Proposition 2.9) and 2.10, for all $0 \le j \le b$ and the AR coefficients in (2.4), we can show that there exist smooth functions $\phi_j(\cdot)$'s (c.f. (2.16)) which only depend on the covariance function $\gamma(\cdot,\cdot)$, such that $\phi_j(i/n)$ approximates ϕ_{ij} well when $i > b$. Consequently, we can replace the AR coefficients using the function $\phi_j(t)$ with ϕ_{ij} in the approximation equation (2.4).

REMARK 2.12. As we can see from Theorem 2.11, the approximation error for the locally stationary time series comprises two parts: the truncation error, i.e., using an AR(b) approximation instead of all the predecessors, and the error caused by using smooth AR coefficients. The first part of the error depends on b and the decay rate (c.f. (2.2)) of the covariance structure, whereas the second part depends on b and the normalization that $t_i = i/n$ with the Lipschitz regularity. Moreover, to assure the AR approximation in (2.20) holds with $o_{\mathbb{P}}(1)$ error for each data point, we need to require that $\tau > 3.5$ and an elementary calculation shows that the choice of b should satisfy that

$$\frac{b}{\log b} = O(n^{\frac{1}{\tau-1}}). \tag{2.22}$$

Finally, we can get similar results as in Remark 2.6 when the covariance structure is of exponential decay.

Before concluding this subsection, we provide an example to illustrate the generality of Definition 2.8 .

EXAMPLE 2.13. *We consider the physical form of locally stationary time series as in Example 2.7. Corresponding to (2.11), in [36, 37], the authors define the locally stationary time series* $\{x_i\}$ *as follows*

$$x_i = G(\frac{i}{n}, \mathcal{F}_i), \tag{2.23}$$

where $G : [0,1] \times \mathbb{R}^\infty \to \mathbb{R}$ *is a measurable function such that* $\xi_i(t) := G(t, \mathcal{F}_i)$ *is a properly defined random variable for all* $t \in [0,1]$. *In (2.23), by allowing the data generating mechanism* G *depending on the time index* t *in such a way that* $G(t, \mathcal{F}_i)$ *changes smoothly with respect to* t, *one has local stationarity in the sense that the subsequence* $\{x_i, ..., x_{i+j-1}\}$ *is approximately stationary if its length* j *is sufficiently small compared to* n. *Analogous to (2.12), we can define the physical dependence measure for (2.23) as follows*

$$\delta(j,q) := \sup_{t\in[0,1]} ||G(t,\mathcal{F}_0) - G(t,\mathcal{F}_{0,j})||_q. \tag{2.24}$$

Moreover, the following assumptions are needed to ensure local stationarity.

ASSUMPTION 2.14. $G(\cdot,\cdot)$ *defined in (2.23) satisfies the property of stochastic Lipschitz continuity, i.e., for some* $q > 2$ *and* $C > 0$,

$$||G(t_1,\mathcal{F}_i) - G(t_2,\mathcal{F}_i)||_q \le C|t_1 - t_2|, \tag{2.25}$$

where $t_1, t_2 \in [0,1]$. *Furthermore,*

$$\sup_{t\in[0,1]} \max_{1\le i\le n} ||G(t,\mathcal{F}_i)||_q < \infty. \tag{2.26}$$

We remark that the physical representation (2.23) with Assumption 2.14 satisfies Definition 2.8. Especially, for each fixed $t \in [0,1]$, we denote the covariance function of the locally stationary time series $\{x_i\}$ as

$$\gamma(t,j) = \mathrm{Cov}(G(t,\mathcal{F}_0), G(t,\mathcal{F}_j)). \tag{2.27}$$

Note that the assumptions (2.25) and (2.26) ensure that $\gamma(t,j)$ is Lipschiz continuous in t. Moreover, for each fixed t, $\gamma(t,\cdot)$ is the autocovariance function of $\{G(t,\cdot)\}$, which is a stationary process.

The physical representation form (2.23) includes many commonly used locally stationary time series models. For example, let $\{z_i\}$ be zero-mean i.i.d. random variables with variance σ^2. We also assume $a_j(\cdot), j = 0, 1, \cdots$ be $C^d([0,1])$ functions such that

$$G(t,\mathcal{F}_i) = \sum_{k=0}^{\infty} a_k(t) z_{i-k}. \tag{2.28}$$

(2.28) is a locally stationary linear process. It is easy to see that Assumptions 2.2, 2.10 and 2.14 will be satisfied if

$$\sup_{t\in[0,1]} |a_j(t)| \le C j^{-\tau},\ j \ge 1;\ \sum_{j=0}^{\infty} \sup_{t\in[0,1]} |a_j'(t)| < \infty,$$

and

$$\sup_{t\in[0,1]} |a_j^{(d)}(t)| \le C j^{-\tau},\ j \ge 1.$$

Further, we note that the local spectral density function of $G(t,\mathcal{F}_i)$ can be written as

$$f(t,w) = \sigma^2 |\psi(t, e^{-\mathrm{i}j\omega})|^2,$$

where $\psi(\cdot,\cdot)$ is defined such that $G(t,\mathcal{F}_i) = \psi(t,B)z_i$ with B being the backshift operator. By Proposition 2.9, the UPDC is satisfied if $\sigma^2|\psi(t,e^{-\mathrm{i}j\omega})|^2 \ge \kappa$ for all t and ω, where $\kappa > 0$ is some universal constant. For more examples of locally stationary time series in the form of (2.23), we refer the readers to [33] and [13, Section 2.1].

Finally, we mention that in [11, 30], the locally stationary time series is defined as follows (see Definition 2.1 of [30]). $\{x_i\}$ is locally stationary time series if for each scaled time point $u \in [0,1]$, there exists a strictly stationary process $\{h_i(u)\}$ such that

$$|x_i - h_i(u)| \le \left(|t_i - u| + \frac{1}{n}\right) U_i(u),\ a.s, \tag{2.29}$$

where $U_i(u) \in \mathcal{L}^q([0,1])$ for some $q > 0$. By a similar argument, we can obtain the connection between this definition and Definition 2.8. We omit the details here.

3. A Test of Stability for AR Approximation. In this section, we study the inference for the AR(b) approximation for locally stationary time series using a high dimensional $\mathcal{L}^2$ test. For the ease of statement, till the end of the paper, we assume the locally stationary time series admits the physical representation as in (2.23). In view of (2.20), we can conduct tests for any hypothesized parametric form on $\phi_j(\cdot)$. For the sake of brevity, we concentrate on the test of stability of the approximation coefficients with respect to time for locally stationary time series (c.f. (3.1)).

3.1. *Problem setup and test statistics.* In this subsection, we formally state the hypothesis testing problems and propose our statistics based on some nonparametric sieve estimators of $\{\phi_j(\cdot)\}$.

Since $\phi_0(\cdot)$ is related to the trend of the time series and in some real applications the trend is removed before performing forecasting, we focus our discussion on the test the stability of $\phi_j(\cdot),\ j \geq 1$. For the testing of the trend, we refer the readers to Remark 3.2 for more detail. Formally, the null hypothesis we would like to test is

$$\widetilde{\mathbf{H}}_0 :\ \phi_j(\cdot) \text{ is a constant function on } [0,1],\ \text{for all } j \geq 1.$$

However, according to (2.5) and (2.18), when b diverges, for instance, in the form of (2.22), we have that $\sup_t |\phi_j(t)| = o(n^{-1/2})$ once τ is reasonably large. From an inferential viewpoint, we can ignore $\phi_j(\cdot)$ for $j > b$. Together with the approximation (2.20), it suffices for us to test

$$\mathbf{H}_0 :\ \phi_j(\cdot) \text{ is a constant function on } [0,1],\ j = 1, 2, \cdots, b. \tag{3.1}$$

Before providing the test statistic for $\mathbf{H}_0$, we shall first investigate the interesting insight that $\mathbf{H}_0$ is asymptotically equivalent to testing whether $\{x_i\}_{i=1}^n$ is correlation stationary, i.e., there exists some function ϱ such that

$$\mathbf{H}_0' :\ \mathrm{Corr}(x_i, x_j) = \varrho(|i-j|), \tag{3.2}$$

where $\mathrm{Corr}(x_i, x_j)$ stands for the correlation between x_i and x_j. We formalize the above statements in Theorem 3.1 below.

THEOREM 3.1. *Suppose Assumptions 2.1, 2.2, 2.10 and 2.14 hold true. For $j \leq b$, on one hand, when $\mathbf{H}_0'$ holds true, then $\phi_j(\frac{i}{n}) = \phi_j$ which is independent of time. On the other hand, when $\phi_j(\frac{i}{n}) = \phi_j, j = 1, 2, \cdots, b$, there exists some smooth function ϱ, such that*

$$\mathrm{Corr}(x_i, x_{i+j}) = \varrho_{|i-j|} + O\left(\frac{b}{n} + (\log b)^{\tau-1} b^{-(\tau-2)}\right).$$

In rest of this subsection, we propose test statistics for $\mathbf{H}_0$ in (3.1). We start with the estimation of the coefficient functions $\phi_j(\cdot), j = 0, 1, 2, \cdots, b$. Since $\phi_j(t) \in C^d([0,1])$, it is natural for us to approximate it via a finite and diverging term basis expansion. Specifically, by [6, Section 2.3], we have that

$$\phi_j(\frac{i}{n}) = \sum_{k=1}^{c} a_{jk}\alpha_k(\frac{i}{n}) + O(c^{-d}),\ 0 \leq j \leq b,\ i > b, \tag{3.3}$$

where $\{\alpha_k(t)\}$ are some pre-chosen basis functions on $[0,1]$ and c is the number of basis functions. For the ease of discussion, in the current paper, we assume that c is of the following form

$$c = O(n^{\mathfrak{a}}),\ 0 < \mathfrak{a} < 1. \tag{3.4}$$

Moreover, for the reader's convenience, in Section E of [14], we collect the commonly used basis functions.

In view of (3.3), we need to estimate the a_{jk}'s in order to get an estimation for $\phi_j(t)$. For $i > b$, by (2.20), write

$$x_i = \sum_{j=0}^{b}\sum_{k=1}^{c} a_{jk} z_{kj} + \epsilon_i + O_{\mathbb{P}}\left((\log b)^{\tau-1} b^{-(\tau-3.5)} + \frac{b^{2.5}}{n} + bc^{-d}\right), \tag{3.5}$$

where $z_{kj} \equiv z_{kj}(i/n) := \alpha_k(i/n)x_{i-j}$ for $j \geq 1$ and $z_{k0} = \alpha_k(i/n)$. By (3.5), we can estimate all the $a'_{jk}s$ using only one ordinary least squares (OLS) regression with diverging number of predictors. In particular, we write all $a_{jk},\ j = 0, 1, 2\cdots, b,\ k = 1, 2, \cdots, c$ as a vector $\boldsymbol{\beta} \in \mathbb{R}^{(b+1)c}$, then the OLS estimator for $\boldsymbol{\beta}$ can be written as $\widehat{\boldsymbol{\beta}} = (Y^*Y)^{-1}Y^*\boldsymbol{x}$, where $\boldsymbol{x} = (x_{b+1}, \cdots, x_n)^* \in \mathbb{R}^{n-b}$ and Y is the design matrix. After estimating $a'_{jk}s$, $\phi_j(i/n)$ is estimated using (3.3) as

$$\widehat{\phi}_j(\frac{i}{n}) = \widehat{\boldsymbol{\beta}}^* \mathbb{B}_j(\frac{i}{n}), \tag{3.6}$$

where $\mathbb{B}_j(i/n) := \mathbb{B}_{j,b}(i/n) \in \mathbb{R}^{(b+1)c}$ has $(b+1)$ blocks and the j-th block is $\mathbf{B}(\frac{i}{n}) = (\alpha_1(i/n), \cdots, \alpha_c(i/n))^* \in \mathbb{R}^c,\ j = 0, 1, 2, \cdots, b$, and zeros otherwise.

With the estimation (3.6), we proceed to provide the $\mathcal{L}^2$ test statistics. To test $\mathbf{H}_0$, we use the following statistic in terms of (3.6)

$$T = \sum_{j=1}^{b} \int_0^1 (\widehat{\phi}_j(t) - \overline{\widehat{\phi}}_j)^2 dt,\ \overline{\widehat{\phi}}_j = \int_0^1 \widehat{\phi}_j(t)dt. \tag{3.7}$$

In fact, when $\mathbf{H}_0$ holds true, heuristically, $\widehat{\phi}_j(t)$ should be close to $\overline{\widehat{\phi}}_j$ for all $t \in [0, 1]$. Consequently, we shall have a small value of T.

REMARK 3.2. We remark that in some cases especially regarding on forecasting, practitioners and researchers may be interested in testing whether all optimal forecast coefficient functions including the trend $\phi_j(\cdot)$, $j = 0, 1, \cdots, b$, do not change over time. That is equivalent to testing whether both the trend and the correlation structure of the time series stay constant over time. In this case, one will test

$$\mathbf{H}_{0,g}:\ \phi_j(\cdot) \text{ is a constant function on } [0, 1],\ j = 0, 1, \cdots, b. \tag{3.8}$$

Similar to (3.7), for the test of $\mathbf{H}_{0,g}$, we shall use

$$T_g = \sum_{j=0}^{b} \int_0^1 (\widehat{\phi}_j(t) - \overline{\widehat{\phi}}_j)^2 dt,\ \overline{\widehat{\phi}}_j = \int_0^1 \widehat{\phi}_j(t)dt. \tag{3.9}$$

3.2. *High dimensional Gaussian approximation and asymptotic normality.* In this subsection, we prove the asymptotic normality for the statistic T. The key ingredient is to establish the Gaussian approximation results for quadratic forms of high dimensional locally stationary time series.

We first show that the study of the statistic T reduces to the investigation of a weighted quadratic form of high dimensional locally stationary time series. We prepare some notations. Denote $\bar{B} = \int_0^1 \mathbf{B}(t)dt$ and $W = I - \bar{B}\bar{B}^*$. Let $\mathbf{W}$ be a $(b+1)c \times (b+1)c$ dimensional diagonal block matrix with diagonal block W and $\mathbf{I}_{bc}$ be a $(b+1)c \times (b+1)c$ dimensional diagonal matrix whose non-zero entries are ones and in the lower $bc \times bc$ major part. Recall $\boldsymbol{x}_i = (1, x_{i-1}, \cdots, x_{i-b})^*$ and set

$$p = (b+1)c. \tag{3.10}$$

Recall ϵ_i in (2.4). We denote the sequence of p-dimensional vectors $\boldsymbol{z}_i$ by

$$\boldsymbol{z}_i = \boldsymbol{h}_i \otimes \mathbf{B}(\frac{i}{n}) \in \mathbb{R}^p,\ \boldsymbol{h}_i = \boldsymbol{x}_i\epsilon_i, \tag{3.11}$$

where $\otimes$ is the Kronecker product. We point out that when $i > b$, it is easy to see that $\boldsymbol{h}_i$ is a locally stationary time series. For notational convenience, we denote

$$\boldsymbol{h}_i = \mathbf{U}(\frac{i}{n}, \mathcal{F}_i),\ i > b. \tag{3.12}$$

Recall (2.27). We also denote the $b \times b$ matrix $\Sigma^b(t) = (\Sigma^b_{ij}(t))$ such that

$$\Sigma^b_{ij}(t) = \gamma(t, |i-j|).$$

LEMMA 3.3. *Denote* $\mathbf{X} = \frac{1}{\sqrt{n}} \sum_{i=b+1}^{n} \boldsymbol{z}_i^*$, *and the* $p \times p$ *matrix* Γ *by* $\Gamma = \overline{\Sigma}^{-1} \mathbf{I}_{bc} \mathbf{W} \overline{\Sigma}^{-1}$, *where*

$$\overline{\Sigma} = \begin{pmatrix} \mathbf{I}_c & \mathbf{0} \\ \mathbf{0} & \Sigma \end{pmatrix},\ \Sigma = \int_0^1 \Sigma^b(t) \otimes (\mathbf{B}(t)\mathbf{B}^*(t)) dt. \tag{3.13}$$

Suppose Assumptions 2.1, 2.10, 2.14 and A.1 of [14] hold true. Moreover, we assume that the physical dependence measure $\delta(j, q), q > 2$, *in (2.24) satisfies*

$$\delta(j, q) \le C j^{-\tau},\ j \ge 1, \tag{3.14}$$

for some constant $C > 0$ *and* $\tau > 4.5 + \varpi$, *where* $\varpi > 0$ *is some fixed small constant. Then for* c *in the form of (3.4) and* b *such that* $C_1 n^{\varsigma_1/\tau} \le b \le C_2 n^{\varsigma_2/\tau}$, *where* $C_1, C_2 > 0$ *are some constants and* $0 < \varsigma_1 < \varsigma_2$ *are some small constants, when* n *is sufficiently large, we have that*

$$nT = \mathbf{X}^* \Gamma \mathbf{X} + o_{\mathbb{P}}(1). \tag{3.15}$$

Based on Lemma 3.3, for the purpose of statistical inference, it suffices to establish the distribution of $\mathbf{X}^*\Gamma\mathbf{X}$. Since both b and c are divergent, in light of (3.10), to this end, we shall establish a Gaussian approximation result for this quadratic form of high-dimensional locally stationary time series $\{\boldsymbol{z}_i\}$. Choose a sequence of centred Gaussian random vectors $\{\boldsymbol{v}_i\}_{i=b+1}^n$ which preserves the covariance structure of $\{\boldsymbol{h}_i\}_{i=b+1}^n$ and define $\boldsymbol{g}_i = \boldsymbol{v}_i \otimes \mathbf{B}(\frac{i}{n})$. Denote

$$\mathbf{Y} = \frac{1}{\sqrt{n}} \sum_{i=b+1}^{n} \boldsymbol{g}_i^*.$$

In Theorem 3.4 below, we establish the Gaussian approximation result by controlling the Kolmogorov distance

$$\mathcal{K}(\mathbf{X}, \mathbf{Y}) = \sup_{x \in \mathbb{R}} \left| \mathbb{P}\Big(\mathbf{X}^*\Gamma\mathbf{X} \le x\Big) - \mathbb{P}\Big(\mathbf{Y}^*\Gamma\mathbf{Y} \le x\Big) \right|. \tag{3.16}$$

Denote

$$\xi_c := \sup_{1 \le i \le c} \sup_{t \in [0,1]} \Big| \alpha_i(t) \Big|. \tag{3.17}$$

THEOREM 3.4. *Under the assumptions of Lemma 3.3, there exist some constant* $C > 0$ *and a small constant* $\delta > 0$, *such that*

$$\mathcal{K}(\mathbf{X}, \mathbf{Y}) \le C\Big(\log n \frac{\xi_c}{M_z} + p^{\frac{7}{4}} n^{-1/2} M_z^3 M^2 + M^{\frac{-q\tau+1}{2q+1}} \xi_c^{(q+1)/(2q+1)} p^{\frac{q+1}{2q+1}} n^{\frac{\delta q}{2q+1}} + p^{1/2} \xi_c^{1/2} \Big((p\xi_c M^{-\tau+1} + p\xi_c^2 n M_z^{-(q-2)}) \Big)^{1/2} + n^{-\delta} \Big),$$

where $M_z, M \to \infty$ *when* $n \to \infty$ *and we recall* p *in (3.10). Here* $q > 2$ *is from (2.25).*

REMARK 3.5. We remark that ξ_c in (3.17) can be well controlled for commonly used basis functions. For instance, $\xi_c = O(1)$ for the Fourier basis and the normalized orthogonal polynomials and $\xi_c = O(\sqrt{c})$ for orthogonal wavelet; see Section E of [14] for more detail. Moreover, it is easy to see that the approximate rate in Theorem 3.4 can be easily made to be of order $o(1)$. In particular, when $\tau > 0$ is large enough, $\xi_c = O(1)$ and $q > 2$ is large enough, we can allow $p = n^{2/7-\delta_1}$, where $\delta_1 > 0$ is some sufficiently small constant. This is claimed to be the best known dimension setting in [17].

Armed with Theorem 3.4, the asymptotic normality of nT can be readily obtained as in Proposition 3.6 below. Denote the long-run covariance matrix for $\{\boldsymbol{h}_i\}$ as

$$\Omega(t) = \sum_{j=-\infty}^{\infty} \mathrm{Cov}\Big(\mathbf{U}(t, \mathcal{F}_0), \mathbf{U}(t, \mathcal{F}_j)\Big), \tag{3.18}$$

and the aggregated covariance matrix as $\Omega = \int_0^1 \Omega(t) \otimes \Big(\mathbf{B}(t)\mathbf{B}^*(t)\Big)\mathrm{d}t$. Ω can be regarded as the integrated long-run covariance matrix of $\{\boldsymbol{z}_i\}$. For $k \in \mathbb{N}$ and Γ in (3.15), we define

$$f_k = \Big(\mathrm{Tr}[\Omega^{1/2}\Gamma\Omega^{1/2}]^k\Big)^{1/k}, \tag{3.19}$$

where $\mathrm{Tr}(\cdot)$ is the trace of the given matrix.

PROPOSITION 3.6. *Under the assumptions of Lemma 3.3, when* $\mathbf{H}_0$ *in (3.1) holds true, we have*

$$\frac{nT - f_1}{f_2} \Rightarrow \mathcal{N}(0, 2).$$

Next, we discuss the power of the test under a certain class of local alternatives. For a given α, set

$$\mathbf{H}_a : \sum_{j=1}^{\infty} \int_0^1 \Big(\phi_j(t) - \bar{\phi}_j\Big)^2 dt > C_\alpha \frac{\sqrt{bc}}{n},$$

where $\bar{\phi}_j = \int_0^1 \phi_j(t)dt$ and $C_\alpha \equiv C_\alpha(n) \to \infty$ as $n \to \infty$. For instance, we can choose $C_\alpha > n^\kappa \mathcal{Z}_{1-\alpha},\ \kappa > 0$, where $\mathcal{Z}_{1-\alpha}$ is the $(1-\alpha)\%$ quantile of the standard Gaussian distribution.

PROPOSITION 3.7. *Under the assumptions of Lemma 3.3, when* $\mathbf{H}_a$ *holds true, we have*

$$\frac{nT - f_1 - n\sum_{j=1}^{\infty}\int_0^1 \Big(\phi_j(t) - \bar{\phi}_j\Big)^2 dt}{f_2} \Rightarrow \mathcal{N}(0, 2).$$

Consequently, under $\mathbf{H}_a$*, the power of our test will asymptotically be 1, i.e.,*

$$\mathbb{P}\Big(\Big|\frac{nT - f_1}{f_2}\Big| \geq \sqrt{2}\mathcal{Z}_{1-\alpha}\Big) \to 1,\ n \to \infty.$$

REMARK 3.8. In this remark, we discuss how to deal with T_g in (3.9). By a discussion similar to Lemma 3.3, T_g can also be expressed as a quadratic form

$$nT_g = \mathbf{X}^*\Gamma_g\mathbf{X} + o_{\mathbb{P}}(1),\ \Gamma_g = \overline{\Sigma}^{-1}\mathbf{W}\overline{\Sigma}^{-1},$$

where we recall (3.13). Consequently, the only difference lies in the deterministic weight matrix. By Theorem 3.4, we can prove similar results to nT_g as in Propositions 3.6 and 3.7. We omit further details.

3.3. *Multiplier bootstrap procedure.* In this subsection, we propose a practical procedure for testing $\mathbf{H}_0$ based on multiplier bootstrap. Similar discussion applies to $\mathbf{H}_{0,g}$.

One on hand, it is difficult to directly use Proposition 3.6 to carry out the stability test since the quantities f_1 and f_2 rely on Ω which is hard to be estimated in general. On the other hand, the high-dimensional Gaussian quadratic form $\mathbf{Y}^*\Gamma\mathbf{Y}$ converges at a slow rate. To overcome these difficulties, we extend the strategy of [35] and use a high-dimensional mulitplier bootstrap statistic to mimic the distributions of nT. Note that (3.15) can be explicitly written as

$$nT = \Big(\frac{1}{\sqrt{n}}\sum_{i=b+1}^{n} \boldsymbol{z}_i^*\Big)\Gamma\Big(\frac{1}{\sqrt{n}}\sum_{i=b+1}^{n} \boldsymbol{z}_i\Big) + o_{\mathbb{P}}(1). \tag{3.20}$$

Recall (3.11). For some positive integer m, denote

$$\Phi = \frac{1}{\sqrt{n-m-b+1}\sqrt{m}}\sum_{i=b+1}^{n-m}\Big[\Big(\sum_{j=i}^{i+m}\boldsymbol{h}_j\Big)\otimes\Big(\mathbf{B}(\frac{i}{n})\Big)\Big]R_i, \tag{3.21}$$

where $R_i, i = b+1, \cdots, n-m$, are i.i.d. standard Gaussian random variables. Φ is an important statistic since the covariance of Φ is close to Ω when conditional on the data; see (B.39) of [14] for a more precise statement.

Since $\{\boldsymbol{h}_i\}$ is based on $\{\epsilon_i\}$ which cannot be observed directly, we shall instead use the residuals

$$\widehat{\epsilon}_i^b := x_i - \widehat{\phi}_0(\frac{i}{n}) - \sum_{j=1}^{b}\widehat{\phi}_j(\frac{i}{n})x_{i-j}. \tag{3.22}$$

Denote $\{\widehat{\boldsymbol{h}}_i\}$ similarly as in (3.11) by replacing $\{\epsilon_i\}$ with $\{\widehat{\epsilon}_i^b\}$. Accordingly, we denote $\widehat{\Phi}$ as in (3.21) using $\{\widehat{\boldsymbol{h}}_i\}$. With the above notations, we denote the bootstrap quadratic form as

$$\widehat{\mathcal{T}} := \widehat{\Phi}^*\widehat{\Gamma}\widehat{\Phi}, \tag{3.23}$$

where $\widehat{\Gamma} := \widehat{\Sigma}^{-1}\mathbf{I}_{bc}\mathbf{W}\widehat{\Sigma}^{-1}$ with $\widehat{\Sigma} = \frac{1}{n}Y^*Y$. Note that $\widehat{\Gamma}$ is a consistent estimator of Γ.

In Theorem 3.9 below, we prove that $\widehat{\mathcal{T}}$ can mimic the distribution of nT asymptotically. Denote

$$\zeta_c := \sup_t \|\mathbf{B}(t)\|. \tag{3.24}$$

THEOREM 3.9. *Suppose the assumptions of Lemma 3.3 hold and*

$$\sqrt{b}\zeta_c^2 c^{-1/2}\Big(\sqrt{\frac{m}{n}} + \frac{1}{m}\Big) = o(1). \tag{3.25}$$

Furthermore, we assume that Assumption 2.14 holds with $q > 4$. When $\mathbf{H}_0$ holds true, there exists some set $\mathcal{A}_n$ such that $\mathbb{P}(\mathcal{A}_n) = 1 - o(1)$ and under the event $\mathcal{A}_n$, we have that conditional on the data $\{x_i\}_{i=b+1}^n$,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\widehat{\mathcal{T}} - f_1}{\sqrt{2}f_2} \le x\right) - \mathbb{P}\left(\Psi \le x\right)\right| = o(1),$$

where $\Psi \sim \mathcal{N}(0,1)$ is a standard normal random variable.

REMARK 3.10. First, ζ_c can be well controlled by the commonly used sieve basis functions. For example, we have $\zeta_c = O(\sqrt{c})$ for the Fourier basis and orthogonal wavelet, and $\zeta_c = O(c)$ for Legendre polynomial; see Section E of [14] for more detail. Second, in the worst scenario when $\zeta_c = O(\sqrt{c})$, the assumption (3.25) reads as

$$\sqrt{p}\Big(\sqrt{\frac{m}{n}} + \frac{1}{m}\Big) = o(1).$$

Hence, in the optimal case when $m = O(n^{1/3})$, we are allowed to choose $p \ll n^{2/3}$. In this regime, Theorem 3.4 still holds true. Third, for the detailed construction of $\mathcal{A}_n$, we refer the reader to (B.40) of [14]. Finally, a theoretical discussion of the accuracy of the bootstrap can be found in Section A of [14] and the choices of the hyperparameters m, c, d, are discussed in Section C of [14].

Based on Theorem 3.9, we can use the following Algorithm 1 for practical implementation to calculate the p-value.

Algorithm 1 Multiplier Bootstrap

Inputs: tuning parameters b, c and m chosen by the data-drive procedure demonstrated in Section C of [14], time series $\{x_i\}$, and sieve basis functions.
Step one: Compute $\widehat{\Sigma}^{-1}$ using $n(Y^*Y)^{-1}$ and the residuals $\{\widehat{\epsilon}_i^b\}_{i=b+1}^n$ according to (3.22).
Step two: Generate B (say 1000) i.i.d. copies of $\{\widehat{\Phi}^{(k)}\}_{k=1}^B$. Compute $\widehat{\mathcal{T}}_k, k = 1, 2, \cdots, B$, correspondingly as in (3.23).
Step three: Let $\widehat{\mathcal{T}}_{(1)} \le \widehat{\mathcal{T}}_{(2)} \le \cdots \le \widehat{\mathcal{T}}_{(B)}$ be the order statistics of $\widehat{\mathcal{T}}_k, k = 1, 2, \cdots, B$. Reject $\mathbf{H}_0$ at the level α if $nT > \widehat{\mathcal{T}}_{(\lfloor B(1-\alpha)\rfloor)}$, where $\lfloor x \rfloor$ denotes the largest integer smaller or equal to x. Let $B^* = \max\{r : \widehat{\mathcal{T}}_r \le nT\}$.
Output: p-value of the test can be computed as $1 - \frac{B^*}{B}$.

4. Applications to globally optimal forecasting. In this section, independent of Section 3, we discuss an application of our AR approximation theory in global forecasting for locally stationary time series. In particular, we consider short-term forecasting of locally stationary time series by estimating the smooth forecasting coefficients using sieve expansion as in (3.6). For definiteness, we focus on the physical form as in (2.23). We first introduce the notion of *asymptotically optimal predictor*.

DEFINITION 4.1. *A linear predictor $\widetilde{z}$ of a random variable z based on $x_1, \cdots, x_n$, is called asymptotically optimal if*

$$\mathbb{E}(z - \widetilde{z})^2 \le \sigma_n^2 + o(1/n), \tag{4.1}$$

where σ_n^2 is the mean squared error (MSE) of the best linear predictor of z based on $x_1, \cdots, x_n$.

The rationale for such definition is that, in practice, the MSE of forecast can only be estimated with a smallest possible error of $O(1/n)$ when time series length is n. It is well-known that the parametric rate for estimating the coefficients of a time series model is $O(n^{-1/2})$. When one uses the estimated coefficients to forecast the future, the corresponding influence on the MSE of forecast is $O(1/n)$ (at best). Therefore, if a linear predictor achieves an MSE

of forecast within $o(1/n)$ range of the optimal one, it is practically indistinguishable from the optimal predictor asymptotically.

In what follows, we shall focus on the discussion of one-step ahead prediction. The general case can be handled similarly. In order to make the forecasting feasible, we assume that the smooth data generating mechanism extends to time $n+1$. That is, we assume $x_{n+1} = G((n+1)/n, \mathcal{F}_{n+1})$. Naturally, we propose the following estimate for $\hat{x}_{n+1}$, the best linear predictor of x_{n+1} based on its predecessors $x_1, \cdots, x_b$,

$$\widehat{x}^b_{n+1} = \phi_0(1) + \sum_{j=1}^{b} \phi_j(1) x_{n+1-j},\ n > b. \tag{4.2}$$

The next theorem shows that $\widehat{x}^b_{n+1}$ is an asymptotic optimal predictor satisfying (4.1) in Definition 4.1.

THEOREM 4.2. *Suppose Assumptions 2.1, 2.2, 2.10 and 2.14 hold true. Then there exists some constant $C > 0$, such that for sufficiently large n,*

$$\mathbb{E}(x_{n+1} - \widehat{x}^b_{n+1})^2 \le \mathbb{E}(x_{n+1} - \widehat{x}_{n+1})^2 + O\left(\left((\log b)^{\tau-1} b^{-(\tau-3.5)} + \frac{b^{2.5}}{n}\right)^2\right). \tag{4.3}$$

Using the choice of b in (2.22), when n is sufficiently large, we find that the error on the right-hand side of (4.3) reads as follows asymptotically

$$n^{-2+\frac{5}{\tau-1}}.$$

In this sense, Theorem 4.2 states that the estimator (4.2) is an asymptotic optimal one-step ahead forecast if $\frac{5}{\tau-1} < 1$, i.e., $\tau > 6$.

To obtain an estimation for the predictor, in light of (3.6), based on (4.2), we shall estimate $\widehat{x}^b_{n+1}$, or equivalently, forecast x_{n+1} using

$$\widehat{\mathrm{x}}^b_{n+1} = \widehat{\phi}_0(1) + \sum_{j=1}^{b} \widehat{\phi}_j(1) x_{n+1-j}. \tag{4.4}$$

Next, we discuss the estimation of the MSE of the forecast, i.e., the variance of $\{\epsilon_{n+1}\}$. Denote the series of estimated forecast error $\{\widehat{\epsilon}^b_i\}$ by $\widehat{\epsilon}^b_i := x_i - \widehat{\phi}_0(i/n) - \sum_{j=1}^{b} \widehat{\phi}_j(i/n) x_{i-j}$ and the variance of $\{\epsilon_i\}$ by $\{\sigma_i^2\}$. According to [13, Lemma 3.11], we find that there exists a smooth function $\varphi(\cdot) \in C^d([0,1])$ such that for some constant $C > 0$,

$$\sup_{i>b} |\sigma_i^2 - \varphi(\frac{i}{n})| \le C\left((\log b)^{\tau-1} b^{-(\tau-3.5)} + \frac{b^{2.5}}{n}\right). \tag{4.5}$$

Therefore, similar to the estimation of the smooth AR coefficients, we can again use the basis expansion to estimate the smooth function $\varphi(\cdot)$. Similar to (3.3), we write

$$\varphi(\frac{i}{n}) = \sum_{k=1}^{c} \mathfrak{b}_k \alpha_k(\frac{i}{n}) + O(c^{-d}).$$

Furthermore, by equation (3.14) of [13], we write

$$(\widehat{\epsilon}^b_i)^2 = \sum_{k=1}^{c} \mathfrak{b}_k \alpha_k(\frac{i}{n}) + \nu_i + O_{\mathbb{P}}\Big(b(\zeta_c \frac{\log n}{\sqrt{n}} + n^{-d\mathfrak{a}})\Big),\ i > b,$$

where $\{\nu_i\}$ is a centered sequence of locally stationary time series satisfying Assumptions 2.1, 2.2, 2.10 and 2.14. Consequently, we can use an OLS with $(\widehat{\epsilon}_i^b)^2$ being the response and $\alpha_k(\frac{i}{n})$, $k=1,\cdots,c$, being the explanatory variables to estimate $\mathfrak{b}_k$'s, which are denoted as $\widehat{\mathfrak{b}}_k, k=1,2,\cdots,c$. Finally, we estimate

$$\widehat{\varphi}(i/n)=\sum_{k=1}^{c}\widehat{\mathfrak{b}}_k\alpha_k(i/n).$$

We are now state the asymptotic behaviour of the MSE of (4.4) in Theorem 4.3 below. Recall (3.4) and (3.24).

THEOREM 4.3. *Suppose Assumptions 2.1, 2.2, 2.10, 2.14 and A.1 of [14] hold true. For some constants $C_1, C_2>0$ and small constants $0<\varsigma_1<\varsigma_2<\frac{2}{5}\tau$ such that $C_1 n^{\frac{\varsigma_1}{\tau}}\le b\le C_2 n^{\frac{\varsigma_2}{\tau}}$. We have*

$$\big|\sigma_{n+1}^2-\widehat{\varphi}(1)\big|=O_{\mathbb{P}}\Big(b(\zeta_c\frac{\log n}{\sqrt{n}}+n^{-d\mathfrak{a}})+(\log b)^{\tau-1}b^{-(\tau-3.5)}+\frac{b^{2.5}}{n}\Big).$$

5. Simulation studies. In this section, we perform extensive Monte Carlo simulations to study the finite-sample accuracy and power of the multiplier bootstrap Algorithm 1 and compare it with some existing methods in the literature.

5.1. *Simulation setup.* We consider four different types of non-stationary time series models: two linear time series models, a two-regime model, a Markov switching model and a bilinear model.

1. Linear AR model: Consider the following time-varying AR(2) model
$$x_i=\sum_{j=1}^{2}a_j(\frac{i}{n})x_{i-j}+\epsilon_i,\ \epsilon_i=\Big(0.4+0.4\Big|\sin(2\pi\frac{i}{n})\Big|\Big)\eta_i,$$
where $\eta_i, i=1,2,\cdots,n$, are i.i.d. random variables whose distributions will be specified when we finish introducing the models. It is elementary to see that when $a_j(\frac{i}{n}), j=1,2$, are constants, the prediction is stable.
2. Linear MA model: Consider the following time-varying MA(2) model
$$x_i=\sum_{j=1}^{2}a_j(\frac{i}{n})\epsilon_{i-j}+\epsilon_i,\ \epsilon_i=\Big(0.4+0.4\Big|\sin(2\pi\frac{i}{n})\Big|\Big)\eta_i.$$
3. Two-regime model: Consider the following self-exciting threshold auto-regressive (SETAR) model [16, 29]
$$x_i=\begin{cases}a_1(\frac{i}{n})x_{i-1}+\epsilon_i,\ x_{i-1}\ge 0,\\ a_2(\frac{i}{n})x_{i-1}+\epsilon_i,\ x_{i-1}<0.\end{cases}\quad \epsilon_i=\Big(0.4+0.4\Big|\sin(2\pi\frac{i}{n})\Big|\Big)\eta_i.$$
It is easy to check that the SETAR model is stable if $a_j(\frac{i}{n}),\ j=1,2$, are constants and bounded by one.
4. Markov two-regime switching model: Consider the following Markov switching AR(1) model
$$x_i=\begin{cases}a_1(\frac{i}{n})x_{i-1}+\epsilon_i,\ s_i=0,\\ a_2(\frac{i}{n})x_{i-1}+\epsilon_i,\ s_i=1.\end{cases}\quad \epsilon_i=\Big(0.4+0.4\Big|\sin(2\pi\frac{i}{n})\Big|\Big)\eta_i,$$

where the unobserved state variable s_i is a discrete Markov chain taking values 0 and 1, with transition probabilities $p_{00} = \frac{2}{3},\ p_{01} = \frac{1}{3},\ p_{10} = p_{11} = \frac{1}{2}$. It is easy to check that the above model is stable if the functions $a_j(\cdot), j = 1, 2$, are constants and bounded by one [24]. In the simulations, the initial state is chosen to be 1.

5. Simple bilinear model: Consider the first order bilinear model

$$x_i = \Big(a_1(\frac{i}{n})\epsilon_{i-1} + a_2(\frac{i}{n})\Big)x_{i-1} + \epsilon_i,\ \epsilon_i = \Big(0.4 + 0.4\Big|\sin(2\pi\frac{i}{n})\Big|\Big)\eta_i.$$

It is known from [16] that when the functions $a_j(\cdot), j = 1, 2$, are constants and bounded by one, x_i has an ARMA representation and hence stable.

In the simulations below, we record our results based on 1,000 repetitions and for Algorithm 1, we choose $B = 1,000$. For the choices of random variables $\eta_i, i = 1, 2, \cdots$, we set η_i to be student-t distribution with degree of 5, i.e., $t(5)$, for models 1-2 and standard normal random variables for models 3-5.

5.2. *Accuracy and power of the stability test.* In this subsection, we study the performance of the proposed test (3.1). First, we study the finite sample accuracy of our test the null hypothesis that

$$a_1(\frac{i}{n}) = a_2(\frac{i}{n}) \equiv 0.4. \tag{5.1}$$

Observe that the simulated time series are not covariance stationary as the marginal variances change smoothly over time. We choose the values of b, c and m according to the methods described in Section C of [14]. It can be seen from Table 1 that our bootstrap testing procedure is reasonably accurate for all three types of sieve basis functions even for a smaller sample size $n = 256$.

Second, we study the power of the tests and report the results in Table 2 when the underlying time series is not correlation stationary, i.e., the AR coefficients are time-varying. Specifically, we use

$$a_1(\frac{i}{n}) \equiv 0.4,\ a_2(\frac{i}{n}) = 0.2 + \delta\sin(2\pi\frac{i}{n}),\ \delta > 0 \text{ is some constant,} \tag{5.2}$$

for the models 1-5 in Section 5.1. It can be seen that the simulated powers are reasonably good even for smaller values of δ and sample size, and the results will be improved when δ and sample size increase. Additionally, the power performances of the three types of sieve basis functions are similar in general.

5.3. *Comparison with tests for covariance stationarity.* In this subsection, we compare our method with some existing works on the tests of covariance stationarity: the $\mathcal{L}^2$ distance method in [11], the discrete Fourier transform method in [15] and the Haar wavelet periodogram method in [22]. The first method is easy to implement; for the second method, we use the codes from the author's website (see `https://www.stat.tamu.edu/~suhasini/test_papers/DFT_covariance_lagl.R`); and for the third method, we employ the R package `locits`, which is contributed by the author. For the purpose of comparison of accuracy, besides the five models considered in Section 5.1, we also consider the following two strictly stationary time series.

6. Linear time series: stationary ARMA(1,1) process. We consider the following process

$$x_i - 0.5x_{i-1} = \epsilon_i + 0.5\epsilon_{i-1},$$

where ϵ_i are i.i.d. $\mathcal{N}(0, 1)$ random variables.

	$\alpha = 0.1$					$\alpha = 0.05$				
Basis/Model	1	2	3	4	5	1	2	3	4	5
	n=256									
Fourier	0.132	0.11	0.12	0.13	0.11	0.067	0.07	0.06	0.04	0.06
Legendre	0.091	0.136	0.13	0.12	0.13	0.06	0.059	0.041	0.07	0.07
Daubechies-9	0.132	0.12	0.11	0.133	0.132	0.063	0.067	0.059	0.068	0.065
	n=512									
Fourier	0.09	0.13	0.11	0.13	0.127	0.05	0.06	0.067	0.068	0.069
Legendre	0.09	0.094	0.092	0.12	0.118	0.04	0.058	0.07	0.043	0.057
Daubechies-9	0.091	0.11	0.098	0.11	0.118	0.048	0.052	0.054	0.053	0.054

TABLE 1

Simulated type I errors using the setup (5.1). The models are listed in Section 5.1 and the basis functions can be found in Section E of [14]. The results are reported based on 1,000 simulations. We can see that our multiplier bootstrap procedure is accurate for both $\alpha = 0.1$ and $\alpha = 0.05$.

	$\delta = 0.2/0.5$					$\delta = 0.35/0.7$				
Basis/Model	1	2	3	4	5	1	2	3	4	5
	n=256									
Fourier	0.84	0.86	0.84	0.837	0.94	0.97	0.97	0.96	0.99	0.98
Legendre	0.8	0.806	0.81	0.84	0.83	0.97	0.968	0.95	0.97	0.91
Daubechies-9	0.81	0.81	0.86	0.81	0.81	0.97	0.96	0.983	0.98	0.98
	n=512									
Fourier	0.91	0.9	0.96	0.9	0.93	0.96	0.97	0.973	0.98	0.97
Legendre	0.9	0.91	0.92	0.893	0.91	0.94	0.95	0.98	0.97	0.96
Daubechies-9	0.87	0.88	0.93	0.91	0.91	0.96	0.99	0.97	0.97	0.96

TABLE 2

Simulated power under the setup (5.2) using nominal level 0.1. For models 1-2, we consider the cases $\delta = 0.2$ and $\delta = 0.35$, whereas for models 3-5, we use $\delta = 0.5$ and $\delta = 0.7$. The results are based on 1,000 simulations. We can see that our Algorithm 1 can achieve a high power even for not so large δ and n.

7. Nonlinear time series: stationary SETAR. We consider the following model

$$x_i = \begin{cases} 0.4x_{i-1} + \epsilon_i, & x_{i-1} \geq 0, \\ 0.5x_{i-1} + \epsilon_i, & x_{i-1} < 0, \end{cases}$$

where ϵ_i are i.i.d. $\mathcal{N}(0,1)$ random variables.

Furthermore, for the comparison of power, we consider the following two non-stationary time series whose errors have constant variances.

$6^{\#}$. Non-stationary linear time series. We consider the following process

$$x_i = \delta \sin(4\pi \frac{i}{n}) x_{i-1} + \epsilon_i,$$

where $\epsilon_i, i = 1, 2, \cdots, n$, are i.i.d. standard normal random variables.

$7^{\#}$. Non-stationary nonlinear time series. We consider the following process

$$x_i = \begin{cases} \delta \sin(4\pi \frac{i}{n}) x_{i-1} + \epsilon_i, & 1 \le i \le 0.75n, \\ 0.4 x_{i-1} + \epsilon_i, & 0.75n < i \le n \text{ and } x_{i-1} \ge 0, \\ 0.3 x_{i-1} + \epsilon_i, & 0.75n < i \le n \text{ and } x_{i-1} < 0, \end{cases}$$

where $\epsilon_i, i = 1, 2, \cdots, n$, are i.i.d. standard normal random variables.

In the simulations below, we report the type I error rates under the nominal levels 0.05 and 0.1 for the above seven models in Table 3, where for models 1-5 we use the setup (5.1). Our simulation results are based on 1,000 repetitions, where $\mathcal{L}^2$ refers to the $\mathcal{L}^2$ distance method, DFT 1-3 refer to the approaches using the imagery part, real part, both imagery and real parts of the discrete Fourier transform method, respectively, HWT is the Haar wavelet periodogram method and MB is our multiplier bootstrap method Algorithm 1 using orthogonal wavelets constructed by (E.2) of [14] with Daubechies-9 wavelet.

Since HWT needs the length to be a power of two, we set the length of time series to be 256 and 512. For the parameters of the $\mathcal{L}^2$ test, we use $M = 8, N = 32$ for $n = 256$, and $M = 8, N = 64$ for $n = 512$. For the DFT, we choose the lag to be 0 as suggested by the authors in [15]. Since the mean of model 5 is non-zero, we test its first order difference for the methods mentioned above. Moreover, we report the power of the above tests under certain alternatives in Table 4 for models $6^{\#} - 7^{\#}$ and models 1-5 under the setup (5.2).

	$\alpha = 0.1$						$\alpha = 0.05$					
Model	$\mathcal{L}^2$	DFT1	DFT2	DFT3	HWT	MB	$\mathcal{L}^2$	DFT1	DFT2	DFT3	HWT	MB
	n=256											
1	0.08	0.148	0.057	0.13	0.18	0.132	0.024	0.067	0.017	0.063	0.083	0.063
2	0.081	0.097	0.068	0.12	0.085	0.12	0.038	0.04	0.07	0.057	0.028	0.067
3	0.171	0.183	0.04	0.137	0.227	0.11	0.087	0.103	0.011	0.033	0.093	0.059
4	0.2	0.163	0.05	0.12	0.176	0.133	0.077	0.087	0.013	0.034	0.113	0.068
5	0.46	0.293	0.077	0.19	0.153	0.132	0.29	0.21	0.03	0.14	0.12	0.065
6	0.11	0.105	0.096	0.09	0.087	0.088	0.047	0.053	0.053	0.039	0.052	0.057
7	0.051	0.097	0.08	0.092	0.085	0.127	0.018	0.04	0.06	0.047	0.038	0.061
	n=512											
1	0.087	0.127	0.03	0.13	0.237	0.091	0.023	0.1	0.02	0.043	0.137	0.048
2	0.051	0.096	0.085	0.093	0.075	0.11	0.026	0.036	0.067	0.044	0.033	0.052
3	0.26	0.16	0.04	0.117	0.243	0.098	0.127	0.1	0.007	0.037	0.14	0.054
4	0.287	0.167	0.027	0.09	0.247	0.11	0.177	0.103	0.013	0.073	0.163	0.053
5	0.64	0.303	0.087	0.283	0.35	0.118	0.413	0.26	0.063	0.167	0.23	0.054
6	0.11	0.093	0.084	0.088	0.088	0.092	0.035	0.046	0.047	0.048	0.053	0.048
7	0.051	0.087	0.113	0.083	0.093	0.092	0.013	0.037	0.047	0.043	0.04	0.051

TABLE 3

Comparison of accuracy for models 1-7 using different methods. We report the results based on 1,000 simulations. It can be concluded that our multiplier bootstrap is accurate regardless of whether the time series is linear or stationary.

We first discuss the results for models 6-7 since they are not only correlation stationary but also covariance stationary. It can be seen from Table 3 that all the methods including our

MB achieve a reasonable level of accuracy for the linear model 6. However, for the nonlinear model 7, we conclude from Table 3 that the $\mathcal{L}^2$ method loses its accuracy due to the fact that the latter test is designed only for linear models. Regarding the power in Table 4, for model $6^{\#}$, when the sample size and δ are smaller, only our MB method is powerful. When $n = 256$ and δ increases, the $\mathcal{L}^2$ test starts to become powerful. Further, when both the sample size and δ increase, the HWT method becomes powerful. Similar discussion holds for model $7^{\#}$. Therefore, we conclude that, when the marginal variance of the time series stays constant, even though other methods in the literature may be accurate for the purpose of testing for correlation stationarity, our MB method is generally more powerful when the sample size is moderate and/or the departure from stationary is small.

	$\delta = 0.2/0.5$						$\delta = 0.35/0.7$					
Model	$\mathcal{L}^2$	DFT1	DFT2	DFT3	HWT	MB	$\mathcal{L}^2$	DFT1	DFT2	DFT3	HWT	MB
	$n = 256$											
1	0.263	0.14	0.03	0.07	0.3	0.81	0.503	0.113	0.053	0.089	0.4	0.97
2	0.183	0.497	0.08	0.092	0.585	0.81	0.68	0.14	0.06	0.047	0.38	0.96
3	0.44	0.153	0.04	0.16	0.393	0.86	0.7	0.14	0.05	0.09	0.64	0.983
4	0.603	0.16	0.04	0.203	0.44	0.81	0.86	0.2	0.07	0.12	0.647	0.98
5	0.92	0.243	0.143	0.24	0.57	0.81	0.997	0.347	0.193	0.397	0.797	0.98
$6^{\#}$	0.697	0.12	0.093	0.11	0.327	0.86	0.923	0.16	0.15	0.15	0.563	0.94
$7^{\#}$	0.463	0.137	0.107	0.133	0.273	0.85	0.81	0.193	0.203	0.223	0.483	0.96
	$n = 512$											
1	0.477	0.173	0.04	0.08	0.52	0.87	0.857	0.137	0.03	0.1	0.75	0.96
2	0.51	0.297	0.082	0.092	0.385	0.88	0.918	0.24	0.06	0.047	0.838	0.99
3	0.657	0.24	0.05	0.083	0.61	0.93	0.96	0.17	0.24	0.113	0.95	0.97
4	0.84	0.23	0.043	0.143	0.773	0.91	0.987	0.293	0.053	0.19	0.97	0.97
5	0.963	0.297	0.127	0.263	0.87	0.91	0.983	0.523	0.24	0.478	0.994	0.96
$6^{\#}$	0.847	0.147	0.087	0.103	0.67	0.88	0.95	0.13	0.09	0.133	0.963	0.95
$7^{\#}$	0.69	0.14	0.13	0.217	0.383	0.91	0.953	0.3	0.313	0.383	0.823	0.943

TABLE 4

Comparison of power at nominal level 0.1 using different methods. We report the results based on 1,000 simulations. It can be concluded that our multiplier bootstrap can have high power regardless of whether the time series is linear or stationary.

Next, we study models 1-5 from Section 5.1. None of these models is covariance stationary. For the type I error rates, we use the setting (5.1) where all the models are correlation stationary. For the power, we use the setup (5.2). We find that DFT-3 is accurate for models 1-4 but with low power across all the models. Moreover, the $\mathcal{L}^2$ test seems to have a high power for models 3-5. But this is at the cost of blown-up type I error rates. This inaccuracy increases when the sample size becomes larger. For the HWT method, even though its power becomes larger when the sample size and δ increase, it also loses its accuracy. Finally, for all the models 1-5, our MB method obtains both high accuracy and power. In summary, most of the existing tests for covariance stationarity are not suitable for the purpose of testing for correlation stationarity. From our simulation studies, our multiplier bootstrap method Algorithm 1 performs well for the latter purpose.

6. An empirical illustration. In this section, we illustrate the usefulness of our results by analyzing a financial data set. We study the stock return data of the Nigerian Breweries (NB) Plc. This stock is traded in Nigerian Stock Exchange (NSE). Regarding on market returns, the brewery industry in Nigerian has done pretty well in outperforming Brazil, Russia, India, and China (BRIC) and emerging markets by a wide margin over the past ten years. Nigerian Breweries Plc is the largest brewing company in Nigeria, which mainly serves the Nigerian market and also exports to other parts of West Africa. The data can be found on the website of morningstar (see http://performance.morningstar.com/stock/performance-return.action?p=price_history_page&t=NIBR&region=nga&culture=en-US). We are interested in understanding the volatility of the NB stock. We shall study the absolute value of the daily log-return of the stock for the latter purpose.

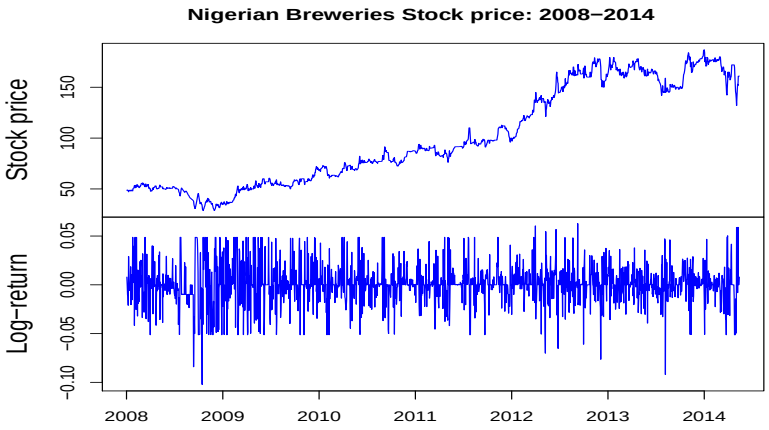

Fig 1: Nigerian Breweries stock return from 2008 to 2014. The upper panel is the original stock price and the lower panel is the log-return.

We perform our analysis on the time period 2008-2014 (Figure 1). This time series contains the data of the 2008 global financial crisis and its post period. As said in the report from the Heritage Foundation [27], "the economy is experiencing the slowest recovery in 70 years" and even till 2014, the economy does not fully recover.

Then we apply the methodologies described in Sections 3 for the absolute values of log-return time series. It is clear that we need to fit a mean curve for this model. Then we test the stability of the AR approximation as described in Section 3 using Algorithm 1. For the sieve basis functions, we use the orthogonal wavelets constructed by (E.2) of [14] with Daubechies-9 wavelet. We choose the parameters b, c and m based on the discussion of Section C of [14] which yields $b = 7$, $J_n = 5$ (i.e., $c = 32$) and $m = 18$. We apply the bootstrap procedure described in Algorithm 1 and find that the p-value is 0.0825. We hence conclude that the prediction is likely to be unstable during this time period.

Next, we use the time series 2008-2014 as the training dataset to study the rolling forecast performance over the first month of 2015 using (4.4) . We employ the data-driven approach from Section C of [14] to choose $b = 5$ and $J_n = 3$. The averaged MSE is 0.189. We point out that this leads to a 20.9% improvement compared to simply fitting a stationary ARMA model using all the time series from 2008 to 2014 where the MSE is 0.239, and leads to a 24.7% improvement compared to simply using the data of the previous trading day where the MSE is 0.251.

Finally, we study the absolute value of the stock return from 2012 to 2014. We apply our bootstrap procedure Algorithm 1 to test correlation stationarity of the time series. We select $b = 6$, $J_n = 4$ (i.e., $c = 16$) and $m = 12$ for this sub-series and find that the p-value is 0.599. We hence conclude that the prediction is stable during this time period. Therefore, we fit a stationary ARMA model to this sub-series and do the prediction. This yields an MSE of 0.192, which is close to the MSE when we use the whole non-stationary time series. The result from this sub-series shows an interesting trade-off between forecasting using a shorter and stationary time series and a longer but non-stationary series. The forecast model of the shorter stationary period can be estimated at a faster rate but at the expense of a smaller sample size. The opposite happens to the longer non-stationary period. Note that 2012-2014 is nearly half as long as 2008-2014 and hence the length of the shorter stationary period is substantial compared to that of the long period. In this case we see that the forecasting accuracy using the short period is comparable to that of the longer period. In many applications where the data generating mechanism is constantly changing, the stable period is typically very short and a nonparametric model for the longer period is preferred. Finally, we emphasize that the correlation stationarity test is an important tool to decide a period of prediction stability.

SUPPLEMENTARY MATERIAL

Supplement to " Auto-Regressive Approximations to Non-stationary Time Series, with Inference and Applications"

The supplementary material [14] contains further explanation, technical proofs and auxiliary lemmas for the main results of the paper.

REFERENCES

[1] BAXTER, G. (1962). An Asymptotic Result for the Finite Predictor. *Mathematica Scandinavica* **10** 137-144.

[2] BAXTER, G. (1963). A norm inequality for a "finite-section" Wiener-Hopf equation. *Illinois J. Math.* **7** 97–103.

[3] BROCKWELL, P. and DAVIS, R. (1987). *Time series: Theory and Methods*. Springer-Verlag.

[4] CAI, T. T., LIU, W. and ZHOU, H. H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Ann. Statist.* **44** 455–488.

[5] CHEN, L. and FANG, X. (2011). Multivariate Normal Approximation by Stein's Method: The Concentration Inequality Approach. *arXiv preprint arXiv:1111.4073*.

[6] CHEN, X. (2007). *Large Sample Sieve Estimation of Semi-nonparametric Models*. Chapter 76 in Handbook of Econometrics, Vol. 6B, James J. Heckman and Edward E. Leamer.

[7] CHEN, X., XU, M. and WU, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist.* **41** 2994–3021.

[8] DAHLHAUS, R. (2012). Locally stationary processes. In *Handbook of statistics*, **30** 351–413. Elsevier.

[9] DAS, S. and POLITIS, D. N. (2021). Predictive Inference for Locally Stationary Time Series With an Application to Climate Data. *J. Amer. Statist. Assoc.* **116** 919–934.

[10] DEMKO, S., MOSS, W. and SMITH, P. (1984). Decay Rates for Inverses of Band Matrices. *Math. Comput.* **43** 491-499.

[11] DETTE, H., PREUBB, P. and VETTER, M. (2011). A Measure of Stationarity in Locally Stationary Processes With Applications to Testing. *J. Am. Stat. Assoc.* **106** 1113-1124.

[12] DETTE, H. and WU, W. (2020). Prediction in locally stationary time series. *Journal of Business & Economic Statistics* 1–12.

[13] DING, X. and ZHOU, Z. (2020). Estimation and inference for precision matrices of nonstationary time series. *Ann. Statist.* **48** 2455–2477.

[14] DING, X. and ZHOU, Z. (2021). Supplement to "Auto-Regressive Approximations to Non-stationary Time Series, with Inference and Applications".

[15] DWIVEDI, Y. and RAO, S. S. (2011). A test for second–order stationarity of a time series based on the discrete Fourier transform. *Journal of Time Series Analysis* **32** 68-91.

[16] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

[17] FANG, X. (2016). A Multivariate CLT for Bounded Decomposable Random Vectors with the Best Known Rate. *Journal of Theoretical Probability* **29** 1510–1523.
[18] FRYZLEWICZ, P., VAN BELLEGEM, S. and VON SACHS, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics* **55** 737–764.
[19] KLEY, T., PREUSS, P. and FRYZLEWICZ, P. (2019). Predictive, finite-sample model choice for time series under stationarity and non-stationarity. *Electron. J. Stat.* **13** 3710–3774.
[20] KOLMOGOROV, A. N. (1941). Stationary Sequences in Hilbert Space. *Bull. Moscow State U. Math.* **2** 1–40.
[21] KOLMOGOROV, A. N. (1941). Interpolation and extrapolation of stationary random sequences. *Izv. Akad. Nauk SSSR., Ser. Mat.* **5** 3–14.
[22] NASON, G. (2013). A test for second–order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *J.R. Statist. Soc. B* **75** 879-904.
[23] PAPARODITIS, E. (2010). Validating Stationarity Assumptions in Time Series Analysis by Rolling Local Periodograms. *Journal of the American Statistical Association* **105** 839-851.
[24] QUANDT, R. (1972). A new approach to estimating switching regressions. *J. Am. Stat. Assoc.* **67** 306-310.
[25] ROSENBLATT, M. (1952). Remarks on a Multivariate Transformation. *Ann. Math. Statist.* **23** 470–472.
[26] ROUEFF, F. and SANCHEZ-PEREZ, A. (2018). Prediction of weakly locally stationary processes by auto-regression. *Lat. Am. J. Probab. Math. Stat.* **15** 1215–1239.
[27] SHERK, J. (2014). *Not Looking for Work: Why Labor Force Participation Has Fallen During the Recovery*. https://www.heritage.org/jobs-and-labor/report/not-looking-work-why-labor-force-participation-has-fallen-during-the-recovery.
[28] STONE, C. J. (1982). Optimal Global Rates of Convergence for Nonparametric Regression. *Ann. Statist.* **10** 1040–1053.
[29] TONG, H. (2011). Threshold models in time series analysis - 30 years on. *Stat. Interface* **4** 107-118.
[30] VOGT, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40** 2601–2633.
[31] WIENER, N. (1949). *Extrapolation, interpolation and smoothing of stationary time series*. New York, Wiley.
[32] WOLD, H. (1954). *A study in the analysis of stationary time series*. Almqvist and Wiksell, Stockholm 2d ed, With an appendix by Peter Whittle.
[33] WU, W. (2005). Nonlinear system theory: Another look at dependence. *Proc Natl Acad Sci U S A.* **40** 14150-14151.
[34] YUAN, M. (2010). High Dimensional Inverse Covariance Matrix Estimation via Linear Programming. *J. Mach. Learn. Res.* **11** 2261–2286.
[35] ZHOU, Z. (2013). Heteroscedasticity and Autocorrelation Robust Structural Change Detection. *J. Am. Stat. Assoc.* **108** 726-740.
[36] ZHOU, Z. and WU, W. (2009). Local linear quantile estimation for non-stationary time series. *Ann. Stat.* **37** 2696-2729.
[37] ZHOU, Z. and WU, W. (2010). Simultaneous inference of linear models with time varying coefficents. *J.R. Statist. Soc. B* **72** 513-531.