# Conditional Independence Testing for High-Dimensional Non-Stationary Nonlinear Time Series

Michael Wieck-Sosa[1], Michel F. C. Haddad[2], and Aaditya Ramdas[1,3]

[1]Department of Statistics and Data Science, Carnegie Mellon University
[2]Department of Business Analytics and Applied Economics, Queen Mary University of London
[3]Machine Learning Department, Carnegie Mellon University

May 22, 2024

**Abstract**

We introduce a conditional independence testing framework for high-dimensional non-stationary time series based on time-varying nonlinear regression. Our framework requires just one realization of the process and does not require sample splitting. Recently, Shah and Peters [SP20] showed that conditional independence testing is fundamentally impossible without making further assumptions even in the idealized iid setting. Our test can be viewed as an extension of their generalized covariance measure to the time series setting, and we extend their arguments to show that it has uniformly asymptotic Type-I error control provided that the time-varying nonlinear regression estimators satisfy modest convergence rate requirements. In practice, our test can be used with any black-box time-varying regression estimator for non-stationary time series. We utilize a bootstrap procedure based on a distribution-uniform version of the strong Gaussian approximation for high-dimensional non-stationary time series from Mies and Steland [MS22]. We discuss how our framework can be used with simultaneous testing procedures to determine *when* conditional independence holds for *which* dimensions at particular leads and lags while controlling the familywise error rate. We use our testing framework to conduct a Granger causality analysis of non-stationary nonlinear epidemiological signals, in particular determining whether the time-varying viral load distribution of a city contains auxiliary information about future case growth rates of COVID-19. Our work also has implications for other problem domains, since it shows how to gain power by using groups of time series, such as sensor networks in the brain, or measurements from nearby field sites, using all of the data instead of ad hoc spatial averaging techniques commonly used in neuroscience and climate science.

## Contents

# 1 Introduction

Many fundamental concepts in statistics can be expressed in terms of conditional independence, such as sufficiency, ancillarity, and the Markov condition [Daw79]. Moreover, conditional independence plays a key role in causal inference [Pea09], causal discovery [SGS00], and graphical modeling [KF09]. See Shah and Peters [SP20], Zhang et al. [Zha+12], Neykov, Balakrishnan, and Wasserman [NBW21], Peters, Janzing, and Schölkopf [PJS17], and Runge et al. [Run+23] for more background.

Within the realm of time series analysis, conditional independence is closely related to Granger causality and transfer entropy [SP11; Con12; AM12; Cha82; QKC15; FM82]. Concepts from graphical modeling, which were originally developed for the iid setting, have been extended to the time series setting [Dah00; Eic12]. Notably, Basu and Rao [BR23] recently introduced a graphical modeling framework for non-stationary time series. Also, conditional independence tests for time series are used in constraint-based and hybrid causal discovery algorithms for time series [Run+19b]. A related problem to ours is (unconditional) independence testing for non-stationary time series [Liu+23; Bru22]. There is a vast literature on conditional independence testing for stationary time series. For example, there are tests based on characteristic functions [SW07] and copulas [BRT12]. Notably, the conditional mutual information-based test by Runge [Run18] was used in Runge et al. [Run+19a].

There has been a lot of recent work on constraint-based causal discovery algorithms for non-stationary time series. For a notable example, see Huang et al. [Hua+20] and the related discussion in Dong et al. [Don+23] in the context of finance. It is clear that the underlying conditional independence test upon which the causal discovery algorithm is based must be well-suited to the data. If the data exhibits temporal dependence or non-stationarity, this could violate the validity of the underlying conditional independence test and lead to erroneous conclusions about the causal structure. This motivates the development of conditional independence tests designed for non-stationary time series.

The literature on conditional independence tests for non-stationary time series is very limited. We discuss the two tests that we are aware of which can be used with non-stationary time series. First, Malinsky and Spirtes [MS19] focus on a type of non-stationarity in which the processes exhibit "stochastic trends". The consistency of their procedure requires assuming that the data were generating by a vector autoregressive linear model with iid Gaussian errors. Second, Flaxman, Neill, and Smola [FNS15] introduce a flexible framework based on pre-whitening for conditional independence testing that can in principle be used with non-stationary time series. However, the theoretical justifications for these tests are unclear in light of the recent hardness of conditional independence testing result from Shah and Peters [SP20]. The validity of all of these conditional independence tests would need to be checked on a case-by-case basis.

**Our contributions and paper outline.** We address these concerns and the potential shortcomings of the existing literature by developing a new conditional independence test for high-dimensional non-stationary nonlinear time series. The rest of the paper is structured as follows. In the rest of Section 1, we introduce the setting and notation which will be used throughout the paper and we discuss the hypothesis of conditional independence. In Section 2, we discuss our conditional independence testing framework. In Section 3, we use our testing framework to conduct a Granger causality analysis to understand epidemiological dynamics. In Section 4, we discuss the simulation results. In Section 5, we discuss promising avenues for future work.

We defer many of the theoretical discussions to the Appendix. In Section A, we conduct a literature review on distribution-uniform inference. In Section B, we discuss hypothesis testing and conditional independence testing for non-stationary time series. In Section C, we discuss stepdown and inheritance procedures for simultaneous testing. In Section D, we discuss the assumptions for our theoretical framework for non-stationary time series. In particular, so that we can control the temporal dependence and non-stationarity uniformly over distributions for the process. In Section E, we introduce a distribution-uniform extension of the theoretical framework for high-dimensional non-stationary time series from Mies and Steland [MS22]. Due to the non-asymptotic nature of the results in Mies and Steland [MS22], we show the results can be almost immediately extended to hold distribution-uniformly. These distribution-uniform results are critical for the theoretical results for our conditional independence test. In Section F, we prove that our test has uniformly asymptotic Type-I error control. In Section G, we provide auxiliary lemmas.

## 1.1 Setting and notation

We work in the practical setting in which we only observe one realization of a non-stationary time series. It is necessary to impose some kinds of restrictions on the dependence and non-stationarity so that inference is possible. We extend the framework for high-dimensional locally stationary nonlinear time series from Mies and Steland [MS22] so that the temporal dependence and non-stationarity can be controlled uniformly over a collection of distributions for the time series. We quantify temporal dependence using the functional dependence measure from Wu [Wu05].

Let us reflect on the purpose for introducing locally stationary time series by recalling Dahlhaus [Dah97]. In regular time series analysis, letting $n$ approach infinity corresponds to getting information about the future. However, if the process of interest is non-stationary, letting $n$ approach infinity does not give us any additional information about the process at earlier points in time. Dahlhaus [Dah97] introduced the idea of rescaling time to the unit interval so that infill asymptotics can be used to study non-stationary processes. In this setting, the sample size $n$ no longer corresponds to getting information about the future, but instead denotes the number of observations we have of a process that changes slowly over time. As $n$ increases, we get more and more data about each *local* structure of the non-stationary process of interest. Zhou and Wu [ZW09] introduced the framework for representing locally stationary time series as nonlinear functions of iid random elements as in Wu [Wu05]. Before introducing the details for the causal representation, we must introduce the necessary notation.

We work in a high-dimensional time series setting under a triangular array framework. We allow the dimensions of each of the processes to grow with $n$ and the collection of distributions of the processes to change with $n$. Thus, we write the dependence of the processes, distributions, and dimensions on $n$ explicitly going forward. Let $(X_{t,n}, Y_{t,n}, Z_{t,n})_{t\in[n]}$, $[n] := \{1, \dots, n\}$, be the observed sequence. Let $d_X := d_{X,n}$, $d_Y := d_{Y,n}$, $d_Z := d_{Z,n}$ denote the dimensions of the processes $X, Y, Z$, respectively. Denote dimension $i \in [d_X]$ of $X_{t,n}$ by $X_{t,n,i}$, dimension $j \in [d_Y]$ of $Y_{t,n}$ by $Y_{t,n,j}$, and dimension $k \in [d_Z]$ of $Z_{t,n}$ by $Z_{t,n,k}$.

Next, we introduce the notation for the time-offsets of $X_{t,n}$, $Y_{t,n}$, and $Z_{t,n}$. Denote by the time-offset $a \in A_i$ of $X_{t,n,i}$ by $X_{t,n,i,a} := X_{t+a,n,i}$, the time-offset $b \in B_j$ of $Y_{t,n,j}$ by $Y_{t,n,j,b} := Y_{t+b,n,j}$, and the time-offset $c \in C_k$ of $Z_{t,n,k}$ by $Z_{t,n,k,c} := Z_{t+c,n,k}$. Negative time-offsets are called *lags* of the process, and positive time-offsets are called *leads* of the process. Here, $A_i, B_j \subset \{-n+1, \dots, n-1\}$ are the set of time-offsets of $X_{t,n,i}$ and $Y_{t,n,j}$ under consideration. Similarly, $C_k \subset \{-n+1, \dots, -1, 0\}$ is the set of time-offsets of $Z_{t,n,k}$ under consideration, which we restrict to be non-positive so that the time-offset is known at time $t$. Denote the set of all time-offsets of $X_{t,n}$ by $A = \bigcup_{i=1}^{d_X} A_i$, all time-offsets of $Y_{t,n}$ by $B = \bigcup_{j=1}^{d_Y} B_j$, all time-lags of $Z_{t,n}$ by $C = \bigcup_{k=1}^{d_Z} C_k$. Denote the vectors with all dimensions and time-offsets by

$$\boldsymbol{X}_{t,n} = (X_{t,n,i,a})_{i\in[d_X],a\in A_i}, \quad \boldsymbol{Y}_{t,n} = (Y_{t,n,j,b})_{j\in[d_Y],b\in B_j}, \quad \boldsymbol{Z}_{t,n} = (Z_{t,n,k,c})_{k\in[d_Z],c\in C_k}$$

and denote their respective dimensions by $\boldsymbol{d_X} = \sum_{i=1}^{d_X} |A_i|$, $\boldsymbol{d_Y} = \sum_{j=1}^{d_Y} |B_j|$, and $\boldsymbol{d_Z} = \sum_{k=1}^{d_Z} |C_k|$. Also, denote the corresponding processes by

$$\boldsymbol{X}_n = (\boldsymbol{X}_{t,n})_{t\in\mathcal{T}_n}, \quad \boldsymbol{Y}_n = (\boldsymbol{Y}_{t,n})_{t\in\mathcal{T}_n}, \quad \boldsymbol{Z}_n = (\boldsymbol{Z}_{t,n})_{t\in\mathcal{T}_n}.$$

We allow the number of time-offsets and the number of dimensions to grow with $n$. To simplify the notation, let

$$\mathcal{M}_n = \{(i,j,a,b) : i \in [d_X], j \in [d_Y], a \in A_i, b \in B_j\}$$

be an index set of the dimensions and time-offsets. Going forward, we will refer to the tuple $(i,j,a,b) \in \mathcal{M}_n$ by $m \in \mathcal{M}_n$. The index set $\mathcal{M}_n$ depends on the sample size $n$ through the dimensions and time-offsets, so its cardinality $M_n = |\mathcal{M}_n|$ may grow with $n$.

We are interested in valid inference for *time-lagged* conditional independence relationships, so in our test we only use the subset of original times $\mathcal{T}_n \subseteq \{1,\ldots,n\}$ in which *all* time-offsets of each dimension of $X_{t,n}$, $Y_{t,n}$, and $Z_{t,n}$ are actually observed, where

$$\mathcal{T}_n = \{1 - \min(\{0\} \cup A \cup B \cup C), n - \max(\{0\} \cup A \cup B \cup C)\}.$$

Similarly, denote the corresponding interval of rescaled times in which all time-offsets are well-defined by

$$\mathcal{U}_n = \left[\frac{1}{n} - \frac{\min(\{0\} \cup A \cup B \cup C)}{n}, 1 - \frac{\max(\{0\} \cup A \cup B \cup C)}{n}\right] \subset [0,1].$$

Note that if no negative time-offsets (i.e. lags) are used then $\min(A \cup B \cup C) = 0$, and if no positive time-offsets (i.e. leads) are used then $\max(A \cup B \cup C) = 0$. We will write $t \in \mathcal{T}_n$ instead of $t \in [n]$ emphasize that we are only using the subset of times in which all time-offsets are observed. Going forward, we will denote the cardinality of $\mathcal{T}_n$ by $T_n = |\mathcal{T}_n|$.

We require that the largest (in magnitude) time-offset grows at a slower rate than $n$ such that as $n \to \infty$ we have $\min(\{0\} \cup A \cup B \cup C)/n \to 0$ and $\max(\{0\} \cup A \cup B \cup C)/n \to 0$ so that $T_n \to \infty$ arbitrarily slowly. In our theoretical analyses, we consider the asymptotics as $n \to \infty$ as usual. In the "best case" scenario in terms of regularity conditions for the temporal dependence and non-stationarity, the fastest that the number of dimensions and time-offsets $M_n$ can grow with $T_n$ is $M_n = O(T_n^{\frac{1}{4}-\delta})$ for some $\delta > 0$. See Subsection E.1 for a more detailed discussion about how quickly $M_n$ can grow with $T_n$.

Now that the necessary notation has been introduced, we introduce the causal representation of the process. For each $n \in \mathbb{N}$, we assume that each dimension of the observed sequence $(X_{t,n}, Y_{t,n}, Z_{t,n})_{t\in[n]}$ can be represented as a nonlinear function of iid random elements. What follows is most similar to the definition of high-dimensional locally stationary time series from Mies and Steland [MS22], which builds on the work of Zhou and Wu [ZW09] and Wu [Wu05]. The idea of representing processes as nonlinear functions of iid processes has a long history going back to at least Rosenblatt [Ros61] and Wiener [Wie66].

**Assumption 1** (Causal representation of the process)**.** *Let $\mathcal{F}_t^X = (\eta_t^X, \eta_{t-1}^X, \ldots)$, $\mathcal{F}_t^Y = (\eta_t^Y, \eta_{t-1}^Y, \ldots)$, $\mathcal{F}_t^Z = (\eta_t^Z, \eta_{t-1}^Z, \ldots)$ where $(\eta_t^X)_{t\in\mathbb{Z}}$, $(\eta_t^Y)_{t\in\mathbb{Z}}$, $(\eta_t^Z)_{t\in\mathbb{Z}}$ are sequences of iid random elements. Assume that we can represent each dimension of the observed sequence as the output of an evolving nonlinear system that was given a sequence of iid inputs:*

$$X_{t,n,i} = G_{n,i}^X(t/n, \mathcal{F}_t^X), \; Y_{t,n,j} = G_{n,j}^Y(t/n, \mathcal{F}_t^Y), \; Z_{t,n,k} = G_{n,k}^Z(t/n, \mathcal{F}_t^Z),$$

*where the systems are defined for all $u \in [0,1]$ by*

$$\tilde{X}_{t,n,i}(u) = G_{n,i}^X(u, \mathcal{F}_t^X), \; \tilde{Y}_{t,n,j}(u) = G_{n,j}^Y(u, \mathcal{F}_t^Y), \; \tilde{Z}_{t,n,k}(u) = G_{n,k}^Z(u, \mathcal{F}_t^Z),$$

*so that we may write $X_{t,n,i} = \tilde{X}_{t,n,i}(t/n)$, $Y_{t,n,j} = \tilde{Y}_{t,n,j}(t/n)$, $Z_{t,n,k} = \tilde{Z}_{t,n,k}(t/n)$.*

*For each $n \in \mathbb{N}$, $m \in \mathcal{M}_n$, $u \in [0,1]$, we have implicitly assumed that $G_{n,i}^X(u,\cdot)$, $G_{n,j}^Y(u,\cdot)$, $G_{n,k}^Z(u,\cdot)$ are measurable functions such that $G_{n,i}^X(u, \mathcal{F}_t^X)$, $G_{n,j}^Y(u, \mathcal{F}_t^Y)$, $G_{n,k}^Z(u, \mathcal{F}_t^Z)$ are well-defined random variables for each $t \in \mathbb{Z}$. Finally, assume that for each fixed $u$ (and $m, n$), we have that $(G_{n,i}^X(u, \mathcal{F}_t^X))_{t\in\mathbb{Z}}$, $(G_{n,j}^Y(u, \mathcal{F}_t^Y))_{t\in\mathbb{Z}}$, $(G_{n,k}^Z(u, \mathcal{F}_t^Z))_{t\in\mathbb{Z}}$ are each stationary time series.*

In light of Assumption 1, we may denote the causal representations for dimensions $i \in [d_X]$, $j \in [d_Y]$, $k \in [d_Z]$ with time-offsets $a \in A_i$, $b \in B_j$, $c \in C_k$, respectively, by

$$\tilde{X}_{t,n,i,a}(u) = G^X_{n,i}\left(u + \frac{a}{n}, \mathcal{F}^X_{t+a}\right), \tilde{Y}_{t,n,j,b}(u) = G^Y_{n,j}\left(u + \frac{b}{n}, \mathcal{F}^Y_{t+b}\right), \tilde{Z}_{t,n,k,c}(u) = G^Z_{n,k}\left(u + \frac{c}{n}, \mathcal{F}^Z_{t+c}\right),$$

so that we may write $X_{t,n,i,a} = \tilde{X}_{t,n,i,a}(t/n)$, $Y_{t,n,j,b} = \tilde{Y}_{t,n,j,b}(t/n)$, $Z_{t,n,k,c} = \tilde{Z}_{t,n,k,c}(t/n)$ for each dimension with time-offset of the observed sequence.

Similarly, denote the causal representations for all dimensions and no time-offsets by

$$\tilde{X}_{t,n}(u) = (G^X_{n,i}(u, \mathcal{F}^X_t))_{i \in [d_X]}, \tilde{Y}_{t,n}(u) = (G^Y_{n,j}(u, \mathcal{F}^Y_t))_{j \in [d_Y]}, \tilde{Z}_{t,n}(u) = (G^Z_{n,k}(u, \mathcal{F}^Z_t))_{k \in [d_Z]},$$

so that we may write the observed sequence as $X_{t,n} = \tilde{X}_{t,n}(t/n)$, $Y_{t,n} = \tilde{Y}_{t,n}(t/n)$, $Z_{t,n} = \tilde{Z}_{t,n}(t/n)$.

Lastly, denote the causal representations for all dimensions with time-offsets by

$$\tilde{\boldsymbol{X}}_{t,n}(u) = (\tilde{X}_{t,n,i,a}(u))_{i \in [d_X], a \in A_i}, \tilde{\boldsymbol{Y}}_{t,n}(u) = (\tilde{Y}_{t,n,j,b}(u))_{j \in [d_Y], b \in B_j}, \tilde{\boldsymbol{Z}}_{t,n}(u) = (\tilde{Z}_{t,n,k,c}(u))_{k \in [d_Z], c \in C_k}$$

so that we may write $\boldsymbol{X}_{t,n} = \tilde{\boldsymbol{X}}_{t,n}(t/n)$, $\boldsymbol{Y}_{t,n} = \tilde{\boldsymbol{Y}}_{t,n}(t/n)$, $\boldsymbol{Z}_{t,n} = \tilde{\boldsymbol{Z}}_{t,n}(t/n)$ for the observed sequence including dimensions and all time-offsets. We assume that for each $n \in \mathbb{N}$, $m \in \mathcal{M}_n$, $u \in \mathcal{U}_n$, $t \in \mathcal{T}_n$ the distribution of $(\tilde{X}_{t,n,i,a}(u), \tilde{Y}_{t,n,j,b}(u), \tilde{\boldsymbol{Z}}_{t,n}(u))$ is absolutely continuous with respect to the Lebesgue measure.

Let $\Omega$ be a sample space, $\mathcal{A}$ the Borel sigma-algebra, and $(\Omega, \mathcal{A})$ a measurable space. For fixed $n \in \mathbb{N}$, let $(\Omega, \mathcal{A})$ be equipped with a family of probability measures $(\mathbb{P}_P)_{P \in \mathcal{P}_n}$ where $\mathcal{P}_n$ is a collection of distributions of $(\tilde{X}_{t,n}(u), \tilde{Y}_{t,n}(u), \tilde{Z}_{t,n}(u))_{u \in [0,1], t \in \mathbb{Z}}$. That is, for a fixed $n \in \mathbb{N}$, the distribution or law of $(\tilde{X}_{t,n}(u), \tilde{Y}_{t,n}(u), \tilde{Z}_{t,n}(u))_{u \in [0,1], t \in \mathbb{Z}}$ under $\mathbb{P}_P$ is $P$. The family of probability measures $(\mathbb{P}_P)_{P \in \mathcal{P}_n}$ is defined with respect to the same measurable space $(\Omega, \mathcal{A})$, but need not have the same dominating measure. Denote the family of probability spaces by $(\Omega, \mathcal{A}, \mathbb{P}_P)_{P \in \mathcal{P}_n}$ and a sequence of such families of probability spaces by $((\Omega, \mathcal{A}, \mathbb{P}_P)_{P \in \mathcal{P}_n})_{n \in \mathbb{N}}$. For a given sample size $n \in \mathbb{N}$ and distribution $P \in \mathcal{P}_n$, let $\mathbb{E}_P(\cdot)$ denote the expectation of a random variable with distribution determined by $P$, and let $\mathbb{P}_P(A)$ denote the probability of an event $A \in \mathcal{A}$.

Lastly, we use the notation $o_{\mathcal{P}}(\cdot)$ and $O_{\mathcal{P}}(\cdot)$ in the same way that Shah and Peters [SP20] do, so we replicate their notation here. Let $(V_{P,n})_{n \in \mathbb{N}, P \in \mathcal{P}_n}$ be a family of sequences of random variables with distributions determined by $P \in \mathcal{P}_n$. We write $V_{P,n} = o_{\mathcal{P}}(1)$ to mean that for all $\epsilon > 0$, we have

$$\sup_{P \in \mathcal{P}_n} \mathbb{P}_P(|V_{P,n}| > \epsilon) \to 0.$$

Also, by $V_{P,n} = O_{\mathcal{P}}(1)$ we mean for all $\epsilon > 0$, there exists a constant $K > 0$ such that

$$\sup_{n \in \mathbb{N}} \sup_{P \in \mathcal{P}_n} \mathbb{P}_P(|V_{P,n}| > K) < \epsilon.$$

Let $(W_{P,n})_{n \in \mathbb{N}, P \in \mathcal{P}_n}$ be another family of sequences of random variables. By $V_{P,n} = o_{\mathcal{P}}(W_{P,n})$ we mean $V_{P,n} = W_{P,n} R_{P,n}$ and $R_{P,n} = o_{\mathcal{P}}(1)$, and by $V_{P,n} = O_{\mathcal{P}}(W_{P,n})$ we mean $V_{P,n} = W_{P,n} R_{P,n}$ and $R_{P,n} = O_{\mathcal{P}}(1)$.

## 1.2 Conditional independence for non-stationary time series

We discuss how to conduct a test for the global null hypothesis

$$H^{\mathrm{CI}}_{0,n} : X_{t,n,i,a} \perp\!\!\!\perp Y_{t,n,j,b} \mid \boldsymbol{Z}_{t,n} \text{ for all } t \in \mathcal{T}_n, m \in \mathcal{M}_n \tag{1}$$

$$H^{\mathrm{CI}}_{1,n} : X_{t,n,i,a} \not\perp\!\!\!\perp Y_{t,n,j,b} \mid \boldsymbol{Z}_{t,n} \text{ for at least one } t \in \mathcal{T}_n, m \in \mathcal{M}_n \tag{2}$$

that is asymptotically valid as $n \to \infty$, uniformly over an large collection of distributions $\mathcal{P}_n$ for which the null holds. If domain knowledge suggests that it is reasonable to restrict $\mathcal{P}_n$ to be those in which either conditional independence *or* conditional dependence holds for *all* times, then we can replace the alternative hypothesis (2) with

$$X_{t,n,i,a} \not\perp\!\!\!\perp Y_{t,n,j,b} \mid \boldsymbol{Z}_{t,n} \text{ for all } t \in \mathcal{T}_n, \text{ for at least one } m \in \mathcal{M}_n. \tag{3}$$

In Appendix C, we discuss simultaneous testing procedures so that we can determine whether conditional independence relationships hold or not during particular time windows and for specific dimension/time-offset indices $m \in \mathcal{M}_n$ all while controlling the familywise error rate. Moreover, we discuss how to gain power by using groups of time series when we don't care about *particular* conditional dependence relationships, but whether there is *some* conditional dependence relationship in the group of time series. For example, different cities in a county in epidemiology, companies in a sector in finance and economics, nodes in a sensor network in a brain or Earth region in neuroscience or climate science, or simply groupings of several leads or lags. Hence, our testing framework offers a large toolkit that can be used for variety of settings when only one observation of a high-dimensional non-stationary time series is made.

In the familiar context of forecasting, the global null hypothesis (1) can be viewed as a null hypothesis of (strong) nonlinear Granger non-causality at all times. In the setting of forecasting, $\boldsymbol{Z}_{t,n}$ would be the vector of covariates known at the time of forecasting $t$, $X_{t,n,i,a}$ would be an auxiliary time series known at the time $t$ of forecasting, so $a$ would be a non-positive integer indicating the current value or $a$-th lag, and $Y_{t,n,j,b}$ would be the forecasting target so $b$ would be a non-negative integer indicating the current value for nowcasting or the $b$-th forecast horizon. However, dropping the restrictions of requiring $a$ to be non-positive and $b$ to be non-negative is more general than the forecasting setting, so we allow for this in the notation as described in Subsection 1.1.

# 2 The Testing Framework

Subsection 2.1 contains definitions and notations that will be used throughout this section. In Subsection 2.2 introduces necessary concepts for our conditional independence test. The theoretical assumptions for our framework are stated in Subsection D.1 due to their length. In Subsection 2.3, we discuss our conditional independence test.

Our testing framework can be considered an extension of the *generalized covariance measure* (GCM) conditional independence test from Shah and Peters [SP20] to the non-stationary time series setting in which the conditional independence relationships can change over time. We first briefly summarize the univariate version of the original GCM test. Let $X$, $Y$ be two random variables and let $Z$ be a random vector. Assume the joint distribution of $(X, Y, Z)$ is absolutely continuous with respect to the Lebesgue measure. The GCM test is based on the "weak" conditional independence criterion of Daudin [Dau80], which states that if $X \perp\!\!\!\perp Y \mid Z$ then $\mathbb{E}_P[\phi(X, Z)\varphi(Y, Z)] = 0$ for all functions $\phi \in L^2_{X,Z}$ and $\varphi \in L^2_{Y,Z}$ such that $\mathbb{E}_P[\phi(X, Z) \mid Z] = 0$ and $\mathbb{E}_P[\varphi(Y, Z) \mid Z] = 0$. Thus, under the null hypothesis of conditional independence, the expectation of the products of errors $\mathbb{E}_P(\varepsilon\xi)$ from the regressions $X = \phi(Z) + \varepsilon$ and $Y = \varphi(Z) + \xi$, or equivalently the expected conditional covariance $\mathbb{E}_P[\mathrm{Cov}_P(X, Y|Z)]$, is equal to zero. The GCM test statistic is based on the normalized sum of the products of residuals from the regressions of $X$ on $Z$ and $Y$ on $Z$.

We can translate this into our setting as follows. Assume that for each $n \in \mathbb{N}$, $u \in \mathcal{U}_n$, $m \in \mathcal{M}_n$, $t \in \mathcal{T}_n$ the joint distribution of $(\tilde{X}_{t,n,i,a}(u), \tilde{Y}_{t,n,j,b}(u), \tilde{\boldsymbol{Z}}_{t,n}(u))$ is absolutely continuous with respect to the Lebesgue measure. For some $n \in \mathbb{N}$, $u \in \mathcal{U}_n$, $m \in \mathcal{M}_n$, $t \in \mathcal{T}_n$, if $\tilde{X}_{t,n,i,a}(u) \perp\!\!\!\perp \tilde{Y}_{t,n,j,b}(u) \mid \tilde{\boldsymbol{Z}}_{t,n}(u)$ then $\mathbb{E}_P[\phi(\tilde{X}_{t,n,i,a}(u), \tilde{\boldsymbol{Z}}_{t,n}(u))\varphi(\tilde{Y}_{t,n,j,b}(u), \tilde{\boldsymbol{Z}}_{t,n}(u)] = 0$ for all functions $\phi \in L^2_{\tilde{X}_{t,n,i,a}(u),\tilde{\boldsymbol{Z}}_{t,n}(u)}$ and $\varphi \in L^2_{\tilde{Y}_{t,n,j,b}(u),\tilde{\boldsymbol{Z}}_{t,n}(u)}$ such that $\mathbb{E}_P[\phi(\tilde{X}_{t,n,i,a}(u), \tilde{\boldsymbol{Z}}_{t,n}(u)) \mid \tilde{\boldsymbol{Z}}_{t,n}(u)] = 0$ and $\mathbb{E}_P[\varphi(\tilde{Y}_{t,n,j,b}(u), \tilde{\boldsymbol{Z}}_{t,n}(u))) \mid \tilde{\boldsymbol{Z}}_{t,n}(u))] = 0$ and hence the corresponding *local* expected conditional covariance

$$\rho_{P,t,n,m}(u) = \mathbb{E}_P[\mathrm{Cov}_P(\tilde{X}_{t,n,i,a}(u), \tilde{Y}_{t,n,j,b}(u)|\tilde{\boldsymbol{Z}}_{t,n}(u))]$$

is equal to zero. Hence, the process of error products from the time-varying nonlinear regressions of $(X_{t,n,i,a})_{t\in\mathcal{T}_n}$ on $(\boldsymbol{Z}_{t,n})_{t\in\mathcal{T}_n}$ and $(Y_{t,n,j,b})_{t\in\mathcal{T}_n}$ on $(\boldsymbol{Z}_{t,n})_{t\in\mathcal{T}_n}$ has mean zero. The exact notation will be introduced in Subsection 2.1.

Whereas other tests for conditional independence for stochastic processes treat the processes as entire objects, as in the formulation (8), our testing framework is based on the perspective of detecting *conditional flows of information* between the processes via their local expected conditional covariances. In particular, under the global null hypothesis of conditional independence (1), all of the expected conditional covariances $\rho_{P,t,n,m}(t/n)$ are equal to zero. Hence, we aim to detect conditional dependencies by determining whether the local expected conditional covariances $\rho_{P,t,n,m}(u)$ deviate from zero *at any point in time* for any index $m \in \mathcal{M}_n$.

Note that while our test is based on the expected conditional covariance functional, our framework can be easily adapted to be used with any functional that is equal to zero under the null of conditional independence. Zhang and Janson [ZJ20] discuss some of the shortcomings of the expected conditional covariance function, in particular that it lacks sensitivity to nonlinear relationships and interactions. We leave further explorations and comparisons with other functionals for future work.

Our test statistic is based on the products of residuals from the time-varying nonlinear regressions of $(X_{t,n,i,a})_{t\in\mathcal{T}_n}$ on $(\boldsymbol{Z}_{t,n})_{t\in\mathcal{T}_n}$ and $(Y_{t,n,j,b})_{t\in\mathcal{T}_n}$ on $(\boldsymbol{Z}_{t,n})_{t\in\mathcal{T}_n}$, respectively. We require *time-uniform* convergence rates on the time-varying nonlinear regression estimators, see Ding and Zhou [DZ21] and Zhang and Wu [ZW15] for instance. We reject the null hypothesis of conditional independence (1) if the magnitude of the partial sum process of the products of residuals ever becomes "too large" *at some point in time* over the entire time period. The limiting distribution of our test statistic is therefore based on the *strong Gaussian approximation* for high-dimensional non-stationary time from Mies and Steland [MS22], as opposed to a central limit theorem.

Analogously to the GCM test from Shah and Peters [SP20], our test only has power against alternatives in which the local expected conditional covariances are non-zero for at least *some points in time*. In other words, if the local expected conditional covariances are *always* zero we cannot hope to detect whether conditional dependence holds at some or all times. Note that for non-stationary time series the local expected conditional covariances can be zero at some times and non-zero at other times even if the corresponding conditional dependence relationships hold at all times. Hence, our test is useful for both alternative hypotheses (2) and (3).

## 2.1 Basic setup

For a fixed sample size $n \in \mathbb{N}$, distribution $P \in \mathcal{P}_n$, time $t \in \mathcal{T}_n$ and dimension/time-offset index tuple $m = (i, j, a, b) \in \mathcal{M}_n$, we can always decompose

$$X_{t,n,i,a} = f_{P,t,n,i,a}(t/n, \boldsymbol{Z}_{t,n}) + \varepsilon_{P,t,n,i,a}, \;\; Y_{t,n,j,b} = g_{P,t,n,j,b}(t/n, \boldsymbol{Z}_{t,n}) + \xi_{P,t,n,j,b},$$

where $f_{P,t,n,i,a}(u, \boldsymbol{z}) = \mathbb{E}_P(\tilde{X}_{t,n,i,a}(u)|\tilde{\boldsymbol{Z}}_{t,n}(u) = \boldsymbol{z})$ and $g_{P,t,n,j,b}(u, \boldsymbol{z}) = \mathbb{E}_P(\tilde{Y}_{t,n,j,b}(u)|\tilde{\boldsymbol{Z}}_{t,n}(u) = \boldsymbol{z})$ are the time-varying regression functions which map rescaled time $u \in \mathcal{U}_n$ and $\boldsymbol{z} \in \mathbb{R}^{\boldsymbol{d_z}}$ to $\mathbb{R}$, with the error processes satisfying $\mathbb{E}_P(\varepsilon_{P,t,n,i,a}|\boldsymbol{Z}_{t,n}) = 0$ and $\mathbb{E}_P(\xi_{P,t,n,j,b}|\boldsymbol{Z}_{t,n}) = 0$. Also, define $u_{P,t,n,i,a}(\boldsymbol{z}) = \mathbb{E}_P(\varepsilon^2_{P,t,n,i,a}|\boldsymbol{Z}_{t,n} = \boldsymbol{z})$ and $v_{P,t,n,j,b}(\boldsymbol{z}) = \mathbb{E}_P(\xi^2_{P,t,n,j,b}|\boldsymbol{Z}_{t,n} = \boldsymbol{z})$.

Define the process of error products by

$$R_{P,t,n,m} = \varepsilon_{P,t,n,i,a}\xi_{P,t,n,j,b}.$$

Similarly, denote the process of the residual products by

$$\hat{R}_{t,n,m} = \hat{\varepsilon}_{t,n,i,a}\hat{\xi}_{t,n,j,b},$$

where $\hat{\varepsilon}_{t,n,i,a} = X_{t,n,i,a} - \hat{f}_{t,n,i,a}(t/n, \boldsymbol{Z}_{t,n})$ and $\hat{\xi}_{t,n,j,b} = Y_{t,n,j,b} - \hat{g}_{t,n,j,b}(t/n, \boldsymbol{Z}_{t,n})$ and $\hat{f}_{t,n,i,a}$, $\hat{g}_{t,n,j,b}$ are estimates of $f_{P,t,n,i,a}$, $g_{P,t,n,j,b}$ created by time-varying nonlinear regressions of $(X_{t,n,i,a})_{t\in\mathcal{T}_n}$ on $(\boldsymbol{Z}_{t,n})_{t\in\mathcal{T}_n}$ and $(Y_{t,n,j,b})_{t\in\mathcal{T}_n}$ on $(\boldsymbol{Z}_{t,n})_{t\in\mathcal{T}_n}$, respectively.

In Subsection 2.3, we state the rates of convergence required of the time-varying regression estimators $\hat{f}_{t,n,i,a}$ and $\hat{g}_{t,n,j,b}$. In practice, our test can be used with any black-box time-varying regression estimator for non-stationary time series, such as the time-varying regression sieve estimator from Ding and Zhou [DZ21], the $L_2$ boosting method from Yousuf and Ng [YN21], the kernel estimator from Zhang and Wu [ZW15], or the M-estimator from Liu and Zhou [LZ23]. When conducting the test, one may use *all* the available times $\{1, \dots, n\}$ to estimate the time-varying regression functions for each of the (finitely many) time-offsets of each dimension and then only consider the products of residuals at the subset of times $\mathcal{T}_n$.

The covariate process $(\boldsymbol{Z}_{t,n})_{t\in\mathcal{T}_n}$ is a non-stationary time series and each of the dimensions can depend on one another. We can, for instance, include any lags of any of the dimensions of the original time series $Z_{t,n}$ since these lags are known at time $t$. The error processes $(\varepsilon_{P,t,n,i,a})_{t\in\mathcal{T}_n}$, $(\xi_{P,t,n,j,b})_{t\in\mathcal{T}_n}$ can also be non-stationary time series that depend on $(\boldsymbol{Z}_{t,n})_{t\in\mathcal{T}_n}$ and $(X_{t,n,i,a})_{t\in\mathcal{T}_n}$, $(Y_{t,n,j,b})_{t\in\mathcal{T}_n}$, respectively. In Subsection 2.2, we discuss concepts from our theoretical framework that are necessary for understanding our test. For expository purposes, we state the detailed assumptions about the nonlinear locally stationary time series framework in Subsection D.1.

## 2.2 Theoretical framework

Analogously to Assumption 1 regarding the causal representations of the processes, we will now introduce the causal representations of the error processes from Subsection 2.1.

**Assumption 2** (Causal representations of the error processes)**.** *For each $n \in \mathbb{N}$, $m \in \mathcal{M}_n$, $t \in \mathcal{T}_n$, we can represent the error processes as*

$$\varepsilon_{P,t,n,i,a} = G^{\varepsilon}_{P,n,i}\left(\frac{t}{n} + \frac{a}{n}, \mathcal{F}^{\varepsilon}_{t+a}\right) = X_{t,n,i,a} - \mathbb{E}_P(X_{t,n,i,a}|\boldsymbol{Z}_{t,n}),$$

$$\xi_{P,t,n,j,b} = G^{\xi}_{P,n,j}\left(\frac{t}{n} + \frac{b}{n}, \mathcal{F}^{\xi}_{t+b}\right) = Y_{t,n,j,b} - \mathbb{E}_P(Y_{t,n,j,b}|\boldsymbol{Z}_{t,n}),$$

*where $\mathcal{F}^{\varepsilon}_{t+a} = (\eta^{\varepsilon}_{t+a}, \eta^{\varepsilon}_{t+a-1}, \ldots)$, $\mathcal{F}^{\xi}_{t+b} = (\eta^{\xi}_{t+b}, \eta^{\xi}_{t+b-1}, \ldots)$ and $(\eta^{\varepsilon}_{t+a})_{t\in\mathbb{Z}}$, $(\eta^{\xi}_{t+b})_{t\in\mathbb{Z}}$ are sequences of iid random elements. In particular, $\eta^{\varepsilon}_{t+a} = (\eta^X_{t+a}, \eta^Z_t)'$, $\eta^{\xi}_{t+b} = (\eta^Y_{t+b}, \eta^Z_t)'$ for each $t \in \mathbb{Z}$ so that the error processes each depend on the inputs for the covariate processes and their respective response process. Also, we have $\mathbb{E}_P(\varepsilon_{P,t,n,i,a}|\mathcal{F}^Z_t) = 0$ and $\mathbb{E}_P(\xi_{P,t,n,j,b}|\mathcal{F}^Z_t) = 0$. For a given $n \in \mathbb{N}$, $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, we have that $G^{\varepsilon}_{P,n,i}(u, \cdot)$, $G^{\xi}_{P,n,j}(u, \cdot)$ are measurable functions such that $G^{\varepsilon}_{P,n,i}(u, \mathcal{F}^{\varepsilon}_t)$, $G^{\xi}_{P,n,j}(u, \mathcal{F}^{\xi}_t)$ are well-defined random variables for each $t \in \mathbb{Z}$ and $(G^{\varepsilon}_{P,n,i}(u, \mathcal{F}^{\varepsilon}_t))_{t\in\mathbb{Z}}$, $(G^{\xi}_{P,n,j}(u, \mathcal{F}^{\xi}_t))_{t\in\mathbb{Z}}$ are stationary time series. Also, denote*

$$\tilde{\varepsilon}_{P,t,n,i,a}(u) = G^{\varepsilon}_{P,n,i}\left(u + \frac{a}{n}, \mathcal{F}^{\varepsilon}_{t+a}\right), \tilde{\xi}_{P,t,n,j,b}(u) = G^{\xi}_{P,n,j}\left(u + \frac{b}{n}, \mathcal{F}^{\xi}_{t+b}\right),$$

*so that we may write $\varepsilon_{P,t,n,i,a} = \tilde{\varepsilon}_{P,t,n,i,a}(t/n)$, $\xi_{P,t,n,j,b} = \tilde{\xi}_{P,t,n,j,b}(t/n)$.*

In light of the causal representations of $\varepsilon_{P,t,n,i,a}$ and $\xi_{P,t,n,j,b}$ from Assumption 2, for each $P \in \mathcal{P}_n$, $t \in \mathcal{T}_n$, $n \in \mathbb{N}$, $u \in \mathcal{U}_n$ we have the following causal representation of the high-dimensional non-stationary vector-valued error processes

$$\boldsymbol{\varepsilon}_{P,t,n} = \boldsymbol{G}^{\boldsymbol{\varepsilon}}_{P,n}\left(\frac{t}{n}, \mathcal{F}^{\boldsymbol{\varepsilon}}_t\right) = (G^{\varepsilon}_{P,n,i}\left(\frac{t}{n} + \frac{a}{n}, \mathcal{F}^{\varepsilon}_{t+a}\right))_{i\in[d_X],a\in A_i},$$

where $\mathcal{F}^{\boldsymbol{\varepsilon}}_t = (\eta^{\boldsymbol{\varepsilon}}_t, \eta^{\boldsymbol{\varepsilon}}_{t-1}, \ldots)$ and $\eta^{\boldsymbol{\varepsilon}}_t = (\eta^X_{t+a_{\max}}, \eta^Z_t)$ for each $t \in \mathbb{Z}$, where $a_{\max} = \max(A)$. In light of the previous discussion, for a fixed $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, and $n \in \mathbb{N}$ we have that $\boldsymbol{G}^{\boldsymbol{\varepsilon}}_{P,n}(u, \mathcal{F}^{\boldsymbol{\varepsilon}}_t)$ is a well-defined high-dimensional random vector for each $t \in \mathbb{Z}$ and $(\boldsymbol{G}^{\boldsymbol{\varepsilon}}_{P,n}(u, \mathcal{F}^{\boldsymbol{\varepsilon}}_t))_{t\in\mathbb{Z}}$ is a high-dimensional stationary vector-valued time series, where we denote

$$\tilde{\boldsymbol{\varepsilon}}_{P,t,n}(u) = \boldsymbol{G}^{\boldsymbol{\varepsilon}}_{P,n}\left(u, \mathcal{F}^{\boldsymbol{\varepsilon}}_t\right) = (G^{\varepsilon}_{P,n,i}\left(u + \frac{a}{n}, \mathcal{F}^{\varepsilon}_{t+a}\right))_{i\in[d_X],a\in A_i},$$

so that we may write $\boldsymbol{\varepsilon}_{P,t,n} = \tilde{\boldsymbol{\varepsilon}}_{P,t,n}(t/n)$. Similarly, for $\boldsymbol{\xi}_{P,t,n}$ we write

$$\boldsymbol{\xi}_{P,t,n} = \boldsymbol{G}^{\boldsymbol{\xi}}_{P,n}\left(\frac{t}{n}, \mathcal{F}^{\boldsymbol{\xi}}_t\right) = (G^{\xi}_{P,n,j}\left(\frac{t}{n} + \frac{b}{n}, \mathcal{F}^{\xi}_{t+b}\right))_{j\in[d_Y],b\in B_j},$$

where $\mathcal{F}^{\boldsymbol{\xi}}_t = (\eta^{\boldsymbol{\xi}}_t, \eta^{\boldsymbol{\xi}}_{t-1}, \ldots)$ and $\eta^{\boldsymbol{\xi}}_t = (\eta^Y_{t+b_{\max}}, \eta^Z_t)$ for each $t \in \mathbb{Z}$, where $b_{\max} = \max(B)$, and

$$\tilde{\boldsymbol{\xi}}_{P,t,n}(u) = \boldsymbol{G}^{\boldsymbol{\xi}}_{P,n}\left(u, \mathcal{F}^{\boldsymbol{\xi}}_t\right) = (G^{\xi}_{P,n,j}\left(u + \frac{b}{n}, \mathcal{F}^{\xi}_{t+b}\right))_{j\in[d_Y],b\in B_j},$$

so that we may write $\boldsymbol{\xi}_{P,t,n} = \tilde{\boldsymbol{\xi}}_{P,t,n}(t/n)$.

Moreover, for each $m \in \mathcal{M}_n$ the process of error products can be represented as

$$R_{P,t,n,m} = G^R_{P,n,m}\left(\frac{t}{n}, \mathcal{F}^R_{m,t}\right) = G^{\varepsilon}_{P,n,i}\left(\frac{t}{n} + \frac{a}{n}, \mathcal{F}^{\varepsilon}_{t+a}\right) G^{\xi}_{P,n,j}\left(\frac{t}{n} + \frac{b}{n}, \mathcal{F}^{\xi}_{t+b}\right),$$

where $\mathcal{F}^R_{m,t} = (\eta^R_{m,t}, \eta^R_{m,t-1}, \ldots)$ and $(\eta^R_{m,t})_{t\in\mathbb{Z}}$ is a sequence of iid random elements with $\eta^R_{m,t} = (\eta^X_{t+a}, \eta^Y_{t+b}, \eta^Z_t)'$ for each $t \in \mathbb{Z}$. For a fixed rescaled time $u \in \mathcal{U}_n$ and $m \in \mathcal{M}_n$, $G^R_{P,n,m}(u, \cdot)$ is a measurable function such that $G^R_{P,n,m}(u, \mathcal{F}^R_{m,t})$ is a well-defined random variable for each $t \in \mathbb{Z}$ and $(G^R_{P,n,m}(u, \mathcal{F}^R_{m,t}))_{t\in\mathbb{Z}}$ is a stationary time series, where we denote

$$\tilde{R}_{P,t,n,m}(u) = G^R_{P,n,m}\left(u, \mathcal{F}^R_{m,t}\right) = G^{\varepsilon}_{P,n,i}\left(u + \frac{a}{n}, \mathcal{F}^{\varepsilon}_{t+a}\right) G^{\xi}_{P,n,j}\left(u + \frac{b}{n}, \mathcal{F}^{\xi}_{t+b}\right),$$

so that we may write $R_{P,t,n,m} = \tilde{R}_{P,t,n,m}(t/n)$.

For each $P \in \mathcal{P}_n$, $t \in \mathcal{T}_n$, $n \in \mathbb{N}$, and $u \in \mathcal{U}_n$ we have the following causal representation of the high-dimensional non-stationary $\mathbb{R}^{M_n}$-valued process of all the products of errors $\boldsymbol{R}_{P,t,n}$ by

$$\boldsymbol{R}_{P,t,n} = \boldsymbol{G}^{\boldsymbol{R}}_{P,n}\left(\frac{t}{n}, \mathcal{F}^{\boldsymbol{R}}_t\right) = \left(G^R_{P,n,m}\left(\frac{t}{n}, \mathcal{F}^R_{m,t}\right)\right)_{m\in\mathcal{M}_n}$$

where $\mathcal{F}^{\boldsymbol{R}}_t = (\eta^{\boldsymbol{R}}_t, \eta^{\boldsymbol{R}}_{t-1}, \ldots)$ and $\eta^{\boldsymbol{R}}_t = (\eta^X_{t+a_{\max}}, \eta^Y_{t+b_{\max}}, \eta^Z_t)$ for each $t \in \mathbb{Z}$, where $a_{\max} = \max(A)$ and $b_{\max} = \max(B)$. For a fixed $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, and $n \in \mathbb{N}$ we have that $\boldsymbol{G}^{\boldsymbol{R}}_{P,n}(u, \mathcal{F}^{\boldsymbol{R}}_t)$ is a well-defined high-dimensional random vector for each $t \in \mathbb{Z}$ and $(\boldsymbol{G}^{\boldsymbol{R}}_{P,n}(u, \mathcal{F}^{\boldsymbol{R}}_t))_{t\in\mathbb{Z}}$ is a high-dimensional stationary $\mathbb{R}^{M_n}$-valued time series, where we denote

$$\tilde{\boldsymbol{R}}_{P,t,n}(u) = \boldsymbol{G}^{\boldsymbol{R}}_{P,n}\left(u, \mathcal{F}^{\boldsymbol{R}}_t\right) = \left(G^R_{P,n,m}\left(u, \mathcal{F}^R_{m,t}\right)\right)_{m\in\mathcal{M}_n},$$

so that we may write $\boldsymbol{R}_{P,t,n} = \tilde{\boldsymbol{R}}_{P,t,n}(t/n)$.

We will now define the local long-run covariance matrices and variances of the high-dimensional process of error products defined in Assumption 2.

**Definition 1** (Local long-run covariance matrices and variances of process of error products)**.** *For each $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, $n \in \mathbb{N}$, define the long-run covariance matrix of the $\mathbb{R}^{M_n}$-valued stationary process $(\boldsymbol{G}^{\boldsymbol{R}}_{P,n}(u, \mathcal{F}^{\boldsymbol{R}}_t))_{t\in\mathbb{Z}}$ by*

$$\boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(u) = \sum_{h\in\mathbb{Z}} \mathrm{Cov}_P(\boldsymbol{G}^{\boldsymbol{R}}_{P,n}(u, \mathcal{F}^{\boldsymbol{R}}_0), \boldsymbol{G}^{\boldsymbol{R}}_{P,n}(u, \mathcal{F}^{\boldsymbol{R}}_h)).$$

*Similarly, for each $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, $n \in \mathbb{N}$, and $m \in \mathcal{M}_n$, denote the long-run variance of the $\mathbb{R}$-valued stationary process $(G^R_{P,n,m}(u, \mathcal{F}^R_t))_{t\in\mathbb{Z}}$ as*

$$\Sigma^R_{P,n,m}(u) = \sum_{h\in\mathbb{Z}} \mathrm{Cov}_P(G^R_{P,n,m}(u, \mathcal{F}^R_{m,0}), G^R_{P,n,m}(u, \mathcal{F}^R_{m,h})).$$

### 2.3 The test

The key to our bootstrap-based testing procedure is a consistent estimator for the local long-run covariance matrices of the products of errors based on the products of residuals. We use the same cumulative covariance estimator from Mies and Steland [MS22]. However, the theoretical results in Mies and Steland [MS22] for covariance estimation are for using the estimator with the original time series. Hence, we extend the covariance estimation results so that it can be used with products of residuals. In Section E, we introduce distribution-uniform extensions of many of the theoretical results from Mies and Steland [MS22].

The estimator for the local long-run covariance matrix $\boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(t/n)$ is given by

$$\hat{\boldsymbol{\Sigma}}^{\boldsymbol{R}}_n(t/n) = \hat{Q}^{\boldsymbol{R}}_{t,n} - \hat{Q}^{\boldsymbol{R}}_{t-1,n},$$

where

$$\hat{Q}^{\boldsymbol{R}}_{k,n} = \sum_{r=L_n}^{k} \frac{1}{L_n}\left(\sum_{s=r-L_n+l_n}^{r-1+l_n} \hat{\boldsymbol{R}}_{s,n}\right)\left(\sum_{s=r-L_n+l_n}^{r-1+l_n} \hat{\boldsymbol{R}}_{s,n}\right)^T$$

is the estimator of the cumulative covariance $Q^{\boldsymbol{R}}_{P,k,n} = \sum_{t=1}^{k} \boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(t/n)$, where $l_n = \min(\mathcal{T}_n)$ and $L_n \in \mathbb{N}$ is the lag window size parameter where $L_n = O(T_n^{\zeta})$ for some $\zeta \in (0, \frac{1}{2})$. Going forward, we will denote $\hat{Q}^{\boldsymbol{R}}_n := (\hat{Q}^{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}$ and $Q^{\boldsymbol{R}}_{P,n} := (Q^{\boldsymbol{R}}_{P,t,n})_{t\in\mathcal{T}_n}$.

Define the max-type and sum of squares-type test statistics by

$$S_{n,\infty}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) = \max_{s\in\mathcal{T}_n} \left\| \frac{1}{\sqrt{T_n}} \sum_{t\leq s} \hat{\boldsymbol{R}}_{t,n} \right\|_\infty, \quad S_{n,2}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) = \max_{s\in\mathcal{T}_n} \left\| \frac{1}{\sqrt{T_n}} \sum_{t\leq s} \hat{\boldsymbol{R}}_{t,n} \right\|_2. \tag{4}$$

Under the assumptions from Subsection D.1, we will show that tests based on $S_{n,\infty}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n})$ or $S_{n,2}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n})$ have uniformly asymptotic Type-I error control. It is well known that max-type test statistics have good power against sparse alternatives while sum of squares-type statistics have good power against dense alternatives. Note that we may use any test statistic $S_{n,p}$ with $p \geq 2$.

Since in practice we might not know which test statistic to use, we suggest using a Bonferroni combination test, as in Zhang and Shao [ZS24] and Gao, Wang, and Shao [GWS23]. This way we can achieve good power against both dense and sparse alternatives. That is, we reject the null with the Bonferroni combination test at significance level $\alpha$ if the p-value of either the test using $S_{n,\infty}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n})$ or $S_{n,2}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n})$ drops below $\alpha/2$. We leave the theoretical results for the asymptotic independence of $S_{n,\infty}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n})$ and $S_{n,2}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n})$ as well as the simulation results for the Bonferroni combination test for future work.

For some increasing process of symmetric, positive semidefnite matrices $Q_n^{\boldsymbol{V}} := (Q_{t,n}^{\boldsymbol{V}})_{t\in\mathcal{T}_n}$, let $\boldsymbol{V}_{t,n} \in \mathcal{N}(0, Q_{t,n}^{\boldsymbol{V}} - Q_{t-1,n}^{\boldsymbol{V}})$ for $t \in \mathcal{T}_n$. For some $\alpha \in (0,1)$, denote the $(1-\alpha)$ quantile of $S_n((\boldsymbol{V}_{t,n})_{t\in\mathcal{T}_n})$ by $a_\alpha(Q_n^{\boldsymbol{V}})$ for some test statistic $S_n$ satisfying

$$|S_n((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) - S_n((\boldsymbol{V}_{t,n})_{t\in\mathcal{T}_n})| \leq \max_{s\in\mathcal{T}_n} \left\| \frac{1}{\sqrt{T_n}} \sum_{t\leq s} (\hat{\boldsymbol{R}}_{t,n} - \boldsymbol{V}_{t,n}) \right\|_2. \tag{5}$$

Turning back to our test, we will calculate the cumulative covariance process $\hat{Q}_n^{\boldsymbol{R}}$ based on the process of products of residuals. In practice, we can approximate the quantile $a_\alpha(\hat{Q}_n^{\boldsymbol{R}})$ via Monte Carlo. We reject the null hypothesis of conditional independence at level $\alpha$ if $S_n((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \tau_n$ for some offsets $\nu_n \to 0$ and $\tau_n \to 0$ so that

$$\nu_n \gg \log(T_n) M_n \left[ \left( \frac{M_n}{T_n} \right)^{2\xi(\bar{q}^R, \bar{\beta}^R)} + \tau_n^{-2} \left( \varphi_{n,1} + \varphi_{n,2} \right) \right], \tag{6}$$

where

$$\varphi_{n,1} = T_n^{-\frac{1}{2}} M_n^{\frac{1}{4}} L_n^{\frac{1}{4}} + T_n^{-\frac{1}{4}} M_n^{\frac{1}{4}} L_n^{\frac{1}{4}} + L_n^{-\frac{1}{2}} + L_n^{1-\frac{\bar{\beta}^R}{2}} + T_n^{-1},$$

comes from the covariance estimation error and

$$\varphi_{n,2} = L_n^{\frac{1}{4}} T_n^{-\frac{1}{4}} \tau_n^{\frac{1}{2}} M_n^{-\frac{1}{4}} + \tau_n^{\frac{1}{2}} M_n^{-\frac{1}{4}} + T_n^{-\frac{1}{2}} M_n^{-\frac{1}{2}} L_n^{\frac{1}{2}} \tau_n + L_n^{\frac{1}{4}} T_n^{-\frac{1}{4}} M_n^{-\frac{1}{2}} \tau_n + M_n^{-\frac{1}{2}} \tau_n,$$

comes from the prediction errors since we use the products of residuals.

The constants $\bar{\beta}^R > 2$, $\bar{q}^R > 4$, are related to regularity conditions controlling the temporal dependence and non-stationarity uniformly over the collection of distributions $\mathcal{P}_n$ in Subsection D.1 and $\xi(\bar{q}^R, \bar{\beta}^R)$ is a rate defined in Subsection E.1. Note that we control the temporal dependence and non-stationarity for each component (i.e. dimension and time-offset) of the high-dimensional process of the product of errors $(\boldsymbol{R}_{P,t,n})_{t\in\mathcal{T}_n}$ and hence for the terms $\Theta_n$, $\Gamma_n$ from Mies and Steland [MS22] we have $\Theta_n \asymp M_n^{\frac{1}{2}}$ and $\Gamma_n \asymp M_n^{\frac{1}{2}}$. This is discussed in more detail in Subsection D.1. In practice, we discuss reasonable choices of offsets by conducting extensive simulations in Section 4.

The following result shows that if the time-varying regression estimators satisfy modest time-uniform and distribution-uniform convergence rate requirements, then the quantile $a_\alpha(\hat{Q}_n^{\boldsymbol{R}})_{t\in\mathcal{T}_n})$ closely approximates the $(1-\alpha)$ quantile of any test statistic $S_n$ satisfying (5) and thus we can use it to calibrate a test. See Ding and Zhou [DZ21] and Zhang and Wu [ZW15] for examples of time-varying regression estimators with time-uniform convergence rates. For the theorem below, define

$$\hat{w}^f_{P,t,n,i,a}(t/n, \boldsymbol{z}) = f_{P,t,n,i,a}(t/n, \boldsymbol{z}) - \hat{f}_{t,n,i,a}(t/n, \boldsymbol{z}),$$

$$\hat{w}^g_{P,t,n,j,b}(t/n, \boldsymbol{z}) = g_{P,t,n,j,b}(t/n, \boldsymbol{z}) - \hat{g}_{t,n,j,b}(t/n, \boldsymbol{z}),$$

where $t \in \mathcal{T}_n$ and $\boldsymbol{z} \in \mathbb{R}^{\boldsymbol{d_z}}$.

**Theorem 2.1.** *Suppose all the assumptions of Subsection D.1 used to control the temporal dependence and non-stationarity of the processes uniformly over the collection of distributions $\mathcal{P}_n$ hold. Moreover, suppose that the regression estimators satisfy*

$$\max_{m\in\mathcal{M}_n}\max_{t\in\mathcal{T}_n}\sup_{\boldsymbol{z}\in\mathbb{R}^{\boldsymbol{d_z}}}|\hat{w}^f_{P,t,n,i,a}(t/n,\boldsymbol{z})||\hat{w}^g_{P,t,n,j,b}(t/n,\boldsymbol{z})| = o_{\mathcal{P}}(M_n^{-\frac{1}{2}}T_n^{-\frac{1}{2}}\tau_n),$$

$$\max_{i\in[d_X]}\max_{a\in A_i}\max_{t\in\mathcal{T}_n}\sup_{\boldsymbol{z}\in\mathbb{R}^{\boldsymbol{d_z}}}|\hat{w}^f_{P,t,n,i,a}(t/n,\boldsymbol{z})| = o_{\mathcal{P}}(M_n^{-\frac{1}{2}}\tau_n),$$

$$\max_{j\in[d_Y]}\max_{b\in B_j}\max_{t\in\mathcal{T}_n}\sup_{\boldsymbol{z}\in\mathbb{R}^{\boldsymbol{d_z}}}|\hat{w}^g_{P,t,n,j,b}(t/n,\boldsymbol{z})| = o_{\mathcal{P}}(M_n^{-\frac{1}{2}}\tau_n).$$

*For any test statistic $S_n$ satisfying condition (5), if the offsets $\tau_n \to 0$ and $\nu_n \to 0$ are chosen such that condition (6) holds, we have that*

$$\lim_{n\to\infty}\sup\sup_{P\in\mathcal{P}_n}\mathbb{P}_P(S_n((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \tau_n) \leq \alpha.$$

Since the fastest $M_n$ can grow with $T_n$ is $M_n = O(T_n^{\frac{1}{4}-\delta})$ for some $\delta > 0$, we can satisfy the convergence rate requirements with

$$\max_{i\in[d_X]}\max_{a\in A_i}\max_{t\in\mathcal{T}_n}\sup_{\boldsymbol{z}\in\mathbb{R}^{\boldsymbol{d_z}}}|\hat{w}^f_{P,t,n,i,a}(t/n,\boldsymbol{z})| = o_{\mathcal{P}}(M_n^{-\frac{1}{4}}T_n^{-\frac{1}{4}}\tau_n),$$

$$\max_{j\in[d_Y]}\max_{b\in B_j}\max_{t\in\mathcal{T}_n}\sup_{\boldsymbol{z}\in\mathbb{R}^{\boldsymbol{d_z}}}|\hat{w}^g_{P,t,n,j,b}(t/n,\boldsymbol{z})| = o_{\mathcal{P}}(M_n^{-\frac{1}{4}}T_n^{-\frac{1}{4}}\tau_n).$$

# 3 COVID-19 Viral Load Dynamics

In this section, we apply our conditional independence test from Section 2 to a unique COVID-19 dataset. We determine whether a population's level of infectivity, as measured by cycle threshold (Ct) values from PCR tests, contains *additional* information about future case counts of COVID-19 beyond the information already contained in auxiliary signals. We use the nonparametric time-varying sieve regression estimator from Ding and Zhou [DZ21]. We use the sieve regression estimator because it performs well in the simulations reported in Section 4 . However, our test can be used with any time-varying regression estimator for non-stationary time series, such as the $L_2$ boosting method from Yousuf and Ng [YN21], the kernel estimator from Zhang and Wu [ZW15], or the M-estimator from Liu and Zhou [LZ23].

We will briefly describe the sieve regression estimator from Ding and Zhou [DZ21]. We simplify the notation for the processes and time-varying regression estimators in this section. We assume an additive form for the regression functions so that we may write

$$X(t/n) = \sum_{j=1}^{d_Z} f_j(t/n, Z_j(t/n)) + \varepsilon(t/n), \quad Y(t/n) = \sum_{j=1}^{d_Z} g_j(t/n, Z_j(t/n)) + \xi(t/n)$$

where $t \in [n]$, $d_Z \in \mathbb{N}$ is the dimension of the covariate process, and the error processes satisfy $\mathbb{E}(\varepsilon(t/n)|Z_j(t/n)) = 0$ and $\mathbb{E}(\xi(t/n)|Z_j(t/n)) = 0$. All of the processes $X$, $Y$, $Z$, $\epsilon$, $\xi$ can be non-stationary time series as described in Subsections 1.1, 2.2, and D.1. Each of the functions $m_j \in \{f_j, g_j\}$ for $j \in [d_Z]$ are *smooth* functions of rescaled time $u \in [0,1]$ and the value of the $j$-th covariate process $z \in \mathbb{R}$ and each $m_j$ can be approximated by $m_{j,c,d}(u,z) = \sum_{\ell_1}^{c}\sum_{\ell_2=1}^{d}\beta_{j,\ell_1,\ell_2}b_{\ell_1,\ell_2}(u,z)$, where $\{b_{\ell_1,\ell_2}(u,z)\}$ are sieve basis functions (e.g. Legendre polynomials, orthogonal wavelet basis functions, etc) and $\{\beta_{j,\ell_1,\ell_2}\}$ are coefficients which we can estimate with OLS.

We use our conditional independence test from Subsection 2.3 with the residuals obtained from regressing $X$ on $Z$ and $Y$ on $Z$ using the nonparametric time-varying sieve regression estimator for non-stationary time series explained above. Our model specification is based on the dynamics of a Susceptible-Infectious-Recovered-Deceased (SIRD) compartmental model, which is a system of ODEs describing infection dynamics. Following Chernozhukov, Kasahara, and Schrimpf [CKS21], we can manipulate the SIRD system of ODEs to obtain the following equation, which is the inspiration for our regression model specification:

$$\frac{\ddot{C}(u)}{\dot{C}(u)} = \frac{S(u)}{N}\beta(u) - \gamma + \frac{\dot{\tau}(u)}{\tau(u)}$$

where $u \in [0,1]$, $C(u)$ is the cumulative case count at time $u$, $S(u)$ is the susceptible population size at time $u$, $\beta(u)$ is the infection rate at time $u$, $\gamma$ is the rate of death, and $\tau(u)$ is the detection rate of new infections at time $u$. Crucially, this equation does not rely on the unobserved infected population.

Our covariates are related to the discrete-time analogue of the right hand side of the equation above $\frac{S(u)}{N}\beta(u) - \gamma + \frac{\dot{\tau}(u)}{\tau(u)}$. We relate the infection rate $\beta(u)$ and size of the susceptible population $S(u)$ to the following covariates known at the time of forecasting. First, we use the 3 and 4 week lags of the growth rates of new cases. To calculate the growth rates of new cases we use the 1 week difference of the log of the number of new cases in the last 2 weeks. Second, we use the 3 week lag of the log of new cases in the last 2 weeks. Third, we use the 3 week lag of the test positivity rates, which we define as the proportion of tests that are positive in the last 2 weeks. We also relate the logarithmic derivative of the detection rate $\frac{\dot{\tau}(u)}{\tau(u)}$ to the growth rate of new tests known at the time of forecasting. In particular, we use the 3 week lag in the growth rate of new tests, which we define as the 1 week difference of log of the number of new tests in the last 2 weeks. We emphasize that these covariates are 3 or 4 week lags and so they are known at the time of forecasting. Going forward, denote these covariates known at time $t$ by $Z(t/n)$.

Consider the discrete time analogue of the logarithmic derivative of new cases $\frac{\ddot{C}(u)}{\dot{C}(u)}$, which is on the left hand side of the equation above. To calculate the growth rates of new cases, we use the 1 week difference of the log of the number of new cases in the last 2 weeks. Going forward, denote the growth rates of new daily cases at time $t \in [n]$ by $Y(t/n)$. Also, denote the 3 week lag of the proportion of cycle threshold values under 25 at time $t$ by $X(t/n)$. $Y(t/n)$ is the target time series that we aim to forecast, and we want to know whether $X(t/n)$ contains auxiliary information about $Y(t/n)$ that is not already contained in the covariate processes $Z(t/n)$.

The null hypothesis is $X(t/n) \perp\!\!\!\perp Y(t/n) \mid Z(t/n)$ for all $t \in [n]$. Under the null, $X(t/n)$ the current proportion of cycle threshold values under 25 does not contain any auxiliary information about $Y(t/n)$ the 3-week ahead case growth rates of COVID-19 that is not already contained in the other signals $Z(t/n)$. The test statistic exceeds the 0.975 quantile. Informed by our extensive simulations reported in Section 4, we use the 0.975 instead of the 0.95 quantile for a 0.05 level test. Hence, we reject the null at $\alpha = 0.05$. We do not assume that the conditional independence relationship is static over time, so we only conclude that $X(t/n) \not\perp\!\!\!\perp Y(t/n) \mid Z(t/n)$ for *some* $t \in [n]$.

In general, if domain knowledge suggests that it is reasonable to assume that the conditional independence structure does not change over time, then we could have set the alternative hypothesis as $X(t/n) \not\perp\!\!\!\perp Y(t/n) \mid Z(t/n)$ for *all* $t \in [n]$. However, we avoid this assumption that the conditional independence structure remains static over time. In our companion paper we will determine during which time windows particular lags of certain quantiles of the time-varying viral load distribution contain auxiliary information about future case growth rates of COVID-19 at different forecasting horizons. In Section C, we discuss our proposed approaches for simultaneous testing using our bootstrap-based testing procedure.

Overall, our analysis supports the theory put forth by Tom and Mina [TM20] and Hay et al. [Hay+21] that cycle threshold values contain unique information about future case growth rates of COVID-19. In particular, we conclude that cycle threshold values contain a non-zero amount of information about future case growth rates of COVID-19 after accounting for the information contained in other signals. The result of our conditional independence test should only be understood as a form of variable significance testing or "screening". Our test does *not* tell us the extent to which cycle threshold values are important or useful for forecasting.

# 4 Simulations

In all of our tests, we use the time-varying sieve estimator from Ding and Zhou [DZ21]. We currently only report simulation results for univariate processes, adapting the data generating processes (DGPs) from the simulations in Ding and Zhou [DZ21]. Thus, it is not entirely surprising that similar parameter choices for the sieve estimator perform well. Denote by $c$ and $d$ the number of basis functions as discussed in Ding and Zhou [DZ21]. We use Legendre basis functions, setting $d = 2$ and alternating between $c = 7$ and $c = 10$. It appears as though $c = 7$ works better with Linear, Nonlinear, and TV-Linear, and $c = 10$ works better with TV-Nonlinear (where TV stands for time-varying). This makes sense because $c$ should increase as the regression function becomes more complex.

For the sake of clarity, we simplify the notation for processes in this section and denote the process of interest by $X(t/n)$ for $t \in [n]$. In all simulations, we test for the null hypothesis of

$$X(t/n) \perp\!\!\!\perp X(t/n - 3/n) \mid X(t/n - 1/n) \text{ for all times } t \in [n]$$

versus the alternative hypothesis of

$$X(t/n) \not\perp\!\!\!\perp X(t/n - 3/n) \mid X(t/n - 1/n) \text{ for all times } t \in [n].$$

We consider the following DGPs for the power simulations:

$$X(t/n) = 0.6X(t/n - 1/n) - 0.4X(t/n - 3/n) + \varepsilon(t/n) \tag{P1: Linear}$$

$$X(t/n) = \exp[-X(t/n - 1/n)^2] + \exp[-X(t/n - 3/n)^2] + \varepsilon(t/n) \tag{P2: Nonlinear}$$

$$\begin{aligned} X(t/n) = {} & 0.7\sin(2\pi t/n)X(t/n - 1/n) \\ & + 0.7\cos(2\pi t/n)X(t/n - 3/n) + \varepsilon(t/n) \end{aligned} \tag{P3: TV-Linear}$$

$$\begin{aligned} X(t/n) = {} & \sin(2\pi t/n)\exp[-X(t/n - 1/n)^2/2] \\ & + \cos(2\pi t/n)\exp[-X(t/n - 3/n)^2/2] + \varepsilon(t/n) \end{aligned} \tag{P4: TV-Nonlinear}$$

We consider the following DGPs for the size simulations:

$$X(t/n) = 0.6X(t/n - 1/n) + \varepsilon(t/n) \tag{S1: Linear}$$

$$X(t/n) = \exp[-X(t/n - 1/n)^2] + \varepsilon(t/n) \tag{S2: Nonlinear}$$

$$X(t/n) = 0.7\sin(2\pi t/n)X(t/n - 1/n) + \varepsilon(t/n) \tag{S3: TV-Linear}$$

$$X(t/n) = \sin(2\pi t/n)\exp[-X(t/n - 1/n)^2/2] + \varepsilon(t/n) \tag{S4: TV-Nonlinear}$$

We consider the following noise processes for the power and size simulations, where $Z(t/n) \sim N(0, 1)$ for $t \in [n]$:

$$\varepsilon(t/n) = 0.1Z(t/n) \tag{E1: IID Normal}$$

$$\varepsilon(t/n) = 0.6\sin(2\pi t/n)\varepsilon(t/n - 1/n) + 0.1Z(t/n) \tag{E2: TV-Linear v1}$$

$$\varepsilon(t/n) = 0.6\sin(2\pi t/n)\varepsilon(t/n - 1/n) + \cos(2\pi t/n)0.1Z(t/n) \tag{E3: TV-Linear v2}$$

The order of "difficulty" is $P1 < P2 < P3 < P4$, $S1 < S2 < S3 < S4$, and $E1 < E2 < E3$. That is, the Linear DGPs $P1$ and $S1$, when paired with the IID Gaussian error process $E1$, are the "easiest settings". The TV-Nonlinear DGPs $P4$ and $S4$, when paired with either of the TV-Linear error processes $E2$ or $E3$ are the "hardest settings". The following tables are structured as follows. The columns (P1,..., S4) indicate the model used for the power (P) or size (S) simulation. The rows (E1, E2, E3) indicate the error process used for the simulation. Each cell contains the empirical rejection rates out of 100 simulations.

The next two tables are the simulations with Legendre basis functions, $d = 2$, and alternating between $c = 7$ and $c = 10$, respectively. Observe that in the first table when we set $c = 7$ for S4: TV-Nonlinear, E1: IID Normal the empirical rejection rate is 22/100 and S4: TV-Nonlinear, E3: TV-Linear v2 it is 23/100. However, when we set $c = 10$ in the second table the empirical rejection rate is 6/100 and 10/100, respectively. This makes sense because higher values of $c$ should work better for more complex regression functions. Also, we can see the bias discussed in Subsection 2.3 which leads to our the next set of simulations in which we use offsets to account for this bias.

| $n = 500$, nsim = 100, Legendre basis, $c = 7$, $d = 2$, $\alpha = 0.05$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| DGP | P1 | P2 | P3 | P4 | S1 | S2 | S3 | S4 |
| E1 | 100/100 | 100/100 | 97/100 | 96/100 | 10/100 | 6/100 | 7/100 | 22/100 |
| E1 | 100/100 | 100/100 | 98/100 | 93/100 | 13/100 | 10/100 | 9/100 | 18/100 |
| E2 | 100/100 | 100/100 | 85/100 | 100/100 | 9/100 | 6/100 | 1/100 | 23/100 |

| $n = 500$, nsim = 100, Legendre basis, $c = 10$, $d = 2$, $\alpha = 0.05$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| DGP | P1 | P2 | P3 | P4 | S1 | S2 | S3 | S4 |
| E1 | 100/100 | 100/100 | 100/100 | 100/100 | 13/100 | 15/100 | 7/100 | 10/100 |
| E2 | 100/100 | 100/100 | 96/100 | 92/100 | 14/100 | 18/100 | 11/100 | 3/100 |
| E3 | 100/100 | 100/100 | 84/100 | 100/100 | 5/100 | 4/100 | 3/100 | 10/100 |

In the next tables, we study the empirical rejection rates with $\alpha = 0.025$ to correct for the bias in the test.

| $n = 500$, nsim = 100, Legendre basis, $c = 7$, $d = 2$, $\alpha = 0.025$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| DGP | P1 | P2 | P3 | P4 | S1 | S2 | S3 | S4 |
| E1 | 88/100 | 100/100 | 93/100 | 91/100 | 3/100 | 1/100 | 2/100 | 13/100 |
| E2 | 100/100 | 100/100 | 88/100 | 90/100 | 3/100 | 2/100 | 4/100 | 7/100 |
| E3 | 100/100 | 100/100 | 79/100 | 99/100 | 4/100 | 4/100 | 2/100 | 15/100 |

| $n = 500$, nsim = 100, Legendre basis, $c = 10$, $d = 2$, $\alpha = 0.025$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| DGP | P1 | P2 | P3 | P4 | S1 | S2 | S3 | S4 |
| E1 | 100/100 | 100/100 | 91/100 | 99/100 | 13/100 | 3/100 | 1/100 | 1/100 |
| E2 | 100/100 | 100/100 | 88/100 | 90/100 | 3/100 | 2/100 | 4/100 | 7/100 |
| E3 | 100/100 | 100/100 | 86/100 | 90/100 | 9/100 | 10/100 | 2/100 | 3/100 |

In future work, we will report more comprehensive simulation results. There are currently several limitations to the scope of our reported simulation results. We state these limitations clearly now.

First, we only use univariate DGPs. While it is good to see that our test performs well on univariate DGPs, it is necessary to understand the different ways in which our test may not work well with high-dimensional time series since that is what our test is designed for. In future work, we will report comprehensive simulation results with more complex multivariate DGPs. We will use time-varying vector autoregressions for these multivariate DGPs for computational reasons and so that the relationships between the processes are relatively simple to analyze.

Second, we currently do not compare with other conditional independence tests. We leave these comparisons for future work. Given the complex nature of the non-stationarity and nonlinearity in some of the DGPs we use, it does not seem fair to the conditional independence tests designed for stationary or linear time series to apply them to "stationarized" versions of these complex DGPs.

Third, we find that a conservative "rule-of-thumb" choice for the offsets discussed in Subsection 2.3 would be to use $\alpha = 0.025$ for a level 0.05 test. However, we have not developed an entirely data-driven approach to select offsets to account for this bias. Future work will theoretically investigate these offsets and develop a more data-driven approach for selecting offsets.

# 5 Discussion and Future Work

The greater purpose of our work is to improve the practice of epidemiological forecasting. In particular, the motivation for developing this conditional independence test was to create a large-scale testing framework for Granger causality analysis of epidemiological signals while avoiding unrealistic assumptions of independence, stationarity, linearity, and Gaussianity. We were originally interested in determining *when* specific quantiles of the viral load distribution of a city contain auxiliary information about future case growth rates of COVID-19 at different forecasting horizons.

Based on the initial results reported in Section 3, which only uses data from one city, it is clear that we will need to group together signals from multiple nearby cities to gain enough power to say more about the dynamics of viral load and future case growth rates of COVID-19. Since our dataset consists of dozens of signals for more than 1,000 cities, this will require much more computation and the justification of many context-dependent decisions for the hierarchical testing procedure described in Section C. Hence, we separate this part of our work into a companion paper written for epidemiological forecasting practitioners that are not necessarily interested in all of the details about the underlying time series theory.

There are several exciting avenues for future work. First, as mentioned above, we will use our testing framework to conduct a large-scale Granger causality analysis of non-stationary nonlinear epidemiological signals from many cities. Second, we will develop a causal discovery framework for high-dimensional non-stationary time series based on our testing framework. Third, we will develop a sample splitting-based methodology for non-stationary time series so that we can infer expected conditional covariance curves and other functionals that are of interest in the emerging field of causal inference for time series.

## References


[AM12] Pierre-Olivier Amblard and Olivier J. J. Michel. "The Relation between Granger Causality and Directed Information Theory: A Review". In: *Entropy* (2012).

[AB12] D. W. K. Andrews and P. J. Barwick. "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure". In: *Econometrica* (2012).

[AG09] D. W. K. Andrews and P. Guggenberger. "Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities". In: *Econometrica* (2009).

[AS10] D. W. K. Andrews and G. Soares. "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection". In: *Econometrica* (2010).

[BW23] Lujia Bai and Weichi Wu. "Time-varying correlation network analysis of non-stationary multivariate time series with complex trends". In: (2023).

[BR23] Sumanta Basu and Suhasini Subba Rao. "Graphical models for nonstationary time series". In: *The Annals of Statistics* (2023).

[BRT12] Taoufik Bouezmarni, Jeroen VK Rombouts, and Abderrahim Taamouti. "Nonparametric copula-based test for conditional independence with applications to Granger causality". In: *Journal of Business and Economic Statistics* (2012).

[Bru22] Guy-Niklas Brunotte. "A test of independence under local stationarity based on the local characteristic function". In: (2022).

[Cha82] Gary Chamberlain. "The general equivalence of Granger and Sims causality". In: *Econometrica* (1982).

[CCK13] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors". In: *The Annals of Statistics* 41.6 (2013), pp. 2786–2819.

[CKS21] Victor Chernozhukov, Hiroyuki Kasahara, and Paul Schrimpf. "Causal impact of masks, policies, behavior on early covid-19 pandemic in the US". In: *Journal of Econometrics* (2021).

[CPH22] Alexander Mangulad Christgau, Lasse Petersen, and Niels Richard Hansen. "Nonparametric Conditional Local Independence Testing". In: *arXiv preprint arXiv:2203.13559* (2022).

[Con12] Nonparametric Copula-Based Test for Conditional Independence with Applications to Granger Causality. "Taoufik Bouezmarni and Jeroen V.K. Rombouts and Abderrahim Taamouti". In: *Journal of Business Economic Statistics* (2012).

[Dah97] Rainer Dahlhaus. "Fitting time series models to nonstationary processes". In: *The Annals of Statistics* (1997).

[Dah00] Rainer Dahlhaus. "Graphical interaction models for multivariate time series". In: *Metrika* (2000).

[Dau80] J. J. Daudin. "Partial association measures and an application to qualitative regression". In: *Biometrika* 67.3 (1980), pp. 581–590.

[Daw79] Philip A. Dawid. "Conditional independence in statistical theory". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 41.1 (1979), pp. 1–15.

[DW21] Holger Dette and Weichi Wu. "Confidence surfaces for the mean of locally stationary functional time series". In: (2021).

[DW19] Holger Dette and Weichi Wu. "Detecting relevant changes in the mean of nonstationary processes—A mass excess approach". In: *The Annals of Statistics* (2019).

[DZ20] Xiucai Ding and Zhou Zhou. "Estimation and inference for precision matrices of nonstationary time series". In: *The Annals of Statistics* (2020).

[DZ21] Xiucai Ding and Zhou Zhou. "Simultaneous Sieve Inference for Time-Inhomogeneous Nonlinear Time Series Regression". In: (2021).

[Don+23] Xinshuai Dong et al. “On the Three Demons in Causality in Finance: Time Resolution, Nonstationarity, and Latent Factors”. In: (2023).

[Eic12] Michael Eichler. “Graphical modelling of multivariate time series”. In: *Probability Theory Related Fields* (2012).

[FNS15] Seth R. Flaxman, Daniel B. Neill, and Alexander J. Smola. “Gaussian processes for independence tests with non-iid data in causal inference”. In: *ACM Transactions on Intelligent Systems and Technology* 7.2 (2015), pp. 1–23.

[FM82] J. P. Florens and M. Mouchart. “A note on noncausality”. In: *Econometrica* (1982).

[GWS23] Hanjia Gao, Runmin Wang, and Xiaofeng Shao. “Dimension-agnostic Change Point Detection”. In: (2023).

[GF12] Jelle J. Goeman and Livio Finos. “The inheritance procedure: multiple testing of tree-structured hypotheses”. In: *Statistical applications in genetics and molecular biology* (2012).

[GS10] Jelle J. Goeman and Aldo Solari. “The sequential rejection principle of familywise error control”. In: *The Annals of Statistics* (2010).

[Hay+21] James Hay et al. “Estimating epidemiologic dynamics from cross-sectional viral load distributions”. In: *Science* (2021).

[Hoc+23] Tom Hochsprung et al. “Increasing effect sizes of pairwise conditional independence tests between random vectors”. In: *Uncertainty in Artificial Intelligence* (2023).

[Hua+20] Biwei Huang et al. “Causal discovery from heterogeneous/nonstationary data”. In: *Journal of Machine Learning Research* (2020).

[IM04] G. W. Imbens and C. F. Manski. “Confidence Intervals for Partially Identified Parameters”. In: *Econometrica* (2004).

[Kas18] Maximilian Kasy. “Uniformity and the delta method”. In: *Journal of Econometric Methods* (2018).

[KF09] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[KKK22] Arun K. Kuchibhotla, John E. Kolassa, and Todd A. Kuffner. “Post-selection inference”. In: *Annual Review of Statistics and Its Application* (2022).

[KBW23] Arun Kumar Kuchibhotla, Sivaraman Balakrishnan, and Larry Wasserman. “Median regularity and honest inference”. In: *Biometrika* (2023).

[KKS23] Daisuke Kurisu, Kengo Kato, and Xiaofeng Shao. “Gaussian approximation and spatially dependent wild bootstrap for high-dimensional spatial data”. In: *Journal of the American Statistical Association* (2023).

[LLZ22] Jia Li, Zhipeng Liao, and Wenyu Zhou. “A general test for functional inequalities”. In: (2022).

[Li89] Ker-Chau Li. “Honest confidence regions for nonparametric regression”. In: *The Annals of Statistics* (1989).

[LZ23] Miaoshiqi Liu and Zhou Zhou. “Simultaneous Nonparametric Inference of M-regression under Complex Temporal Dynamics”. In: (2023).

[Liu+23] Zhaolu Liu et al. “Kernel-based Joint Independence Tests for Multivariate Stationary and Nonstationary Time-Series”. In: *arXiv* (2023).

[Lun+22] Anton Rask Lundborg et al. “The Projected Covariance Measure for assumption-lean variable significance testing”. In: (2022).

[LSP22] Anton Rask Lundborg, Rajen D. Shah, and Jonas Peters. “Conditional independence testing in Hilbert spaces with applications to functional data analysis”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.5 (2022), pp. 1821–1850.

[LW23] Tianpai Luo and Weichi Wu. “Simultaneous inference for monotone and smoothly time varying functions under complex temporal dynamics”. In: (2023).

[MS19] Daniel Malinsky and Peter Spirtes. “Learning the structure of a nonstationary vector autoregression”. In: *The 22nd International Conference on Artificial Intelligence and Statistics* (2019).

[Mei08] Nicolai Meinshausen. “Hierarchical testing of variable importance”. In: *Biometrika* (2008).

[Mie23] Fabian Mies. “Functional estimation and change detection for nonstationary time series”. In: *Journal of the American Statistical Association* (2023).

[MS24] Fabian Mies and Ansgar Steland. “Projection inference for high-dimensional covariance matrices with structured shrinkage targets”. In: *Electronic Journal of Statistics* (2024).

[MS22] Fabian Mies and Ansgar Steland. “Sequential Gaussian approximation for nonstationary time series in high dimensions”. In: *arXiv preprint arXiv:2203.03237* (2022).

[MH21] Pablo Montero-Manso and Rob J. Hyndman. “Principles and algorithms for forecasting groups of time series: Locality and globality”. In: *International Journal of Forecasting* (2021).

[NBW21] Matey Neykov, Sivaraman Balakrishnan, and Larry Wasserman. “Minimax Optimal Conditional Independence Testing”. In: *The Annals of Statistics* (2021).

[Pea09] Judea Pearl. *Causality.* Cambridge University Press, 2009.

[PJS17] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms.* MIT Press, 2017.

[QKC15] Christopher J. Quinn, Negar Kiyavash, and Todd P. Coleman. “Directed Information Graphs”. In: *IEEE Transactions on Information Theory* (2015).

[Reb80] Rolando Rebolledo. “Central limit theorems for local martingales”. In: (1980).

[RWG19] Alessandro Rinaldo, Larry Wasserman, and Max G’Sell. “Bootstrapping and sample splitting for high-dimensional, assumption-lean inference”. In: *The Annals of Statistics* (2019).

[RS08] J. P. Romano and A. M. Shaikh. “Inference for Identifiable Parameters in Partially Identified Econometric Models”. In: *Journal of Statistical Planning and Inference* (2008).

[RSW14] J. P. Romano, A. M. Shaikh, and M. Wolf. “A Practical Two-step Method for Testing Moment Inequalities”. In: *Econometrica* (2014).

[RW05] Joseph P. Romano and Michael Wolf. “Exact and approximate stepdown methods for multiple hypothesis testing”. In: *Journal of the American Statistical Association* (2005).

[Ros61] Murray Rosenblatt. “Independence and dependence”. In: *Proc. 4th Berkeley sympos. math. statist. and prob* (1961).

[Run18] Jakob Runge. “Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information”. In: *International Conference on Artificial Intelligence and Statistics* (2018).

[Run+23] Jakob Runge et al. “Causal inference for time series”. In: *Nature Reviews Earth and Environment* (2023).

[Run+19a] Jakob Runge et al. “Detecting and quantifying causal associations in large nonlinear time series datasets”. In: *Science advances* (2019).

[Run+19b] Jakob Runge et al. “Inferring causation from time series in Earth system sciences”. In: *Nature communications* 10.1 (2019).

[SP11] Sohan Seth and Jose C. Principe. “Assessing Granger non-causality using nonparametric measure of conditional independence”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2011).

[SP20] Rajen D. Shah and Jonas Peters. “The hardness of conditional independence testing and the generalised covariance measure”. In: *Annals of Statistics* 48.3 (2020), pp. 1514–1538.

[SGS00] Peter Spirtes, Clark N. Glymour, and Richard Scheines. *Causation, Prediction, and Search.* MIT Press, 2000.

[SW07] Liangjun Su and Halbert White. “A consistent characteristic function-based test for conditional independence”. In: *Journal of Econometrics* (2007).

[Tib+18] Ryan J Tibshirani et al. "Uniform asymptotic inference and the bootstrap after model selection". In: *The Annals of Statistics* (2018).

[TM20] Michael R. Tom and Michael J. Mina. "To interpret the SARS-CoV-2 test, consider the cycle threshold value". In: *Clinical Infectious Diseases* (2020).

[WNR23a] Jonas Wahl, Urmi Ninad, and Jakob Runge. "Foundations of Causal Discovery on Groups of Variables". In: (2023).

[WNR23b] Jonas Wahl, Urmi Ninad, and Jakob Runge. "Vector causal inference between two groups of variables". In: *Proceedings of the AAAI Conference on Artificial Intelligence* (2023).

[WR23] Ian Waudby-Smith and Aaditya Ramdas. "Distribution-uniform anytime-valid inference". In: (2023).

[Wau+21] Ian Waudby-Smith et al. "Time-uniform central limit theory and asymptotic confidence sequences". In: (2021).

[Wie66] Norbert Wiener. "Nonlinear problems in random theory". In: (1966).

[Wu05] Wei Biao Wu. "Nonlinear system theory: Another look at dependence". In: *Proceedings of the National Academy of Sciences* 102.40 (2005), pp. 14150–14154.

[WX12] Wei Biao Wu and Han Xiao. "Covariance matrix estimation in time series". In: *Handbook of Statistics* 30 (2012), pp. 187–209.

[YN21] Kashif Yousuf and Serena Ng. "Boosting high dimensional predictive regressions with time varying parameters". In: *Journal of Econometrics* (2021).

[Zha+12] Kun Zhang et al. "Kernel-based conditional independence test and application in causal discovery". In: (2012).

[ZJ20] Lu Zhang and Lucas Janson. "Floodgate: inference for model-free variable importance". In: (2020).

[ZW15] Ting Zhang and Wei Biao Wu. "Time-varying nonlinear regression models: nonparametric estimation and model selection". In: *Annals of Statistics* (2015), pp. 741–768.

[ZS24] Yi Zhang and Xiaofeng Shao. "Another look at bandwidth-free inference: a sample splitting approach". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* (2024).

[Zho13] Zhou Zhou. "Heteroscedasticity and Autocorrelation Robust Structural Change Detection". In: *Journal of the American Statistical Association* (2013).

[ZW09] Zhou Zhou and Wei Biao Wu. "Local linear quantile estimation for nonstationary time series". In: *The Annals of Statistics* (2009).

# A Additional Discussions

## A.1 Literature review of distribution-uniform inference

First, we discuss the conditional independence testing literature. There has been a lot of recent work on distribution-uniform conditional independence testing frameworks because of the hardness result and subsequent testing framework developed by Shah and Peters [SP20]. For instance, Lundborg, Shah, and Peters [LSP22] introduced many distribution-uniform convergence results for separable Banach and Hilbert spaces. Recently, Christgau, Petersen, and Hansen [CPH22] introduced a distribution-uniform "conditional local independence" testing framework for the setting where $n$ realizations of a point process are observed. Christgau, Petersen, and Hansen [CPH22] also introduce a distribution-uniform extension of Rebolledo's martingale central limit theorem [Reb80] and extend many distribution-uniform convergence results from Lundborg, Shah, and Peters [LSP22] to metric spaces.

Second, we discuss relevant developments in the anytime-valid inference literature. Recently, Waudby-Smith and Ramdas [WR23] introduced a distribution-uniform strong (almost-sure) Gaussian approximation for the full sum of iid random variables, which appears to be the first such result in the literature. The work in Waudby-Smith and Ramdas [WR23] is motivated by prior work on asymptotic anytime-valid inference from Waudby-Smith et al. [Wau+21], in which the authors defined the concept of an "asymptotic confidence sequence" (AsympCS). Moreover, Waudby-Smith et al. [Wau+21] introduced an AsympCS for iid random variables and a Lindeberg-type AsympCS which can capture time-varying means under martingale dependence. To briefly compare, the result from Waudby-Smith and Ramdas [WR23] is a strong (almost-sure) Gaussian approximation for the full sum of iid random variables in a sequential setting, whereas our strong (in probability) Gaussian approximation is for the max of partial sums of a non-stationary high-dimensional random vector in a fixed-$n$ setting.

Third, we mention other areas in which distribution-uniform inference is studied under different names. There is a vast literature discussing the importance of distribution-uniform inference under the name of "honest" or "uniform" inference, see Li [Li89], Kasy [Kas18], Tibshirani et al. [Tib+18], Rinaldo, Wasserman, and G'Sell [RWG19], and Kuchibhotla, Balakrishnan, and Wasserman [KBW23]. Also, there is a plethora of literature on distribution-uniform moment inequality testing Imbens and Manski [IM04], Romano and Shaikh [RS08], Andrews and Guggenberger [AG09], Andrews and Soares [AS10], Andrews and Barwick [AB12], and Romano, Shaikh, and Wolf [RSW14]. Most recently, Li, Liao, and Zhou [LLZ22] developed a distribution-uniform test for general functional inequalities which admits conditional moment inequalities as a special case. Also, in the supplemental appendix of Li, Liao, and Zhou [LLZ22] introduce a distribution-uniform strong Gaussian approximation for the full sum of a high-dimensional mixingale.

# B Hypothesis Testing and Conditional Independence Testing

## B.1 Hypothesis testing for locally stationary time series

In this subsection, we introduce notation for distribution-uniform simultaneous hypothesis testing for high-dimensional non-stationary time series. For each $n \in \mathbb{N}$, $m \in \mathcal{M}_n$, $u \in \mathcal{U}_n$, and $t \in \mathcal{T}_n$, define a potentially composite null hypothesis $\mathcal{P}_{0,n,m,u,t} \subset \mathcal{P}_n$. We denote the global null hypothesis for the family of null hypotheses $(\mathcal{P}_{0,n,m,u,t})_{m\in\mathcal{M}_n,u\in\mathcal{U}_n,t\in\mathcal{T}_n}$ by

$$\mathcal{P}_{0,n} = \bigcap_{t\in\mathcal{T}_n} \bigcap_{u\in\mathcal{U}_n} \bigcap_{m\in\mathcal{M}_n} \mathcal{P}_{0,n,m,u,t}.$$

For each $n \in \mathbb{N}$, $m \in \mathcal{M}_n$, $u \in \mathcal{U}_n$, and $t \in \mathcal{T}_n$, let $\psi_{n,m,u,t}$ be a potentially randomized test that can be applied to the data such that

$$\psi_{n,m,u,t} : \mathbb{R}^{(d_X+d_Y+d_Y)\cdot T_n} \times [0,1] \to \{0,1\}$$

is a measurable function where 1 indicates rejecting the null hypothesis $H_{0,n,m,u,t}$ and where $T_n = |\mathcal{T}_n|$. The last argument is for a uniform random variable $U_{n,m,u,t} \sim U[0,1]$ that is independent of the data. Note that the joint distribution of the $(U_{n,m,u,t})_{m\in\mathcal{M}_n,u\in\mathcal{U}_n,t\in\mathcal{T}_n}$ can be very complicated. For some $n \in \mathbb{N}$ and $P \in \mathcal{P}_n$, denote the subset of indices at rescaled time $u \in \mathcal{U}_n$ and inputs $t \in \mathcal{T}_n$ in which the null hypothesis is true by

$$\tilde{\mathcal{M}}_{P,n}(u,t) = \{m \in \mathcal{M}_n : H_{0,n,m,u,t} \text{ is true under } P\}.$$

For some sample size $n \in \mathbb{N}$, level $\alpha \in (0,1)$, global null hypothesis $\mathcal{P}_{0,n}$, and collection of distributions $\mathcal{P}_n$, we say that the family of tests $(\psi_{n,m,u,t})_{m\in\mathcal{M}_n,u\in\mathcal{U}_n,t\in\mathcal{T}_n}$ has *valid FWER control in the weak sense at sample size* $n$ if

$$\sup_{P\in\mathcal{P}_{0,n}} \mathbb{P}_P\left(\bigcup_{t\in\mathcal{T}_n}\bigcup_{u\in\mathcal{U}_n}\bigcup_{m\in\mathcal{M}_n}\{\psi_{n,m,u,t}=1\}\right) \le \alpha,$$

and *valid FWER control in the strong sense at sample size* $n$ if

$$\sup_{P\in\mathcal{P}_{n}} \mathbb{P}_P\left(\bigcup_{t\in\mathcal{T}_n}\bigcup_{u\in\mathcal{U}_n}\bigcup_{m\in\tilde{\mathcal{M}}_{P,n}(u,t)}\{\psi_{n,m,u,t}=1\}\right) \le \alpha.$$

For some level $\alpha \in (0,1)$, sequence of global null hypotheses $(\mathcal{P}_{0,n})_{n\in\mathbb{N}}$, and sequence of collections of distributions $(\mathcal{P}_n)_{n\in\mathbb{N}}$, we say that the sequence of families of tests $((\psi_{n,m,u,t})_{m\in\mathcal{M}_n,u\in\mathcal{U}_n,t\in\mathcal{T}_n})_{n\in\mathbb{N}}$ has *uniformly asymptotic FWER control in the weak sense* if

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_{0,n}} \mathbb{P}_P\left(\bigcup_{t\in\mathcal{T}_n}\bigcup_{u\in\mathcal{U}_n}\bigcup_{m\in\mathcal{M}_n}\{\psi_{n,m,u,t}=1\}\right) \le \alpha,$$

and *uniformly asymptotic FWER control in the strong sense* if

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_{n}} \mathbb{P}_P\left(\bigcup_{t\in\mathcal{T}_n}\bigcup_{u\in\mathcal{U}_n}\bigcup_{m\in\tilde{\mathcal{M}}_{P,n}(u,t)}\{\psi_{n,m,u,t}=1\}\right) \le \alpha.$$

Let us briefly reflect on these definitions. FWER control in the weak sense is only useful for tests against the global null hypothesis, not for simultaneous testing of multiple hypotheses. Ideally, we would like to have FWER control in the strong sense. We would like to have valid FWER control at sample size $n$, but in many situations asymptotic FWER control will suffice. Thus, in many practical situations it may be reasonable to only have uniformly asymptotic FWER control in the strong sense.

Turning to the subject of conditional independence testing, consider the formulation of the null hypotheses of conditional independence for some $t \in \mathcal{T}_n$ and $m \in \mathcal{M}_n$ as

$$H^{\mathrm{CI}}_{0,n,m,\frac{t}{n},t} : X_{t,n,i,a} \perp\!\!\!\perp Y_{t,n,j,b} \mid \boldsymbol{Z}_{t,n}$$

where $X_{t,n,i,a} = \tilde{X}_{t,n,i,a}(t/n)$, $Y_{t,n,j,b} = \tilde{Y}_{t,n,j,b}(t/n)$, and $\boldsymbol{Z}_{t,n} = \tilde{\boldsymbol{Z}}_{t,n}(t/n)$. We can write then write global null hypothesis of conditional independence, with respect to the family of null hypotheses $(H^{\mathrm{CI}}_{0,n,m,\frac{t}{n},t})_{t\in\mathcal{T}_n,m\in\mathcal{M}_n}$, as

$$H^{\mathrm{CI}}_{0,n} : X_{t,n,i,a} \perp\!\!\!\perp Y_{t,n,j,b} \mid \boldsymbol{Z}_{t,n} \text{ for all } t \in \mathcal{T}_n, m \in \mathcal{M}_n.$$

Denote by $(\mathcal{P}^{\mathrm{CI}}_{0,n})_{n\in\mathbb{N}}$ the sequence of global null hypotheses such that for each $n \in \mathbb{N}$, $\mathcal{P}^{\mathrm{CI}}_{0,n}$ consists of a collection of distributions such that $H^{\mathrm{CI}}_{0,n}$ is true. In this work, we show how to construct a sequence of families of tests $((\psi^{\mathrm{CI}}_{n,m,\frac{t}{n},t})_{m\in\mathcal{M}_n,t\in\mathcal{T}_n})_{n\in\mathbb{N}}$ with uniformly asymptotic FWER control in the *weak* sense such that

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}^{\mathrm{CI}}_{0,n}} \mathbb{P}_P\left(\bigcup_{t\in\mathcal{T}_n}\bigcup_{m\in\mathcal{M}_n}\{\psi^{\mathrm{CI}}_{n,m,\frac{t}{n},t}=1\}\right) \le \alpha$$

for some level $\alpha \in (0,1)$. In future work, we will introduce a sample splitting procedure so that we can attack the problem of actual (i.e. strong) familywise error rate control.

## B.2 Remarks on the hypothesis and test statistic

We formulate the null hypothesis of conditional independence (1) in this way because it makes explicit how conditional dependencies could arise among the different leads and lags of each dimension at different points in time. In other words, this formulation emphasizes the *dynamic flow of information* between two time series (at different leads and lags of each dimension) conditional on another time series. To put this statements into context, let us consider different ways that we could have formulated the null hypothesis of conditional independence and discuss why we opted for the formulation (1). Similar to the discussion in Hochsprung et al. [Hoc+23], for each $n \in \mathbb{N}$, the null hypothesis that

$$\boldsymbol{X}_{t,n} \perp\!\!\!\perp \boldsymbol{Y}_{t,n} \mid \boldsymbol{Z}_{t,n} \text{ for all } t \in \mathcal{T}_n \tag{7}$$

implies our formulation of the null hypothesis (1), where the vector processes with all time-offsets of each dimension $\boldsymbol{X}_{t,n} = \tilde{\boldsymbol{X}}_{t,n}(t/n)$, $\boldsymbol{Y}_{t,n} = \tilde{\boldsymbol{Y}}_{t,n}(t/n)$, $\boldsymbol{Z}_{t,n} = \tilde{\boldsymbol{Z}}_{t,n}(t/n)$ were defined in Subsection 1.1. Also, for each $n \in \mathbb{N}$, the null hypothesis

$$\boldsymbol{X}_n \perp\!\!\!\perp \boldsymbol{Y}_n \mid \boldsymbol{Z}_n \tag{8}$$

implies both (1) and (7), where $\boldsymbol{X}_n = (\boldsymbol{X}_{t,n})_{t \in \mathcal{T}_n}$, $\boldsymbol{Y}_n = (\boldsymbol{Y}_{t,n})_{t \in \mathcal{T}_n}$, $\boldsymbol{Z}_n = (\boldsymbol{Z}_{t,n})_{t \in \mathcal{T}_n}$. Hence, to test for (7) or (8), one can test for the hypothesis (1) as we do and then reject (7) and (8) if and only if we reject (1). By the same arguments of Lemma 1 from Hochsprung et al. [Hoc+23], these induced tests for (7) and (8) both have valid level $\alpha$ if the test for the hypothesis (1) has valid level $\alpha$.

Lastly, let us briefly reflect on the assumption of stationarity, which we completely avoid. If one is willing to assume stationarity, then it suffices to use a very similar test statistic to that of Theorem 9 of Shah and Peters [SP20] with suitable (i.e. for stationary time series) Gaussian approximations, regression estimators, and covariance matrix estimators. Moreover, under the assumption of stationarity, one could replace our alternative hypothesis (2) with $X_{t,n,i,a} \not\perp\!\!\!\perp Y_{t,n,j,b} \mid \boldsymbol{Z}_{t,n}$ for all $t \in \mathcal{T}_n$ for at least one $m \in \mathcal{M}_n$ since the conditional independence relationships would not change over time. However, we completely avoid the assumption of stationarity because it is unrealistic in many practical settings when dealing with time series data.

Let us briefly reflect on our choice to use these test statistics from Subsection 2.3 to test for the global null hypothesis of conditional independence (1). We could have also formulated the global null hypothesis directly in terms of the causal representations

$$\tilde{X}_{t,n,i,a}(u) \perp\!\!\!\perp \tilde{Y}_{t,n,j,b}(u) \mid \tilde{\boldsymbol{Z}}_{t,n}(u) \text{ for all } u \in \mathcal{U}_n, t \in \mathcal{T}_n, m \in \mathcal{M}_n \tag{9}$$

for each $n \in \mathbb{N}$, which also implies our formulation (1) since $X_{t,n,i,a} = \tilde{X}_{t,n,i,a}(t/n)$, $Y_{t,n,j,b} = \tilde{Y}_{t,n,j,b}(t/n)$, and $\boldsymbol{Z}_{t,n} = \tilde{\boldsymbol{Z}}_{t,n}(t/n)$. As noted in the previous discussion about (7) and (8), to test for (9), one can test for the hypothesis (1) and then reject (9) if and only if we reject (1). By the same arguments of Lemma 1 from Hochsprung et al. [Hoc+23], this induced test for (9) will have valid level $\alpha$ if the test for the hypothesis (1) has valid level $\alpha$. To test for (9) so that we consider *all* the rescaled times $\mathcal{U}_n$ and not just $t/n$ for $t \in \mathcal{T}_n$, we could have, for example, formulated the test statistic as

$$\max_{m \in \mathcal{M}_n} \sup_{u \in \mathcal{U}_n} \left| \frac{1}{\sqrt{T_n h_n}} \sum_{t \in \mathcal{T}_n} K\left(\frac{t/n - u}{h_n}\right) \hat{R}_{t,n,m} \right| \Big/ \hat{\Sigma}^R_{n,m}(u)$$

where $K$ is a kernel function, $h_n$ is a bandwidth parameter, and $\hat{\Sigma}^R_{n,m}(u)$ is an estimate of the local long-run variances $\Sigma^R_{P,n,m}(u)$ of the process of error products.

However, there are several reasons to prefer our test statistics even if we were to formulate the hypothesis as (9) directly in terms of the causal representations. First, to use this test statistic, we would require the stronger stochastic Lipschitz Assumption 5 instead of the more general total variation Assumption 4. Second, we would require a time-uniformly and distribution-uniformly consistent estimator $\hat{\Sigma}^R_{n,m}(u)$ of the local long-run variances $\Sigma^R_{P,n,m}(u)$ of the process of error products, which requires choosing a bandwidth parameter. For instance, one could develop a distribution-uniform extension of the estimator from Theorem 4.4 of Dette and Wu [DW19]. Third, this test statistic would require kernel smoothing and thus would requires slightly faster convergence rates for the regression functions because of *another* bandwidth that must be chosen. In comparison, our test statistics allow for both abrupt and smooth non-stationarity, do not require an estimate of each of the local long run variances, do not require kernel smoothing, and circumvent choosing *two* bandwidth parameters.

Lastly, let us briefly reflect on another approach to inferring the time-varying conditional independence structure that was deliberately not taken in this paper. Along the lines of Theorem 8 from Shah and Peters [SP20], it is also possible to test for conditional independence by creating simultaneous confidence bands (SCBs) for the expected conditional covariance curves and determining when zero is included in the SCBs. A very similar approach to that taken in Bai and Wu [BW23] for inferring time-varying correlation networks can be used. If zero is not included in the confidence interval for the local expected conditional covariance at a particular point in time, then we can reject the corresponding local null hypothesis of conditional independence for that dimension/time-offset index at that point in time. However, this approach would either require multiple independent realizations of the time series or a sophisticated sample splitting procedure for non-stationary time series given only one realization. See Kuchibhotla, Kolassa, and Kuffner [KKK22] for a recent review of post-selection inference, and see Shah and Peters [SP20], Lundborg, Shah, and Peters [LSP22], and Christgau, Petersen, and Hansen [CPH22] for discussions of sample splitting in the context of conditional independence testing. We do not discuss any kind of sample splitting or consider the possibility of having access to multiple independent realizations of the time series so that this manuscript can focus on what can be done given only *one realization* of a non-stationary time series. We will introduce our sample splitting procedure in future work.

# C Simultaneous Testing

In this section, we will describe two approaches for simultaneous testing. In Subsection C.1, we introduce a stepdown procedure based on Romano and Wolf [RW05]. In Subsection C.2, we describe an inheritance procedure based on Goeman and Finos [GF12]. We apply the inheritance procedure in the companion paper.

## C.1 Stepdown procedure

We discuss how to simultaneously test for the hypotheses

$$H_{0,n,m}^{\mathrm{CI},w} : X_{t,n,i,a} \perp\!\!\!\perp Y_{t,n,j,b} \mid \boldsymbol{Z}_{t,n} \text{ for all } t \in \mathcal{T}_n^w$$
$$H_{1,n,m}^{\mathrm{CI},w} : X_{t,n,i,a} \not\perp\!\!\!\perp Y_{t,n,j,b} \mid \boldsymbol{Z}_{t,n} \text{ for at least one } t \in \mathcal{T}_n^w$$

for each $n \in \mathbb{N}$, for *each* pair $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$ of dimension/time-offset indices and window numbers where $\mathcal{W}_n = \{1, \ldots, W_n\}$ for some $W_n \in \mathbb{N}$. Let $(\mathcal{U}_n^w)_{w \in \mathcal{W}_n}$ be time windows consisting of possibly overlapping sub-intervals of $\mathcal{U}_n$ such that we have $\mathcal{U}_n = \bigcup_{w \in \mathcal{W}_n} \mathcal{U}_n^w$. Similarly, let $\mathcal{T}_n^w = \mathcal{T}_n \cap \mathcal{U}_n^w$ be subsets of $\mathcal{T}_n$ consisting of consecutive positive integers contained in the corresponding time windows $\mathcal{U}_n^w$ such that we have $\mathcal{T}_n = \bigcup_{w \in \mathcal{W}_n} \mathcal{T}_n^w$. Denote the cardinality $T_n^w := |\mathcal{T}_n^w|$ for some $w \in \mathcal{W}_n$. Mainly for the sake of convenience, we will require that $T_n^{w_1} = T_n^{w_2}$ for all $w_1, w_2 \in \mathcal{W}_n$. In the simplest case, we could choose $W_n = 1$ for all $n \in \mathbb{N}$ and $m \in \mathcal{M}_n$ so that $\mathcal{U}_n^w = \mathcal{U}_n$. However, it is often of interest to study how conditional independence relationships evolve over time.

Several remarks are in order. First, note that the time-varying regression functions for each $m \in \mathcal{M}_n$ are estimated using the data at *all* of the observed times. Thus, to be abundantly clear, choosing more windows does not affect the amount of data we have to estimate the time-varying regression functions. Second, the number of windows can grow with the sample size $n \in \mathbb{N}$. Third, note that there is a balance between choosing the time windows $\mathcal{U}_n^w$ to be large enough so that each test has enough power to detect conditional dependence, but small enough so that each time window is of scientific interest. In Subsection C.2, we discuss how to overcome this dilemma using the inheritance procedure.

For each pair $(m, w) \in \mathcal{M}_n \times \mathcal{W}_n$, denote

$$\mathcal{P}_{0,n,m}^{\mathrm{CI},w} = \bigcap_{t \in \mathcal{T}_n^w} \mathcal{P}_{0,n,m,\frac{t}{n},t}^{\mathrm{CI}}$$

where $\mathcal{P}_{0,n,m,\frac{t}{n},t}^{\mathrm{CI}} \subset \mathcal{P}_n$ is a collection of distributions such that $X_{t,n,i,a} \perp\!\!\!\perp Y_{t,n,j,b} \mid \boldsymbol{Z}_{t,n}$. That is, each

$\mathcal{P}_{0,n,m}^{\mathrm{CI},w}$ is a collection of distributions for which conditional independence holds for all rescaled times $t/n$ for $t \in \mathcal{T}_n^w$ in window $w \in \mathcal{W}_n$ for index $m \in \mathcal{M}_n$.

Let $\mathcal{M}_n^* \times \mathcal{W}_n^* \subset \mathcal{M}_n \times \mathcal{W}_n$ be a subset of pairs of indices and time windows. Consider the intersection of null hypotheses of the form

$$\mathcal{P}_{0,n}^{\mathrm{CI}}(\mathcal{M}_n^* \times \mathcal{W}_n^*) = \bigcap_{(m,w)\in\mathcal{M}_n^*\times\mathcal{W}_n^*} \mathcal{P}_{0,n,m}^{\mathrm{CI},w}.$$

In words, $\mathcal{P}_{0,n}^{\mathrm{CI}}(\mathcal{M}_n^* \times \mathcal{W}_n^*)$ is a collection of distributions such that conditional independence always holds during *certain* time windows for *particular* indices.

Inspired by Chernozhukov, Chetverikov, and Kato [CCK13] and Kurisu, Kato, and Shao [KKS23], we combine our bootstrap procedure with the stepdown procedure for strong FWER control from Romano and Wolf [RW05]. We will show that for some $\alpha \in (0,1)$ and sequence of collections of null hypotheses $(\mathcal{P}_{0,n}^{\mathrm{CI}}(\mathcal{M}_n^* \times \mathcal{W}_n^*))_{n\in\mathbb{N}}$ our testing procedure has the following uniformly asymptotic FWER control

$$\limsup_{n\to\infty} \sup_{\mathcal{M}_n^*\times\mathcal{W}_n^*\subset\mathcal{M}_n\times\mathcal{W}_n} \sup_{P\in\mathcal{P}_{0,n}^{\mathrm{CI}}(\mathcal{M}_n^*\times\mathcal{W}_n^*)} \mathbb{P}_P\left(\bigcup_{(m,w)\in\mathcal{M}_n^*\times\mathcal{W}_n^*} \{\text{reject } H_{0,n,m}^{\mathrm{CI},w}\}\right) \leq \alpha.$$

Now, we will explain the bootstrap stepdown procedure. The key idea is stacking the times from all windows so that each window is treated as a separate index. This is equivalent to considering a time-lead equal to the start of each window as another index. Hence, instead of considering the max over the times $\mathcal{T}_n$ as in Subsection 2.3, we take the max over the times $1,\ldots,T_n^w$ corresponding to the window size.

For each pair $(m,w) \in \mathcal{M}_n \times \mathcal{W}_n$ define

$$S_{n,m,w}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n^w}) = \max_{s=1,\ldots,T_n^w} \left| \frac{1}{\sqrt{T_n}} \sum_{t\leq s} \hat{R}_{t,n,m} \right|.$$

At the first step $\ell = 1$, set $\mathcal{M}_n^{(\ell)} = \mathcal{M}_n$ and $\mathcal{W}_n^{(\ell)} = \mathcal{W}_n$. At each step $\ell$, we reject any of the hypotheses $H_{0,n,m}^{\mathrm{CI},w}$ in which the corresponding $S_{n,m,w}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n^w})$ is greater than the $(1-\alpha)$ quantile of $S_{n,m,w,\infty}^{(\ell)}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n})$ where

$$S_{n,m,w,\infty}^{(\ell)}((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) = \max_{(m,w)\in\mathcal{M}_n^{(\ell)}\times\mathcal{W}_n^{(\ell)}} \max_{s=1,\ldots,T_n^w} \left| \frac{1}{\sqrt{T_n}} \sum_{t\leq s} \hat{R}_{t,n,m} \right|.$$

If we did not reject any hypotheses, stop. If we did reject at least one hypothesis, continue. At the next step $\ell$, denote $(\mathcal{M}_n \times \mathcal{W}_n)^{(\ell)} \subset \mathcal{M}_n \times \mathcal{W}_n$ by the subset of pairs of indices and window numbers in which we did not reject the corresponding hypotheses $H_{0,n,m}^{\mathrm{CI},w}$. Repeat the previous bootstrap procedure until no additional hypotheses are rejected at a given step, or until all hypotheses have been rejected. Let $(\mathcal{M}_n \times \mathcal{W}_n)^{\mathrm{reject}} \subset \mathcal{M}_n \times \mathcal{W}_n$ be the subset of pairs indices and window numbers in which the corresponding hypotheses $H_{0,n,m}^{\mathrm{CI},w}$ were rejected. Similarly, let $(\mathcal{M}_n \times \mathcal{W}_n)^{\mathrm{retain}} \subset \mathcal{M}_n \times \mathcal{W}_n$ be the subset of pairs of indices and window numbers in which the corresponding hypotheses $H_{0,n,m}^{\mathrm{CI},w}$ were not rejected.

Let us state the conclusions of our simultaneous testing procedure would be. For each particular pair $(m,w) \in (\mathcal{M}_n \times \mathcal{W}_n)^{\mathrm{retain}}$, we fail to reject the null hypothesis that conditional independence holds at all rescaled times $t/n$ for $t \in \mathcal{T}_n^w$ for window number $w$ for index $m$. Analogously, for each particular pair $(m,w) \in (\mathcal{M}_n \times \mathcal{W}_n)^{\mathrm{reject}}$, we reject the null hypothesis that conditional independence holds at all rescaled times $t/n$ for $t \in \mathcal{T}_n^w$ for the window number and index pair $(m,w)$. That is, we conclude that conditional dependence holds for at least one rescaled time $t/n$ for $t \in \mathcal{T}_n^w$ for the window number and index pair $(m,w)$. However, the *exact* dynamics for how the conditional independence relationships evolve during this time window are still hidden to us. This is the price we pay for making few assumptions (e.g. without assumptions of linearity, Gaussianity, etc) when we only have access to one realization of a non-stationary time series. In future work, we will return to the topic of inferring the exact time-varying conditional independence structure of a non-stationary time series given just one realization.

## C.2 Inheritance procedure

The procedure described in Subsection C.1 requires users to preselect the time windows and to separately test for each of the dimension/time-offset indices. This approach may be satisfactory in many settings. However, several problems can arise in noisy, high-dimensional settings. We will focus on just two. First, it is challenging for a practitioner to know how to choose the time windows to be large enough so that the test has enough power, but small enough so the conclusions are scientifically interesting. Second, it may be difficult to test for conditional independence for all signals at once due to the inherent multiplicity of the problem in high dimensions. We propose using the inheritance procedure of Goeman and Finos [GF12], which is based on and the hierarchical testing procedure from Meinshausen [Mei08] and the sequential rejection principle from Goeman and Solari [GS10].

To deal with first problem of selecting windows, we introduce a procedure for automatically selecting time windows at appropriate temporal resolutions while controlling the familywise error rate. The main idea is to utilize the temporal hierarchy of the hypotheses when choosing the time windows from Subsection C.1 as follows. We will start by testing whether conditional independence holds at all times or not at the *coarsest* resolution - that is, by choosing the first time window to be $\mathcal{U}_n$ itself. If we detect conditional dependencies at some point in time during this time window, we will then split the window and conduct the test on the first half and on the second half of the window. The procedure will continue attempting to identify finer and finer time windows during which conditional independence does not always hold by successively splitting time windows in which the respective hypothesis of conditional independence was rejected.

To deal with second problem of testing multiplicity, we again propose starting at the coarsest resolution of the signals. The procedure will utilize a given hierarchical grouping structure of signals for which conditional independence holds during a particular time window. The procedure will continue trying to identify smaller sub-groups in the given hierarchy of signals down to the level of individual signals. This solution naturally encourages using groups of time series that are highly correlated and related to the same underlying phenomena. For instance, consider the setting in which we have a hierarchical grouping of cities (e.g. at the country, state, and county levels) or a sensor network corresponding to different regions in the brain or the Pacific Ocean. We might not care about identifying the *particular* city or sensor for which conditional dependencies were discovered, only that there were *some* conditional dependencies discovered by the group of time series. Hence, ad hoc dimensionality reduction and spatial averaging techniques commonly used in neuroscience and climate science to deal with the problem of high-dimensionality can be completely avoided using our framework. To be clear, we envision that the hierarchy of groups of signals will be given by the expert based on domain knowledge. Hierarchical clustering approaches for non-stationary time series could also be explored using *extreme* caution. See Montero-Manso and Hyndman [MH21] and Wahl, Ninad, and Runge [WNR23a; WNR23b] for more information about groups of time series and frameworks for conditional independence testing, causal discovery, and causal inference within this paradigm.

Combining these ideas, we are left with a trade-off between having a higher resolution for signals and time windows. The exact tree structure of the hypotheses is to be determined by the expert. This requires justifying highly context-dependent decisions. As discussed previously, we will apply this inheritance procedure for simultaneous testing in a companion paper so that this manuscript can be focused on the testing framework. The companion paper will describe in detail how the tree of hypotheses is constructed based on epidemiological domain knowledge. Our preliminary analysis in Section 3 suggests that it is necessary to group signals together so that we have enough power to detect time-varying conditional dependencies and gain insight about the dynamics of viral load and future case growth rates of COVID-19.

# D Theoretical Properties

## D.1 Assumptions

For the sake of brevity, going forward we will denote the set of well-defined tuples of processes, dimensions, and time-offsets by

$$(W, l, d) \in \mathbb{W} = \{(X, i, a), (Y, j, b), (Z, k, c)\}$$

and similarly denote the set of well-defined tuples of error processes, dimensions, and time-offsets by

$$(e, l, d) \in \mathbb{E} = \{(\varepsilon, i, a), (\xi, j, b)\}.$$

The assumptions that follow can be considered *distribution-uniform* extensions of the assumptions for high-dimensional nonlinear locally stationary time series described by the functional dependence measure as introduced in Mies and Steland [MS22]. First, we introduce the framework for quantifying temporal dependence via the functional dependence measure of Wu [Wu05].

**Definition 2** (Functional dependence measure). *For any tuple $(\phi, l, d) \in \mathbb{W} \cup \mathbb{E}$ let $(\tilde{\eta}_t^{\phi})_{t\in\mathbb{Z}}$ be an iid copy of $(\eta_t^{\phi})_{t\in\mathbb{Z}}$. Define*

$$\tilde{\mathcal{F}}_{t,t-h}^{\phi} = (\eta_t^{\phi}, \ldots, \eta_{t-h+1}^{\phi}, \tilde{\eta}_{t-h}^{\phi}, \eta_{t-h-1}^{\phi}, \ldots)$$

*to be $\mathcal{F}_t^{\phi}$ with the $(t-h)$-th element $\eta_{t-h}^{\phi}$ replaced with $\tilde{\eta}_{t-h}^{\phi}$. Moreover, for the product of errors $R$ define $\tilde{\mathcal{F}}_{m,t,t-h}^{R}$ as $\mathcal{F}_{m,t}^{R}$ with the $(t-h)$-th element $\eta_{m,t-h}^{R}$ replaced with the iid copy $\tilde{\eta}_{m,t-h}^{R}$. Now, we define the functional dependence measures of the processes.*

*For each $h \in \mathbb{N}_0$ and $q > 2$, we define the functional dependence measures of $G_{n,l}^{W}(u, \mathcal{F}_t^{W})$ for $(W, l, \cdot) \in \mathbb{W}$, $n \in \mathbb{N}$, $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, $t \in \mathcal{T}_n$ as*

$$\theta_{P,u,n,l,t}^{W}(h, q) = [\mathbb{E}_P(|G_{n,l}^{W}(u, \mathcal{F}_t^{W}) - G_{n,l}^{W}(u, \tilde{\mathcal{F}}_{t,t-h}^{W})|^q)]^{1/q}.$$

*Similarly, we define the functional dependence measures for $G_{P,n,l}^{e}(u, \mathcal{F}_t^{e})$ for $(e, l, \cdot) \in \mathbb{E}$, $n \in \mathbb{N}$, $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, $t \in \mathcal{T}_n$ as*

$$\theta_{P,u,n,l,t}^{e}(h, q) = [\mathbb{E}_P(|G_{P,n,l}^{e}(u, \mathcal{F}_t^{e}) - G_{P,n,l}^{e}(u, \tilde{\mathcal{F}}_{t,t-h}^{e})|^q)]^{1/q}$$

*and*

$$\theta_{P,u,n,l,t}^{e,\infty}(h) = \inf\{K \geq 0 : \mathbb{P}_P(|G_{P,n,l}^{e}(u, \mathcal{F}_t^{e}) - G_{P,n,l}^{e}(u, \tilde{\mathcal{F}}_{t,t-h}^{e})| > K) = 0\}$$

*and for the vector-valued $\boldsymbol{G}_{P,n}^{\boldsymbol{e}}(u, \mathcal{F}_t^{\boldsymbol{e}})$ for $\boldsymbol{e} \in \{\boldsymbol{\varepsilon}, \boldsymbol{\xi}\}$, $n \in \mathbb{N}$, $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, $t \in \mathcal{T}_n$ as*

$$\theta_{P,u,n,t}^{\boldsymbol{e}}(h, q, r) = [\mathbb{E}_P(||\boldsymbol{G}_{P,n}^{\boldsymbol{e}}(u, \mathcal{F}_t^{\boldsymbol{e}}) - \boldsymbol{G}_{P,n}^{\boldsymbol{e}}(u, \tilde{\mathcal{F}}_{t,t-h}^{\boldsymbol{e}})||_r^q)]^{1/q}.$$

*Lastly, define the functional dependence measures for $G_{P,n,m}^{R}(u, \cdot)$ for $n \in \mathbb{N}$, $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, $t \in \mathcal{T}_n$, $m \in \mathcal{M}_n$ as*

$$\theta_{P,u,n,m,t}^{R}(h, q) = [\mathbb{E}_P(|G_{P,n,m}^{R}(u, \mathcal{F}_{m,t}^{R}) - G_{P,n,m}^{R}(u, \tilde{\mathcal{F}}_{m,t,t-h}^{R})|^q)]^{1/q}$$

*and for the vector-valued $\boldsymbol{G}_{P,n}^{\boldsymbol{R}}(u, \cdot)$ for $n \in \mathbb{N}$, $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, $t \in \mathcal{T}_n$ as*

$$\theta_{P,u,n,t}^{\boldsymbol{R}}(h, q, r) = [\mathbb{E}_P(||\boldsymbol{G}_{P,n}^{\boldsymbol{R}}(u, \mathcal{F}_t^{\boldsymbol{R}}) - \boldsymbol{G}_{P,n}^{\boldsymbol{R}}(u, \tilde{\mathcal{F}}_{t,t-h}^{\boldsymbol{R}})||_r^q)]^{1/q}$$

We impose the following regularity conditions to control the temporal dependence and non-stationarity uniformly over a collection of distributions $\mathcal{P}_n$. We impose these conditions on each dimension of the processes so that it is easy to verify the conditions even in high-dimensional settings. First, we introduce an assumption imposing a uniform polynomial decay of the temporal dependence.

**Assumption 3** (Distribution-uniform decay of temporal dependence). *We assume that there exist $\bar{\Theta} > 0$, $\bar{\beta} > 2$, $\bar{q} > 4$, such that for all $P \in \mathcal{P}_n$, $n \in \mathbb{N}$, $t \in \mathcal{T}_n$, $u \in \mathcal{U}_n$, $(W, l, d) \in \mathbb{W}$, it holds that*

$$[\mathbb{E}_P(|G_{n,l}^{W}(u, \mathcal{F}_0^{W})|^{\bar{q}})]^{1/\bar{q}} \leq \bar{\Theta}, \qquad \theta_{P,u,n,l,t}^{W}(h, \bar{q}) \leq \bar{\Theta} \cdot h^{-\bar{\beta}}, \qquad h \geq 1.$$

*Next, we assume that there exist $\bar{\Theta}^* > 0$, $\bar{\beta}^* > 2$, $\bar{q}^* > 4$ such that for all $P \in \mathcal{P}_n$, $n \in \mathbb{N}$, $t \in \mathcal{T}_n$, $u \in \mathcal{U}_n$, $(e, l, d) \in \mathbb{E}$, it holds that*

$$[\mathbb{E}_P(|G_{P,n,l}^{e}(u, \mathcal{F}_0^{e})|^{\bar{q}^*})]^{1/\bar{q}^*} \leq \bar{\Theta}^*, \qquad \theta_{P,u,n,l,t}^{e,\infty}(h) \leq \bar{\Theta}^* \cdot h^{-\bar{\beta}^*}, \qquad h \geq 1.$$

*For additional control in terms of $R$ alone, we also assume that there exist $\bar{\Theta}^R > 0$, $\bar{\beta}^R > 2$, $\bar{q}^R > 4$, such that for all $P \in \mathcal{P}_n$, $n \in \mathbb{N}$, $t \in \mathcal{T}_n$, $u \in \mathcal{U}_n$, $m \in \mathcal{M}_n$, it holds that*

$$[\mathbb{E}_P(|G_{P,n,m}^{R}(u, \mathcal{F}_{m,0}^{R})|^{\bar{q}^R})]^{1/\bar{q}^R} \leq \bar{\Theta}^R, \qquad \theta_{P,u,n,m,t}^{R}(h, \bar{q}^R) \leq \bar{\Theta}^R \cdot h^{-\bar{\beta}^R}, \qquad h \geq 1.$$

We impose the following regularity conditions to control the non-stationarity uniformly over a collection of distributions $\mathcal{P}_n$ which combine ideas related to total variation control from [Mie23; MS22; MS24]. Again, we impose these regularity conditions on each dimension so that the conditions can be easily verified even in high-dimensional settings. To control the non-stationarity, we can use either the total variation control from Assumption 4 or the stochastic Lipschitz control from Assumption 5. The total variation control is more general since it encompasses piecewise locally stationary time series [Zho13], but it is less well-studied. Since the total variation assumption for controlling the non-stationarity is so new, there do not currently exist time-varying regression estimators with theoretical guarantees under this assumption. Nevertheless, we find it useful to state the convergence rate requirements for our test in general terms because we expect that more work will be done using this total variation control of the non-stationarity from Assumption 4 in the near future.

**Assumption 4** (Distribution-uniform total variation condition for non-stationarity)**.** *Recall $\bar{\Theta} > 0$, $\bar{\Theta}^* > 0$ from Assumption 3. We assume that there exist constants $\bar{\Gamma} > 0$, $\bar{\Gamma}^* > 0$ such that for all $P \in \mathcal{P}_n$, $n \in \mathbb{N}$, $t \in \mathcal{T}_n$, $(W, l, d) \in \mathbb{W}$, $(e, l, d) \in \mathbb{E}$ it holds that*

$$\sup_{s\in\mathbb{N}} \sup_{0=u_0<u_1<\ldots<u_s=1} \left( \sum_{i=1}^{s} |G^W_{n,l}(u_i, \mathcal{F}^W_t) - G^W_{n,l}(u_{i-1}, \mathcal{F}^W_t)|^2 \right)^{1/2} \leq \bar{\Theta} \cdot \bar{\Gamma}$$

*and*

$$\sup_{s\in\mathbb{N}} \sup_{0=u_0<u_1<\ldots<u_s=1} \left( \sum_{i=1}^{s} |G^e_{P,n,l}(u_i, \mathcal{F}^e_t) - G^e_{P,n,l}(u_{i-1}, \mathcal{F}^e_t)|^2 \right)^{1/2} \leq \bar{\Theta}^* \cdot \bar{\Gamma}^*.$$

*Similarly, recall $\bar{\Theta}^R > 0$ from Assumption 3. For additional control in terms of $R$ alone, we also assume that there exists a constant $\bar{\Gamma}^R > 0$ such that for all $P \in \mathcal{P}_n$, $n \in \mathbb{N}$, $t \in \mathcal{T}_n$, $m \in \mathcal{M}_n$ it holds that*

$$\sup_{s\in\mathbb{N}} \sup_{0=u_0<u_1<\ldots<u_s=1} \left( \sum_{i=1}^{s} |G^R_{P,n,m}(u_i, \mathcal{F}^R_{m,t}) - G^R_{P,n,m}(u_{i-1}, \mathcal{F}^R_{m,t})|^2 \right)^{1/2} \leq \bar{\Theta}^R \cdot \bar{\Gamma}^R.$$

The less general stochastic Lipschitz Assumption 5 can be made instead of the Assumption 4.

**Assumption 5** (Distribution-uniform stochastic Lipschitz condition)**.** *For each $P \in \mathcal{P}_n$, $n \in \mathbb{N}$, $(W, l, d) \in \mathbb{W}$, $(e, l, d) \in \mathbb{E}$, and $t \in \mathbb{Z}$, we assume that each $G^W_{n,l}(\cdot, \mathcal{F}^W_t)$ and $G^e_{P,n,l}(\cdot, \mathcal{F}^e_t)$ are stochastic Lipschitz function of rescaled time $u \in \mathcal{U}_n$. In particular, we assume that there exist $\bar{\Theta} > 0$, $\bar{q} > 4$ and $\bar{\Theta}^* > 0$, $\bar{q}^* > 4$, such that for all $P \in \mathcal{P}_n$, $n \in \mathbb{N}$, $t \in \mathcal{T}_n$, $u, v \in \mathcal{U}_n$, $(W, l, d) \in \mathbb{W}$, $(e, l, d) \in \mathbb{E}$, it holds that*

$$\sup_{t\in\mathbb{Z}} [\mathbb{E}_P(|G^W_{n,l}(u, \mathcal{F}^W_t) - G^W_{n,l}(v, \mathcal{F}^W_t)|^{\bar{q}})]^{1/\bar{q}} \leq \bar{\Theta}|u - v|$$

*and*

$$\sup_{t\in\mathbb{Z}} [\mathbb{E}_P(|G^e_{P,n,l}(u, \mathcal{F}^e_t) - G^e_{P,n,l}(v, \mathcal{F}^e_t)|^{\bar{q}^*})]^{1/\bar{q}^*} \leq \bar{\Theta}^*|u - v|.$$

*Similarly, for additional control in terms of $R$ alone, we also assume that there exist $\bar{\Theta}^R > 0$, $\bar{q}^R > 4$, such that for all $P \in \mathcal{P}_n$, $n \in \mathbb{N}$, $t \in \mathcal{T}_n$, $u, v \in \mathcal{U}_n$, $m \in \mathcal{M}_n$ it holds that*

$$\sup_{t\in\mathbb{Z}} [\mathbb{E}_P(|G^R_{P,n,m}(u, \mathcal{F}^R_{m,t}) - G^R_{P,n,m}(v, \mathcal{F}^R_{m,t})|^{\bar{q}^R})]^{1/\bar{q}^R} \leq \bar{\Theta}^R|u - v|.$$

We now impose upper and lower bounds on the eigenvalues of the long-run covariance matrices for the high-dimensional non-stationary process of error products from 1, which is in line with the related literature [WX12; DZ20; DZ21; LW23].

**Assumption 6** (Regularity conditions for local long-run covariance matrix and variances)**.** *We assume that the eigenvalues of the local long-run covariance matrices $\boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(u)$ are time-uniformly and distribution-uniformly bounded between $\lambda_{\inf}$ and $\lambda_{\sup}$*

$$0 < \lambda_{\inf} \leq \inf_{P\in\mathcal{P}_n} \lambda_{\min}(\boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(u)) \leq \sup_{P\in\mathcal{P}_n} \lambda_{\max}(\boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(u)) \leq \lambda_{\sup} < \infty$$

*for all $u \in \mathcal{U}_n$, and $n \in \mathbb{N}$. Also, we assume that the local long-run variances are time-uniformly and distribution-uniformly positive*

$$\inf_{P \in \mathcal{P}_n} \Sigma^R_{P,n,m}(u) > 0$$

*for all $u \in \mathcal{U}_n$, $m \in \mathcal{M}_n$, and $n \in \mathbb{N}$.*

As in [DW21; LW23], we impose the following assumption to allow for sub-exponential tails. If $d_X$ and $d_Y$ are fixed, then this assumption can be relaxed to allow for polynomial tails.

**Assumption 7.** *We assume that there exists a constant $t_0 > 0$ such that*

$$\mathbb{E}_P(\exp(t_0 |G^R_{P,n,m}(u, \mathcal{F}^R_{m,0})|)) < \infty$$

*for all $P \in \mathcal{P}_n$, $u \in \mathcal{U}_n$, and $m \in \mathcal{M}_n$.*

# E Distribution-Uniform Theory

## E.1 Distribution-uniform strong Gaussian approximation

We introduce a distribution-uniform extension of the strong Gaussian approximation for high-dimensional non-stationary nonlinear time series from Mies and Steland [MS22]. The notation used for the theoretical framework for non-stationary time series in this subsection is based on Mies and Steland [MS22]. We we slightly change some notation by adding subscripts $P$ and $n$ for the sake of clarity, such that the dependence on the distribution and sample size explicit. Most results can be read in exactly the same way as they were originally stated in Mies and Steland [MS22] with the difference being that certain bounds hold uniformly for all distributions $P \in \mathcal{P}_n$. The notable exception is the strong Gaussian approximation, which requires slightly more care. All of the results in this subsection are non-asymptotic and can be applied to triangular arrays. For instance, see Examples 2 and 3 in Mies and Steland [MS22] for discussion of locally stationary time series. Note that the distribution-uniform assumptions stated in Subsection D.1 satisfy the more general conditions 8 and 9 below. Thus, we can use the results in this subsection for our test as discussed in Subsection 2.3.

For each sample size $n \in \mathbb{N}$, consider a $\mathbb{R}^{d_n}$-valued time series $(X_{t,n})_{t\in[n]}$ and dimension $d_n \in \mathbb{N}$. Let $(\epsilon_i)_{i\in\mathbb{Z}}$, $(\tilde{\epsilon}_i)_{i\in\mathbb{Z}}$ be two iid sequences of random elements. Denote

$$\boldsymbol{\epsilon}_t = (\epsilon_t, \epsilon_{t-1}, \ldots),$$

$$\tilde{\boldsymbol{\epsilon}}_{t,j} = (\epsilon_t, \ldots, \epsilon_{j+1}, \tilde{\epsilon}_j, \epsilon_{j-1}, \ldots),$$
$$\bar{\boldsymbol{\epsilon}}_{t,j} = (\epsilon, \ldots, \epsilon_{j+1}, \tilde{\epsilon}_j, \tilde{\epsilon}_{j-1}, \ldots).$$

We assume that for each $n \in \mathbb{N}$ and $t \in [n]$, $X_{t,n}$ can be represented as a function of these iid random elements

$$X_{t,n} = G_{t,n}(\boldsymbol{\epsilon}_t)$$

where $G_{t,n}$ is a measurable function such that $X_{t,n}$ is a well-defined random vector. Let $\mathcal{P}_n$ be a collection of distributions of $(X_{t,n})_{t\in[n]}$. We make the following distribution-uniform assumptions on the temporal dependence and non-stationarity of $(X_{t,n})_{t\in[n]}$. In the special case of high-dimensional locally stationary time series $G_{w,n}$ is defined for all $w \in [0,1]$ for each $n \in \mathbb{N}$.

**Assumption 8** (Distribution-uniform decay of temporal dependence)**.** *We assume that there exist $\beta > 0$, $q \geq 2$ and a constant $\Theta_n > 0$ for each $n \in \mathbb{N}$, such that for all distributions $P \in \mathcal{P}_n$ and times $t \in [n]$ it holds*

$$(\mathbb{E}_P ||G_{t,n}(\boldsymbol{\epsilon}_t) - G_{t,n}(\tilde{\boldsymbol{\epsilon}}_{t,t-j})||^q_r)^{\frac{1}{q}} \leq \Theta_n \cdot (j \vee 1)^{-\beta}$$

*for $j \geq 0$, and that*

$$(\mathbb{E}_P ||G_{t,n}(\boldsymbol{\epsilon}_s)||^q_2)^{1/q} \leq \Theta_n.$$

*for any $s \in \mathbb{Z}$.*

**Assumption 9** (Distribution-uniform total variation condition for non-stationarity)**.** *Recall $\Theta_n$ from Assumption 8. For each $n \in \mathbb{N}$, we also assume that there exists some $\Gamma_n \geq 1$ such that for all $P \in \mathcal{P}_n$ it holds*

$$\sum_{t=2}^{n} (\mathbb{E}_P ||G_{t,n}(\boldsymbol{\epsilon}_s) - G_{t-1,n}(\boldsymbol{\epsilon}_s)||_2^2)^{\frac{1}{2}} \leq \Gamma_n \cdot \Theta_n$$

*for any $s \in \mathbb{Z}$.*

Let us introduce the setting rigorously. Let $\Omega$ be a sample space, $\mathcal{A}$ the Borel sigma-algebra, and $(\Omega, \mathcal{A})$ a measurable space. For fixed $n \in \mathbb{N}$, let $(\Omega, \mathcal{A})$ be equipped with a family of probability measures $(\mathbb{P}_P)_{P\in\mathcal{P}_n}$ where $\mathcal{P}_n$ is a collection of distributions of $(X_{t,n})_{t\in[n]}$. That is, for a fixed $n \in \mathbb{N}$, the distribution of $(X_{t,n})_{t\in[n]}$ under $\mathbb{P}_P$ is $P$. The family of probability measures $(\mathbb{P}_P)_{P\in\mathcal{P}_n}$ is defined with respect to the same measurable space $(\Omega, \mathcal{A})$, but need not have the same dominating measure. Denote a family of probability spaces by $(\Omega, \mathcal{A}, \mathbb{P}_P)_{P\in\mathcal{P}_n}$ and a sequence of such families of probability spaces by $((\Omega, \mathcal{A}, \mathbb{P}_P)_{P\in\mathcal{P}_n})_{n\in\mathbb{N}}$. When we say that the process $(X_{t,n})_{t\in[n]}$ is defined on the collection of probability spaces $(\Omega, \mathcal{A}, \mathbb{P}_P)_{P\in\mathcal{P}_n}$ for some $n \in \mathbb{N}$, we mean that $(X_{t,n})_{t\in[n]}$ is defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P}_P)$ for each $P \in \mathcal{P}_n$.

Define the two rates

$$\chi(q, \beta) = \begin{cases} \frac{q-2}{6q-4}, \ \beta \geq \frac{3}{2}, \\ \frac{(\beta-1)(q-2)}{q(4\beta-3)-2}, \ \beta \in (1, \frac{3}{2}), \end{cases}$$

and

$$\xi(q, \beta) = \begin{cases} \frac{q-2}{6q-4}, \ \beta \geq 3, \\ \frac{(\beta-2)(q-2)}{(4\beta-6)q-4}, \ \frac{3+\frac{2}{q}}{1+\frac{2}{q}} < \beta < 3, \\ \frac{1}{2} - \frac{1}{\beta}, \ 2 < \beta \leq \frac{3+\frac{2}{q}}{1+\frac{2}{q}}, \end{cases}$$

which will appear in the theorem below. Note that we allow $M_n = O(T_n^{\frac{1-\delta}{1+\frac{1}{2\xi(q,\beta)}}})$ for some $\delta > 0$ so that the error of the strong Gaussian approximation for high-dimensional non-stationary time series from Mies and Steland [MS22] is negligible. As discussed in Mies and Steland [MS22], in the limiting case when $\beta \geq 3$ and $q \to \infty$ we have $M_n = O(T_n^{\frac{1}{4}-\delta})$ for some $\delta > 0$. Also, in this limiting case, any choice of $\zeta \in (0, \frac{1}{2})$ for the lag window $L_n = n^\zeta$ satisfies the conditions of the theorem in Subsection 2.3.

The following theorem is a distribution-uniform version of the strong Gaussian approximation from Theorem 3.1 in Mies and Steland [MS22].

**Lemma E.1.** *For some $n \in \mathbb{N}$, let the process $(X_{t,n})_{t\in[n]}$ be defined on the collection of probability spaces $(\Omega, \mathcal{A}, \mathbb{P}_P)_{P\in\mathcal{P}_n}$ such that Assumption 8 is satisfied for $\mathcal{P}_n$ with some $q > 2$, $\beta > 1$ and constant $\Theta_n > 0$. For all $P \in \mathcal{P}_n$ and $t \in [n]$, let $X_{t,n} = G_{t,n}(\boldsymbol{\epsilon}_t)$ with $\mathbb{E}_P(X_{t,n}) = 0$ and suppose $d_n < cn$ for some $c > 0$. Let $(\Omega', \mathcal{A}', \mathbb{P}'_P)_{P\in\mathcal{P}_n}$ be a new collection of probability spaces on which there exist random vectors $(X'_{t,n})_{t\in[n]}$ such that $(X'_{t,n})_{t\in[n]} \overset{d}{=} (X_{t,n})_{t\in[n]}$ for each $P \in \mathcal{P}_n$ and independent, mean zero, Gaussian random vectors $Y'_{t,n}$ such that*

$$\sup_{P\in\mathcal{P}_n} \left( \mathbb{E}_P \max_{k\leq n} \left|\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{k} (X'_{t,n} - Y'_{t,n}) \right|\right|_2^2 \right)^{\frac{1}{2}} \leq C\Theta_n \sqrt{\log(n)} \left( \frac{d_n}{n} \right)^{\chi(q,\beta)}$$

*for some universal constant $C$ depending only on $q$, $c$, and $\beta$.*

*If $\beta > 2$, then the local long-run covariance matrix $\Sigma_{P,t,n} = \sum_{h=-\infty}^{\infty} \mathrm{Cov}_P(G_{t,n}(\boldsymbol{\epsilon}_0), G_{t,n}(\boldsymbol{\epsilon}_h))$ is well defined for all $P \in \mathcal{P}_n$ and $n \in \mathbb{N}$. If Assumption 9 is also satisfied, then there exist random vectors $(X'_{t,n})_{t\in[n]}$ such that $(X'_{t,n})_{t\in[n]} \overset{d}{=} (X_{t,n})_{t\in[n]}$ for each $P \in \mathcal{P}_n$ and independent, mean zero, Gaussian random vectors $Y^*_{t,n} \sim \mathcal{N}(0, \Sigma_{P,t,n})$ such that*

$$\sup_{P\in\mathcal{P}_n} \left( \mathbb{E}_P \max_{k\leq n} \left|\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{k} (X'_{t,n} - Y^*_{t,n}) \right|\right|_2^2 \right)^{\frac{1}{2}} \leq C\Theta_n \Gamma_n^{\frac{1}{2}\frac{\beta-2}{\beta-1}} \sqrt{\log(n)} \left( \frac{d_n}{n} \right)^{\xi(q,\beta)}$$

*for some universal constant $C$ depending only on $q$, $c$, and $\beta$.*

*Proof of Lemma E.1*: The result follows by the same steps in the proof from Theorem 3.1 in Mies and Steland [MS22] while carrying the supremum over $\mathcal{P}_n$ and by using the distribution-uniform control of the temporal dependence from Assumption 8 and the distribution-uniform control of the non-stationarity from Assumption 9. □

The following result is a distribution-uniform version of Theorem 3.2 from Mies and Steland [MS22].

**Lemma E.2.** *For some $n \in \mathbb{N}$, let the process $X_{t,n} = G_{t,n}(\boldsymbol{\epsilon}_t)$ for $t \in [n]$ be defined on the collection of probability spaces $(\Omega, \mathcal{A}, \mathbb{P}_P)_{P \in \mathcal{P}_n}$. There exists a universal constant $C = C(q, r)$ such that for all $n \in \mathbb{N}$ we have*

$$\sup_{P \in \mathcal{P}_n} \left( \mathbb{E}_P \max_{k \leq n} \left|\left| \sum_{t=1}^{k} (X_{t,n} - \mathbb{E}_P(X_{t,n})) \right|\right|_r^q \right)^{\frac{1}{q}} \leq \sup_{P \in \mathcal{P}_n} C\, n^{\frac{1}{2} - \frac{1}{q}} \sum_{j=1}^{\infty} \left( \sum_{t=1}^{n} \theta_{P,t,j,q,r}^q \right)^{\frac{1}{q}} \tag{10}$$

$$\leq \sup_{P \in \mathcal{P}_n} C\, n^{\frac{1}{2}} \sum_{j=1}^{\infty} \sum_{t=1}^{n} \max_{t \leq n} \theta_{P,t,j,q,r}, \tag{11}$$

*where $\theta_{P,t,j,q,r} = (\mathbb{E}_P ||G_{t,n}(\boldsymbol{\epsilon}_t) - G_{t,n}(\tilde{\boldsymbol{\epsilon}}_{t,t-j})||_r^q)^{\frac{1}{q}}$.*

*In the special case $r = 2$, the inequality may be improved to*

$$\sup_{P \in \mathcal{P}_n} \left( \mathbb{E}_P \max_{k \leq n} \left|\left| \sum_{t=1}^{k} (X_{t,n} - \mathbb{E}_P(X_{t,n})) \right|\right|_2^q \right)^{\frac{1}{q}} \tag{12}$$

$$\leq \sup_{P \in \mathcal{P}_n} C \sum_{j=1}^{\infty} (j \wedge n)^{\frac{1}{2} - \frac{1}{q}} \left( \sum_{t=1}^{n} \theta_{P,t,j,q,2}^q \right)^{\frac{1}{q}} + \sup_{P \in \mathcal{P}_n} C \sum_{j=1}^{n} \left( \sum_{t=1}^{n} \theta_{P,t,j,2,2} \right)^{\frac{1}{2}}. \tag{13}$$

*Proof of Lemma E.2*: The result follows by the same steps in the proof from Theorem 3.2 in Mies and Steland [MS22] while carrying the supremum over $\mathcal{P}_n$. □

## E.2 Distribution-uniform feasible Gaussian approximation

In this subsection, we introduce distribution-uniform extensions of Theorem 4.1 and Proposition 4.2 from Mies and Steland [MS22] so that the distribution-uniform strong Gaussian approximation from Subsection E.1 can be used for statistical inference. The key is a distribution-uniform cumulative covariance estimator $\hat{Q}(k)$ of the cumulative long-run covariance process $Q(k) = \sum_{t=1}^{k} \Sigma_{P,t,n}$ where $\Sigma_{P,t,n} = \sum_{h=-\infty}^{\infty} \mathrm{Cov}_P(G_{t,n}(\boldsymbol{\epsilon}_0), G_{t,n}(\boldsymbol{\epsilon}_h))$ and $X_{t,n} = G_{t,n}(\boldsymbol{\epsilon}_t)$. We will prove these guarantees for the same estimator from Mies and Steland [MS22], namely

$$\hat{Q}(k) = \sum_{t=b_n}^{k} \frac{1}{b_n} \left( \sum_{s=t-b_n+1}^{t} X_{s,n} \right) \left( \sum_{s=t-b_n+1}^{t} X_{s,n} \right)^T$$

for some window size $b_n = O(n^{\zeta})$ for some $\zeta \in (0, \frac{1}{2})$.

The following theorem is a distribution-uniform extension of Theorem 4.1 from Mies and Steland [MS22].

**Lemma E.3.** *For some $n \in \mathbb{N}$, let the process $(X_{t,n})_{t \in [n]}$ be defined on the collection of probability spaces $(\Omega, \mathcal{A}, \mathbb{P}_P)_{P \in \mathcal{P}_n}$ and let $X_{t,n} = G_{t,n}(\boldsymbol{\epsilon}_t)$ satisfy Assumption 9 and Assumption 8 for $\mathcal{P}_n$ with $q \geq 4$ and $\beta > 2$. Then*

$$\sup_{P \in \mathcal{P}_n} \mathbb{E}_P \max_{k = b_n, \ldots, n} \left|\left| \hat{Q}(k) - \sum_{t=1}^{k} \Sigma_{P,t,n} \right|\right|_{\mathrm{tr}} \leq C \Theta_n^2 \left( \Gamma_n \sqrt{b_n} + \sqrt{n d_n b_n} + n b_n^{-1} + n b_n^{2-\beta} \right)$$

*for some universal constant $C$ depending only on $\beta$ and $q$.*

*Proof of Lemma E.3*: The result follows by the same steps in the proof from Theorem 4.1 in Mies and Steland [MS22] while carrying the supremum over $\mathcal{P}_n$ and by using the distribution-uniform control of the temporal dependence from Assumption 8 and the distribution-uniform control of the non-stationarity from Assumption 9. □

The following theorem is a distribution-uniform extension of Proposition 4.2 from Mies and Steland [MS22].

**Lemma E.4.** *For each $n \in \mathbb{N}$ and $P \in \mathcal{P}_n$, let $\Sigma_{P,t,n}, \Sigma'_{P,t,n} \in \mathbb{R}^{d_n \times d_n}$ be symmetric, positive definite matrices for $t \in [n]$. For some $n \in \mathbb{N}$, let the independent, mean zero, Gaussian random vectors $Y_{t,n} \sim \mathcal{N}(0, \Sigma_{P,t,n})$ for $t \in [n]$ be defined on the collection of probability spaces $(\Omega, \mathcal{A}, \mathbb{P}_P)_{P \in \mathcal{P}_n}$. Let $(\Omega', \mathcal{A}', \mathbb{P}'_P)_{P \in \mathcal{P}_n}$ be a new collection of probability spaces on which there exist independent, mean zero, Gaussian random vectors $Y'_{t,n} \sim \mathcal{N}(0, \Sigma'_{P,t,n})$ for $t \in [n]$ such that*

$$\sup_{P \in \mathcal{P}_n} \mathbb{E}_P \max_{k=1,\ldots,n} \left|\left| \sum_{t=1}^{k} Y_{t,n} - \sum_{t=1}^{k} Y'_{t,n} \right|\right|_2^2 \leq C \log(n) \, [\sqrt{n\delta_n \rho_n} + \rho_n],$$

*where*

$$\delta_n = \sup_{P \in \mathcal{P}_n} \max_{k=1,\ldots,n} \left|\left| \sum_{t=1}^{k} \Sigma_{P,t,n} - \sum_{t=1}^{k} \Sigma'_{P,t,n} \right|\right|_{\mathrm{tr}},$$

*and*

$$\rho_n = \sup_{P \in \mathcal{P}_n} \max_{t=1,\ldots,n} ||\Sigma_{P,t,n}||_{\mathrm{tr}}.$$

*Proof of Lemma E.4*: The result follows by the same steps in the proof from Proposition 4.2 in Mies and Steland [MS22] while carrying the supremum over $\mathcal{P}_n$. □

# F Proof of Theorem 2.1

Recall the terms

$$\hat{w}^f_{P,t,n,i,a}(t/n, \boldsymbol{z}) = f_{P,t,n,i,a}(t/n, \boldsymbol{z}) - \hat{f}_{t,n,i,a}(t/n, \boldsymbol{z}),$$
$$\hat{w}^g_{P,t,n,j,b}(t/n, \boldsymbol{z}) = g_{P,t,n,j,b}(t/n, \boldsymbol{z}) - \hat{g}_{t,n,j,b}(t/n, \boldsymbol{z}),$$

for $t \in \mathcal{T}_n$ and $\boldsymbol{z} \in \mathbb{R}^{\boldsymbol{d_z}}$.

Similarly, define

$$\hat{w}^f_{P,t,n,i,a} = f_{P,t,n,i,a}(t/n, \boldsymbol{Z}_{t,n}) - \hat{f}_{t,n,i,a}(t/n, \boldsymbol{Z}_{t,n}),$$
$$\hat{w}^g_{P,t,n,j,b} = g_{P,t,n,j,b}(t/n, \boldsymbol{Z}_{t,n}) - \hat{g}_{t,n,j,b}(t/n, \boldsymbol{Z}_{t,n}),$$

as the prediction errors with the covariates at time $t \in \mathcal{T}_n$.

Then we may write the vectors with all dimensions and time-offsets as

$$\hat{\boldsymbol{w}}^f_{P,t,n} = (\hat{w}^f_{P,t,n,i,a})_{i \in [d_X], a \in A_i},$$

$$\hat{\boldsymbol{w}}^g_{P,t,n} = (\hat{w}^g_{P,t,n,j,b})_{j \in [d_Y], b \in B_j}.$$

Further, denote the three bias terms by

$$\hat{\boldsymbol{w}}^{f,g}_{P,t,n} = (\hat{w}^f_{P,t,n,i,a} \hat{w}^g_{P,t,n,j,b})_{m \in \mathcal{M}_n},$$
$$\hat{\boldsymbol{w}}^{g,\varepsilon}_{P,t,n} = (\hat{w}^g_{P,t,n,j,b} \varepsilon_{P,t,n,i,a})_{m \in \mathcal{M}_n},$$
$$\hat{\boldsymbol{w}}^{f,\xi}_{P,t,n} = (\hat{w}^f_{P,t,n,i,a} \xi_{P,t,n,j,b})_{m \in \mathcal{M}_n}.$$

**Step 1 (Bias Terms):** First, we decompose the products of residuals into the products of errors and the three bias terms, then apply the triangle inequality and subadditivity, yielding

$$
\begin{aligned}
&\sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \tau_n) \\
&\leq \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\boldsymbol{R}_{P,t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \frac{\tau_n}{2}) \\
&+ \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\hat{\boldsymbol{w}}^{f,g}_{P,t,n})_{t\in\mathcal{T}_n}) > \frac{\tau_n}{6}) \\
&+ \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\hat{\boldsymbol{w}}^{g,\varepsilon}_{P,t,n})_{t\in\mathcal{T}_n}) > \frac{\tau_n}{6}) \\
&+ \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\hat{\boldsymbol{w}}^{f,\xi}_{P,t,n})_{t\in\mathcal{T}_n}) > \frac{\tau_n}{6}).
\end{aligned}
$$

We will handle each of the three bias terms separately.

**Step 1.1:** Observe that

$$
\begin{aligned}
\tau_n^{-1} S_n((\hat{\boldsymbol{w}}^{f,g}_{P,t,n})_{t\in\mathcal{T}_n}) &\overset{(1)}{\leq} \tau_n^{-1} \max_{s\in\mathcal{T}_n} \left\| \frac{1}{\sqrt{T_n}} \sum_{t\leq s} \hat{\boldsymbol{w}}^{f,g}_{P,t,n} \right\|_2 \\
&\overset{(2)}{\leq} \tau_n^{-1} \frac{1}{\sqrt{T_n}} \left( \sum_{m\in\mathcal{M}_n} \left( \sum_{t\in\mathcal{T}_n} |\hat{w}^f_{P,t,n,i,a}||\hat{w}^g_{P,t,n,j,b}| \right)^2 \right)^{\frac{1}{2}} \\
&\overset{(3)}{\leq} \tau_n^{-1} \sqrt{M_n}\sqrt{T_n} \max_{m\in\mathcal{M}_n} \max_{t\in\mathcal{T}_n} \sup_{\boldsymbol{z}\in\mathbb{R}^{\boldsymbol{d_z}}} |\hat{w}^f_{P,t,n,i,a}(t/n,\boldsymbol{z})||\hat{w}^g_{P,t,n,j,b}(t/n,\boldsymbol{z})| \\
&\overset{(4)}{=} o_{\mathcal{P}}(1),
\end{aligned}
$$

where (1) follows from our basic assumption about the form test statistic as discussed in Subsection 2.3, (2) and (3) follow from elementary calculations, and (4) follows by the assumptions on the convergence rates of the time-varying regression estimators.

**Step 1.2:** Conditional on $\boldsymbol{Y}_n$ and $\boldsymbol{Z}_n$, $(\hat{\boldsymbol{w}}^{g,\varepsilon}_{P,t,n})_{t\in\mathcal{T}_n}$ is a mean-zero process since for each $n\in\mathbb{N}$, $P\in\mathcal{P}_n$, $m\in\mathcal{M}_n$, $t\in\mathcal{T}_n$ we have

$$
\mathbb{E}_P(\hat{w}^g_{P,t,n,j,b}\varepsilon_{P,t,n,i,a}|\boldsymbol{Y}_n,\boldsymbol{Z}_n) = \hat{w}^g_{P,t,n,j,b}\mathbb{E}_P(\varepsilon_{P,t,n,i,a}|\boldsymbol{Z}_n) = 0.
$$

Observe that

$$
\begin{aligned}
&\sup_{P\in\mathcal{P}_n} \mathbb{P}_P(\tau_n^{-1} S_n((\hat{\boldsymbol{w}}^{g,\varepsilon}_{P,t,n})_{t\in\mathcal{T}_n}) > \delta) \\
&\overset{(1)}{\leq} \sup_{P\in\mathcal{P}_n} \mathbb{P}_P\left(\tau_n^{-2}\max_{s\in\mathcal{T}_n}\left\|\frac{1}{\sqrt{T_n}}\sum_{t\leq s}\hat{\boldsymbol{w}}^{g,\varepsilon}_{P,t,n}\right\|_2^2 \wedge \delta^2 \geq \delta^2\right) \\
&\overset{(2)}{\leq} \delta^{-2}\sup_{P\in\mathcal{P}_n} \mathbb{E}_P\left(\tau_n^{-2}\max_{s\in\mathcal{T}_n}\left\|\frac{1}{\sqrt{T_n}}\sum_{t\leq s}\hat{\boldsymbol{w}}^{g,\varepsilon}_{P,t,n}\right\|_2^2 \wedge \delta^2\right) \\
&\overset{(3)}{=} \delta^{-2}\sup_{P\in\mathcal{P}_n} \mathbb{E}_P\left[\mathbb{E}_P\left(\tau_n^{-2}\max_{s\in\mathcal{T}_n}\left\|\frac{1}{\sqrt{T_n}}\sum_{t\leq s}\hat{\boldsymbol{w}}^{g,\varepsilon}_{P,t,n}\right\|_2^2 \wedge \delta^2 \mid \boldsymbol{Y}_n, \boldsymbol{Z}_n\right)\right] \\
&\overset{(4)}{\leq} \delta^{-2}\tau_n^{-2}T_n^{-1}\sup_{P\in\mathcal{P}_n} \mathbb{E}_P\left[\mathbb{E}_P\left(\max_{s\in\mathcal{T}_n}\left\|\sum_{t\leq s}\hat{\boldsymbol{w}}^{g,\varepsilon}_{P,t,n}\right\|_2^2 \mid \boldsymbol{Y}_n, \boldsymbol{Z}_n\right) \wedge \delta^2\right] \\
&\overset{(5)}{\leq} \delta^{-2}\tau_n^{-2}T_n^{-1} \\
&\sup_{P\in\mathcal{P}_n} \mathbb{E}_P\left[\left(2C\sum_{h=1}^{\infty}\left(\sum_{t\in\mathcal{T}_n}(\max_{j\in[d_Y]}\max_{b\in B_j}\max_{t\in\mathcal{T}_n}\sup_{\boldsymbol{z}\in\mathbb{R}^{\boldsymbol{d_z}}}|\hat{w}^{g}_{P,t,n,j,b}(t/n,\boldsymbol{z})|M_n^{\frac{1}{2}}\bar{\Theta}^* h^{\bar{\beta}^*})^2\right)^{\frac{1}{2}}\right)^2 \wedge \delta^2\right] \\
&\overset{(6)}{\leq} \delta^{-2}\tau_n^{-2}4C^2M_n\bar{\Theta}^{*^2}\sup_{P\in\mathcal{P}_n} \mathbb{E}_P\left[\max_{j\in[d_Y]}\max_{b\in B_j}\max_{t\in\mathcal{T}_n}\sup_{\boldsymbol{z}\in\mathbb{R}^{\boldsymbol{d_z}}}|\hat{w}^{g}_{P,t,n,j,b}(t/n,\boldsymbol{z})|^2\left(\sum_{h=1}^{\infty}h^{\bar{\beta}^*}\right)^2 \wedge \delta^2\right] \\
&\overset{(7)}{\leq} \delta^{-2}\tau_n^{-2}4C^2M_n\bar{\Theta}^{*^2}H^2\sup_{P\in\mathcal{P}_n} \mathbb{E}_P\left[\max_{j\in[d_Y]}\max_{b\in B_j}\max_{t\in\mathcal{T}_n}\sup_{\boldsymbol{z}\in\mathbb{R}^{\boldsymbol{d_z}}}|\hat{w}^{g}_{P,t,n,j,b}(t/n,\boldsymbol{z})|^2 \wedge \delta^2\right] \\
&\overset{(8)}{\leq} C'_\delta\tau_n^{-2}M_n\sup_{P\in\mathcal{P}_n} \mathbb{E}_P\left[\max_{j\in[d_Y]}\max_{b\in B_j}\max_{t\in\mathcal{T}_n}\sup_{\boldsymbol{z}\in\mathbb{R}^{\boldsymbol{d_z}}}|\hat{w}^{g}_{P,t,n,j,b}(t/n,\boldsymbol{z})|^2 \wedge \delta^2\right] \\
&\overset{(9)}{=} o(1),
\end{aligned}
$$

where the previous lines follow by (1) the assumption about the form of the test statistic, squaring, and the definition of minimum, (2) Markov's inequality, (3) the law of iterated expectation, (4) linearity and monotonicity of expectations and conditional expectations, (5) Lemma E.2, Hölder's inequality, and bounding the functional dependence measure by Assumption 3, (6) summing over $t$, taking powers, factoring terms, and linearity of expectation, (7) convergence of the p-series for $p > 1$ since $\beta^* > 2$ by Assumption 3 where we denote the positive constant $H := \sum_{h=1}^{\infty} h^{\bar{\beta}^*}$ and linearity of expectation, (8) simplifying the expression by letting $C'_\delta = 4\delta^{-2}C^2H^2\bar{\Theta}^{*^2}$, and (9) the assumption on the time-uniform convergence rate for the time-varying regression estimator and Lemma G.2.

**Step 1.3:** The same arguments as **Step 1.2** can be used to show that

$$
\sup_{P\in\mathcal{P}_n} \mathbb{P}_P(\tau_n^{-1} S_n((\hat{\boldsymbol{w}}^{f,\xi}_{P,t,n})_{t\in\mathcal{T}_n}) > \delta) = o_{\mathcal{P}}(1).
$$

Next, we turn to the products of errors $(\boldsymbol{R}_{P,t,n})_{t\in\mathcal{T}_n}$.

**Step 2 (Strong Gaussian Approximation):** Denote the Gaussian random vectors associated with the strong Gaussian approximation of the product of errors by $\boldsymbol{R}^{\dagger}_{t,n} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(t/n))$ for $t \in \mathcal{T}_n$.

Observe that

$$
\begin{aligned}
&\sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\boldsymbol{R}_{P,t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \frac{\tau_n}{2}) \\
&\overset{(1)}{\leq} \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\boldsymbol{R}^\dagger_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \frac{\tau_n}{4}) \\
&+ \sup_{P\in\mathcal{P}_n} \mathbb{P}_P\left(\max_{s\in\mathcal{T}_n}\left|\left|\frac{1}{\sqrt{T_n}}\sum_{t\leq s}(\boldsymbol{R}_{P,t,n} - \boldsymbol{R}^\dagger_{t,n})\right|\right|_2 > \frac{\tau_n}{4}\right) \\
&\overset{(2)}{\leq} \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\boldsymbol{R}^\dagger_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \frac{\tau_n}{4}) \\
&+ 4\tau_n^{-1}\sup_{P\in\mathcal{P}_n} \mathbb{E}_P\left(\max_{s\in\mathcal{T}_n}\left|\left|\frac{1}{\sqrt{T_n}}\sum_{t\leq s}(\boldsymbol{R}_{P,t,n} - \boldsymbol{R}^\dagger_{t,n})\right|\right|_2\right) \\
&\overset{(3)}{\leq} \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\boldsymbol{R}^\dagger_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \frac{\tau_n}{4}) \\
&+ 4\tau_n^{-1}C\bar{\Theta}^R(\bar{\Gamma}^R)^{\frac{1}{2}\frac{\bar{\beta}^R-2}{\bar{\beta}^R-1}} M_n^{\frac{1}{2}+\frac{1}{4}\frac{\bar{\beta}^R-2}{\bar{\beta}^R-1}}\sqrt{\log(T_n)}\left(\frac{M_n}{T_n}\right)^{\xi(\bar{q}^R,\bar{\beta}^R)},
\end{aligned}
$$

where (1) follows from the triangle inequality, subadditivity, and the assumption about the form of the test statistic, (2) follows by Markov's inequality, and (3) follows by the distribution-uniform strong Gaussian approximation for high-dimensional non-stationary time series from Lemma E.1.

By subadditivity and monotonicity, we have

$$
\begin{aligned}
&\sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\boldsymbol{R}^\dagger_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \frac{\tau_n}{4}) \\
&\leq \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\boldsymbol{R}^\dagger_{t,n})_{t\in\mathcal{T}_n}) > a_\alpha(Q^{\boldsymbol{R}}_{P,n})) \\
&+ \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(a_\alpha(Q^{\boldsymbol{R}}_{P,n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \frac{\tau_n}{4}) \\
&= \alpha + \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(a_\alpha(Q^{\boldsymbol{R}}_{P,n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \frac{\tau_n}{4}).
\end{aligned}
$$

**Step 3 (Covariance Approximation):** Now, we focus on upper bounding

$$
\sup_{P\in\mathcal{P}_n} \mathbb{P}_P(a_\alpha(Q^{\boldsymbol{R}}_{P,n}) > a_{\alpha-\nu_n}(\hat{Q}_n^{\boldsymbol{R}}) + \frac{\tau_n}{4}).
$$

**Step 3.1:** Let us reflect on the implications of Proposition 4.2 of Mies and Steland [MS22]. Proposition 4.2 states that for each $n \in \mathbb{N}$ and $P \in \mathcal{P}_n$, for some cumulative covariance process $\bar{Q}^{\boldsymbol{R}}_{P,n}$, there exist independent Gaussian random vectors $\bar{\boldsymbol{R}}_{t,n} \sim \mathcal{N}(0, \bar{\boldsymbol{\Sigma}}^{\boldsymbol{R}}_{P,n}(t/n))$ for $t \in \mathcal{T}_n$ with $\bar{\boldsymbol{\Sigma}}^{\boldsymbol{R}}_{P,n}(t/n) = \bar{Q}^{\boldsymbol{R}}_{P,t,n} - \bar{Q}^{\boldsymbol{R}}_{P,t-1,n}$ that are coupled with the Gaussian random vectors from the strong Gaussian approximation of the product of errors $\boldsymbol{R}^\dagger_{t,n} \sim \mathcal{N}(0, \boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(t/n))$ for $t \in \mathcal{T}_n$, such that

$$
\mathbb{E}_P \max_{k\in\mathcal{T}_n}\left|\left|\sum_{t\leq k}\boldsymbol{R}^\dagger_{t,n} - \sum_{t\leq k}\bar{\boldsymbol{R}}_{t,n}\right|\right|_2^2 \leq C\,\log(T_n)\,[\sqrt{T_n\bar{\delta}_{P,n}\rho_{P,n}} + \rho_{P,n}] = \bar{\Delta}_{P,n},
$$

where

$$
\bar{\delta}_{P,n} = \max_{k\in\mathcal{T}_n}\left|\left|\sum_{t\leq k}\boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(t/n) - \sum_{t\leq k}\bar{\boldsymbol{\Sigma}}^{\boldsymbol{R}}_{P,n}(t/n)\right|\right|_{\mathrm{tr}}
$$

and

$$
\rho_{P,n} = \max_{t\in\mathcal{T}_n}||\boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(t/n)||_{\mathrm{tr}}.
$$

Then for each $n \in \mathbb{N}$ and $P \in \mathcal{P}_n$, we have

$$\begin{aligned}
&\mathbb{P}_P(S_n((\boldsymbol{R}^\dagger_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\bar{Q}^{\boldsymbol{R}}_{P,n}) + \frac{\tau_n}{4}) \\
&\overset{(1)}{\leq} \mathbb{P}_P(S_n((\bar{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\bar{Q}^{\boldsymbol{R}}_{P,n})) \\
&+ \mathbb{P}_P\left(\max_{s\in\mathcal{T}_n}\left|\left|\frac{1}{\sqrt{T_n}}\sum_{t\leq s}(\boldsymbol{R}^\dagger_{t,n} - \bar{\boldsymbol{R}}_{t,n})\right|\right|_2 > \frac{\tau_n}{4}\right) \\
&\overset{(2)}{=} \mathbb{P}_P(S_n((\bar{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\bar{Q}^{\boldsymbol{R}}_{P,n})) \\
&+ \mathbb{P}_P\left(\max_{s\in\mathcal{T}_n}\left|\left|\frac{1}{\sqrt{T_n}}\sum_{t\leq s}(\boldsymbol{R}^\dagger_{t,n} - \bar{\boldsymbol{R}}_{t,n})\right|\right|_2^2 > \frac{\tau_n^2}{16}\right) \\
&\overset{(3)}{\leq} \mathbb{P}_P(S_n((\bar{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\bar{Q}^{\boldsymbol{R}}_{P,n})) \\
&+ 16\tau_n^{-2}T_n^{-1}\mathbb{E}_P\left(\max_{s\in\mathcal{T}_n}\left|\left|\sum_{t\leq s}(\boldsymbol{R}^\dagger_{t,n} - \bar{\boldsymbol{R}}_{t,n})\right|\right|_2^2\right) \\
&\overset{(4)}{\leq} (\alpha - \nu_n) + 16\tau_n^{-2}\bar{\Delta}_{P,n}T_n^{-1} \overset{(5)}{=} \alpha + \left[16\tau_n^{-2}\bar{\Delta}_{P,n}T_n^{-1} - \nu_n\right],
\end{aligned}$$

where the previous lines follow by (1) the triangle inequality, subadditivity, the assumption about the form of the test statistic, (2) squaring, (3) Markov's inequality, (4) Proposition 4.2 from Mies and Steland [MS22], and (5) rearranging terms.

We see that if $\left[16\tau_n^{-2}\bar{\Delta}_{P,n}T_n^{-1} - \nu_n\right] < 0$ then

$$\mathbb{P}_P(S_n((\boldsymbol{R}^\dagger_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\bar{Q}^{\boldsymbol{R}}_{P,n}) + \frac{\tau_n}{4}) < \alpha,$$

which implies that $a_{\alpha-\nu_n}(\bar{Q}^{\boldsymbol{R}}_{P,n}) + \frac{\tau_n}{4}$ is greater than the $(1-\alpha)$ quantile of $S_n((\boldsymbol{R}^\dagger_{t,n})_{t\in\mathcal{T}_n}$ which we denote by $a_\alpha(\tilde{Q}^{\boldsymbol{R}}_n)$. Hence, if $a_\alpha(\tilde{Q}^{\boldsymbol{R}}_n) \geq a_{\alpha-\nu_n}(\bar{Q}^{\boldsymbol{R}}_{P,n}) + \frac{\tau_n}{4}$ then $\left[16\tau_n^{-2}\bar{\Delta}_{P,n}T_n^{-1} - \nu_n\right] \geq 0$, or equivalently $\bar{\Delta}_{P,n} \geq \frac{1}{16}T_n\nu_n\tau_n^2$.

**Step 3.2:** Now, we apply this idea with the cumulative covariance process of the products of residuals $\hat{Q}^{\boldsymbol{R}}_n$. By the implication stated at the end of **Step 3.1** and monotonicity, we have

$$\begin{aligned}
&\sup_{P\in\mathcal{P}_n} \mathbb{P}_P(a_\alpha(Q^{\boldsymbol{R}}_{P,n}) > a_{\alpha-\nu_n}(\hat{Q}^{\boldsymbol{R}}_n) + \frac{\tau_n}{4}) \\
&\leq \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(\hat{\Delta}_{P,n} \geq \frac{1}{16}T_n\nu_n\tau_n^2),
\end{aligned}$$

where we have replaced $\bar{\Delta}_{P,n}$, $\bar{\delta}_{P,n}$ with $\hat{\Delta}_{P,n}$, $\hat{\delta}_{P,n}$ which are defined by

$$\hat{\Delta}_{P,n} = C\,\log(T_n)\;[\sqrt{T_n\hat{\delta}_{P,n}\rho_{P,n}} + \rho_{P,n}],$$

$$\hat{\delta}_{P,n} = \max_{k=L_n,\ldots,T_n}\left|\left|\sum_{t\leq k}\boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(t/n) - \hat{Q}^{\boldsymbol{R}}_{k,n}\right|\right|_{\mathrm{tr}},$$

and $\rho_{P,n}$ is defined in the same way as

$$\rho_{P,n} = \max_{t=L_n,\ldots,T_n}||\boldsymbol{\Sigma}^{\boldsymbol{R}}_{P,n}(t/n)||_{\mathrm{tr}}.$$

Thus, if we can find $\varphi_n$ such that $\hat{\Delta}_{P,n} = O_{\mathcal{P}}(\varphi_n)$ and if we select the offsets so that $\nu_n\tau_n^2 \gg T_n^{-1}\varphi_n$, or equivalently $\nu_n \gg \tau_n^{-2}T_n^{-1}\varphi_n$, then we will have

$$\sup_{P\in\mathcal{P}_n}\mathbb{P}_P(\hat{\Delta}_{P,n} \geq \frac{1}{16}T_n\nu_n\tau_n^2) \to 0.$$

By Lemma G and Assumption 3, we have

$$\sup_{P\in\mathcal{P}_n} \rho_{P,n} \le C_\rho M_n (\bar{\Theta}^R)^2,$$

so we obtain $\hat{\Delta}_{P,n} = O_{\mathcal{P}}(\varphi_n)$ with

$$\varphi_n = \log(T_n) M_n \left[ T_n^{\frac{1}{2}} M_n^{-\frac{1}{2}} \hat{\delta}_{P,n}^{\frac{1}{2}} + 1 \right].$$

Plugging $\varphi_n$ into the offset condition $\nu_n \gg \tau_n^{-2} T_n^{-1} \varphi_n$ that we wish to satisfy, if we have

$$\nu_n \gg \log(T_n) M_n \left( \tau_n^{-2} \left( T_n^{-\frac{1}{2}} M_n^{-\frac{1}{2}} \hat{\delta}_{P,n}^{\frac{1}{2}} + T_n^{-1} \right) \right),$$

then

$$\sup_{P\in\mathcal{P}_n} \mathbb{P}_P(\hat{\Delta}_{P,n} \ge \frac{1}{16} T_n \nu_n \tau_n^2) \to 0.$$

It remains to analyze $\hat{\delta}_{P,n}$.

**Step 3.3:** By the triangle inequality, we have

$$\begin{aligned}
\hat{\delta}_{P,n} = & \max_{k=L_n,\dots,T_n} \left|\left| \sum_{t\le k} \boldsymbol{\Sigma}_{P,n}^{\boldsymbol{R}}(t/n) - \hat{Q}_{k,n}^{\boldsymbol{R}} \right|\right|_{\mathrm{tr}} \\
\le & \max_{k=L_n,\dots,T_n} \left|\left| \sum_{t\le k} \boldsymbol{\Sigma}_{P,n}^{\boldsymbol{R}}(t/n) - Q_{P,k,n}^{\boldsymbol{R}} \right|\right|_{\mathrm{tr}} \\
+ & \max_{k=L_n,\dots,T_n} \left|\left| \hat{Q}_{k,n}^{\boldsymbol{R}} - Q_{P,k,n}^{\boldsymbol{R}} \right|\right|_{\mathrm{tr}}.
\end{aligned}$$

By Lemma E.3, Assumption 3, and Assumption 4, the covariance estimation error can be bounded as

$$\begin{aligned}
&\sup_{P\in\mathcal{P}_n} \mathbb{E}_P \left( \max_{k=L_n,\dots,T_n} \left|\left| \sum_{t\le k} \boldsymbol{\Sigma}_{P,n}^{\boldsymbol{R}}(t/n) - Q_{P,k,n}^{\boldsymbol{R}} \right|\right|_{\mathrm{tr}} \right) \\
&\le C(\bar{\Theta}^R)^2 M_n \left( \bar{\Gamma}^R M_n^{\frac{1}{2}} L_n^{\frac{1}{2}} + T_n^{\frac{1}{2}} M_n^{\frac{1}{2}} L_n^{\frac{1}{2}} + T_n L_n^{-1} + T_n L_n^{2-\bar{\beta}^R} \right) \\
&= O(r_{n,1}^{\delta}),
\end{aligned}$$

where

$$r_{n,1}^{\delta} = M_n \left( M_n^{\frac{1}{2}} L_n^{\frac{1}{2}} + T_n^{\frac{1}{2}} M_n^{\frac{1}{2}} L_n^{\frac{1}{2}} + T_n L_n^{-1} + T_n L_n^{2-\bar{\beta}^R} \right).$$

Next, we must handle the prediction error due using the products of residuals. Plugging in the definitions of $Q_{P,k,n}^{\boldsymbol{R}}$, $\hat{Q}_{k,n}^{\boldsymbol{R}}$ and applying the triangle inequality yields

$$\begin{aligned}
&\max_{k=L_n,\dots,T_n} \left|\left| \hat{Q}_{k,n}^{\boldsymbol{R}} - Q_{P,k,n}^{\boldsymbol{R}} \right|\right|_{\mathrm{tr}} \\
&= \max_{k=L_n,\dots,T_n} \left|\left| \sum_{r=L_n}^{k} \frac{1}{L_n} \left( \sum_{s=r-L_n+l_n}^{r-1+l_n} \hat{\boldsymbol{R}}_{s,n} \right)^{\otimes 2} - \sum_{r=L_n}^{k} \frac{1}{L_n} \left( \sum_{s=r-L_n+l_n}^{r-1+l_n} \boldsymbol{R}_{P,s,n} \right)^{\otimes 2} \right|\right|_{\mathrm{tr}} \\
&\le \frac{1}{L_n} \sum_{r=L_n}^{T_n} \left|\left| \left( \sum_{s=r-L_n+l_n}^{r-1+l_n} \hat{\boldsymbol{R}}_{s,n} \right)^{\otimes 2} - \left( \sum_{s=r-L_n+l_n}^{r-1+l_n} \boldsymbol{R}_{P,s,n} \right)^{\otimes 2} \right|\right|_{\mathrm{tr}}.
\end{aligned}$$

For vectors $\hat{v}, v \in \mathbb{R}^d$, we have

$$\begin{aligned}
&||\hat{v}\hat{v}^T - vv^T||_{\mathrm{tr}} \\
&\overset{(1)}{=} ||(\hat{v}-v)v^T + v(\hat{v}-v)^T + (\hat{v}-v)(\hat{v}-v)^T||_{\mathrm{tr}} \\
&\overset{(2)}{\le} 2||(\hat{v}-v)v^T||_{\mathrm{tr}} + ||(\hat{v}-v)(\hat{v}-v)^T||_{\mathrm{tr}} \\
&\overset{(3)}{=} 2||\hat{v}-v||_2 ||v||_2 + ||\hat{v}-v||_2^2,
\end{aligned}$$

where (1) follows from adding and subtracting terms, (2) follows from the triangle inequality, and (3) follows by the properties of outer products and the definition of the trace norm.

Applying this inequality to the quantity above and rearranging terms yields

$$
\begin{aligned}
&\max_{k=L_n,\dots,T_n} \left\|\hat{Q}^{\boldsymbol{R}}_{k,n} - Q^{\boldsymbol{R}}_{P,k,n}\right\|_{\mathrm{tr}} \\
&\leq \frac{2}{L_n}\sum_{r=L_n}^{T_n}\left\|\sum_{s=r-L_n+l_n}^{r-1+l_n}\left(\hat{\boldsymbol{R}}_{s,n} - \boldsymbol{R}_{P,s,n}\right)\right\|_2 \left\|\sum_{s=r-L_n+l_n}^{r-1+l_n}\boldsymbol{R}_{P,s,n}\right\|_2 \\
&+ \frac{1}{L_n}\sum_{r=L_n}^{T_n}\left\|\sum_{s=r-L_n+l_n}^{r-1+l_n}\left(\hat{\boldsymbol{R}}_{s,n} - \boldsymbol{R}_{P,s,n}\right)\right\|_2^2 \\
&\leq 2\sum_{r=1}^{T_n}\left\|\frac{1}{\sqrt{L_n}}\sum_{s=r-L_n+l_n}^{r-1+l_n}\left(\hat{\boldsymbol{R}}_{s,n} - \boldsymbol{R}_{P,s,n}\right)\right\|_2 \left\|\frac{1}{\sqrt{L_n}}\sum_{s=r-L_n+l_n}^{r-1+l_n}\boldsymbol{R}_{P,s,n}\right\|_2 \\
&+ \sum_{r=1}^{T_n}\left\|\frac{1}{\sqrt{L_n}}\sum_{s=r-L_n+l_n}^{r-1+l_n}\left(\hat{\boldsymbol{R}}_{s,n} - \boldsymbol{R}_{P,s,n}\right)\right\|_2^2.
\end{aligned}
$$

We can decompose the quantity concerning the difference of the residual products and error products into the same three bias terms as **Step 1**. Applying the triangle inequality yields

$$
\begin{aligned}
\left\|\frac{1}{\sqrt{L_n}}\sum_{s=r-L_n+l_n}^{r-1+l_n}\left(\hat{\boldsymbol{R}}_{s,n} - \boldsymbol{R}_{P,s,n}\right)\right\|_2 \leq &\left\|\frac{1}{\sqrt{L_n}}\sum_{s=r-L_n+l_n}^{r-1+l_n}\hat{\boldsymbol{w}}^{f,g}_{P,t,n}\right\|_2 \\
&+ \left\|\frac{1}{\sqrt{L_n}}\sum_{s=r-L_n+l_n}^{r-1+l_n}\hat{\boldsymbol{w}}^{g,\varepsilon}_{P,t,n}\right\|_2 \\
&+ \left\|\frac{1}{\sqrt{L_n}}\sum_{s=r-L_n+l_n}^{r-1+l_n}\hat{\boldsymbol{w}}^{f,\xi}_{P,t,n}\right\|_2.
\end{aligned}
$$

Using the same arguments as in **Step 1.1**, we obtain

$$
\left\|\frac{1}{\sqrt{L_n}}\sum_{s=r-L_n+l_n}^{r-1+l_n}\hat{\boldsymbol{w}}^{f,g}_{P,t,n}\right\|_2 = o_{\mathcal{P}}(L_n^{\frac{1}{2}}T_n^{-\frac{1}{2}}\tau_n).
$$

Moreover, using the same arguments as in **Step 1.2**, we obtain

$$
\left\|\frac{1}{\sqrt{L_n}}\sum_{s=r-L_n+l_n}^{r-1+l_n}\hat{\boldsymbol{w}}^{g,\varepsilon}_{P,t,n}\right\|_2 = o_{\mathcal{P}}(\tau_n),
$$

and

$$
\left\|\frac{1}{\sqrt{L_n}}\sum_{s=r-L_n+l_n}^{r-1+l_n}\hat{\boldsymbol{w}}^{f,\xi}_{P,t,n}\right\|_2 = o_{\mathcal{P}}(\tau_n).
$$

Next, by Lemma E.2 and Assumption 3, we have

$$
\begin{aligned}
&\sup_{P\in\mathcal{P}_n} \mathbb{E}_P \left( \left\| \frac{1}{\sqrt{L_n}} \sum_{s=r-L_n+l_n}^{r-1+l_n} \boldsymbol{R}_{P,s,n} \right\|_2^2 \right)^{\frac{1}{2}} \\
&= \frac{1}{\sqrt{L_n}} \sup_{P\in\mathcal{P}_n} \mathbb{E}_P \left( \left\| \sum_{s=r-L_n+l_n}^{r-1+l_n} \boldsymbol{R}_{P,s,n} \right\|_2^2 \right)^{\frac{1}{2}} \\
&\leq 2\frac{1}{\sqrt{L_n}} L_n^{\frac{1}{2}} M_n^{\frac{1}{2}} \bar{\Theta}^R C \sum_{h=1}^{\infty} h^{-\bar{\beta}} \\
&= 2M_n^{\frac{1}{2}} \bar{\Theta}^R C H \\
&= O(M_n^{\frac{1}{2}})
\end{aligned}
$$

where $H = \sum_{h=1}^{\infty} h^{-\bar{\beta}} < \infty$ since $\bar{\beta} > 2$.

Collecting together the rates and simplifying the expression, we have

$$
\max_{k=L_n,\dots,T_n} \left\| \hat{Q}^{\boldsymbol{R}}_{k,n} - Q^{\boldsymbol{R}}_{P,k,n} \right\|_{\mathrm{tr}} = O_{\mathcal{P}}(r^{\delta}_{n,2}),
$$

where

$$
r^{\delta}_{n,2} = L_n^{\frac{1}{2}} T_n^{\frac{1}{2}} \tau_n M_n^{\frac{1}{2}} + T_n \tau_n M_n^{\frac{1}{2}} + L_n \tau_n^2 + L_n^{\frac{1}{2}} T_n^{\frac{1}{2}} \tau_n^2 + T_n \tau_n^2.
$$

Therefore, we have

$$
\hat{\delta}_{P,n} = O_{\mathcal{P}}(r^{\delta}_{n,1} + r^{\delta}_{n,2}),
$$

so the original offset condition will be met if

$$
\nu_n \gg \log(T_n) M_n \left( \tau_n^{-2} \left( T_n^{-\frac{1}{2}} M_n^{-\frac{1}{2}} (r^{\delta}_{n,2})^{\frac{1}{2}} + T_n^{-\frac{1}{2}} M_n^{-\frac{1}{2}} (r^{\delta}_{n,1})^{\frac{1}{2}} + T_n^{-1} \right) \right).
$$

Observe that

$$
\begin{aligned}
&T_n^{-\frac{1}{2}} M_n^{-\frac{1}{2}} (r^{\delta}_{n,1})^{\frac{1}{2}} + T_n^{-1} \\
&\leq T_n^{-\frac{1}{2}} M_n^{-\frac{1}{2}} \left( M_n^{\frac{1}{2}} (M_n^{\frac{1}{4}} L_n^{\frac{1}{4}} + T_n^{\frac{1}{4}} M_n^{\frac{1}{4}} L_n^{\frac{1}{4}} + T_n^{\frac{1}{2}} L_n^{-\frac{1}{2}} + T_n^{\frac{1}{2}} L_n^{1-\frac{\bar{\beta}^R}{2}}) \right) + T_n^{-1} \\
&= T_n^{-\frac{1}{2}} M_n^{\frac{1}{4}} L_n^{\frac{1}{4}} + T_n^{-\frac{1}{4}} M_n^{\frac{1}{4}} L_n^{\frac{1}{4}} + L_n^{-\frac{1}{2}} + L_n^{1-\frac{\bar{\beta}^R}{2}} + T_n^{-1} \\
&= \varphi_{n,1},
\end{aligned}
$$

which comes from the covariance estimation error.

Next, to deal with the prediction errors due to using the products of residuals, we have

$$
\begin{aligned}
&T_n^{-\frac{1}{2}} M_n^{-\frac{1}{2}} (r^{\delta}_{n,2})^{\frac{1}{2}} \\
&\leq T_n^{-\frac{1}{2}} M_n^{-\frac{1}{2}} \left( L_n^{\frac{1}{4}} T_n^{\frac{1}{4}} \tau_n^{\frac{1}{2}} M_n^{\frac{1}{4}} + T_n^{\frac{1}{2}} \tau_n^{\frac{1}{2}} M_n^{\frac{1}{4}} + L_n^{\frac{1}{2}} \tau_n + L_n^{\frac{1}{4}} T_n^{\frac{1}{4}} \tau_n + T_n^{\frac{1}{2}} \tau_n \right) \\
&= L_n^{\frac{1}{4}} T_n^{-\frac{1}{4}} \tau_n^{\frac{1}{2}} M_n^{-\frac{1}{4}} + \tau_n^{\frac{1}{2}} M_n^{-\frac{1}{4}} + T_n^{-\frac{1}{2}} M_n^{-\frac{1}{2}} L_n^{\frac{1}{2}} \tau_n + L_n^{\frac{1}{4}} T_n^{-\frac{1}{4}} M_n^{-\frac{1}{2}} \tau_n + M_n^{-\frac{1}{2}} \tau_n \\
&= \varphi_{n,2}
\end{aligned}
$$

The assumption on the offset condition (6) implies that

$$
\nu_n \gg \log(T_n) M_n \left( \tau_n^{-2} \left( \varphi_{n,1} + \varphi_{n,2} \right) \right),
$$

and therefore

$$
\sup_{P\in\mathcal{P}_n} \mathbb{P}_P(\hat{\Delta}_{P,n} \geq \frac{1}{16} T_n \nu_n \tau_n^2) \to 0.
$$

Combining the results from all of the previous steps, we obtain the final result

$$
\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} \mathbb{P}_P(S_n((\hat{\boldsymbol{R}}_{t,n})_{t\in\mathcal{T}_n}) > a_{\alpha-\nu_n}(\hat{Q}^{\boldsymbol{R}}_n) + \tau_n) \leq \alpha.
$$

□

# G Auxiliary Lemmas

The following result is a distribution-uniform version of Proposition 5.4 from Mies and Steland [MS22].

**Lemma G.1.** *Let $G_{t,n}$ satisfy Assumption 8 with $q \geq 2$. Denote*

$$\gamma_{P,t,n}(h) = \mathrm{Cov}_P[G_{t,n}(\boldsymbol{\epsilon}_0), G_{t,n}(\boldsymbol{\epsilon}_h)] \in \mathbb{R}^{d_n \times d_n}.$$

*Then*

$$\sup_{P \in \mathcal{P}_n} ||\gamma_{P,t,n}(h)||_{\mathrm{tr}} \leq \Theta_n^2 \sum_{j=h}^{\infty} j^{-\beta},$$

*where $|| \cdot ||_{\mathrm{tr}}$ denotes the trace norm. Hence, if $\beta > 2$, then the long-run covariance matrix $\gamma_{P,t,n} = \sum_{h=-\infty}^{\infty} \gamma_{P,t,n}(h)$ is well-defined for all $P \in \mathcal{P}_n$.*

*Proof of Lemma* : The result follows by the same steps in the proof from Proposition 5.4 in Mies and Steland [MS22] while carrying the supremum over $\mathcal{P}_n$. □

The following result is similar to the bounded convergence lemmas appearing as Lemma S1 in Lundborg et al. [Lun+22] and Lemma 25 in Shah and Peters [SP20].

**Lemma G.2.** *For each $n \in \mathbb{N}$, let $X_n$ be a real-valued random variable with distribution determined by $P \in \mathcal{P}_n$ where the collection of distributions $\mathcal{P}_n$ can change with $n$. Let $C > 0$ and suppose that $|X_n| \leq C$ for all $n \in \mathbb{N}$ and $X_n = o_{\mathcal{P}}(1)$. Then $\sup_{P \in \mathcal{P}_n} \mathbb{E}_P(|X_n|) = o(1)$.*

*Proof of Lemma G.2*: For any given $\epsilon > 0$,

$$|X_n| = |X_n| \mathbb{1}_{\{|X_n| > \epsilon\}} + |X_n| \mathbb{1}_{\{|X_n| \leq \epsilon\}} \leq C \mathbb{1}_{\{|X_n| > \epsilon\}} + \epsilon.$$

By the assumption that $X_n = o_{\mathcal{P}}(1)$, we can find some $N \in \mathbb{N}$ such that $\sup_{P \in \mathcal{P}_n} \mathbb{P}_P(|X_n| > \epsilon) < \epsilon / C$ for $n \geq N$. Hence, for $n \geq N$ we have

$$\sup_{P \in \mathcal{P}_n} \mathbb{E}_P(|X_n|) \leq C \sup_{P \in \mathcal{P}_n} \mathbb{P}_P(|X_n| > \epsilon) + \epsilon < 2\epsilon.$$

Since $\epsilon > 0$ was arbitrary, we obtain the desired result. □