

# Conditional Superior Predictive Ability\*

Jia Li<sup>†</sup>

Department of Economics, Duke University

Zhipeng Liao<sup>‡</sup>

Department of Economics, UCLA

Rogier Quaadvlieg<sup>§</sup>

Department of Business Economics, Erasmus School of Economics

February 11, 2020

## Abstract

This paper proposes a test for the conditional superior predictive ability (CSPA) of a family of forecast methods with respect to a benchmark. The test is functional in nature: Under the null hypothesis, the benchmark's conditional expected loss is no more than those of the competitors, uniformly across all conditioning states. By inverting the CSPA tests for a set of benchmarks, we obtain confidence sets for the uniformly most superior method. The econometric inference pertains to testing a system of conditional moment inequalities for time series data with general serial dependence, and we justify its asymptotic validity using a uniform nonparametric inference method based on a new strong approximation theory for mixingales. The usefulness of the method is demonstrated in empirical applications on volatility and inflation forecasting.

**Keywords:** conditional moment inequality, forecast evaluation, inflation, intersection bounds, machine learning, volatility.

**JEL Code:** C14, C22.

---

\*We thank Raffaella Giacomini, Jinyong Hahn, Peter Reinhard Hansen, Francesca Molinari, Hyungsik Roger Moon and conference and seminar participants at CREST, the 2018 Triangle Econometrics Conference, Southern California Winter Econometrics Day, and the 2019 Toulouse Financial Econometrics Conference for their comments. Liao's research was partially supported by National Science Foundation Grant SES-1628889. Quaadvlieg was financially supported by the Netherlands Organisation for Scientific Research (NWO) Grant 451-17-009.

<sup>†</sup>Corresponding author: Box 90097 Duke University, Durham, NC 27707, USA; e-mail: jl410@duke.edu.

<sup>‡</sup>Department of Economics, UCLA, Log Angeles, CA 90095, USA; e-mail: zhipeng.liao@econ.ucla.edu.

<sup>§</sup>P.O. Box 1738, 3000DR Rotterdam, Netherlands; e-mail: quaadvlieg@ese.eur.nl.

# 1 Introduction

A central problem in time-series econometrics is forecasting economic quantities, such as GDP growth, inflation, stock returns, and volatility. Empiricists often face an extensive list of “reasonable” candidate forecast methods that are invariably backed by influential prior studies. For example, classical ARMA- and ARCH-type models contain numerous variants, and the recent trend of using machine-learning algorithms—such as LASSO, random forests, support vector machines, and deep neural nets—can make a forecaster’s choice even more difficult. Rigorously evaluating the relative performance of these methods, and identifying superior ones, is thus of great and ever-growing importance.

The most popular forecast evaluation method is, arguably, Diebold and Mariano’s (1995) test. Under the null hypothesis of the Diebold–Mariano test, two competing methods have the same unconditional expected loss, and the test can be carried out using a simple t-test. More generally, a multivariate extension of the Diebold–Mariano test can be used to test *unconditional equal predictive ability (UEPA)* concerning multiple competing forecasts, which amounts to testing a system of unconditional moments (i.e., expected loss differentials) being zero.

Two significant extensions of the Diebold–Mariano test have been developed in the literature. The first is the test for *unconditional superior predictive ability (USPA)*, which is first studied by White (2000) and later refined by Hansen (2005). The null hypothesis states that a benchmark method weakly outperforms a collection of competing alternatives, as formalized by a system of unconditional moment inequalities. In his seminal work, White (2000) proposes critical values under the least favorable null (i.e., all inequalities are binding), which effectively reduces the USPA test into a multivariate version of the Diebold–Mariano test. Hansen makes an important observation that White’s test can be very conservative when there are competing models that are clearly inferior to the benchmark. To remedy this issue, Hansen proposes a preliminary selection based on studentized moment conditions to remove these clearly inferior methods, and shows that this can significantly improve the test’s statistical power.

The second extension is the *conditional equal predictive ability (CEPA)* test proposed by Giacomini and White (2006). The authors eloquently argue that, in practice, a forecast evaluator is not only interested in knowing whether one method is better than another on average, but also interested in when this occurs. This consideration is particularly relevant when the methods on the forecaster’s “shortlist” appear similar on average, but can actually behave very differently con-

ditional on certain economic states. The CEPA null hypothesis states that conditional expected loss functions of different forecast methods are identically the same across all conditioning states. The corresponding econometric inference in principle concerns global features of the conditional expected loss functions. Giacomini and White, however, do not directly attack this functional inference problem. Instead, they propose a practical method based on a fixed number of (instrumented) unconditional moments implied by the original conditional ones. As such, the Giacomini–White test is operationally the same as a finite-dimensional UEPA test.

Set against this background, we extend these existing evaluation paradigms by proposing a test for *conditional superior predictive ability (CSPA)*, which synergizes the key insights of White (2000), Hansen (2005), and Giacomini and White (2006). Specifically, the CSPA null hypothesis asserts the uniform (weak) superiority of the benchmark method, in the sense that the conditional expected loss of the benchmark does not exceed those of the competing forecasts across all conditioning states. On the other hand, a rejection of the CSPA null hypothesis suggests that some competing alternative method outperforms the benchmark in certain states, which are revealed as a by-product of the testing procedure, providing useful diagnostic information.

Our CSPA test formally pertains to testing inequalities for forecast methods’ conditional expected loss functions. To implement the test, we nonparametrically estimate the conditional mean function using the series method (Andrews (1991a), Newey (1997), Chen (2007)), and then adopt the intersection-bound methodology (Chernozhukov, Lee, and Rosen (2013)) to conduct inference regarding the functional inequalities. It is well-known that the underlying uniform inference problem is non-Donsker, for which conventional functional central limit theorems based on the weak convergence concept are not applicable. In a setting with independent data, Chernozhukov, Lee, and Rosen (2013) address this issue by using Yurinskii’s coupling, which provides a strong Gaussian approximation for the growing-dimensional moment conditions in the series estimation. As a result, the t-statistic process (indexed by the conditioning state variable) can be strongly approximated by a divergent Gaussian process that can be used to construct critical values.

In this paper, we also adopt the strong approximation strategy to make inference on the conditional moment inequalities, but in a more general time-series setting. One possible way to achieve this is to invoke the strong approximation theory recently developed by Li and Liao (2019), which provides a Yurinskii coupling for general dependent data modeled as mixingales, and then proceed as in Chernozhukov, Lee, and Rosen (2013). However, a drawback of this approach is that Yurinskii’s coupling concerns the approximation for the entire sample moment vector (under Euclidean distance), and hence, occurs at a slow rate of convergence. This in turn leads to relatively restrictive conditions on how fast the number of approximating series terms,  $m_n$ , can grow with

the sample size  $n$ . Specifically, in both the independent-data setting of Chernozhukov, Lee, and Rosen (2013) and the time-series setting of Li and Liao (2019), Yurinskii’s coupling is available when  $m_n = o(n^{1/5})$ . This issue can be more severe in the time-series context (which is also the setting here), as the requisite restriction on  $m_n$  becomes more stringent when the data is serially “more dependent.”

Motivated by this issue, in this paper we develop a new strong approximation theory in the time-series setting which substantially improves the theory of Li and Liao (2019) for making uniform series inference in the time-series context. Rather than strongly approximating the entire vector of moment conditions, we instead establish a strong approximation for the “maximum projection” of this growing-dimensional statistic along a large number of directions. This new result is weaker than Yurinskii’s coupling, and it can be established under notably weaker restrictions on the number of series terms. The “cost” of not coupling the entire vector of sample moments is that one can no longer construct a strong approximation for the t-statistic process associated with the (functional) series estimator, which is needed as a high-level condition in Chernozhukov, Lee, and Rosen’s (2013) theory (see Condition C.2 of that paper). That being said, we show that the general framework of intersection-bound inference can nevertheless be adapted to accommodate this weaker notion of coupling. Our theory thus extends that of Chernozhukov, Lee, and Rosen (2013) by both allowing for time-series dependence in the data and a weaker notion of strong approximation. This theory may be further extended to settings with spatial dependence, but that extension is beyond the scope of the present paper on forecast evaluation.

These new econometric ingredients suggest that the proposed (nonparametric) CSPA test differs from the conventional forecast evaluation methods not only in concept, but also in econometric technicality: The unconditional tests of Diebold and Mariano (1995), White (2000), and Hansen (2005) concern a fixed number of unconditional expected losses; Giacomini and White’s (2006) CEPA hypothesis is nonparametric in nature, but they only test a fixed number of implied unconditional moments. In contrast, the CSPA test directly addresses the functional inference by adopting and generalizing recent results from the partial identification literature (see Molinari (2019) for a recent review).<sup>1</sup> CSPA unifies these prior evaluation paradigms in the following theoretical sense: USPA is its special case with empty conditioning information set, and CEPA corresponds to the least favorable null hypothesis of CSPA. It is interesting to note that, like Hansen’s (2005) USPA

---

<sup>1</sup>Our theory is a complement, instead of substitute, of Giacomini and White’s (2006) test. In fact, the strong approximation theory developed here can be used to provide a nonparametric interpretation of Giacomini and White’s test, by allowing the number of instruments to grow with the sample size. We do not develop this explicitly here so as to be focused on the CSPA test.

test, our CSPA test also involves a preliminary selection that removes clearly inferior forecasts in the computation of critical values and, in the same spirit, it also removes regions in the state space on which the competing forecasts are clearly inferior, which is unique to our functional inference problem.

We apply the CSPA test in two important empirical settings to demonstrate its usefulness. The first pertains to the evaluation of volatility forecasts, which is one of the most important topics in financial econometrics. We consider a variety of autoregressive (AR) models for realized volatility, including: AR(1), AR(2) with or without adaptive variable selection, fractionally integrated AR, HAR (Corsi (2009)) and its HARQ extension (Bollerslev, Patton, and Quaedvlieg (2016, 2018)). Consistent with prior studies, we find that HARQ is generally superior to the other methods using Hansen’s (2005) USPA test. But the CSPA test provides further useful diagnostic information. We find that in many cases, we cannot reject the CSPA null hypothesis that HARQ weakly dominates the other methods uniformly across different states. Interestingly, the conditional test also reveals cases in which alternative methods—particularly the fractionally integrated model and, somewhat surprisingly, the simple AR(1) model—significantly outperforms HARQ over certain regions of the state space. With the CSPA criterion, we pose a new challenge for the empirical search of “uniformly” superior volatility forecasting methods, for which the proposed test can be used to formally run the horse race.

In the second empirical application, we evaluate inflation forecasts in a macroeconomic setting. We consider eight forecasting methods for monthly inflation based on the recent work of Medeiros, Vasconcelos, Veiga, and Zilberman (2019). Four of these methods are traditional inflation models such as factor model and Bayesian vector autoregression. The other four are machine-learning algorithms. Using Hansen’s (2005) USPA test, we find that the traditional models are typically beaten by at least one of the four machine-learning methods, but the latter methods appear to be virtually indistinguishable judged by (unconditional) average loss. In contrast, the CSPA test clearly distinguishes them for a wide variety of economically important conditioning variables. In particular, the bagging (Breiman (1996)) forecasts turn out to perform extremely poorly in the tail regions of many of our conditioning variables, such as the macro uncertainty measure of Jurado, Ludvigson, and Ng (2015).

The paper is organized as follows. Section 2 describes our test and asymptotic theory. Section 3 reports simulation results. Sections 4 and 5 present the empirical applications on volatility and inflation forecasting, respectively. Section 6 concludes. The appendix contains proofs for our main theoretical results. Additional technical results and proofs are detailed in the Supplemental Appendix to this paper.

## 2 Testing for conditional superior predictive ability

We present the theory for the CSPA test in this section. Section 2.1 introduces the hypotheses of interest. In Section 2.2, we describe the CSPA test and establish its asymptotic validity under high-level conditions. Section 2.3 further shows how to invert the CSPA test to obtain confidence sets for the most superior forecast method. Our econometric inference relies on a new coupling theory established in Section 2.4, which may be skipped by readers who are mainly interested in applications. Below, for two real sequences  $a_n$  and  $b_n$ , we write  $a_n \asymp b_n$  if  $b_n/C \leq a_n \leq Cb_n$  for some constant  $C > 1$ .

### 2.1 Forecast evaluation hypotheses

Let  $(F_t^\dagger)_{t \geq 1}$  be the time series to be forecast. We consider a benchmark forecast sequence  $(F_{0,t})_{1 \leq t \leq n}$  and  $J$  competing forecast sequences  $(F_{j,t})_{1 \leq t \leq n}$ ,  $1 \leq j \leq J$ . With a loss function  $L(\cdot, \cdot)$ , we measure the performance of the benchmark method relative to the  $j$ th competing alternative by the loss differential sequence defined as

$$Y_{j,t} \equiv L(F_t^\dagger, F_{j,t}) - L(F_t^\dagger, F_{0,t}). \quad (2.1)$$

In particular,  $Y_{j,t} \geq 0$  indicates that the benchmark (weakly) outperforms method  $j$  in period  $t$ .

Two types of null hypotheses are commonly tested in forecast evaluation. One is the hypothesis of unconditional equal predictive ability (UEPA):

$$H_0^{UEPA} : \mathbb{E}[Y_{j,t}] = 0, \quad 1 \leq j \leq J, \quad (2.2)$$

that is, the benchmark has the same expected performance as all competing alternatives. The other is the hypothesis of unconditional superior predictive ability (USPA):

$$H_0^{USPA} : \mathbb{E}[Y_{j,t}] \geq 0, \quad 1 \leq j \leq J, \quad (2.3)$$

meaning that the benchmark weakly outperforms the others. UEPA and USPA are clearly connected, in that the former is the least-favorable null of the latter.

The unconditional tests are informative about the relative performance of forecast methods on average. As such, they have a “blind spot:” Two methods may appear to have (statistically) identical performance on average, but can behave very differently given certain economic conditions. Giacomini and White (2006) advocate testing the hypothesis of conditional equal predictive ability (CEPA), that is,

$$H_0^{CEPA} : \mathbb{E}[Y_{j,t}|X_t = x] = 0, \quad x \in \mathcal{X}, \quad 1 \leq j \leq J, \quad (2.4)$$

where  $X_t$  is a conditioning state variable chosen by the evaluator, and  $\mathcal{X}$  specifies the conditioning region as a subset of the domain of  $X$ . For example, one can track forecast methods' performance through business cycles by setting  $X_t$  to be a cyclical indicator (e.g., GDP growth). The CEPA null hypothesis then states that the benchmark performs equally well as all competing alternatives, not only on average, but also through the ups-and-downs of the economy. This hypothesis is rejected if some competing forecast method performs differently than the benchmark in some states (say, expansion or recession).

A rejection of the CEPA hypothesis is not directly informative about whether the competing methods is better or worse than the benchmark—it only signifies their difference. In contrast, we consider the *conditional superior predictive ability (CSPA)* hypothesis. The null hypothesis asserts

$$H_0^{CSPA} : \mathbb{E}[Y_{j,t}|X_t = x] \geq 0, \quad x \in \mathcal{X}, \quad 1 \leq j \leq J. \quad (2.5)$$

This imposes a very stringent requirement on the benchmark, that is, it needs to weakly dominate all competing methods across the conditioning region  $\mathcal{X}$ . Therefore, “passing” the CSPA test should be deemed a highly desirable feature of a forecasting method.

Introducing CSPA to the forecast evaluation literature seems to be rather natural and conceptually straightforward: CSPA is to CEPA simply as USPA is to UEPA. However, testing this new hypothesis is fundamentally different from—and econometrically much more complicated than—those in the prior forecast evaluation literature. To see why, note that unconditional tests only concern the finite-dimensional vector  $(\mathbb{E}[Y_{j,t}])_{1 \leq j \leq J}$ . In contrast, conditional tests involve functional inference for conditional expectation functions of loss differentials given by

$$h_j(x) \equiv \mathbb{E}[Y_{j,t}|X_t = x], \quad 1 \leq j \leq J.$$

The related functional inference is theoretically nontrivial because it requires knowledge about the global behavior of the  $h_j$  functions. In their pioneering work, however, Giacomini and White (2006) bypassed the functional inference by instead testing certain implications of CEPA. These authors take as given a finite-dimensional instrument  $W_t$  that is measurable with respect to the  $\sigma$ -field generated by  $X_t$ , and derive from (2.4) the following unconditional moment equalities:

$$\mathbb{E}[Y_{j,t}W_t] = 0, \quad 1 \leq j \leq J.$$

These can then be tested by using a conventional Wald test.

Unlike Giacomini and White (2006), we attack the functional inference problem directly in our study of CSPA. Our approach relies on inference methods recently developed in the partial identification literature, particularly those concerning conditional moment inequalities. We adopt

the intersection-bound approach originally proposed by Chernozhukov, Lee, and Rosen (2013) for microeconomic applications, and extend it to a general time-series setting. More precisely, we rewrite the CSPA hypothesis as

$$H_0^{CSPA} : \eta^* \equiv \min_{1 \leq j \leq J} \inf_{x \in \mathcal{X}} h_j(x) \geq 0. \quad (2.6)$$

For some significance level  $\alpha \in (0, 1/2)$ , we can construct a  $1 - \alpha$  upper confidence bound  $\hat{\eta}_{n,1-\alpha}$  for  $\eta^*$  such that

$$\liminf_{n \rightarrow \infty} \mathbb{P}(\eta^* \leq \hat{\eta}_{n,1-\alpha}) \geq 1 - \alpha. \quad (2.7)$$

Consequently, a test that rejects the CSPA null hypothesis when  $\hat{\eta}_{n,1-\alpha} < 0$  has probability of type-I error bounded by  $\alpha$  in large samples. The feasible computation of  $\hat{\eta}_{n,1-\alpha}$  and the theoretical properties of the test are detailed in Section 2.2 below.

## 2.2 The CSPA test and its asymptotic properties

In this subsection, we describe how to implement the CSPA test and establish its theoretical validity (see Proposition 1). For readability, we present the theory under high-level conditions, most of which are standard for series-based nonparametric inference and are well understood in the literature. The key exception is a condition for coupling a maximum projection of a growing-dimensional sample moment constructed using dependent data, for which we develop a new theory in Section 2.4.

To perform the CSPA test, we first estimate the  $h_j(\cdot)$  functions nonparametrically by the least-square series regression. Let  $P(x) = (p_1(x), \dots, p_{m_n}(x))^\top$  be an  $m_n$ -dimensional vector of approximating basis functions, such as polynomial, Fourier, spline, and wavelet series; see Chen (2007) for a comprehensive review. By convention, we suppose that  $P(\cdot)$  always contains the constant function by setting  $p_1(\cdot) \equiv 1$ . To conduct series estimation, for each  $j$ , we regress  $Y_{j,t}$  on  $P(X_t)$  and obtain the regression coefficient

$$\hat{b}_{j,n} \equiv \hat{Q}_n^{-1} \left( n^{-1} \sum_{t=1}^n P(X_t) Y_{j,t} \right), \quad \text{where} \quad \hat{Q}_n \equiv n^{-1} \sum_{t=1}^n P(X_t) P(X_t)^\top.$$

The functional estimator for  $h_j(\cdot)$  is then given by

$$\hat{h}_{j,n}(\cdot) \equiv P(\cdot)^\top \hat{b}_{j,n}.$$

The series regression coefficient  $\hat{b}_{j,n}$  formally resembles the conventional least-square estimator, but corresponds to a very different theory. We require the number of series terms  $m_n \rightarrow \infty$  asymptotically so that the unknown  $h_j(\cdot)$  function can be approximated sufficiently well by a



large number of approximating functions. The growing dimension of  $\hat{b}_{j,n}$  renders the conventional weak-convergence-based characterization of asymptotic normality inappropriate. This is why we shall need a strong approximation theory for growing-dimensional statistics to construct uniform inference, as in Chernozhukov, Lee, and Rosen (2013), Belloni, Chernozhukov, Chetverikov, and Kato (2015), and Li and Liao (2019).

We now proceed to describe the inference procedure. Let  $u_t \equiv (u_{1,t}, \dots, u_{J,t})^\top$ , where  $u_{j,t}$  is the nonparametric regression error term defined as

$$u_{j,t} \equiv Y_{j,t} - h_j(X_t).$$

We further set  $A_n$  to be the  $Jm_n \times Jm_n$  covariance matrix of  $n^{-1/2} \sum_{t=1}^n u_t \otimes P(X_t)$ , that is,

$$A_n \equiv \text{Var} \left( n^{-1/2} \sum_{t=1}^n u_t \otimes P(X_t) \right), \quad (2.8)$$

where  $\otimes$  denotes the Kronecker product. To conduct feasible inference, we suppose that a heteroskedasticity and autocorrelation consistent (HAC) estimator  $\hat{A}_n$  for  $A_n$  is available and satisfies the following condition, where  $\|\cdot\|_S$  denotes the matrix spectral norm.

**Assumption 1.**  $\|\hat{A}_n - A_n\|_S = O_p(\delta_{A,n})$  for some sequence  $\delta_{A,n} \rightarrow 0$  at polynomial rate.<sup>2</sup>

Assumption 1 is high-level and in fact nonstandard, because it concerns the convergence rate of HAC estimators with growing dimensions (i.e.,  $Jm_n \rightarrow \infty$ ), whereas classical HAC estimation theory (e.g., Newey and West (1987) and Andrews (1991b)) are developed in settings with fixed dimensions. In the present growing-dimensional setting, the consistency of the HAC estimator is not enough for feasible inference, and we need a stronger form of consistency (i.e., with polynomial rate) as stated by the assumption above.

A theoretically valid choice of  $\hat{A}_n$  that verifies Assumption 1 is the Newey–West type HAC estimator (see Theorem 6 of Li and Liao (2019)). However, Newey–West estimators may lead to nontrivial size distortions in finite samples. This is well-known in the HAC estimation literature, and we also document a similar issue in our Monte Carlo experiments. To remedy this finite-sample distortion, in this paper we analyze a more general class of pre-whitened HAC estimators in the spirit of Andrews and Monahan (1992), and characterize their convergence rates in the growing-dimensional setting. We find that the pre-whitened estimator indeed leads to better size control. The theory on the pre-whitened estimator, however, is somewhat tangential to our main result on CSPA testing, and it also requires a technical setup that is not used elsewhere in the paper. To

---

<sup>2</sup>That is,  $\delta_{A,n} \asymp n^{-a}$  for some fixed constant  $a > 0$  that may be arbitrarily small.

remain focused, we relegate all details on the pre-whitened estimator to Supplemental Appendix SC.

Equipped with the estimator  $\widehat{A}_n$ , we can estimate the  $Jm_n \times Jm_n$  covariance matrix of the normalized estimators  $(n^{1/2}(\widehat{b}_{j,n} - b_{j,n}^*))_{1 \leq j \leq J}$  via

$$\widehat{\Omega}_n \equiv \left( I_J \otimes \widehat{Q}_n \right)^{-1} \widehat{A}_n \left( I_J \otimes \widehat{Q}_n \right)^{-1},$$

where  $I_J$  denotes the  $J \times J$  identity matrix, and  $b_{j,n}^*$  is the ‘‘population analogue’’ of  $\widehat{b}_{j,n}$  that is formally introduced in Assumption 2 below. We further partition  $\widehat{\Omega}_n$  into  $J \times J$  blocks of  $m_n \times m_n$  submatrices  $\widehat{\Omega}_n(j, k)$ ,  $1 \leq j, k \leq J$ . Note that  $\widehat{\Omega}_n(j, k)$  is the estimator of the covariance matrix between  $n^{1/2}(\widehat{b}_{j,n} - b_{j,n}^*)$  and  $n^{1/2}(\widehat{b}_{k,n} - b_{k,n}^*)$ . The standard deviation function of  $n^{1/2}(\widehat{h}_{j,n}(x) - h_j(x))$  is then estimated by

$$\widehat{\sigma}_{j,n}(x) \equiv (P(x)^\top \widehat{\Omega}_n(j, j) P(x))^{1/2}.$$

Algorithm 1, below, provides the implementation details of the CSPA test, which is based on the intersection-bound inference of Chernozhukov, Lee, and Rosen (2013).

**Algorithm 1 (Implementation of the CSPA Test).**

Step 1. Simulate a  $Jm_n$ -dimensional random vector  $(\xi_1^{*\top}, \dots, \xi_J^{*\top})^\top \sim \mathcal{N}(0, \widehat{\Omega}_n)$ , where each  $\xi_j^*$  is  $m_n$ -dimensional. Set  $\widehat{t}_{j,n}^*(x) \equiv P(x)^\top \xi_j^* / \widehat{\sigma}_{j,n}(x)$ .

Step 2. Repeat step 1 many times. Set  $\tilde{\gamma}_n \equiv 1 - 0.1/\log(n)$ . Let  $\widehat{K}_n$  be the  $\tilde{\gamma}_n$ -quantile of  $\max_{1 \leq j \leq J} \sup_{x \in \mathcal{X}} \widehat{t}_{j,n}^*(x)$  in the simulated sample and then set

$$\widehat{\mathcal{V}}_n \equiv \left\{ (j, x) : \widehat{h}_{j,n}(x) \leq \min_{1 \leq j \leq J} \inf_{x \in \mathcal{X}} \left( \widehat{h}_{j,n}(x) + n^{-1/2} \widehat{K}_n \widehat{\sigma}_{j,n}(x) \right) + 2n^{-1/2} \widehat{K}_n \widehat{\sigma}_{j,n}(x) \right\}.$$

Step 3. Set  $\widehat{k}_{n,1-\alpha}$  as the  $(1 - \alpha)$ -quantile of  $\sup_{(j,x) \in \widehat{\mathcal{V}}_n} \widehat{t}_{j,n}^*(x)$  and set

$$\widehat{\eta}_{n,1-\alpha} = \min_{1 \leq j \leq J} \inf_{x \in \mathcal{X}} \left[ \widehat{h}_{j,n}(x) + n^{-1/2} \widehat{k}_{n,1-\alpha} \widehat{\sigma}_{j,n}(x) \right].$$

Reject the CSPA null hypothesis at significance level  $\alpha$  if  $\widehat{\eta}_{n,1-\alpha} < 0$ . □

It is instructive to provide some intuition for this procedure. The Gaussian variables  $(\xi_j^*)_{1 \leq j \leq J}$  approximate the distribution of  $(n^{1/2}(\widehat{b}_{j,n} - b_{j,n}^*))_{1 \leq j \leq J}$  and  $(\widehat{t}_{j,n}^*(\cdot))_{1 \leq j \leq J}$  mimic the law of the t-statistic processes associated with  $(\widehat{h}_{j,n}(\cdot))_{1 \leq j \leq J}$ . Step 2 of the algorithm implements the adaptive inequality selection: We jointly select  $j \in \{1, \dots, J\}$  and  $x \in \mathcal{X}$  into the set  $\widehat{\mathcal{V}}_n$  such that, with probability approaching one,  $h_j(x)$  is minimized on  $\widehat{\mathcal{V}}_n$ . For example, if the entire function  $\widehat{h}_{j,n}(\cdot)$  is ‘‘far above’’ the infimum  $\eta^*$  for some  $j$ , then the corresponding inequality is removed from our

subsequent inference. Similarly, for the remaining inequalities, we also remove subsets of  $\mathcal{X}$  on which  $\hat{h}_{j,n}(\cdot)$  is “far above”  $\eta^*$ . Step 3 of the algorithm produces the critical value  $\hat{\eta}_{n,1-\alpha}$ .

We are now ready to present the asymptotic theory that justifies the validity of the CSPA test described in Algorithm 1 above. As mentioned in the introduction, our theory differs from Chernozhukov, Lee, and Rosen (2013) because we allow for general time series dependence and use a less restrictive notion of strong approximation. For ease of discussion, we collect the key ingredients of the theorem in the following two high-level assumptions. Below, we denote  $\zeta_n \equiv \max_{1 \leq j \leq m_n} \sup_{x \in \mathcal{X}} |p_j(x)|$  and  $\zeta_n^L \equiv \sup_{x_1, x_2 \in \mathcal{X}, x_1 \neq x_2} \|P(x_1) - P(x_2)\| / \|x_1 - x_2\|$ .

**Assumption 2.** *Suppose: (i) for each  $j = 1, \dots, J$ ,  $h_j(\cdot)$  is a continuous function on a compact subset  $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ ; (ii) there exist sequences  $(b_{j,n}^*)_{n \geq 1}$  of  $m_n$ -dimensional constant vectors, with  $m_n \rightarrow \infty$  at polynomial rate, such that*

$$\max_{1 \leq j \leq J} \sup_{x \in \mathcal{X}} n^{1/2} \left| h_j(x) - P(x)^\top b_{j,n}^* \right| = o_p((\log n)^{-1}),$$

*(iii) the eigenvalues of  $Q_n$  and  $A_n$  are bounded from above and away from zero uniformly over  $n$ ; (iv)  $\|\hat{Q}_n - Q_n\|_S = O_p(\delta_{Q,n})$  for  $\delta_{Q,n} = o(m_n^{-1/2}(\log n)^{-1})$ ; (v)  $\zeta_n m_n n^{-1/2} = o(1)$ ; and (vi)  $\log(\zeta_n^L) = O(\log n)$ .*

The conditions in Assumption 2 are fairly standard for series estimation; see, for example, Andrews (1991b), Newey (1997), Chen (2007), Chernozhukov, Lee, and Rosen (2013), and Belloni, Chernozhukov, Chetverikov, and Kato (2015). In particular, condition (ii) specifies the precision for approximating the unknown function  $h_j(\cdot)$  via approximating functions. This condition implicitly requires that the function  $h_j(\cdot)$  is sufficiently smooth, for which well-known results are available from numerical approximation theory. Condition (iv) imposes a mild convergence rate condition on  $\hat{Q}_n$ , which can be verified under primitive conditions.<sup>3</sup>

**Assumption 3.** *For any sequence of integers  $L_n = O((\log n)^2 m_n^{1/2} \zeta_n^L)^{d_x}$  and any collection of uniformly bounded vectors  $(\alpha_l)_{1 \leq l \leq L_n} \subseteq \mathbb{R}^{J m_n}$ , there exists a sequence of random variables  $\tilde{U}_n$  such that*

$$\left| \max_{1 \leq l \leq L_n} \alpha_l^\top \left( n^{-1/2} \sum_{t=1}^n u_t \otimes P(X_t) \right) - \tilde{U}_n \right| = o_p((\log n)^{-1}), \quad (2.9)$$

*and  $\tilde{U}_n$  has the same distribution as  $\max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{N}_n$  for some generic Gaussian vector  $\tilde{N}_n \sim \mathcal{N}(0, A_n)$ .*

<sup>3</sup>See, for example, Lemma 2.2 of Chen and Christensen (2015) and Lemma B5 in the supplemental appendix of Li and Liao (2019).

Assumption 3 is the key to the uniform functional inference underlying the CSPA test, and is nontrivial to establish. Note that our series estimation is based on the  $Jm_n$ -dimensional moment condition  $\mathbb{E}[u_t \otimes P(X_t)] = 0$ . The assumption above states that the maximum projection of the normalized growing-dimensional sample moment  $n^{-1/2} \sum_{t=1}^n u_t \otimes P(X_t)$  can be approximated by  $\tilde{U}_n$ , which has the same distribution as  $\max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{N}_n$ . In contrast, Yurinskii's coupling provides a strong approximation for the entire vector in Euclidean norm, namely,

$$\left\| n^{-1/2} \sum_{t=1}^n u_t \otimes P(X_t) - \tilde{N}_n \right\| = o_p\left(\log(n)^{-1}\right), \quad (2.10)$$

which is clearly stronger than (2.9), but it invariably also demands more restrictive regularity conditions. An important part of our theoretical analysis (see Section 2.4) is to construct the coupling in (2.9) for general heterogeneous mixingales under substantially weaker conditions on the growth rate of  $m_n$  than those employed in Li and Liao (2019) for establishing Yurinskii's coupling in a similar time series setting.

A ‘‘cost’’ of using the weaker coupling condition (2.9), instead of (2.10), is that we do not have a strong Gaussian approximation for the entire t-statistic process

$$\left( \frac{n^{1/2} \hat{h}_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} \right)_{1 \leq j \leq J, x \in \mathcal{X}},$$

which is required in Chernozhukov, Lee, and Rosen's (2013) intersection-bound theory as a high-level condition (see their Condition C.2). Consequently, we cannot directly invoke the theory from that prior work. Nevertheless, we show that under Assumption 3, one can still construct strong approximations for the supremum of the t-statistic process over all subsets of  $\mathcal{X}$  (see Theorem A1 in the appendix), which turns out to be enough for establishing the validity of the testing procedure.

The asymptotic properties of the CSPA test are described by the following proposition.

**Proposition 1.** *Suppose that Assumptions 1, 2, and 3 hold. Then, the CSPA test at significance level  $\alpha \in (0, 1/2)$  satisfies the following:*

(a) *Under the null hypothesis with  $\eta^* \geq 0$ , the test has asymptotic size  $\alpha$ , that is,*

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\hat{\eta}_{n,1-\alpha} < 0) \leq \alpha;$$

(b) *Under the alternative hypothesis with  $\eta^* < 0$ , the test has asymptotic power one, that is,*

$$\mathbb{P}(\hat{\eta}_{n,1-\alpha} < 0) \rightarrow 1.$$

COMMENTS. (i) Part (a) of Proposition 1 shows that the CSPA test controls size under the null hypothesis. As is common for testing inequalities, the test may be conservative, that is, the asymptotic rejection probability may be less than  $\alpha$ .

(ii) Part (b) shows that the test is consistent against fixed alternatives. As shown in our proof, this result follows from  $\hat{\eta}_{n,1-\alpha} - \eta^* = O_p(\zeta_n m_n n^{-1/2}) = o_p(1)$ . The proof can be straightforwardly adapted to show that the test is consistent against local alternatives with  $\eta^* < 0$  drifting to zero at rate strictly slower than  $\zeta_n m_n n^{-1/2}$ .  $\square$

### 2.3 Confidence sets for the most superior forecast method

The CSPA test described in the previous subsection concerns the comparison of a benchmark method with the other competing alternatives. In many applications, however, it may be a priori unclear which forecast method should be chosen as the benchmark, and the empirical researcher may naturally experiment with different choices. This practice can be formalized as constructing a model confidence set for the most superior forecast, as we discuss in this subsection.

Formally, we define a partial order  $\succeq$  between two forecast methods indexed by  $j$  and  $k$  as

$$j \succeq k \iff \mathbb{E} \left[ L(F_t^\dagger, F_{j,t}) | X_t = x \right] \leq \mathbb{E} \left[ L(F_t^\dagger, F_{k,t}) | X_t = x \right] \text{ for all } x \in \mathcal{X}.$$

That is, the expected forecast loss of method  $j$  is less than that of method  $k$  across all conditioning states. The set of the most superior methods is then defined as

$$\mathcal{M} \equiv \{0 \leq j \leq J : j \succeq k \text{ for all } 0 \leq k \leq J\}. \quad (2.11)$$

Note that the CSPA null hypothesis with method  $j$  being the benchmark can then be written as  $j \in \mathcal{M}$ . Since the  $\succeq$  order—which is defined using conditional expectation functions—is generally not complete, the set  $\mathcal{M}$  may be empty (i.e., there exists no method that weakly dominates all others).

It is interesting to contrast  $\mathcal{M}$  with its unconditional special case, that is,

$$\mathcal{M}_U \equiv \{0 \leq j \leq J : \mathbb{E}[L(F_t^\dagger, F_{j,t})] \leq \mathbb{E}[L(F_t^\dagger, F_{k,t})] \text{ for all } 0 \leq k \leq J\}.$$

Since  $\mathcal{M}_U$  relies on ordering the scalar-valued expected losses, it is guaranteed to be nonempty. It is also obvious that  $\mathcal{M} \subseteq \mathcal{M}_U$  and, in general, we expect the inclusion to be strict. By imposing a stronger notion of dominance,  $\mathcal{M}$  provides a refinement relative to its unconditional counterpart  $\mathcal{M}_U$ .

An Anderson–Rubin type confidence set for the most superior method can be constructed by inverting the CSPA test. We set

$$\widehat{\mathcal{M}}_{n,1-\alpha} = \{0 \leq j \leq J : \text{the } \alpha\text{-level CSPA test} \tag{2.12}$$

with method  $j$  as the benchmark does not reject\}.

By the duality between tests and confidence sets, Proposition 1 directly implies that for each  $j^* \in \mathcal{M}$ ,

$$\liminf_{n \rightarrow \infty} \mathbb{P} \left( j^* \in \widehat{\mathcal{M}}_{n,1-\alpha} \right) \geq 1 - \alpha.$$

We refer to  $\widehat{\mathcal{M}}_{n,1-\alpha}$  as the *confidence set for the most superior (CSMS)*.

We stress that the CSMS is distinct from the model confidence set (MCS) proposed by Hansen, Lunde, and Nason (2011) in two ways. First, the CSMS is based on conditional tests, while the MCS is based on unconditional ones; note that the unconditional test is a special case of the conditional test with  $X_t$  being empty. Second, the CSMS,  $\widehat{\mathcal{M}}_{n,1-\alpha}$ , is designed to cover each element  $j^*$  in  $\mathcal{M}$ , instead of the whole set  $\mathcal{M}$ . In contrast, the MCS provides coverage for the  $\mathcal{M}_U$  set. Of course, this distinction is only relevant when  $\mathcal{M}$  contains more than one method. While the unconditional expected losses of two distinct forecast methods might be identical (as real numbers) and result in a non-singleton  $\mathcal{M}_U$ , it is hard to conceive a scenario in which two different forecast sequences share exactly the same conditional expected loss across all states in  $\mathcal{X}$  (as functions). For this reason, we argue that covering each most superior method serves essentially the same empirical goal as covering the whole set  $\mathcal{M}$  in the context of conditional testing. The CSMS may thus be considered as the conditional extension of Hansen, Lunde, and Nason’s (2011) MCS.

## 2.4 Strong approximation for the maximum projection

In this subsection, we establish a strong approximation that can be used to verify the high-level Assumption 3. Since this type of coupling result is of independent theoretical interest, and is broadly useful for other types of nonparametric uniform inference in time-series analysis, we present the theory in a general setting. This subsection may be skipped by readers who are mainly interested in the application of the CSPA test.

We now turn to the setting. Let  $\|\cdot\|_q$  denote the  $L_q$ -norm of a random variable for  $q \geq 1$ . We consider an  $M_n$ -dimensional  $L_q$ -mixingale array  $(X_{n,t})$  with respect to a filtration  $(\mathcal{F}_{n,t})$ . That is,  $X_{n,t}$  satisfies, for each  $1 \leq l \leq M_n$  and  $k \geq 0$ ,

$$\|\mathbb{E}[X_{l,n,t} | \mathcal{F}_{n,t-k}]\|_q \leq \bar{c}_n \psi_k, \quad \|X_{l,n,t} - \mathbb{E}[X_{l,n,t} | \mathcal{F}_{n,t+k}]\|_q \leq \bar{c}_n \psi_{k+1}, \tag{2.13}$$

where  $X_{l,n,t}$  denotes the  $l$ th component of  $X_{n,t}$ , and the constants  $\bar{c}_n$  and  $\psi_k$  control the magnitude and the dependence of the  $X_{n,t}$  variables, respectively. Recall that mixingales form a very general class of time series models, including martingale differences, linear processes, and various types of mixing and near-epoch dependent processes as special cases, and naturally allow for data heterogeneity; we refer the reader to Davidson (1994) for a comprehensive review. To verify Assumption 3, we can set  $X_{n,t} = u_t \otimes P(X_t)$  and  $M_n = Jm_n$ .

Denote

$$S_n \equiv n^{-1/2} \sum_{t=1}^n X_{n,t}, \quad \Sigma_n \equiv \text{Var}(S_n).$$

For bounded  $M_n$ -dimensional vectors  $(\alpha_l)_{1 \leq l \leq L_n}$ , we aim to construct a sequence of random variables  $\tilde{U}_n$  such that  $\tilde{U}_n$  has the same distribution as  $\max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n$  for  $\tilde{S}_n \sim \mathcal{N}(0, \Sigma_n)$  and

$$\left| \max_{1 \leq l \leq L_n} \alpha_l^\top S_n - \tilde{U}_n \right| = o_p((\log n)^{-1}). \quad (2.14)$$

In a recent paper, Li and Liao (2019) establish a Yurinskii-type coupling in a similar mixingale setting, which has the form  $\|S_n - \tilde{S}_n\| = O_p(\delta_n)$  for some  $\delta_n = o(1)$ . The Yurinskii-type coupling is stronger than the coupling concept formulated in (2.14), but with a slower rate of convergence than the latter as discussed below. For ease of comparison, we impose the same regularity conditions as in Li and Liao (2019).

**Assumption 4.** (i) For some  $q \geq 3$ , there exists a martingale difference array  $X_{n,t}^*$  such that  $\max_{1 \leq l \leq M_n} \|X_{l,n,t}^*\|_q \leq \bar{c}_n \psi^*$  for some finite constant  $\psi^* > 0$  and

$$\|S_n - S_n^*\| = O_p(\bar{c}_n M_n^{1/2} n^{-1/2})$$

for  $S_n^* \equiv n^{-1/2} \sum_{t=1}^n X_{n,t}^*$ ; (ii) the eigenvalues of  $\mathbb{E}[X_{n,t}^* X_{n,t}^{*\top}]$  are uniformly bounded from above and away from zero; (iii) uniformly for any integer sequence  $k_n$  that satisfies  $n - o(1) \leq k_n \leq n$ ,

$$\left\| \sum_{t=1}^{k_n} (V_{n,t}^* - \mathbb{E}[V_{n,t}^*]) \right\|_S = O_p(r_n) \quad (2.15)$$

where  $V_{n,t}^* \equiv n^{-1} \mathbb{E}[X_{n,t}^* X_{n,t}^{*\top} | \mathcal{F}_{n,t-1}]$  and  $r_n = o(1)$  is a real positive sequence; (iv) the largest eigenvalue of  $\Sigma_n$  is uniformly bounded.

A few remarks on Assumption 4 are in order. Condition (i) directly imposes a martingale approximation for the mixingale array, which is a well-known and very useful property of mixingales.<sup>4</sup> This condition effectively reduces the task of constructing a strong approximation for  $S_n$

<sup>4</sup>A sufficient condition is  $\sum_{q \geq 0} \psi_q < \infty$ , under which the martingale difference is defined as  $X_{n,t}^* \equiv \sum_{s=-\infty}^{\infty} \{\mathbb{E}[X_{n,t+s} | \mathcal{F}_{n,t}] - \mathbb{E}[X_{n,t+s} | \mathcal{F}_{n,t-1}]\}$ ; see Lemma A4 in the supplemental appendix of Li and Liao (2019).

to a simpler one for the approximating martingale  $S_n^*$ . The other conditions are needed for analyzing the latter, which can also be established under primitive conditions. In particular, we note that condition (iii) can be generally derived by using a matrix law of large numbers, and it holds trivially with  $r_n = 0$  if  $V_{n,t}^*$  is deterministic (while allowed to be time-varying). As a concrete illustration of this assumption, we consider the following example.

**EXAMPLE (MARTINGALE APPROXIMATION).** Suppose that  $X_{n,t}$  is formed as a linear process with the form  $X_{n,t} = \sum_{|j|<\infty} \theta_j \varepsilon_{n,t-j}$  where  $\varepsilon_{n,t}$  is a triangular array of independent variables with finite  $q$ th moments. Under the condition  $\sum_{|j|<\infty} |j\theta_j| < \infty$ , Assumption 4(i) can be verified with the approximating martingale difference array defined explicitly as  $X_{n,t}^* = (\sum_{|j|<\infty} \theta_j) \varepsilon_{n,t}$ , where the filtration is given by  $\mathcal{F}_{n,t} = \sigma(\varepsilon_{n,s} : s \leq t)$ . In this case,  $V_{n,t}^* = n^{-1} (\sum_{|j|<\infty} \theta_j) \mathbb{E} [\varepsilon_{n,t} \varepsilon_{n,t}^\top] (\sum_{|j|<\infty} \theta_j)^\top$  is deterministic and may be time-varying if  $\mathbb{E} [\varepsilon_{n,t} \varepsilon_{n,t}^\top]$  depends on  $t$ . Condition (ii) is satisfied if  $\mathbb{E} [\varepsilon_{n,t} \varepsilon_{n,t}^\top]$  and its inverse have bounded eigenvalues. Condition (iii) is automatically satisfied with  $r_n = 0$ . Although the volatility of the approximating martingale is deterministic, it is interesting to note that  $X_{n,t}$  can have stochastic conditional volatility because  $\mathbb{E}[X_{n,t}^2 | \mathcal{F}_{n,t-1}]$  depends on the realizations of lagged innovations  $(\varepsilon_s)_{s<t}$ .  $\square$

We are now ready to state our main theorem on strong approximation.

**Theorem 1.** *Let  $(\alpha_l)_{1 \leq l \leq L_n}$  be  $M_n$ -dimensional real vectors with uniformly bounded Euclidean norm, and  $\tilde{S}_n$  be a generic  $M_n$ -dimensional random vector with distribution  $\mathcal{N}(0, \Sigma_n)$ . We set*

$$B_{1,n} \equiv n^{-3/2} \sum_{t=1}^n \mathbb{E} \left[ (\log L_n)^{3/2} \max_{1 \leq l \leq L_n} \left( \mathbb{E} \left[ (\alpha_l^\top X_{n,t}^*)^2 | \mathcal{F}_{n,t-1} \right] \right)^{3/2} + \max_{1 \leq l \leq L_n} |\alpha_l^\top X_{n,t}^*|^3 \right],$$

$$B_{2,n} \equiv \min \left\{ L_n^{1/q} \max_{1 \leq t \leq n, 1 \leq l \leq L_n} \|\alpha_l^\top X_{n,t}^*\|_q, M_n^{1/2} \right\} + (\log L_n)^{1/2}.$$

*Suppose that  $\bar{c}_n (\log L_n)^{1/2} M_n^{1/2} n^{-1/2} + B_{1,n}^{1/3} (\log L_n)^{2/3} + B_{2,n} r_n^{1/2} = o((\log n)^{-1})$  and Assumption 4 holds. Then, there exists a sequence of random variables  $\tilde{U}_n$  such that  $\tilde{U}_n$  has the same distribution as  $\max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n$  and*

$$\left| \max_{1 \leq l \leq L_n} \alpha_l^\top S_n - \tilde{U}_n \right| = o_p((\log n)^{-1}).$$

Theorem 1 establishes the strong approximation for the maximum statistic  $\max_{1 \leq l \leq L_n} \alpha_l^\top S_n$ . There are two components in the approximation error. The first is related to the martingale approximation and is of order  $O_p(\bar{c}_n (\log L_n)^{1/2} M_n^{1/2} n^{-1/2})$ . The other error term captures the strong approximation error for the maximum statistic  $\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^*$  formed using the approximating martingale  $S_n^*$ , and has order  $O_p(B_{1,n}^{1/3} (\log L_n)^{2/3} + B_{2,n} r_n^{1/2})$ . In a setting with independent data,



Chernozhukov, Chetverikov, and Kato (2014) prove a similar coupling result. Their proof heavily relies on symmetrization-based empirical process techniques that are not available in a general time series setting. We establish the coupling using very different techniques, which is a key theoretical contribution of the current paper.

Theorem 1 can be used to verify Assumption 3. Since the theorem is somewhat complicated in its general form, we provide a corollary in a special case that is much easier to understand.

**Corollary 1.** *Let  $(\alpha_l)_{1 \leq l \leq L_n}$  and  $\tilde{S}_n$  be defined as in Theorem 1. Suppose that (i) Assumption 4 holds with  $\bar{c}_n = O(1)$  and  $V_{n,t}^*$  being deterministic; (ii)  $M_n \asymp n^a$  for some  $a \in (0, 1/3)$ ; and (iii)  $L_n$  has polynomial growth as  $n \rightarrow \infty$ . Then, there exists a sequence of random variables  $\tilde{U}_n$  such that  $\tilde{U}_n$  has the same distribution as  $\max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n$  and*

$$\left| \max_{1 \leq l \leq L_n} \alpha_l^\top S_n - \tilde{U}_n \right| = o_p((\log n)^{-1}).$$

COMMENT. To apply this result in the context of Section 2.2, we set  $M_n = Jm_n$  and note that  $m_n$  and  $L_n$  have polynomial growth as  $n \rightarrow \infty$ . Under the simple setting stated by Corollary 1, Assumption 3 is verified provided that  $m_n = O(n^a)$  for some  $a \in (0, 1/3)$ . In contrast, Li and Liao's (2019) Yurinskii-type coupling has the form  $\|S_n - \tilde{S}_n\| = O_p(m_n^{5/6} n^{-1/6})$ , with the approximation error shrinking to zero when  $m_n = O(n^a)$  for  $a \in (0, 1/5)$ , which is notably more restrictive.  $\square$

### 3 Monte Carlo study

In this section, we examine the finite-sample performance of the CSPA test in Monte Carlo experiments. Section 3.1 presents the setting and Section 3.2 reports the results.

#### 3.1 The data generating process

We consider a setting with  $J$  conditional moments for  $J = 1, 3, \text{ or } 5$ . The data are simulated according to the following data generating process (DGP):

$$\begin{aligned} Y_{j,t} &= 1 - a e^{-(X_t - c)^2} + u_{j,t}, & 1 \leq j \leq J, \\ X_t &= 0.5X_{t-1} + \epsilon_t, & \text{with } \epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, 0.75), \\ u_{j,t} &= \rho_u u_{j,t-1} + v_{j,t}, & \text{with } v_{j,t} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_v^2). \end{aligned}$$

We consider  $c \in \{0, 0.5, 1\}$  and  $\rho_u \in \{0, 0.4, 0.8\}$ . We also set  $\sigma_v^2 = 3(1 - \rho_u^2)$  so that the variance of  $u_{j,t}$  is kept constant at 3.

A few remarks on this DGP are in order. First, the  $X_t$  process is a centered Gaussian AR(1) process with its variance normalized to unity. Second, the conditional expectation function  $h_j(x) = \mathbb{E}[Y_{j,t}|X_t = x]$  has the form

$$h_j(x) = 1 - a e^{-(x-c)^2},$$

and it attains its minimum  $1 - a$  at  $x = c$ . The corresponding unconditional expectation is

$$\mathbb{E}[Y_{j,t}] = \mathbb{E}[h_j(X_t)] = 1 - \frac{a e^{-c^2/3}}{\sqrt{3}}. \quad (3.1)$$

The  $c$  parameter plays a useful role in our experiments: Since the distribution of  $X_t$  is concentrated around zero, we can explore the effect of data density on the estimation accuracy for different parts of the  $h_j(\cdot)$  function by varying  $c$ . Third, as we increase  $\rho_u$  from 0 to 0.8, the error series  $u_{j,t}$  become more persistent, rendering time-series inference more difficult.

We impose the null and alternative hypotheses for CSPA as follows. The null hypothesis described in (2.5) is satisfied when  $a = 1$ , that is,  $h_j(c) = 0$  and  $h_j(x) > 0$  when  $x \neq c$ . On the other hand, when  $a > 1$ , the conditional moment violates the CSPA null hypothesis because  $h_j(x) < 0$  when  $x$  falls in the  $(c - \sqrt{\log a}, c + \sqrt{\log a})$  interval. The deviation of  $a$  from 1 thus quantifies the “distance” between the null and alternative hypotheses.

In addition to the proposed CSPA test, we also implement Hansen’s (2005) USPA test for comparison. Although these tests concern different hypotheses, it is interesting to concretely demonstrate how their difference manifests in the present numerical setting. From equation (3.1), we see that

$$\mathbb{E}[Y_{j,t}] \geq 0 \quad \text{if and only if} \quad a \leq \sqrt{3}e^{c^2/3}.$$

Therefore, when  $1 < a \leq \sqrt{3}e^{c^2/3}$ , the CSPA null hypothesis will be violated, whereas the USPA null hypothesis still holds. This corresponds to a situation in which a competing forecast model strictly outperforms the benchmark in certain regions of the conditioning state space, but, at the same time, underperforms the benchmark on average. By design, the USPA test should not reject, and the CSPA test will reject, providing useful additional diagnostic information.

We consider three sample sizes,  $n = 250, 500$  or  $1,000$ . The small sample speaks to typical macroeconomic applications with quarterly data, the medium sample corresponds to financial application with monthly data, and the large sample is easily attainable for analysis based on daily observations. The simulation consists of 10,000 Monte Carlo replications.

Finally, we describe the implementation details of the CSPA test. The significance level is set to be 5%. We employ the Legendre polynomials  $P(z) = (1, z, 0.5(z^2 - 1), \dots)$  as basis functions and use the Akaike Information Criteria (AIC) to select the basis up to five series terms.<sup>5</sup> These

---

<sup>5</sup>In results not presented here, we include up to seven series terms, which yields very similar results.

basis functions are orthogonal on the  $[-1, 1]$  interval under Lebesgue measure, and tend to be less affected by multicollinearity issues in finite samples. To obtain estimates of the long-run variance  $\widehat{\Omega}_n$ , we consider both the standard Newey–West type and pre-whitened HAC estimators. For the former, we use the Bartlett kernel (Newey and West (1987)) and, following Andrews (1991b), set the bandwidth to  $\lfloor 0.75n^{1/3} \rfloor$ . For the pre-whitened estimator that is described in detail in Supplemental Appendix SC, we pre-whiten the data using an autoregressive filter adaptively tuned via the AIC up to four lags.<sup>6</sup>

### 3.2 Results

To set the stage, we discuss results from the univariate setting (i.e.,  $J = 1$ ). We first examine the size property of the CSPA test, for which we impose the null hypothesis by setting  $a = 1$ . Table 1 reports the rejection rates of the test under the null for various values of  $\rho_u$  and  $c$ . The left and right panels report results based on the Newey–West HAC estimator and the pre-whitened estimator, respectively.

We summarize the results as follows. First, looking at the “quarterly” case (Panel A), we see that the test based on the standard HAC estimator controls size well when the error terms are moderately persistent (i.e.,  $\rho_u = 0$  or  $0.4$ ). However, the test can be nontrivially oversized when  $\rho_u = 0.8$ . For example, the rejection rate is 8.4% when  $c = 1$ , which is notably higher than the 5% nominal level. In contrast, tests based on the pre-whitened estimator show satisfactory size control even in the presence of high persistence. We note that overrejection resulted from the standard HAC estimator is mainly a small-sample issue, which dampens as we increase the sample size, and is no longer present for the “daily” sample displayed in Panel C. But to be cautious, we focus only on results based on the pre-whitened HAC estimator in all our discussion below.

Second, we note that the CSPA test is generally conservative, as its rejection rate is often below the 5% nominal level. This is expected from the asymptotic theory, as we can see from (2.7) that the probability of type-I error is asymptotically bounded by  $\alpha$ . The intuition for the conservativeness is as follows. In our simulation design, the inequality  $h_j(x) \geq 0$  is binding at  $x = c$ . If this information were known a priori, we could compute the critical value by concentrating on the singleton  $x = c$ . However, in finite samples, we need to conservatively uncover the “binding region” using a nonparametric estimator. To the extent that this preliminary estimation is coarse, the resulting critical value is conservative.

We next turn to power properties of the CSPA test, again for the univariate case with  $J = 1$ .

---

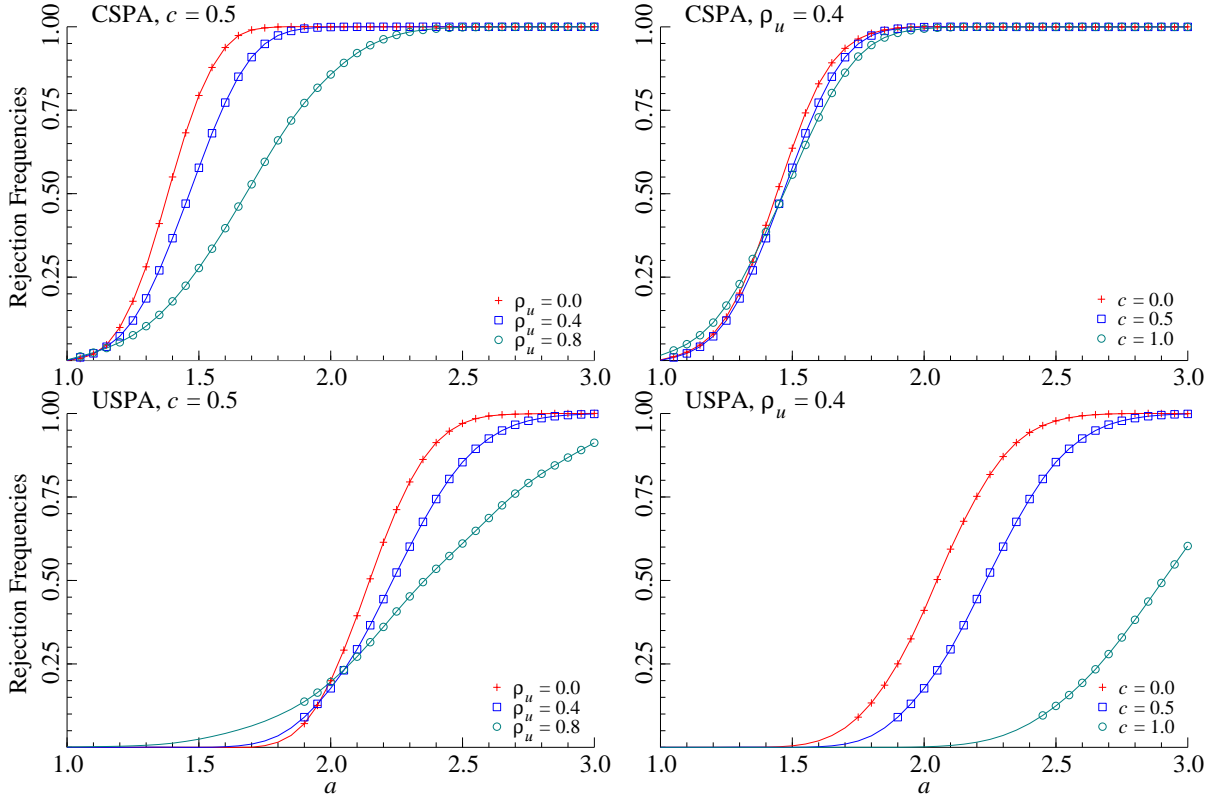
<sup>6</sup>As in Andrews and Monahan (1992), the validity of the pre-whitened HAC estimator does not require the time series to actually follow a vector autoregressive model.

Table 1: Rejection Rates under the Null Hypothesis

	Newey–West HAC			Pre-whitened HAC		
	$\rho_u = 0.0$	$\rho_u = 0.4$	$\rho_u = 0.8$	$\rho_u = 0.0$	$\rho_u = 0.4$	$\rho_u = 0.8$
<i>Panel A: Small-sample Case (n = 250)</i>						
$c = 0.0$	0.007	0.014	0.057	0.008	0.013	0.025
$c = 0.5$	0.008	0.016	0.065	0.008	0.014	0.027
$c = 1.0$	0.030	0.036	0.084	0.030	0.034	0.048
<i>Panel B: Medium-sample Case (n = 500)</i>						
$c = 0.0$	0.003	0.008	0.033	0.003	0.007	0.012
$c = 0.5$	0.005	0.008	0.032	0.005	0.006	0.011
$c = 1.0$	0.021	0.024	0.057	0.021	0.021	0.030
<i>Panel C: Large-sample Case (n = 1,000)</i>						
$c = 0.0$	0.003	0.003	0.017	0.003	0.003	0.006
$c = 0.5$	0.003	0.003	0.018	0.003	0.003	0.006
$c = 1.0$	0.015	0.018	0.030	0.015	0.016	0.015

*Note:* This table presents rejection rates of the CSPA test under the univariate null hypothesis (i.e.,  $J = 1$  and  $a = 1$ ). The test is implemented using either the Newey–West or the pre-whitened HAC estimator. We consider different data generating processes by varying the sample size  $n$ , residual autocorrelation  $\rho_u$ , and the location parameter  $c$ .

Figure 1: Simulation Results: Power Curves



*Note:* This figure presents the power curves of the CSPA test (top row) and Hansen’s (2005) USPA test (bottom row). In the left (resp. right) column we vary the  $\rho_u$  (resp.  $c$ ) parameter while keeping the  $c$  (resp.  $\rho_u$ ) parameter fixed. To highlight whether the value of  $a$  corresponds to the null or the alternative hypothesis, we signify the latter with a marker.

As discussed above, the alternative hypothesis can be imposed by setting  $a > 1$ . In the top panels of Figure 1, we plot the CSPA test’s rejection rates as functions of  $a$  while varying the  $\rho_u$  and  $c$  parameters in the DGP. For brevity, we only show results for the  $n = 500$  case.

The top-left panel of Figure 1 plots the CSPA power curves for different levels of persistence quantified by the  $\rho_u$  parameter, while fixing  $c = 0.5$ . As  $a$  increases, the rejection rate approaches one, which is consistent with the asymptotic theory. In addition, we see that the test has higher power when the error terms are less persistent. On the top-right panel, we plot power curves for different  $c$  values while fixing  $\rho_u = 0.4$ . When  $a$  is small, the test tends to be more conservative when  $c$  is closer to zero, which mirrors the findings in Table 1. Interestingly, this ranking is reversed when  $a$  is large.

It is instructive to compare the power properties of CSPA test with those of Hansen’s (2005) USPA test. The USPA power curves are plotted in the bottom panels of Figure 1, and computed under the same DGPs. We stress that this comparison should be interpreted cautiously, because CSPA and USPA tests are designed for different hypotheses. Specifically, the USPA null hypothesis is violated when  $a > \sqrt{3}e^{c^2/3}$ ; the threshold is 1.73, 1.88, and 2.42 for the cases with  $c = 0, 0.5,$  and 1, respectively.

Looking at the USPA power curves, we see that the test essentially does not reject when  $a$  is less than the  $\sqrt{3}e^{c^2/3}$  threshold; this is particularly evident in the bottom-right panel as we vary the value of  $c$ . When  $a$  exceeds this threshold, we observe increasingly more rejections, with rejection rate approaching one as  $a$  becomes larger. Although the USPA test starts to have nontrivial power under its own alternative hypothesis, its rejection rate is notably lower than that of the CSPA test. This comparison thus concretely illustrates scenarios in which the conditional test can reveal useful information above and beyond its unconditional counterpart.

Finally, we present simulation results for the multivariate setting with  $J = 3$  or 5. Similar to the univariate setting above, we impose the null hypothesis by setting  $a = 1$ , and set  $a = 1.5$  to obtain the alternative. This alternative is chosen to have nondegenerate local power in view of Figure 1. Table 2 reports the rejection rates. From the table, we see that the CSPA test controls size in the multivariate setting and is slightly more conservative. Meanwhile, the power of the test is higher as more conditional moment inequalities are violated under the alternative.

## 4 Empirical application on volatility forecast

As a first empirical application of the CSPA test, we consider the conditional evaluation of forecasting models of realized volatility ( $RV$ ). Ex-post measures of daily volatility based on high-frequency data, and the reduced-form modeling of their dynamics, are popularized by Andersen, Bollerslev, Diebold, and Labys (2003), whose seminal work has since spurred a large and burgeoning literature in financial econometrics. We apply the CSPA test on a collection of prominent forecasting methods. Section 4.1 introduces the forecasting models and reports baseline results from unconditional evaluation tests. Section 4.2 presents results from the CSPA test.

### 4.1 Volatility forecasting methods and their unconditional evaluation

Our analysis is based on the publicly available dataset from Bollerslev, Patton, and Quaedvlieg (2016), which contains daily realized volatility and other high-frequency measures for the S&P 500 index and 27 stocks in Dow Jones Industrial Average that are traded over the full sample period

Table 2: Simulation Results: Multivariate Test

	Null Hypothesis			Alternative Hypothesis		
	$\rho_u = 0$	$\rho_u = 0.4$	$\rho_u = 0.8$	$\rho_u = 0$	$\rho_u = 0.4$	$\rho_u = 0.8$
<i>Panel A: J = 3</i>						
$c = 0.0$	0.004	0.007	0.014	0.983	0.885	0.467
$c = 0.5$	0.003	0.006	0.011	0.958	0.806	0.396
$c = 1.0$	0.013	0.016	0.027	0.915	0.773	0.467
<i>Panel B: J = 5</i>						
$c = 0.0$	0.003	0.007	0.015	0.997	0.961	0.561
$c = 0.5$	0.004	0.007	0.013	0.988	0.911	0.491
$c = 1.0$	0.014	0.019	0.027	0.958	0.877	0.556

*Note:* This table presents rejection rates of the CSPA test when  $J = 3$  or  $J = 5$ . We set  $a = 1$  and  $1.5$  for the null and the alternative hypotheses, respectively. We consider different data generating processes by varying the sample size  $n$ , residual autocorrelation  $\rho_u$ , and the location parameter  $c$ .

from April 1997 to December 2013. Realized volatility is computed as the sum of squared 5-minute returns within regular trading hours. We focus on one-day-ahead forecasts that are formed using rolling-window estimation with 1,000 daily observations. This results in over 3,000 daily forecasts for each series.<sup>7</sup>

We study 6 competing forecasting methods in total. The first is Corsi’s (2009) HAR model, which is arguably the most popular model in the recent financial econometrics literature for volatility forecasting. This model is a restricted AR(22), in which common coefficients are imposed across “daily,” “weekly,” and “monthly” lags. We consider three alternative autoregressive specifications, including AR(1), AR(22), and an adaptive AR(22) model with LASSO-based variable selection.<sup>8</sup> Note that the HAR model is designed to capture the well-known long-memory feature of volatil-

<sup>7</sup>The S&P500 Realized volatility is based on futures data from Tick Data and ends in August 2013, resulting in a total of 4,096 observations. The individual stocks, whose data comes from the TAQ database, spans the full sample with 4,202 observations. See Bollerslev, Patton, and Quaedvlieg (2016) for details.

<sup>8</sup>The regularization parameter of the LASSO is obtained using 10-fold cross-validation.

ity. In this vein, we also include an ARFIMA(1,  $d$ , 0) model in our analysis, which is a classical alternative for capturing long memory via fractional integration. Lastly, we include one of many recent augmentations to HAR—the HARQ model proposed by Bollerslev, Patton, and Quaedly (2016). HARQ allows the first autoregressive parameter to vary over time in response to the level of measurement error in the high-frequency estimate of daily  $RV$ . The level of measurement error is quantified by the daily realized quarticity ( $RQ$ ), defined as the (scaled) sum of the 4th power of high-frequency returns. More precisely, the specifications of these forecast models are given below:

$$\begin{aligned}
\mathbf{AR(1)} \quad & RV_t = \phi_0 + \phi_1 RV_{t-1} + \epsilon_t, \\
\mathbf{AR(22)} \quad & RV_t = \phi_0 + \sum_{i=1}^{22} \phi_i RV_{t-i} + \epsilon_t, \\
\mathbf{HAR} \quad & RV_t = \phi_0 + \phi_1 RV_{t-1} + \phi_2 RV_{t-1|t-5} + \phi_3 RV_{t-1|t-22} + \epsilon_t, \\
\mathbf{HARQ} \quad & RV_t = \phi_0 + (\phi_1 + \phi_{1Q} RQ_{t-1}^{1/2}) RV_{t-1} + \phi_2 RV_{t-1|t-5} + \phi_3 RV_{t-1|t-22} + \epsilon_t, \\
\mathbf{ARFIMA} \quad & (1 - \mathbb{L})^d RV_t = \phi_0 + \phi_1 RV_{t-1} + \epsilon_t,
\end{aligned}$$

where  $RV_{t-1|t-k} = k^{-1} \sum_{i=1}^k RV_{t-i}$  and  $\mathbb{L}$  denotes the lag operator.

To set the stage, we first conduct an unconditional comparison of these six methods using Hansen’s (2005) USPA test. The test is implemented under two schemes: one-versus-one or one-versus-all. Under one-versus-one, we compare each benchmark with one alternative for all model pairs. Under one-versus-all, we perform the USPA test to compare each benchmark model with all the other five competing models jointly. To mitigate the impact of rare but large volatility spikes, we follow Patton (2011) and employ the QLIKE loss function, defined as

$$L(F_t^\dagger, F_{j,t}) = \frac{F_{j,t}}{F_t^\dagger} - \log \left( \frac{F_{j,t}}{F_t^\dagger} \right) - 1.$$

The tests are implemented separately for each asset. As a simple summary of these testing results, we report the number of assets (out of 28 in total) for which the USPA test (one-versus-one or one-versus-all) rejects at 5% significance level in Table 3.<sup>9</sup> From the top panel, we see considerable heterogeneity in the models’ average loss: The average loss of the worst model, AR(1), is more than twice as high as that of the best model, HARQ. HAR appears to easily outperform AR(22)-LASSO. Interestingly, the latter adaptive method actually underperforms the AR(22) model with unrestricted coefficients.

We next turn to formal testing results. Under the one-versus-one scheme, we find that the USPA null hypothesis is almost never rejected when HARQ is the benchmark (see the fifth column). The only exception occurs when the competing method is ARFIMA, and the null hypothesis is rejected

---

<sup>9</sup>We do not intend to interpret the testing results jointly across different assets.



Table 3: Unconditional Superior Predictive Ability for Volatility Forecasts

	<i>Benchmark Methods</i>					
	AR(1)	AR(22)	AR(22) LASSO	HAR	HARQ	ARFIMA
<i>Panel A: Average QLIKE Loss</i>						
	0.401	0.229	0.298	0.213	0.185	0.198
<i>Panel B: One-versus-one USPA tests against different competing methods</i>						
AR(1)		0	3	0	0	0
AR(22)	28		27	0	0	0
AR(22) LASSO	25	0		0	0	0
HAR	28	26	28		0	1
HARQ	28	28	28	28		22
ARFIMA	28	27	28	26	2	
<i>Panel C: One-versus-all USPA tests against all competing methods</i>						
	28	28	28	28	2	22

*Note:* Panel A reports the average QLIKE loss of each of the six models, where the averaging is both over time and across assets. Panel B (resp. C) reports the number of assets for which the one-versus-one (resp. one-versus-all) USPA null hypothesis is rejected at 5% significance level.

for 2 out of the 28 assets. But when ARFIMA is the benchmark, the associated USPA null is rejected for most of the assets against HARQ. Judged in similar fashion, the other four methods can be ranked as follows: HAR, AR(22), AR(22)-LASSO, and AR(1). The one-versus-all tests, as shown in Panel C, more clearly confirms HARQ’s superior performance. In particular, the joint test always rejects when each of the four underperforming models is used as the benchmark.

Overall, these unconditional evaluation results largely confirm prior findings in the literature. However, this does not preclude, for instance, the possibility that AR(1) can outperform HARQ and ARFIMA in *some* states of the world. The CSPA test is designed to investigate such issue, to which we now turn.

## 4.2 CSPA of volatility forecasts

We implement the CSPA test with the CBOE Volatility Index (VIX) as the conditioning state variable. The VIX is an option-based implied volatility measure, and is often deemed to be the “fear gauge” of investors. We stress that the testing methodology does not depend on this specific choice. In our analysis below, we use each of the aforementioned models as a benchmark, and test whether  $\mathbb{E}[Y_{j,t}|\text{VIX}_{t-1}] \geq 0$ , where  $Y_{j,t}$  denotes the loss differential between the  $j$ th competing model and the benchmark. We follow the same implementation procedure using the pre-whitened HAC estimator as described in the simulation study.

To concretely illustrate how the CSPA test works in practice, we start with a case study for Johnson & Johnson (NYSE: JNJ). The evaluation scheme is one-versus-one: The benchmark is HAR and the competing alternative is either AR(1) or HARQ. In Figure 2, we plot the estimated conditional expected loss differential functions, along with the 95% upper confidence bounds given by  $\hat{h}_{j,n}(\cdot) + n^{-1/2} \hat{k}_{n,1-\alpha} \hat{\sigma}_{j,n}(\cdot)$ . Recall that the critical value  $\hat{\eta}_{n,1-\alpha}$  of the CSPA test is the infimum of the latter function and the CSPA test rejects the null hypothesis if some part of this function is below zero.

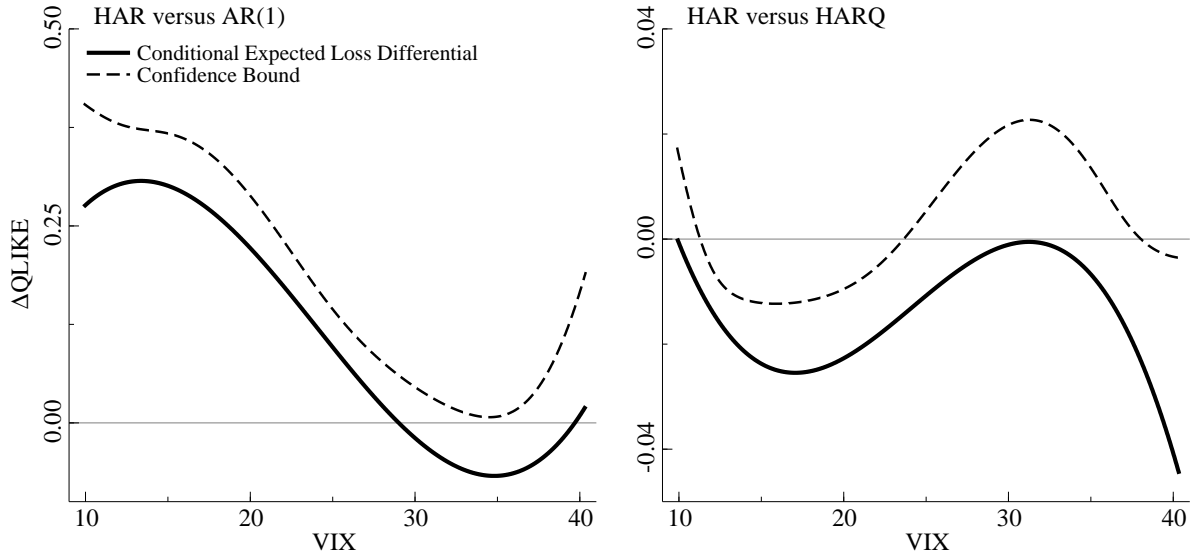
The left panel illustrates the comparison between HAR and AR(1). As shown by the conditional expected loss differential curve, the simple AR(1) forecast underperforms HAR when VIX is below 30, but appear to be more accurate in extremely volatile states. That being said, the CSPA null hypothesis is not rejected at the 5% level (because the confidence bound function is always above zero), although this non-rejection is marginal.

Turning to the right panel of Figure 2, we see that HARQ outperforms the benchmark HAR model not only on average but also uniformly across different states, in that the conditional expectation function is always below zero.<sup>10</sup> The improvement stems from both quiescent (say,

---

<sup>10</sup>A direct consequence is that the CSPA null hypothesis is not rejected if we instead take CSPA as the benchmark

Figure 2: Forecasting Volatility: One-versus-one CSPA Tests



*Note:* This figure plots the estimated conditional expected loss differential functions (solid), along with its 95% confidence bound (dashed), for the realized variance series of Johnson & Johnson, against the conditioning variable VIX. The HAR model is used as the benchmark, with the AR(1) (left panel) and the HARQ (right panel) as competing alternatives. A negative value of the conditional expected loss differential indicates that the HAR model is outperformed by the competing alternative. The CSPA test rejects the null hypothesis (i.e., HAR is weakly superior) if the confidence bound is below zero over some region of the conditioning state space.

VIX  $\approx$  15) and high-uncertainty states. These results are in line with Bollerslev, Patton, and Quaadvlieg’s (2016) finding that the benefit of HARQ is twofold: The model allows for more persistence in quiet times, and sharply down-weights past observations under crisis-like market conditions. Looking at the upper confidence bound, we see that HARQ’s better performance is highly statistically significant when VIX ranges between, for example, 11.3 and 23.6, resulting in a rejection of the CSPA null hypothesis.

We now summarize all testing results. Panel A of Table 4 reports the number of assets, out of 28, for which the one-versus-one CSPA null hypothesis is rejected for each benchmark-competitor pair. These results are markedly different from those of the unconditional test (cf. Table 3), in that the CSPA test, not surprisingly, yields many more rejections than its unconditional counterpart. This is best illustrated by the case in which HARQ is the benchmark (see the fifth columns of Tables and HAR as the competing alternative.

Table 4: Conditional Superior Predictive Ability for Volatility Forecasts

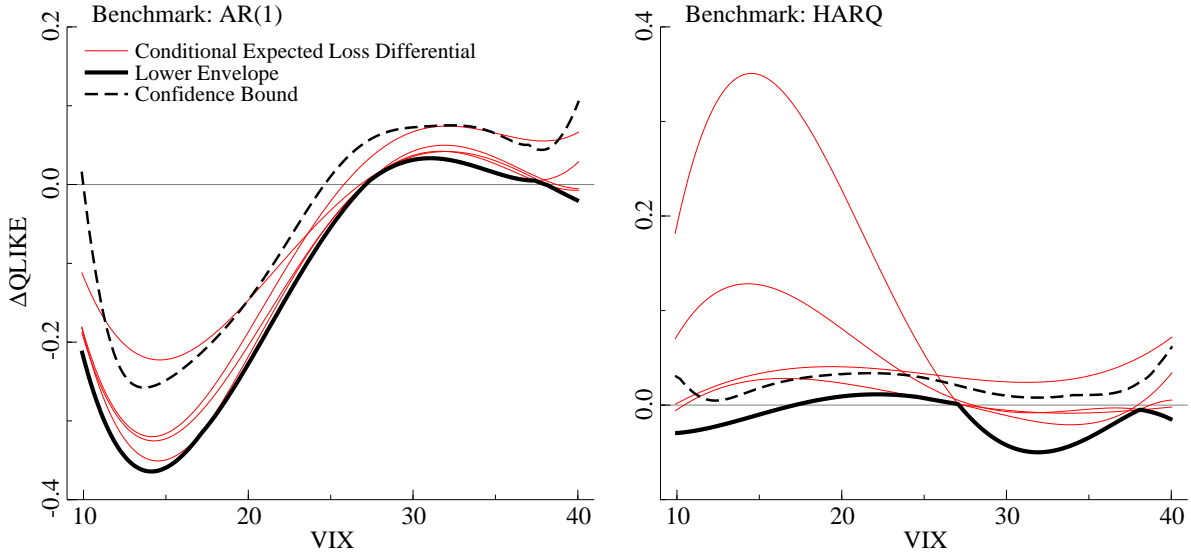
	<i>Benchmark Methods</i>					
	AR(1)	AR(22)	AR(22) LASSO	HAR	HARQ	ARFIMA
<i>Panel A: One-versus-one CSPA tests against different competing models</i>						
AR(1)		13	9	7	7	8
AR(22)	28		28	0	0	1
AR(22)+LASSO	27	15		4	4	9
HAR	28	23	28		3	7
HARQ	28	28	28	28		23
ARFIMA	28	27	28	27	13	
<i>Panel B: One-versus-all CSPA tests against all competing models</i>						
	28	28	28	28	8	22

*Note:* Panel A (resp. B) reports the number of assets, out of 28, for which the one-versus-one (resp. one-versus-all) CSPA null hypothesis is rejected at 5% significance level. Each column corresponds to a different benchmark.

3 and 4): While the USPA null hypothesis is essentially never rejected, we see many rejections from the CSPA test. In other words, although HARQ almost always outperforms the other competing models on average, its dominance is sometimes not uniform. Indeed, HARQ’s close competitor, ARFIMA, leads to 13 CSPA rejections, which can be contrasted with the 2 rejections from the corresponding USPA test.

Another interesting, and somewhat surprising, finding from Table 4 concerns the conditional performance of the simple AR(1). Looking at the large rejection numbers in the first column of the table, we see clearly that AR(1) does not show any uniform superiority with respect to the other methods. But from the first row of the table, which shows the results with AR(1) being the competitor, we also see a nontrivial number of rejections, suggesting that the AR(1) model cannot be easily dominated by the others uniformly across different states, either. This finding mirrors

Figure 3: Forecasting Volatility: One-versus-all CSPA Tests



*Note:* In the left (resp. right) panel, we plot the estimated conditional expected loss differential functions,  $\hat{h}_{j,n}(\cdot)$ , using the AR(1) (resp. HARQ) model as the benchmark, and each of the other five models as the competing alternative. We also plot their lower envelope (solid dark),  $\min_j \hat{h}_{j,n}(\cdot)$ , along with its 95% upper confidence bound (dashed). The one-versus-all CSPA test rejects the null hypothesis if the confidence bound is below zero over some region of the conditioning state space.

the pattern seen in Figure 2: The AR(1) model, which has “short memory” and is fast-updating, can outperform those with long memory during extremely volatile periods. The moral is, when crisis hits, “amnesia” could be a bliss.

Panel B of Table 4 reports results from the one-versus-all CSPA tests. From here, we see more clearly that HARQ reigns supreme in the conditional evaluation, albeit not uniformly dominating. To provide further insight on the one-versus-all CSPA test, we make a visualization in Figure 3 for the case of Johnson & Johnson, where the left and right panels feature AR(1) and HARQ as benchmarks, respectively. In each panel, we plot the estimated conditional expected loss differential function for each competing alternative, their lower envelope, and the confidence bound of the latter. The left panel shows a clear CSPA rejection of the AR(1) benchmark, mainly due to the method’s severe underperformance in low-VIX states. On the right panel, we see that, compared with HARQ, the best-performing competitor (which varies across different states) can reduce the conditional expected loss slightly in some states, but the reduction is statistically insignificant.

In sum, the application on volatility forecasting clearly highlights the usefulness of the CSPA test. We intentionally emphasize how nonparametric tools can provide richer diagnostic information regarding the state-dependent performance of different forecasting models, above and beyond conventional unconditional evaluation methods. As a by-product, we show that the recently developed HARQ model performs well not only on average but, quite often, also uniformly across different states. That being said, the stringent CSPA criterion reveals that HARQ still needs improvement, and the search for uniformly superior volatility forecast should remain to be an important, and challenging, task on financial econometricians' research agenda.

## 5 Empirical application on inflation forecast

Our second application concerns inflation, which is notoriously difficult to forecast (Stock and Watson (2010); Faust and Wright (2013)) as evidenced by the fact that over extended periods of time a simple random walk forecast outperformed the official Greenbook inflation forecasts (Atkeson and Ohanian (2001), Faust and Wright (2009)). Meanwhile, in various academic and commercial applications, machine-learning methods have increasingly shown their potential for improving classical prediction methods. In a recent paper, Medeiros, Vasconcelos, Veiga, and Zilberman (2019) experiment with a variety of machine-learning methods including shrinkage method and random forests, among others. In this section, we apply the CSPA test to examine the conditional performance of these forecasts, along with traditional ones. This is a particularly relevant application of the CSPA test in a macroeconomic context in light of the considerable instability in the performance of inflation forecasting models (Stock and Watson (2009, 2010)), which our test may serve to help understand.

### 5.1 CSPA of inflation forecasts

We apply the CSPA test to evaluate eight inflation forecasting methods, which are selected from those constructed by Medeiros, Vasconcelos, Veiga, and Zilberman (2019) for the one-month-ahead forecast of Consumer Price Index (CPI).<sup>11</sup> This collection includes four traditional methods: random walk, AR(1), Bayesian vector-autoregression (BVAR), and a factor model. The other four methods rely on machine-learning techniques, such as random-forest regression (RF-OLS), LASSO, elastic net (EINet), and bagging. Specifically, RF-OLS implements a standard linear regression

---

<sup>11</sup>We are grateful to Marcelo C. Medeiros for kindly providing their forecasts and data. The forecasts are constructed using a rolling window of 360 months. The estimation sample starts from January 1960 and is based on the 2016 vintage of the Fred-MD database (McCracken and Ng (2016)). For brevity, we refer the reader to Appendix B of Medeiros, Vasconcelos, Veiga, and Zilberman (2019) for the complete description of their models.

based on variables selected by random forest (Breiman (2001)); LASSO is proposed by Tibshirani (1996) and has been used in inflation forecasting by Bai and Ng (2008); ElNet combines the classical ridge regression with the LASSO method (Zou and Hastie (2005)); Bagging, or bootstrap aggregation (Breiman (1996)), is first used to forecast inflation by Inoue and Kilian (2008).<sup>12</sup> The evaluation sample period is between January 1990 and December 2015, consisting of 312 monthly observations in total. All tests below are based on the quadratic loss function and are at the 5% significance level.

When implementing the CSPA test on these inflation forecasts, it is a priori unclear which conditioning state variable would be the most revealing, largely because of the intrinsic difficulty in inflation forecasting and many potentially relevant macroeconomic variables. We thus consider a number of important—and conceptually distinct—conditioning variables, including: average inflation over the past three months (INFL), industrial production growth ( $\Delta$ IP), unemployment rate (UR), 12-month macro uncertainty (MU), economic policy uncertainty (EPU), equity market volatility (EMV), and VIX.<sup>13</sup>

Table 5 presents results of the USPA and CSPA tests. The first two columns summarize the unconditional performance of the eight forecasting methods. The first column reports each method’s relative mean-squared-error (RMSE), defined as the ratio between its average quadratic loss and that of the random-walk forecast. Note that the methods are ordered according to their RMSEs in the table. The second column shows the number of one-versus-one USPA rejections for the method in each row (as the benchmark) against each one of the other seven methods (as the competitor), and we use asterisks to signify rejections from one-versus-all tests. From here, we see that the four machine-learning methods cannot be “separated” by the USPA test, in that the test never rejects. That being said, it does provide strong statistical evidence for the underperformance of the “traditional” models, such as AR(1).

We further turn to the CSPA test to examine the conditional performance of these methods, particularly the ones based on machine learning. In columns 3–9, we present analogous rejection numbers from the CSPA tests based on the aforementioned conditioning variables separately. For ease of discussion, we also report the total number of rejections in the last column. As expected, the conditional test indeed helps discriminating these forecast methods. The most striking finding

---

<sup>12</sup>These four methods are selected because they have the smallest average mean-square-errors among all 18 methods studied by Medeiros, Vasconcelos, Veiga, and Zilberman (2019). In results not reported here, we instead analyze the full list. But the larger-scale study does not change our main empirical findings.

<sup>13</sup>The INFL,  $\Delta$ IP, and UR series are obtained from McCracken and Ng’s (2016) FRED-MD database. The MU index is proposed by Jurado, Ludvigson, and Ng (2015). The EPU and EMV indexes are constructed by Baker, Bloom, and Davis (2016).

Table 5: USPA and CSPA Tests for Inflation Forecasts

	RMSE	USPA	CSPA							Total
			INFL	$\Delta$ IP	UR	MU	EPU	EMV	VIX	
RF-OLS	0.65	0	0*	2*	0	0*	0	0	0	2
ElNet	0.69	0	0	0*	0	2*	1	0	0	3
LASSO	0.69	0	1*	0*	0	2*	1*	0	0	4
Bagging	0.69	0	3	1*	1*	5*	3*	1	3*	17
BVAR	0.74	1*	1	0*	0	0*	3*	0	3*	7
Factor	0.76	3*	4*	3*	3*	5*	3*	3*	4*	25
AR(1)	0.81	3*	3*	3*	3*	4*	2	4*	3*	22
Random Walk	1.00	7*	6*	7*	7*	7*	7*	7*	7*	48

*Note:* The first column reports the relative mean-squared-error (RMSE) of each forecast method’s average quadratic loss with respect to that of the random walk model. The second column reports the number of times the one-versus-one USPA test rejects the benchmark in each row against each of the other seven competing methods. The next seven columns present similar summary statistics for the CSPA tests using the seven conditioning variables separately, with the total number of rejections shown in the last column. For each of the entries, the \* symbol signifies a one-versus-all rejection by the USPA or CSPA test. Forecast methods are sorted based on their RMSEs displayed in the first column.

concerns the bagging method. Conditioning on the macroeconomic uncertainty index, the one-versus-one CSPA test rejects the null hypothesis with the bagging benchmark against five (out of seven) competing alternatives. Bagging’s total number of rejections reported in the last column is also much higher than those of RF-OLS, ElNet, and LASSO. The performance of the RF-OLS method is particularly impressive, in that it not only has the smallest average loss, but also generally exhibits uniform conditional superiority as postulated by the CSPA null hypothesis.

Overall, the analysis above further demonstrates the usefulness of the CSPA test in a macroeconomic context. Unlike our previous application on volatility, this example presents a concrete empirical scenario in which unconditional evaluation is completely silent on the relative performance of certain forecasting methods, and the CSPA test—by imposing a more stringent uniform requirement under the null—can be used to rule out seemingly indistinguishable methods.



## 5.2 Additional diagnostic visualizations

The USPA test cannot distinguish RF-OLS, EInet, LASSO, and bagging, but the CSPA test can tell them apart. This is interesting particularly because the latter three methods have virtually the same average loss, rendering any unconditional evaluation a fruitless effort. In this subsection, we provide further diagnostic information for these forecasts so as to better understand their differential state-dependent performances.

To do so, we visualize one-versus-one CSPA tests in Figure 4, where the top, middle, and bottom panels corresponds to cases with EInet, LASSO, and bagging as the benchmark, respectively. The competing alternative method is fixed to be RF-OLS. For brevity, we focus on only two conditioning state variables: the average inflation of the past three months INFL (left column), and the macroeconomic uncertainty index MU (right column). In each case, we plot the conditional expectation function of the loss differential and the 95% upper confidence bound.

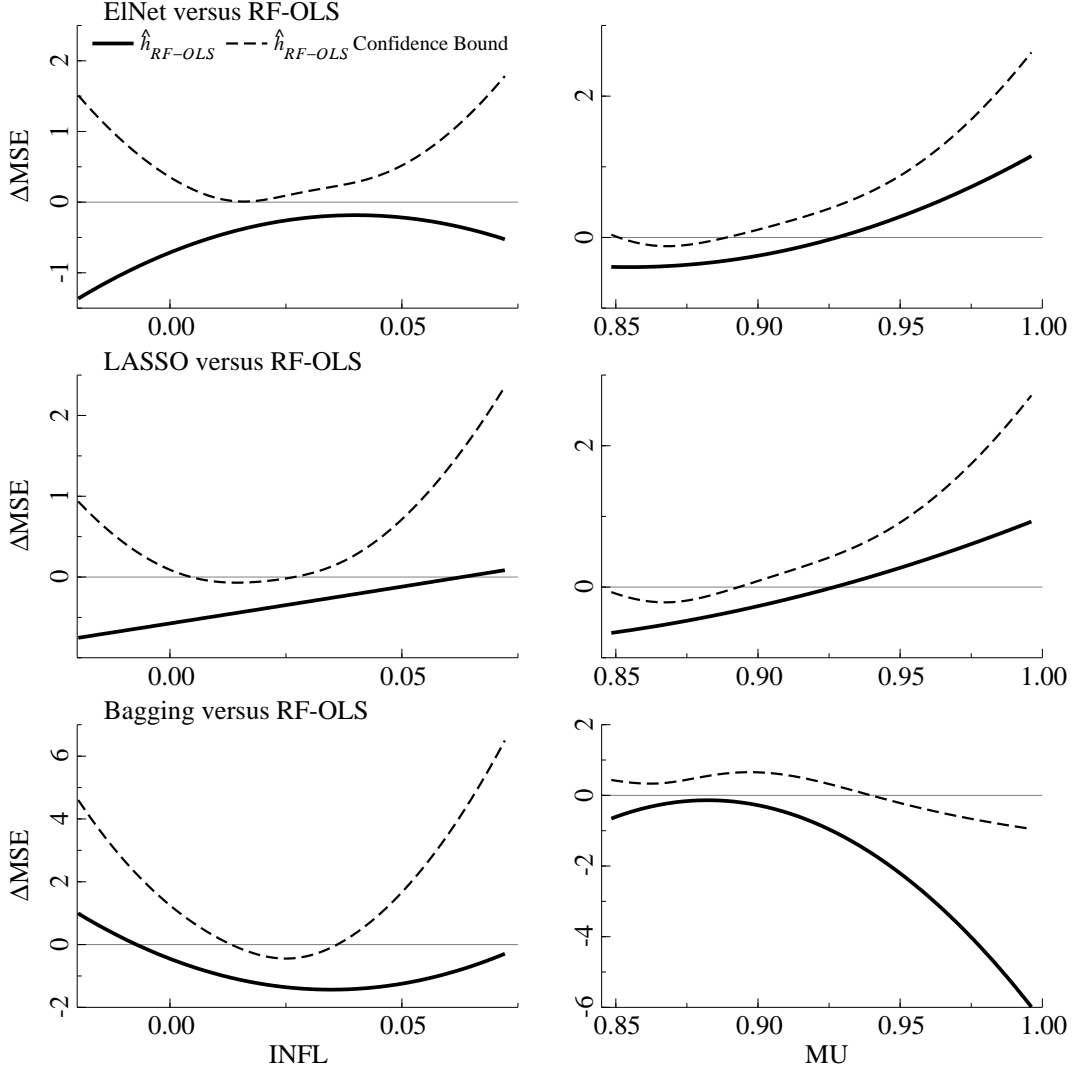
Looking at the left column, we see that when INFL is the conditioning variable, RF-OLS’s conditional expected loss is generally lower than EInet, LASSO, and bagging. This difference is marginally insignificant for the EInet benchmark, but is significant at the 5% level for LASSO and bagging, for which the rejections occur around states with  $\text{INFL} \approx 2\%$ .

Using MU as the conditioning variable reveals a quite different, and perhaps more interesting, pattern. While all three benchmarks are rejected, the underlying reasons are different. On one hand, EInet and LASSO behave quite similarly, as they both underperform RF-OLS in low-uncertainty states, and outperform the latter in high-uncertainty states. These shrinkage-based methods are rejected mainly due to their inaccuracy at low-uncertainty times. On the other hand, bagging appears to be uniformly dominated by the random-forest methods across all states. In contrast to EInet and LASSO, bagging’s underperformance mostly stems from the high-uncertainty states. This finding provides some empirical validation for the concern that some machine learning techniques may over-fit the data during “normal times,” but behave poorly in rare—but important—economic environments that are not seen in the training sample. The CSPA test, and conditional tests in general, is useful to shed light on this issue.

## 6 Conclusion

Motivated by the ever-increasing variety of forecasting methods in many areas of research, we introduce a new test for conditional superior predictive accuracy, or CSPA. This test examines the conditional state-dependent performance of competing forecast methods, and imposes a more stringent uniform weak dominance requirement on the benchmark. Two empirical applications

Figure 4: Forecasting Inflation: One-versus-one CSPA Tests



*Note:* This figure plots the conditional expected loss differential functions  $\hat{h}_{j,n}(\cdot)$  (solid) with lagged inflation (INFL) and macro uncertainty (MU) as the conditioning state variable on the left and right columns, respectively. Elastic Net, LASSO, and bagging methods are used separately as benchmark models, and the competing alternative model is fixed to be RF-OLS. The one-versus-one CSPA test rejects the null hypothesis at 5% significance level if some part of the 95% upper confidence bound (dashed) is below zero.

from financial and macroeconomic settings demonstrate the discriminating power of the CSPA test relative to its conventional unconditional counterpart. Econometrically, this is achieved by introducing and extending recently developed theoretical tools for the uniform nonparametric inference in the time-series setting, as CSPA concerns functional inequalities defined by conditional expected loss differentials. To the best of our knowledge, this is the first application of conditional-moment-inequality methods in time-series econometrics, and the theoretical tools developed here are broadly useful for other types of inference problems involving partial identification and dependent data.

## References

- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71(2), 529–626.
- ANDREWS, D. W. K. (1991a): “Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models,” *Econometrica*, 59(2), 307–345.
- (1991b): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59(3), 817–858.
- ANDREWS, D. W. K., AND J. C. MONAHAN (1992): “An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator,” *Econometrica*, 60(4), 953–966.
- ATKESON, A., AND L. E. OHANIAN (2001): “Are Phillips Curves Useful for Forecasting Inflation?,” *Federal Reserve bank of Minneapolis Quarterly Review*, 25(1), 2–11.
- BAI, J., AND S. NG (2008): “Forecasting Economic Time Series using Targeted Predictors,” *Journal of Econometrics*, 146(2), 304–317.
- BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2016): “Measuring Economic Policy Uncertainty,” *The Quarterly Journal of Economics*, 131(4), 1593–1636.
- BELLONI, A., V. CHERNOZHUKOV, D. CHETVERIKOV, AND K. KATO (2015): “Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results,” *Journal of Econometrics*, 186(2), 345 – 366.
- BOLLERSLEV, T., A. J. PATTON, AND R. QUAEDEVLIIEG (2016): “Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting,” *Journal of Econometrics*, 192(1), 1–18.

- BOLLERSLEV, T., A. J. PATTON, AND R. QUAEDVLIEG (2018): “Modeling and Forecasting (un) Reliable Realized Covariances for more Reliable Financial Decisions,” *Journal of Econometrics*, 207(1), 71–91.
- BREIMAN, L. (1996): “Bagging Predictors,” *Machine learning*, 24(2), 123–140.
- (2001): “Random Forests,” *Machine learning*, 45(1), 5–32.
- CHEN, X. (2007): “Large Sample Sieve Estimation of Semi-Nonparametric Models,” in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, chap. 76. Elsevier, 1 edn.
- CHEN, X., AND T. M. CHRISTENSEN (2015): “Optimal Uniform Convergence Rates and Asymptotic Normality for Series Estimators under Weak Dependence and Weak Conditions,” *Journal of Econometrics*, 188(2), 447–465.
- CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2014): “Gaussian Approximation of Suprema of Empirical Processes,” *Annals of Statistics*, 42(4), 1564–1597.
- CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): “Intersection Bounds: Estimation and Inference,” *Econometrica*, 81(2), 667–737.
- CORSI, F. (2009): “A Simple Approximate Long-Memory Model of Realized Volatility,” *Journal of Financial Econometrics*, 7(2), 174–196.
- DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press.
- DIEBOLD, F. X., AND R. S. MARIANO (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 13(3), 253–263.
- FAUST, J., AND J. H. WRIGHT (2009): “Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset,” *Journal of Business & Economic Statistics*, 27(4), 468–479.
- (2013): “Forecasting Inflation,” in *Handbook of economic forecasting*, vol. 2, pp. 2–56. Elsevier.
- GIACOMINI, R., AND H. WHITE (2006): “Tests of Conditional Predictive Ability,” *Econometrica*, 74(6), 1545–1578.
- HANSEN, P. R. (2005): “A Test for Superior Predictive Ability,” *Journal of Business & Economic Statistics*, 23(4), 365–380.

- HANSEN, P. R., A. LUNDE, AND J. M. NASON (2011): “The Model Confidence Set,” *Econometrica*, 79(2), 453–497.
- INOUE, A., AND L. KILIAN (2008): “How Useful is Bagging in Forecasting Economic Time Series? A Case Study of US Consumer Price Inflation,” *Journal of the American Statistical Association*, 103(482), 511–522.
- JURADO, K., S. C. LUDVIGSON, AND S. NG (2015): “Measuring uncertainty,” *American Economic Review*, 105(3), 1177–1216.
- LI, J., AND Z. LIAO (2019): “Uniform Nonparametric Inference for Time Series,” *Journal of Econometrics*, Forthcoming.
- MCCRACKEN, M. W., AND S. NG (2016): “FRED-MD: A Monthly Database for Macroeconomic Research,” *Journal of Business & Economic Statistics*, 34(4), 574–589.
- MEDEIROS, M. C., G. VASCONCELOS, A. VEIGA, AND E. ZILBERMAN (2019): “Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods,” *Journal of Business & Economic Statistics*, forthcoming.
- MOLINARI, F. (2019): “Econometrics with Partial Identification,” Discussion paper.
- NEWBY, W. K. (1997): “Convergence Rates and Asymptotic Normality for Series Estimators,” *Journal of Econometrics*, 79(1), 147 – 168.
- NEWBY, W. K., AND K. D. WEST (1987): “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55(3), 703–708.
- PATTON, A. J. (2011): “Data-based Ranking of Realised Volatility Estimators,” *Journal of Econometrics*, 161(2), 284–303.
- STOCK, J. H., AND M. W. WATSON (2009): “Phillips Curve Inflation Forecasts,” in *Understanding Inflation and the Implications for Monetary Policy*, ed. by J. Fuhrer, Y. Kodrzycki, J. Little, and G. Olivei. MIT Press, Cambridge.
- (2010): “Modeling Inflation after the Crisis,” Discussion paper, National Bureau of Economic Research.
- TIBSHIRANI, R. (1996): “Regression Shrinkage and Selection Via the Lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288.

WHITE, H. (2000): “A Reality Check for Data Snooping,” *Econometrica*, 68(5), 1097–1126.

ZOU, H., AND T. HASTIE (2005): “Regularization and Variable Selection Via the Elastic Net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.

## A Proofs

In this appendix, we prove the theoretical results in the main text. To highlight our new technical contributions, we demonstrate the main steps of our proofs here, and relegate additional technical details to the Supplemental Appendix. Throughout this appendix, we use  $K$  to denote a generic positive finite constant, which may vary from line to line.

### A.1 Proof of Proposition 1

The proof of Proposition 1 shares some similar steps with Chernozhukov, Lee, and Rosen’s (2013) proofs of their Theorem 2 and Lemma 5. To avoid repetition, we only highlight the key difference here and provide a detailed step-by-step proof in the Supplemental Appendix. An important difference stems from the fact that we rely on a strong approximation condition in the form of Assumption 3, instead of the (stronger) Yurinskii-type coupling used in Lemma 5 of Chernozhukov, Lee, and Rosen (2013), which in turn is needed to verify their high-level Condition C.2 for the uniform strong approximation of the t-statistic process. A key step of our analysis is to establish Theorem A1 below. We denote

$$\sigma_{j,n}(x) \equiv \sqrt{P(x)^\top Q_n^{-1} A_n(j, j) Q_n^{-1} P(x)},$$

where  $A_n(j, j)$  is the  $j$ th  $m_n \times m_n$  diagonal block of the matrix  $A_n$ . For two random variables  $X$  and  $Y$ , we write  $X \stackrel{d}{=} Y$  if they have the same distribution.

**Theorem A1.** *Let  $V_n$  be any subset of  $\mathcal{V} = \{(j, x) : 1 \leq j \leq J \text{ and } x \in \mathcal{X}\}$  and  $\tilde{N}_n$  be a generic sequence of random variables such that  $\tilde{N}_n \sim \mathcal{N}(0, A_n)$ . Suppose that Assumptions 2 and 3 hold. Then, there exist random sequences  $\tilde{U}_{1,n}$ ,  $\tilde{U}_{2,n}$ , and  $\tilde{U}_{3,n}$  such that*

$$\tilde{U}_{1,n} \stackrel{d}{=} \tilde{U}_{2,n} \stackrel{d}{=} \sup_{(j,x) \in V_n} \frac{P(x)^\top Q_n^{-1} \tilde{N}_{j,n}}{\sigma_{j,n}(x)}, \quad \tilde{U}_{3,n} \stackrel{d}{=} \sup_{(j,x) \in V_n} \left| \frac{P(x)^\top Q_n^{-1} \tilde{N}_{j,n}}{\sigma_{j,n}(x)} \right|,$$

and

$$\begin{aligned} \sup_{(j,x) \in V_n} \frac{n^{1/2}(\hat{h}_{j,n}(x) - h_j(x))}{\sigma_{j,n}(x)} - \tilde{U}_{1,n} &= o_p((\log n)^{-1}), \\ \sup_{(j,x) \in V_n} \frac{n^{1/2}(h_j(x) - \hat{h}_{j,n}(x))}{\sigma_{j,n}(x)} - \tilde{U}_{2,n} &= o_p((\log n)^{-1}), \\ \sup_{(j,x) \in V_n} \left| \frac{n^{1/2}(h_j(x) - \hat{h}_{j,n}(x))}{\sigma_{j,n}(x)} \right| - \tilde{U}_{3,n} &= o_p((\log n)^{-1}), \end{aligned}$$

where  $\tilde{N}_{j,n}$  is the  $m_n$ -dimensional subvector defined implicitly by the partition  $\tilde{N}_n^\top = (\tilde{N}_{1,n}^\top, \dots, \tilde{N}_{J,n}^\top)$ .

PROOF. Step 1. In this step, we show that

$$\sup_{(j,x) \in \mathcal{V}} \left| \frac{n^{1/2}(\hat{h}_{j,n}(x) - h_j(x))}{\sigma_{j,n}(x)} - \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right| = o_p((\log n)^{-1}). \quad (\text{A.1})$$

Denote  $h_{j,n}^*(\cdot) = P(\cdot)^\top b_{j,n}^*$ , with  $b_{j,n}^*$  given by Assumption 2. By Assumption 2,  $\sigma_{j,n}(\cdot)$  is bounded away from zero. We then deduce that

$$\sup_{(j,x) \in \mathcal{V}} \left| \frac{n^{1/2}(h_j(x) - h_{j,n}^*(x))}{\sigma_{j,n}(x)} \right| \leq K \sup_{(j,x) \in \mathcal{V}} \left| n^{1/2}(h_j(x) - h_{j,n}^*(x)) \right| = o_p((\log n)^{-1}). \quad (\text{A.2})$$

Observe that

$$\hat{b}_{j,n} - b_{j,n}^* = \hat{Q}_n^{-1} \left( n^{-1} \sum_{t=1}^n P(X_t) u_{j,t} + n^{-1} \sum_{t=1}^n P(X_t) (h_j(X_t) - h_{j,n}^*(X_t)) \right). \quad (\text{A.3})$$

Therefore, by the triangle inequality and the Cauchy–Schwarz inequality, we have uniformly in  $x$ ,

$$\begin{aligned} & \left| \frac{n^{1/2}(\hat{h}_{j,n}(x) - h_{j,n}^*(x))}{\sigma_{j,n}(x)} - \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right| \\ & \leq \left| \frac{P(x)^\top (\hat{Q}_n^{-1} - Q_n^{-1})}{\sigma_{j,n}(x)} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right| \\ & \quad + \left| \frac{P(x)^\top \hat{Q}_n^{-1}}{\sigma_{j,n}(x)} n^{-1/2} \sum_{t=1}^n P(X_t) (h_j(X_t) - h_{j,n}^*(X_t)) \right| \\ & \leq K \left\| \hat{Q}_n^{-1} - Q_n^{-1} \right\|_S \left\| n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right\| \\ & \quad + K \left\| \hat{Q}_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) (h_j(X_t) - h_{j,n}^*(X_t)) \right\|. \end{aligned} \quad (\text{A.4})$$

It is easy to see that  $\|n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t}\| = O_p(m_n^{1/2})$ . Then, by Assumption 2,

$$\left\| \hat{Q}_n^{-1} - Q_n^{-1} \right\|_S \left\| n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right\| = O_p(\delta_{Q,n} m_n^{1/2}) = o_p((\log n)^{-1}). \quad (\text{A.5})$$

In addition, we note that

$$\begin{aligned} & \left\| \widehat{Q}_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) (h_j(X_t) - h_{j,n}^*(X_t)) \right\|^2 \\ & \leq O_p(1) \sum_{t=1}^n (h_j(X_t) - h_{j,n}^*(X_t))^2 = o_p((\log n)^{-2}). \end{aligned} \quad (\text{A.6})$$

By (A.5) and (A.6), the majorant side of (A.4) can be further bounded by  $o_p((\log n)^{-1})$ . This estimate and (A.2) imply (A.1) as claimed.

Step 2. Recall that  $\zeta_n^L$  is the Lipschitz coefficient for the  $P(\cdot)$  function. By the triangle inequality and the Cauchy–Schwarz inequality, we have uniformly for  $x_1, x_2 \in \mathcal{X}$ ,

$$|\sigma_{j,n}(x_1) - \sigma_{j,n}(x_2)| = \frac{|\sigma_{j,n}^2(x_1) - \sigma_{j,n}^2(x_2)|}{\sigma_{j,n}(x_1) + \sigma_{j,n}(x_2)} \leq K \zeta_n^L \|x_1 - x_2\|. \quad (\text{A.7})$$

We then observe, for any  $j \in \{1, \dots, J\}$  and  $x_1, x_2 \in \mathcal{X}$ ,

$$\begin{aligned} & \left| \left( \frac{P(x_1)}{\sigma_{j,n}(x_1)} - \frac{P(x_2)}{\sigma_{j,n}(x_2)} \right)^\top Q_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right| \\ & \leq \|P(x_1) - P(x_2)\| \frac{\|Q_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t}\|}{\sigma_{j,n}(x_1)} \\ & \quad + \frac{|\sigma_{j,n}(x_1) - \sigma_{j,n}(x_2)| \|P(x_2)\|}{\sigma_{j,n}(x_1) \sigma_{j,n}(x_2)} \left\| Q_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right\| \\ & \leq K \zeta_n^L \|x_1 - x_2\| \left\| Q_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right\|, \end{aligned} \quad (\text{A.8})$$

where the first inequality is by the triangle inequality and the Cauchy–Schwarz inequality, and the second inequality follows from (A.7).

Let  $\varepsilon_n \equiv ((\log n)^2 m_n^{1/2} \zeta_n^L)^{-1}$ . Since  $\mathcal{X}$  is compact, there exists a set of points  $\{x_i\}_{i=1}^{K_n} \subseteq \mathcal{X}$  with  $K_n = O(\varepsilon_n^{-d_x})$  such that each  $x \in \mathcal{X}$  can be matched with some  $x_i$  satisfying  $\|x - x_i\| \leq \varepsilon_n$ . For each of such matched pairs,  $\{x, x_i\}$ , (A.8) implies

$$\begin{aligned} & \left| \left( \frac{P(x)^\top}{\sigma_{j,n}(x)} - \frac{P(x_i)^\top}{\sigma_{j,n}(x_i)} \right) Q_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right| \\ & \leq K \zeta_n^L \|x - x_i\| \left\| Q_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right\| \\ & \leq K (\log n)^{-2} m_n^{-1/2} \left\| Q_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right\|. \end{aligned} \quad (\text{A.9})$$

In addition, we associate with the set  $V_n$  a finite subset  $\widetilde{V}_n \subseteq \{1, \dots, J\} \times \mathcal{X}$  defined as

$$\widetilde{V}_n = \{(j, \tilde{x}) : \tilde{x} \in \{x_i\}_{1 \leq i \leq K_n} \text{ and } \|\tilde{x} - x\| \leq \varepsilon_n \text{ for some } x \text{ such that } (j, x) \in V_n\}.$$



By (A.9), we deduce

$$\begin{aligned} & \left| \sup_{(j,x) \in V_n} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} - \max_{(j,x) \in \tilde{V}_n} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right| \\ & \leq K (\log n)^{-2} m_n^{-1/2} \max_{1 \leq j \leq J} \left\| Q_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} \right\| = o_p((\log n)^{-1}), \end{aligned} \quad (\text{A.10})$$

where the  $o_p(\cdot)$  statement follows from  $\|Q_n^{-1} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t}\| = O_p(m_n^{1/2})$ .

Define  $L_n$  as the cardinality of  $\tilde{V}_n$ . Consider a generic  $Jm_n$ -vector  $z = (z_1^\top, \dots, z_J^\top)^\top$  with each  $z_j$  component being  $m_n$ -dimensional. It is easy to see that we can find  $Jm_n$ -vectors  $\alpha_l$ ,  $1 \leq l \leq L_n$ , such that for all  $z$ ,

$$\max_{1 \leq l \leq L_n} \alpha_l^\top z = \max_{(j,x) \in \tilde{V}_n} \frac{P(x)^\top Q_n^{-1} z_j}{\sigma_{j,n}(x)},$$

and

$$\sup_{1 \leq l \leq L_n} \|\alpha_l\| \leq \sup_{(j,x) \in \mathcal{V}} \left\| \frac{P(x) Q_n^{-1}}{\sigma_{j,n}(x)} \right\| \leq K.$$

Note that  $L_n = O(K_n) = O(\varepsilon_n^{-d_x})$  satisfies the requirement in Assumption 3, which implies the existence of a random sequence  $\tilde{U}_n$  satisfying  $\tilde{U}_n \stackrel{d}{=} \max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{N}_n$  and

$$\left| \max_{1 \leq l \leq L_n} \alpha_l^\top \left( n^{-1/2} \sum_{t=1}^n u_t \otimes P(X_t) \right) - \tilde{U}_n \right| = o_p((\log n)^{-1}). \quad (\text{A.11})$$

By the definition of  $\alpha_l$ , we can rewrite (A.11) as

$$\left| \max_{(j,x) \in \tilde{V}_n} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} - \tilde{U}_n \right| = o_p((\log n)^{-1}), \quad (\text{A.12})$$

and also note that

$$\tilde{U}_n \stackrel{d}{=} \max_{(j,x) \in \tilde{V}_n} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} \tilde{N}_{j,n}. \quad (\text{A.13})$$

Following the same argument leading to (A.10), we can also show that

$$\left| \sup_{(j,x) \in V_n} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} \tilde{N}_{j,n} - \max_{(j,x) \in \tilde{V}_n} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} \tilde{N}_{j,n} \right| = o_p((\log n)^{-1}). \quad (\text{A.14})$$

By (A.13), (A.14), and Lemma 9 in Chernozhukov, Lee, and Rosen (2013), there exists another random sequence  $\tilde{U}_{1,n}$  such that

$$\tilde{U}_{1,n} \stackrel{d}{=} \sup_{(j,x) \in V_n} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} \tilde{N}_{j,n}, \quad \tilde{U}_n - \tilde{U}_{1,n} = o_p((\log n)^{-1}). \quad (\text{A.15})$$

From (A.10), (A.12), and (A.15), we further deduce

$$\left| \sup_{(j,x) \in V_n} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} n^{-1/2} \sum_{t=1}^n P(X_t) u_{j,t} - \tilde{U}_{1,n} \right| = o_p((\log n)^{-1}). \quad (\text{A.16})$$

The assertion of the lemma concerning  $\tilde{U}_{1,n}$  then follows from (A.1) and (A.16).

Applying the same arguments with  $(\alpha_l)_{1 \leq l \leq L_n}$  replaced by  $(-\alpha_l)_{1 \leq l \leq L_n}$  and  $(\alpha_l, -\alpha_l)_{1 \leq l \leq L_n}$ , we can prove the other two assertions of the lemma, respectively.  $\square$

**PROOF OF PROPOSITION 1.** We prove this proposition by adapting Chernozhukov, Lee, and Rosen's (2013) proof of their Theorem 2. The main change is to use the three types of coupling results in Theorem A1 in place of Chernozhukov, Lee, and Rosen's Condition C.2 for coupling the entire t-statistic process. To avoid repetition, we relegate the (somewhat tedious) details to the Supplemental Appendix.  $\square$

## A.2 Proofs of Theorem 1 and Corollary 1

The proof consists of several steps. In the first step (see Lemma A1), we approximate the sequence  $\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^*$  with  $\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^+$ , where  $S_n^+$  is constructed as a martingale with its predictable quadratic covariation being equal to the deterministic matrix

$$\Sigma_n^* \equiv \sum_{t=1}^n \mathbb{E} [V_{n,t}^*].$$

In the second step (see Lemma A2), we establish a coupling for  $\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^+$  using Lindeberg's method and Strassen's theorem, for which the fact that  $\Sigma_n^*$  is deterministic is crucial. These approximation results can then be used to construct the coupling for the original  $\max_{1 \leq l \leq L_n} \alpha_l^\top S_n$  statistic.

We start with constructing the aforementioned martingale  $S_n^+$ . The construction is based on the same scheme as in Li and Liao (2019), which we recall as follows. Consider the following stopping time:

$$\tau_n \equiv \max \left\{ t \in \{1, \dots, n\} : \Sigma_n^* - \sum_{s=1}^t V_{n,s}^* \text{ is positive semi-definite} \right\},$$

with the convention that  $\max \emptyset = 0$ . Note that  $\tau_n$  is a stopping time because  $V_{n,t}^*$  is  $\mathcal{F}_{n,t-1}$ -measurable for each  $t$  and  $\Sigma_n^*$  is nonrandom. The matrix

$$\Xi_n \equiv \begin{cases} \Sigma_n^* & \text{when } \tau_n = 0, \\ \Sigma_n^* - \sum_{t=1}^{\tau_n} V_{n,t}^* & \text{when } \tau_n \geq 1, \end{cases}$$

is positive semi-definite by construction.

Let  $k_n$  be a sequence of integers such that  $k_n \rightarrow \infty$  and let  $(\eta_{n,t})_{n+1 \leq t \leq n+k_n}$  be independent  $M_n$ -dimensional standard normal vectors. We construct another martingale difference array  $(Z_{n,t}, \mathcal{H}_{n,t})_{1 \leq t \leq n+k_n}$  as follows:

$$Z_{n,t} \equiv \begin{cases} n^{-1/2} X_{n,t}^* \mathbf{1}_{\{t \leq \tau_n\}} & \text{when } 1 \leq t \leq n, \\ k_n^{-1/2} \Xi_n^{1/2} \eta_{n,t} & \text{when } n+1 \leq t \leq n+k_n, \end{cases}$$

and the filtration is given by

$$\mathcal{H}_{n,t} \equiv \begin{cases} \mathcal{F}_{n,t} & \text{when } 1 \leq t \leq n, \\ \mathcal{F}_{n,n} \vee \sigma(\eta_{n,s} : s \leq t) & \text{when } n+1 \leq t \leq n+k_n. \end{cases}$$

Since  $\tau_n$  is a stopping time, it is easy to verify that  $(Z_{n,t}, \mathcal{H}_{n,t})_{1 \leq t \leq n+k_n}$  indeed forms a martingale difference array. We denote

$$V_{n,t}^+ \equiv \mathbb{E} \left[ Z_{n,t} Z_{n,t}^\top \middle| \mathcal{H}_{n,t-1} \right] \quad (\text{A.17})$$

and set

$$S_n^+ \equiv \sum_{t=1}^{n+k_n} Z_{n,t}. \quad (\text{A.18})$$

Note that the predictable quadratic covariation matrix of  $S_n^+$  is exactly  $\Sigma_n^*$ , that is,

$$\sum_{t=1}^{n+k_n} V_{n,t}^+ = \sum_{t=1}^{\tau_n} V_{n,t}^* + \Xi_n = \Sigma_n^*. \quad (\text{A.19})$$

The approximation error between  $\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^*$  and  $\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^+$  is quantified by Lemma A1, below. We recall from Theorem 1 that  $B_{2,n}$  is defined as

$$B_{2,n} \equiv \min \left\{ L_n^{1/q} \max_{l,t} \|\alpha_l^\top X_{n,t}^*\|_q, M_n^{1/2} \right\} + (\log L_n)^{1/2}.$$

**Lemma A1.** *Under Assumption 4,  $|\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^* - \max_{1 \leq l \leq L_n} \alpha_l^\top S_n^+| = O_p(B_{2,n} r_n^{1/2})$ .*

PROOF. See the Supplemental Appendix. □

In the next step, we construct a sequence of Gaussian random vectors  $\tilde{S}_n^* \sim \mathcal{N}(0, \Sigma_n^*)$  such that the distribution of  $\max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n^*$  is “close” to that of  $\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^+$  in the sense stated by Lemma A2 below. Specifically, let  $(\zeta_{n,t})_{1 \leq t \leq n+k_n}$  be independent  $M_n$ -dimensional standard normal vectors that are also independent of  $\mathcal{H}_{n,n+k_n}$ , and then set

$$\tilde{S}_n^* \equiv \sum_{t=1}^{n+k_n} \tilde{\zeta}_{n,t}, \quad \text{where } \tilde{\zeta}_{n,t} \equiv (V_{n,t}^+)^{1/2} \zeta_{n,t}.$$

By (A.19),  $\tilde{S}_n^* \sim \mathcal{N}(0, \Sigma_n^*)$ . The next lemma quantifies the difference between the distributions of  $\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^+$  and  $\max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n^*$ . Below, for any Borel subset  $A \subseteq \mathbb{R}$  and any constant  $\delta > 0$ , we denote the  $\delta$ -enlargement of  $A$  by  $A^\delta$ , that is,

$$A^\delta \equiv \left\{ x \in \mathbb{R} : \inf_{y \in A} \|x - y\| \leq \delta \right\}.$$

We remind the reader that the  $B_{1,n}$  sequence is defined in Theorem 1.

**Lemma A2.** *Suppose that Assumption 4 holds and  $k_n \geq (\log L_n)^3 B_{1,n}^{-2}$ . Then, for each  $C > 5$ ,*

$$\mathbb{P} \left( \max_{1 \leq l \leq L_n} \alpha_l^\top S_n^+ \in A \right) \leq \mathbb{P} \left( \max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n^* \in A^{CB_{1,n}^{1/3}(\log L_n)^{2/3}} \right) + \epsilon_n(C),$$

where  $\epsilon_n(C)$  is a real sequence satisfying  $\sup_{n \geq 1} \epsilon_n(C) \rightarrow 0$  as  $C \rightarrow \infty$ .

PROOF. See the Supplemental Appendix. □

We are now ready to prove Theorem 1 and its corollary.

PROOF OF THEOREM 1. Since  $B_{1,n}^{1/3}(\log L_n)^{2/3} = o((\log n)^{-1})$  by assumption, we can find a real sequence  $C_n \rightarrow \infty$  such that

$$C_n B_{1,n}^{1/3}(\log L_n)^{2/3} = o((\log n)^{-1}).$$

By Lemma A2 (for which the condition on  $k_n$  can be trivially verified by taking  $k_n$  sufficiently large), we have for each  $n \geq 1$ ,

$$\mathbb{P} \left( \max_{1 \leq l \leq L_n} \alpha_l^\top S_n^+ \in A \right) \leq \mathbb{P} \left( \max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n^* \in A^{C_n B_{1,n}^{1/3}(\log L_n)^{2/3}} \right) + \epsilon_n(C_n).$$

By Strassen's theorem, for each  $n$ , we can construct a random variable  $\tilde{U}_n^*$  such that  $\tilde{U}_n^* \stackrel{d}{=} \max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n^*$  and

$$\mathbb{P} \left( \left| \max_{1 \leq l \leq L_n} \alpha_l^\top S_n^+ - \tilde{U}_n^* \right| > o((\log n)^{-1}) \right) \leq \epsilon_n(C_n).$$

Since  $\epsilon_n(C_n) \rightarrow 0$ , we deduce from the above estimate that

$$\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^+ - \tilde{U}_n^* = o_p((\log n)^{-1}).$$

By Lemma A1, we further have

$$\max_{1 \leq l \leq L_n} \alpha_l^\top S_n^* - \tilde{U}_n^* = o_p((\log n)^{-1}). \tag{A.20}$$

Let  $\tilde{S}_n \equiv (\Sigma_n)^{1/2}(\Sigma_n^*)^{-1/2}\tilde{S}_n^*$ . Note that  $\tilde{S}_n \sim \mathcal{N}(0, \Sigma_n)$ . By definition,

$$\alpha_l^\top \tilde{S}_n - \alpha_l^\top \tilde{S}_n^* = \alpha_l^\top ((\Sigma_n)^{1/2} - (\Sigma_n^*)^{1/2})(\Sigma_n^*)^{-1/2} \tilde{S}_n^*$$

where  $(\Sigma_n^*)^{-1/2}\tilde{S}_n^*$  is a standard normal random vector. By the triangle inequality,

$$\left| \max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n - \max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n^* \right| \leq \max_{1 \leq l \leq L_n} \left| \alpha_l^\top \tilde{S}_n - \alpha_l^\top \tilde{S}_n^* \right|.$$

With an appeal to the maximum inequality, we deduce

$$\begin{aligned} \mathbb{E} \left[ \max_{1 \leq l \leq L_n} \left| \alpha_l^\top \tilde{S}_n - \alpha_l^\top \tilde{S}_n^* \right| \right] &\leq K(\log L_n)^{1/2} \max_{1 \leq l \leq L_n} \left\| \alpha_l^\top ((\Sigma_n)^{1/2} - (\Sigma_n^*)^{1/2}) \right\| \\ &\leq K(\log L_n)^{1/2} \|\Sigma_n - \Sigma_n^*\|_{\mathcal{S}} \\ &= O((\log L_n)^{1/2} \bar{c}_n M_n^{1/2} n^{-1/2}), \end{aligned}$$

where the last line is by (A.79) in the Supplemental Appendix of Li and Liao (2019). Combining the above estimates, we have

$$\left| \max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n - \max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n^* \right| = O_p \left( (\log L_n)^{1/2} \bar{c}_n M_n^{1/2} n^{-1/2} \right) = o_p \left( (\log n)^{-1} \right).$$

By this estimate and Lemma 9 in Chernozhukov, Lee, and Rosen (2013), we can construct another random sequence  $\tilde{U}_n$  such that

$$\tilde{U}_n \stackrel{d}{=} \max_{1 \leq l \leq L_n} \alpha_l^\top \tilde{S}_n, \quad \tilde{U}_n - \tilde{U}_n^* = o_p \left( (\log n)^{-1} \right). \quad (\text{A.21})$$

The assertion of the theorem then follows from Assumption 4(i), (A.20), and (A.21).  $\square$

PROOF OF COROLLARY 1. The proof is done by verifying the conditions of Theorem 1. Since  $V_{n,t}^*$  is deterministic, we can set

$$r_n = 0. \quad (\text{A.22})$$

By the boundedness of  $(\alpha_l)$  and the Cauchy–Schwarz inequality,  $|\alpha_l^\top X_{n,t}^*| \leq K \|X_{n,t}^*\|$ . Hence,

$$\begin{aligned} \max_{1 \leq l \leq L_n} \left( \mathbb{E} \left[ (\alpha_l^\top X_{n,t}^*)^2 | \mathcal{F}_{n,t-1} \right] \right)^{3/2} &\leq K \left( \mathbb{E} \left[ \|X_{n,t}^*\|^2 | \mathcal{F}_{n,t-1} \right] \right)^{3/2} \\ &\leq K \mathbb{E} \left[ \|X_{n,t}^*\|^3 | \mathcal{F}_{n,t-1} \right], \end{aligned}$$

where the second inequality is by Jensen's inequality. It is then easy to see that

$$\begin{aligned} B_{1,n} &\equiv n^{-3/2} \sum_{t=1}^n \mathbb{E} \left[ (\log L_n)^{3/2} \max_{1 \leq l \leq L_n} \left( \mathbb{E} \left[ (\alpha_l^\top X_{n,t}^*)^2 | \mathcal{F}_{n,t-1} \right] \right)^{3/2} + \max_{1 \leq l \leq L_n} |\alpha_l^\top X_{n,t}^*|^3 \right] \\ &\leq K n^{-3/2} \sum_{t=1}^n \left( (\log L_n)^{3/2} + 1 \right) \mathbb{E} \left[ \|X_{n,t}^*\|^3 \right] \leq K \bar{c}_n^3 (\log L_n)^{3/2} n^{-1/2} M_n^{3/2}. \end{aligned}$$

Therefore,

$$B_{1,n}^{1/3}(\log M_n)^{2/3} \leq K\bar{c}_n(\log L_n)^{1/2}(\log M_n)^{2/3}M_n^{1/2}n^{-1/6}.$$

Since  $M_n$  and  $L_n$  both have polynomial growth in  $n$ ,  $\log L_n = O(\log n)$  and  $\log M_n = O(\log n)$ .

Therefore,

$$B_{1,n}^{1/3}(\log M_n)^{2/3} = o\left((\log n)^2 n^{\frac{\alpha}{2} - \frac{1}{6}}\right) = o((\log n)^{-1}). \quad (\text{A.23})$$

By a similar argument, it is easy to see that

$$\bar{c}_n(\log L_n)^{1/2}M_n^{1/2}n^{-1/2} = o((\log n)^{-1}). \quad (\text{A.24})$$

With (A.22), (A.23), and (A.24), we readily verify the conditions of Theorem 1, which finishes the proof of this corollary.  $\square$

# Supplemental Appendix to Conditional Superior Predictive Ability

Jia Li

Department of Economics, Duke University

Zhipeng Liao

Department of Economics, UCLA

Rogier Quaadvlieg

Department of Business Economics, Erasmus School of Economics

February 11, 2020

## **Abstract**

This supplemental appendix contains additional technical details. Section SA provides a detailed proof of Proposition 1 in the main text. Section SB provides proofs for Lemma A1 and Lemma A2 that are used for proving Theorem 1 in the main text. In Section SC, we describe the pre-whitened HAC estimator and establish its rate of convergence under the growing-dimensional setting, which can be used to verify the high-level condition Assumption 1.

## SA Detailed proof of Proposition 1

In this section, we provide a detailed proof for Proposition 1 in the main text. Section SA.1 introduces some notation and preliminary technical results. Section SA.2 presents the proof of that proposition.

### SA.1 Notation and preliminary results

We need several technical lemmas. The first lemma concerns the uniform convergence rate of the estimator of the standard error function. We recall that

$$\mathcal{V} \equiv \{(j, x) : 1 \leq j \leq J \text{ and } x \in \mathcal{X}\}$$

and

$$\sigma_{j,n}(x) \equiv \sqrt{P(x)^\top Q_n^{-1} A_n(j, j) Q_n^{-1} P(x)},$$

where  $A_n(j, j)$  is the  $j$ th  $m_n \times m_n$  diagonal block of  $A_n$ .

**Lemma SA1.** *Under Assumptions 1 and 2, we have*

$$\sup_{(j,x) \in \mathcal{V}} \left| \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} - 1 \right| = O_p(\delta_{Q,n} + \delta_{A,n}).$$

PROOF. Denote  $\Omega_n(j, j) = Q_n^{-1} A_n(j, j) Q_n^{-1}$ . By definition,

$$\begin{aligned} \hat{\Omega}_n(j, j) - \Omega_n(j, j) &= \hat{Q}_n^{-1} \hat{A}_n(j, j) \hat{Q}_n^{-1} - Q_n^{-1} A_n(j, j) Q_n^{-1} \\ &= (\hat{Q}_n^{-1} - Q_n^{-1}) \hat{A}_n(j, j) \hat{Q}_n^{-1} \\ &\quad + Q_n^{-1} (\hat{A}_n(j, j) - A_n(j, j)) \hat{Q}_n^{-1} + Q_n^{-1} A_n(j, j) (\hat{Q}_n^{-1} - Q_n^{-1}). \end{aligned}$$

Under the maintained assumptions, the eigenvalues of  $Q_n$ ,  $A_n$ ,  $\hat{Q}_n$ , and  $\hat{A}_n$  are bounded from above and away from zero in probability. Hence, by the triangle inequality and the Cauchy-Schwarz inequality, we have uniformly in  $x$ ,

$$\begin{aligned} |\hat{\sigma}_{j,n}^2(x) - \sigma_{j,n}^2(x)| &= \left| P(x)^\top (\hat{\Omega}_n(j, j) - \Omega_n(j, j)) P(x) \right| \\ &\leq O_p(1) \|P(x)\|^2 (\|\hat{Q}_n - Q_n\|_S + \|\hat{A}_n - A_n\|_S). \end{aligned}$$

We then deduce that, uniformly in  $(j, x)$ ,

$$\left| \frac{\hat{\sigma}_{j,n}^2(x) - \sigma_{j,n}^2(x)}{\hat{\sigma}_{j,n}^2(x)} \right| \leq O_p(1) \cdot (\|\hat{Q}_n - Q_n\|_S + \|\hat{A}_n - A_n\|_S) = O_p(\delta_{Q,n} + \delta_{A,n}),$$

and hence,

$$\left| \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} - 1 \right| \leq \frac{|\hat{\sigma}_{j,n}^2(x) - \sigma_{j,n}^2(x)|}{\hat{\sigma}_{j,n}^2(x)} = O_p(\delta_{Q,n} + \delta_{A,n}).$$



This finishes the proof.  $\square$

To state Lemma SA2, we introduce a selection operator. We partition a  $Jm_n$ -dimensional vector  $z$  as  $z = (z_1^\top, \dots, z_J^\top)^\top$ , where each  $z_j$  is  $m_n$ -dimensional. The selection operator  $[\cdot]_j$  is defined as

$$[z]_j = z_j \text{ for each } j \in \{1, \dots, J\}. \quad (\text{SA.1})$$

**Lemma SA2.** *Let  $\mathcal{D}_n$  denote the  $\sigma$ -field generated by the data and*

$$\Omega_n \equiv (I_J \otimes Q_n)^{-1} A_n (I_J \otimes Q_n)^{-1}.$$

*Under Assumptions 1 and 2, we have*

$$\mathbb{P} \left( \sup_{(j,x) \in \mathcal{V}} \left| \frac{P(x)^\top [\widehat{\Omega}_n^{1/2} \xi_n]_j}{\widehat{\sigma}_{j,n}(x)} - \frac{P(x)^\top [\Omega_n^{1/2} \xi_n]_j}{\sigma_{j,n}(x)} \right| > (\log n)^{-3/2} \middle| \mathcal{D}_n \right) = o_p(1),$$

*where  $\xi_n$  is a  $Jm_n$ -dimensional standard Gaussian random vector independent of  $\mathcal{D}_n$ .*

PROOF. By Assumptions 1 and 2, we have

$$\begin{aligned} \left\| \widehat{\Omega}_n - \Omega_n \right\|_S &\leq \left\| (\widehat{Q}_n^{-1} - Q_n^{-1}) \widehat{A}_n \widehat{Q}_n^{-1} \right\|_S \\ &\quad + \left\| Q_n^{-1} (\widehat{A}_n - A_n) \widehat{Q}_n^{-1} \right\|_S + \left\| Q_n^{-1} A_n (\widehat{Q}_n^{-1} - Q_n^{-1}) \right\|_S \\ &= O_p(\delta_{Q,n} + \delta_{A,n}). \end{aligned}$$

We can then use the same argument as in the proof of Lemma 5 in Chernozhukov, Lee, and Rosen (2013) to get

$$\mathbb{P} \left( \sup_{(j,x) \in \mathcal{V}} \left| \frac{P(x)^\top [\widehat{\Omega}_n^{1/2} \xi_n]_j}{\widehat{\sigma}_{j,n}(x)} - \frac{P(x)^\top [\Omega_n^{1/2} \xi_n]_j}{\sigma_{j,n}(x)} \right| > (\log n)^{-3/2} \middle| \mathcal{D}_n \right) = O_p((\delta_{Q,n} + \delta_{A,n})(\log n)^2).$$

Since both  $\delta_{Q,n}$  and  $\delta_{A,n}$  vanish at polynomial rates (as  $n \rightarrow \infty$ ) by assumption, the assertion of the lemma readily follows from the estimate above.  $\square$

We now need some additional notation. Consider a generic sequence of Gaussian vectors  $\widetilde{N}_n \sim \mathcal{N}(0, A_n)$  and set  $\widetilde{N}_{j,n} = [\widetilde{N}_n]_j$ ; recall that the selection operator  $[\cdot]_j$  is defined in (SA.1). For each subset  $V \subseteq \mathcal{V}$ , we denote

$$\kappa_{n,V}(q) \equiv \text{the } q\text{-quantile of } \sup_{(j,x) \in V} \frac{P(x)^\top Q_n^{-1} \widetilde{N}_{j,n}}{\sigma_{j,n}(x)}.$$

Recall the definition of  $\tilde{\gamma}_n$  from Algorithm 1 and pick a sequence  $\gamma_n$  such that  $\tilde{\gamma}_n \geq \gamma_n \geq \tilde{\gamma}_n - o(1)$ . We then set

$$\mathcal{V}_n \equiv \left\{ (j, x) \in \mathcal{V} : h_j(x) \leq \eta^* + n^{-1/2} \kappa_n \sigma_{j,n}(x) \right\}, \quad \text{where } \kappa_n \equiv \kappa_{n,\mathcal{V}}(\gamma_n).$$

Denote for each  $(j, x) \in \mathcal{V}$ ,

$$Z_{j,n}(x) \equiv \frac{n^{1/2} \left( h_j(x) - \hat{h}_{j,n}(x) \right)}{\sigma_{j,n}(x)}.$$

Applying Theorem A1 for  $V_n = \mathcal{V}_n$  or  $\mathcal{V}$ , we can find random sequences  $\tilde{U}'_{1,n}$ ,  $\tilde{U}_{2,n}$ ,  $\tilde{U}'_{2,n}$ , and  $\tilde{U}'_{3,n}$  such that

$$\left\{ \begin{array}{l} \tilde{U}'_{1,n} \stackrel{d}{=} \sup_{(j,x) \in \mathcal{V}} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} \tilde{N}_{j,n}, \quad \sup_{(j,x) \in \mathcal{V}} (-Z_{j,n}(x)) - \tilde{U}'_{1,n} = o_p((\log n)^{-1}), \\ \tilde{U}_{2,n} \stackrel{d}{=} \sup_{(j,x) \in \mathcal{V}_n} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} \tilde{N}_{j,n}, \quad \sup_{(j,x) \in \mathcal{V}_n} Z_{j,n}(x) - \tilde{U}_{2,n} = o_p((\log n)^{-1}), \\ \tilde{U}'_{2,n} \stackrel{d}{=} \sup_{(j,x) \in \mathcal{V}} \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} \tilde{N}_{j,n}, \quad \sup_{(j,x) \in \mathcal{V}} Z_{j,n}(x) - \tilde{U}'_{2,n} = o_p((\log n)^{-1}), \\ \tilde{U}'_{3,n} \stackrel{d}{=} \sup_{(j,x) \in \mathcal{V}} \left| \frac{P(x)^\top Q_n^{-1}}{\sigma_{j,n}(x)} \tilde{N}_{j,n} \right|, \quad \sup_{(j,x) \in \mathcal{V}} |Z_{j,n}(x)| - \tilde{U}'_{3,n} = o_p((\log n)^{-1}). \end{array} \right. \quad (\text{SA.2})$$

We note that  $\kappa_{n, \mathcal{V}_n}(q)$  is the  $q$ -quantile of  $\tilde{U}_{2,n}$  and  $\kappa_n$  is the  $\gamma_n$ -quantile of  $\tilde{U}'_{1,n}$  and  $\tilde{U}'_{2,n}$ .

**Lemma SA3.** *Under Assumptions 1, 2, and 3, we have*

$$\mathbb{P} \left( \sup_{(j,x) \in \mathcal{V}} \frac{n^{1/2}(\eta^* - \hat{h}_{j,n}(x))}{\hat{\sigma}_{j,n}(x)} \leq z \right) \geq \mathbb{P} \left( \tilde{U}_{2,n} \leq z \right) - o(1),$$

uniformly over  $z \in [0, \infty)$ .

PROOF. Step 1. We collect some technical estimates in this step. By (SA.2),

$$\left\{ \begin{array}{l} \tilde{\eta}_n \equiv \left| \sup_{(j,x) \in \mathcal{V}_n} Z_{j,n}(x) - \tilde{U}_{2,n} \right| = o_p((\log n)^{-1}), \\ \tilde{\eta}'_n \equiv \left| \sup_{(j,x) \in \mathcal{V}} Z_{j,n}(x) - \tilde{U}'_{2,n} \right| = o_p((\log n)^{-1}). \end{array} \right. \quad (\text{SA.3})$$

We set

$$R_n \equiv \sup_{(j,x) \in \mathcal{V}} \left\{ \left| \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} - 1 \right| (|Z_{j,n}(x)| + \kappa_n) \right\}.$$

By Lemma SA1,

$$\sup_{(j,x) \in \mathcal{V}} \left| \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} - 1 \right| = O_p(\delta_{Q,n} + \delta_{A,n}) = o_p((\log n)^{-1}).$$

Moreover, by the fourth line of (SA.2) and a maximal inequality for Gaussian processes,

$$\sup_{(j,x) \in \mathcal{V}} |Z_{j,n}(x)| = \tilde{U}'_{3,n} + o_p(1) = O_p((\log n)^{1/2}).$$

Since  $\kappa_n = O((\log n)^{1/2})$ , we can combine these estimates to deduce that

$$R_n = o_p\left((\log n)^{-1/2}\right). \quad (\text{SA.4})$$

Step 2. By the same argument as in the proof of Lemma 1 of Chernozhukov, Lee, and Rosen (2013), we obtain

$$\sup_{(j,x) \in \mathcal{V}} \frac{n^{1/2}(\eta^* - \hat{h}_{j,n}(x))}{\hat{\sigma}_{j,n}(x)} \leq \sup_{(j,x) \in \mathcal{V}_n} \left\{ \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} Z_{j,n}(x) \right\} \vee \sup_{(j,x) \notin \mathcal{V}_n} \left\{ \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} (Z_{j,n}(x) - \kappa_n) \right\}.$$

The two terms on the right-hand side of the above inequality can be further bounded as

$$\begin{aligned} \sup_{(j,x) \in \mathcal{V}_n} \left\{ \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} Z_{j,n}(x) \right\} &\leq \tilde{U}_{2,n} + R_n + \tilde{\eta}_n, \\ \sup_{(j,x) \notin \mathcal{V}_n} \left\{ \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} (Z_{j,n}(x) - \kappa_n) \right\} &\leq \tilde{U}'_{2,n} - \kappa_n + R_n + \tilde{\eta}'_n. \end{aligned}$$

Hence, by (SA.3) and (SA.4),

$$\sup_{(j,x) \in \mathcal{V}} \frac{n^{1/2}(\eta^* - \hat{h}_{j,n}(x))}{\hat{\sigma}_{j,n}(x)} \leq \tilde{U}_{2,n} \vee (\tilde{U}'_{2,n} - \kappa_n) + o_p\left((\log n)^{-1/2}\right).$$

Consequently, uniformly in  $z$ ,

$$\begin{aligned} &\mathbb{P}\left(\sup_{(j,x) \in \mathcal{V}} \frac{n^{1/2}(\eta^* - \hat{h}_{j,n}(x))}{\hat{\sigma}_{j,n}(x)} > z\right) \\ &\leq \mathbb{P}\left(\tilde{U}_{2,n} + o_p\left((\log n)^{-1/2}\right) > z\right) + \mathbb{P}\left(\tilde{U}'_{2,n} - \kappa_n + o_p\left((\log n)^{-1/2}\right) > 0\right) \\ &\leq \mathbb{P}\left(\tilde{U}_{2,n} > z\right) + \mathbb{P}\left(\tilde{U}'_{2,n} > \kappa_n\right) + o(1) \\ &\leq \mathbb{P}\left(\tilde{U}_{2,n} > z\right) + o(1). \end{aligned}$$

where the last two lines follows from the anti-concentration inequality for Gaussian processes and the fact that  $\kappa_n$  is the  $\gamma_n$ -quantile of  $\tilde{U}'_{2,n}$  with  $\gamma_n \rightarrow 1$ .  $\square$

**Lemma SA4.** *Under Assumptions 1, 2, and 3,  $\mathbb{P}(\mathcal{V}_n \subseteq \hat{\mathcal{V}}_n) \rightarrow 1$ .*

PROOF. Similar to the proof of Lemma 2 in Chernozhukov, Lee, and Rosen (2013), it is easy to show that

$$\mathbb{P}\left(\mathcal{V}_n \subseteq \hat{\mathcal{V}}_n\right) \geq p_{1,n} - p_{2,n}, \quad (\text{SA.5})$$

where

$$\begin{aligned} p_{1,n} &\equiv \mathbb{P}\left(\frac{n^{1/2}(\hat{h}_{j,n}(x) - h_{j,n}(x))}{\hat{\sigma}_{j,n}(x)} \leq 2\hat{K}_n - \kappa_n \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)}, \text{ for all } (j,x) \in \mathcal{V}_n\right), \\ p_{2,n} &\equiv \mathbb{P}\left(\sup_{(j,x) \in \mathcal{V}} \frac{n^{1/2}(\eta^* - \hat{h}_{j,n}(x))}{\hat{\sigma}_{j,n}(x)} > \hat{K}_n\right). \end{aligned}$$

By Lemma SA2 and Chernozhukov, Lee, and Rosen's (2013) Lemma 11, there exists a positive real sequence  $\beta_n = o(1)$  such that with probability approaching 1,

$$\hat{K}_n \geq \kappa_{n,\mathcal{V}}(\tilde{\gamma}_n - \beta_n) - o((\log n)^{-1}). \quad (\text{SA.6})$$

Since  $\tilde{\gamma}_n \rightarrow 1$ , we can assume that  $\tilde{\gamma}_n > 1/2$  without loss of generality, which implies  $\hat{K}_n \geq 0$ . Since  $\kappa_{n,\mathcal{V}}(q) \geq \kappa_{n,\mathcal{V}_n}(q)$  for any  $q \in (0, 1)$ , we also have

$$\hat{K}_n \geq \max \{ \kappa_{n,\mathcal{V}_n}(\tilde{\gamma}_n - \beta_n) - o((\log n)^{-1}), 0 \}. \quad (\text{SA.7})$$

Since  $\kappa_n = O((\log n)^{1/2})$ , by Lemma SA1, we have

$$\kappa_n \sup_{(j,x) \in \mathcal{V}} \left| \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} - 1 \right| = o_p((\log n)^{-1}).$$

Combined with (SA.6), this estimate implies that with probability approaching 1, for  $\gamma_n = \tilde{\gamma}_n - 2\beta_n$ ,

$$\begin{aligned} 2\hat{K}_n - \kappa_n \frac{\sigma_{j,n}(x)}{\hat{\sigma}_{j,n}(x)} &\geq 2\kappa_{n,\mathcal{V}}(\tilde{\gamma}_n - \beta_n) - \kappa_{n,\mathcal{V}}(\gamma_n) - o((\log n)^{-1}) \\ &\geq \kappa_{n,\mathcal{V}}(\tilde{\gamma}_n - \beta_n) - o((\log n)^{-1}). \end{aligned}$$

Hence,

$$p_{1,n} \geq \mathbb{P} \left( \sup_{(j,x) \in \mathcal{V}} \frac{n^{1/2}(\hat{h}_{j,n}(x) - h_{j,n}(x))}{\hat{\sigma}_{j,n}(x)} \leq \kappa_{n,\mathcal{V}}(\tilde{\gamma}_n - \beta_n) - o((\log n)^{-1}) \right) - o(1). \quad (\text{SA.8})$$

By the first line of (SA.2),

$$\sup_{(j,x) \in \mathcal{V}} \frac{n^{1/2}(\hat{h}_j(x) - h_j(x))}{\sigma_{j,n}(x)} - \tilde{U}'_{1,n} = o_p((\log n)^{-1}).$$

By Lemma SA1, we can further deduce

$$\sup_{(j,x) \in \mathcal{V}} \frac{n^{1/2}(\hat{h}_j(x) - h_j(x))}{\hat{\sigma}_{j,n}(x)} - \tilde{U}'_{1,n} = o_p((\log n)^{-1}). \quad (\text{SA.9})$$

Note that  $\kappa_{n,\mathcal{V}}(q)$  is the  $q$ -quantile of  $\tilde{U}'_{1,n}$ . Hence, from (SA.8) and (SA.9), we deduce

$$\begin{aligned} p_{1,n} &\geq \mathbb{P} \left( \tilde{U}'_{1,n} \leq \kappa_{n,\mathcal{V}}(\tilde{\gamma}_n - \beta_n) - o((\log n)^{-1}) \right) - o(1) \\ &\geq \tilde{\gamma}_n - o(1) = 1 - o(1). \end{aligned} \quad (\text{SA.10})$$

Next, we observe

$$\begin{aligned} &\mathbb{P} \left( \sup_{(j,x) \in \mathcal{V}} \left[ \frac{n^{1/2}(\eta^* - \hat{h}_j(x))}{\hat{\sigma}_{j,n}(x)} \right] \leq \hat{K}_n \right) \\ &\geq \mathbb{P} \left( \sup_{(j,x) \in \mathcal{V}} \left[ \frac{n^{1/2}(\eta^* - \hat{h}_j(x))}{\hat{\sigma}_{j,n}(x)} \right] \leq \max \{ \kappa_{n,\mathcal{V}_n}(\tilde{\gamma}_n - \beta_n) - o((\log n)^{-1}), 0 \} \right) - o(1), \\ &\geq \mathbb{P} \left( \tilde{U}_{2,n} \leq \kappa_{n,\mathcal{V}_n}(\tilde{\gamma}_n - \beta_n) - o((\log n)^{-1}) \right) - o(1) \\ &\geq \tilde{\gamma}_n - o(1) = 1 - o(1), \end{aligned}$$

where the first inequality is by (SA.7), and the second inequality is by Lemma SA3. Hence,  $p_{2,n} = o(1)$ . The assertion of the lemma then follows from this estimate, (SA.5), and (SA.10).  $\square$

## SA.2 Proof of Proposition 1

PROOF OF PROPOSITION 1. (a) Recall that  $\hat{k}_{n,1-\alpha}$  is the  $(1-\alpha)$ -quantile of  $\sup_{(j,x) \in \hat{\mathcal{V}}_n} \hat{t}_{j,n}^*(x)$ . Let  $k_{n,1-\alpha}$  denote the  $(1-\alpha)$ -quantile of  $\sup_{(j,x) \in \mathcal{V}_n} \hat{t}_{j,n}^*(x)$ . By Lemma SA4, we have

$$\mathbb{P}\left(\hat{k}_{n,1-\alpha} \geq k_{n,1-\alpha}\right) \rightarrow 1. \quad (\text{SA.11})$$

By Lemma SA2 and Chernozhukov, Lee, and Rosen's (2013) Lemma 11, we have with probability approaching 1,

$$k_{n,1-\alpha} \geq \kappa_{n,\mathcal{V}_n} (1 - \alpha - o(1)) - o((\log n)^{-1}). \quad (\text{SA.12})$$

Note that

$$\begin{aligned} & \mathbb{P}\left(\eta^* \leq \hat{\eta}_{n,1-\alpha}\right) \\ &= \mathbb{P}\left(\eta^* \leq \inf_{(j,x) \in \mathcal{V}} \left[\hat{h}_{j,n}(x) + n^{-1/2} \hat{k}_{n,1-\alpha} \hat{\sigma}_{j,n}(x)\right]\right) \\ &= \mathbb{P}\left(\sup_{(j,x) \in \mathcal{V}} \left[\frac{n^{1/2}(\eta^* - \hat{h}_{j,n}(x))}{\hat{\sigma}_{j,n}(x)}\right] \leq \hat{k}_{n,1-\alpha}\right) \\ &\geq \mathbb{P}\left(\sup_{(j,x) \in \mathcal{V}} \left[\frac{n^{1/2}(\eta^* - \hat{h}_{j,n}(x))}{\hat{\sigma}_{j,n}(x)}\right] \leq \max\{\kappa_{n,\mathcal{V}_n} (1 - \alpha - o(1)) - o((\log n)^{-1}), 0\}\right) - o(1), \\ &\geq \mathbb{P}\left(\tilde{U}_{2,n} \leq \kappa_{n,\mathcal{V}_n} (1 - \alpha - o(1)) - o((\log n)^{-1})\right) - o(1), \end{aligned} \quad (\text{SA.13})$$

where the first inequality is by (SA.12) and the fact that  $\hat{k}_{n,1-\alpha} \geq 0$ , and the second inequality follows from Lemma SA3. By the anti-concentration inequality of Gaussian processes, we see that the lower bound in (SA.13) is  $1 - \alpha + o(1)$ . Hence, under the null hypothesis with  $\eta^* \geq 0$ , we have  $\liminf_n \mathbb{P}(\hat{\eta}_{n,1-\alpha} \geq 0) \geq 1 - \alpha$ . The assertion of part (a) readily follows.

(b) Denote  $h_{j,n}^*(\cdot) = P(\cdot)^\top b_{j,n}^*$ , with  $b_{j,n}^*$  given by Assumption 2. Note that

$$\hat{b}_{j,n} - b_{j,n}^* = \hat{Q}_n^{-1} \left( n^{-1} \sum_{t=1}^n P(X_t) u_{j,t} + n^{-1} \sum_{t=1}^n P(X_t) (h_j(X_t) - h_{j,n}^*(X_t)) \right) = O_p\left(m_n^{1/2} n^{-1/2}\right).$$

We then deduce that

$$\begin{aligned} \sup_{(j,x) \in \mathcal{V}} \left| \hat{h}_{j,n}(x) - h_j(x) \right| &\leq \sup_{x \in \mathcal{X}} \|P(x)\| \cdot \max_{1 \leq j \leq J} \left\| \hat{b}_{j,n} - b_{j,n}^* \right\| + \sup_{(j,x) \in \mathcal{V}} \left| h_j(x) - P(x)^\top b_{j,n}^* \right| \\ &= O_p\left(\zeta_n m_n n^{-1/2}\right). \end{aligned} \quad (\text{SA.14})$$

Note that  $\hat{k}_{n,1-\alpha} = O_p((\log n)^{1/2})$ . In addition, by Lemma SA1,

$$\sup_{(j,x) \in \mathcal{V}} \hat{\sigma}_{j,n}(x) = O_p(1) \quad \sup_{(j,x) \in \mathcal{V}} \sigma_{j,n}(x) = O_p\left(\zeta_n m_n^{1/2}\right). \quad (\text{SA.15})$$

By (SA.14) and (SA.15),

$$\begin{aligned} & \sup_{(j,x) \in \mathcal{V}} \left| \hat{h}_{j,n}(x) + n^{-1/2} \hat{k}_{n,1-\alpha} \hat{\sigma}_{j,n}(x) - h_j(x) \right| \\ &= O_p\left(\zeta_n m_n n^{-1/2}\right) + O_p\left(\zeta_n m_n^{1/2} (\log n)^{1/2} n^{-1/2}\right) = O_p\left(\zeta_n m_n n^{-1/2}\right), \end{aligned}$$

where the second equality holds because  $m_n$  is of polynomial growth in  $n$ . This estimate further implies that

$$\hat{\eta}_{n,1-\alpha} - \eta^* = O_p\left(\zeta_n m_n n^{-1/2}\right).$$

Under condition (v) of Assumption 2,  $\zeta_n m_n n^{-1/2} = o(1)$ . Hence,  $\hat{\eta}_{n,1-\alpha} - \eta^* = o_p(1)$ . The assertion of part (b) readily follows.  $\square$

## SB Proofs of technical lemmas for Theorem 1

PROOF OF LEMMA A1. Step 1. We outline the proof in this step. For ease of notation, we write  $\max_{l,t}$  in place of  $\max_{1 \leq l \leq L_n, 1 \leq t \leq n}$ , and interpret  $\max_l$  and  $\max_t$  in a similar way. Note that

$$\begin{aligned} & \left| \max_l \alpha_l^\top S_n^* - \max_l \alpha_l^\top S_n^+ \right| \leq \max_l \left| \alpha_l^\top (S_n^* - S_n^+) \right| \\ & \leq \max_l \left| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* 1_{\{t > \tau_n\}} \right| + \max_l \left| k_n^{-1/2} \sum_{t=n+1}^{n+k_n} \alpha_l^\top (\Xi_n^{1/2} \eta_{n,t}) \right|. \end{aligned}$$

In step 2 and step 3, respectively, we show that

$$\max_l \left| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* 1_{\{t > \tau_n\}} \right| = O_p\left(\min\left\{L_n^{1/q} \max_{l,t} \|\alpha_l^\top X_{n,t}^*\|_q, M_n^{1/2}\right\} r_n^{1/2}\right), \quad (\text{SB.1})$$

$$\max_l \left| k_n^{-1/2} \sum_{t=n+1}^{n+k_n} \alpha_l^\top (\Xi_n^{1/2} \eta_{n,t}) \right| = O_p((\log L_n)^{1/2} r_n^{1/2}). \quad (\text{SB.2})$$

These estimates clearly imply the assertion of the lemma. Below, since  $\|\alpha_l\|$  are uniformly bounded, we can and will assume that  $\|\alpha_l\| \leq 1$  without loss of generality.

Step 2. In this step, we prove (SB.1). Fix some  $\varepsilon > 0$ . As shown in the proof of Lemma A1 of Li and Liao (2019) (see their supplemental appendix), there exists a finite constant  $C_1 > 0$  such that for  $u_n^* = \lceil C_1 n r_n \rceil$  and  $k_n^* = n - u_n^*$ ,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\tau_n < k_n^*) < \varepsilon/2, \quad (\text{SB.3})$$

where  $\lceil \cdot \rceil$  denotes the ceiling function. For a generic positive real sequence  $a_n$ , it is easy to see that

$$\begin{aligned}
& \mathbb{P} \left( \max_l \left| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* 1_{\{t > \tau_n\}} \right| > a_n \right) \\
& \leq \mathbb{P} \left( \max_l \left| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* 1_{\{t > \tau_n\}} \right| > a_n, \tau_n \geq k_n^* \right) + \mathbb{P}(\tau_n < k_n^*) \\
& \leq \mathbb{P} \left( \max_l \left| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* 1_{\{t > \tau_n \vee k_n^*\}} \right| > a_n \right) + \mathbb{P}(\tau_n < k_n^*). \tag{SB.4}
\end{aligned}$$

By the maximal inequality under  $L_q$ -norm,

$$\mathbb{E} \left[ \max_l \left| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* 1_{\{t > \tau_n \vee k_n^*\}} \right| \right] \leq L_n^{1/q} \max_l \left\| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* 1_{\{t > \tau_n \vee k_n^*\}} \right\|_q. \tag{SB.5}$$

Since  $\tau_n$  is a stopping time and  $k_n^*$  is deterministic,  $\{t > \tau_n \vee k_n^*\} \in \mathcal{F}_{n,t-1}$ . It is then easy to see that  $(X_{n,t}^* 1_{\{t > \tau_n \vee k_n^*\}})_{t \geq 1}$  also forms a martingale difference sequence. We further note that

$$\begin{aligned}
\left\| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* 1_{\{t > \tau_n \vee k_n^*\}} \right\|_q & \leq K \left( n^{-1} \sum_{t=1}^n \left\| \alpha_l^\top X_{n,t}^* 1_{\{t > \tau_n \vee k_n^*\}} \right\|_q^2 \right)^{1/2} \\
& \leq K \left( n^{-1} \sum_{t=1}^n \left\| \alpha_l^\top X_{n,t}^* \right\|_q^2 1_{\{t > k_n^*\}} \right)^{1/2} \\
& \leq K r_n^{1/2} \max_{l,t} \left\| \alpha_l^\top X_{n,t}^* \right\|_q, \tag{SB.6}
\end{aligned}$$

where the first inequality is by Theorem 2.1 of Rio (2009), the second inequality holds because  $1_{\{t > \tau_n \vee k_n^*\}}$  is bounded by the deterministic variable  $1_{\{t > k_n^*\}}$ , and the last line follows from  $n - k_n^* = O(nr_n)$ . Combining the estimates in (SB.5) and (SB.6), we deduce

$$\mathbb{E} \left[ \max_l \left| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* 1_{\{t > \tau_n \vee k_n^*\}} \right| \right] \leq K L_n^{1/q} r_n^{1/2} \max_{l,t} \left\| \alpha_l^\top X_{n,t}^* \right\|_q. \tag{SB.7}$$

We also observe

$$\begin{aligned}
& \mathbb{E} \left[ \max_l \left| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* \mathbf{1}_{\{t > \tau_n \vee k_n^*\}} \right| \right] \\
& \leq \mathbb{E} \left[ \left\| n^{-1/2} \sum_{t=1}^n X_{n,t}^* \mathbf{1}_{\{t > \tau_n \vee k_n^*\}} \right\| \right] \\
& \leq K \left( n^{-1} \sum_{t=1}^n \mathbb{E} \left[ \|X_{n,t}^* \mathbf{1}_{\{t > \tau_n \vee k_n^*\}}\|^2 \right] \right)^{1/2} \\
& \leq K \left( n^{-1} \sum_{t=1}^n \mathbb{E} \left[ \|X_{n,t}^*\|^2 \right] \mathbf{1}_{\{t > k_n^*\}} \right)^{1/2} \\
& \leq K r_n^{1/2} \max_t \left( \mathbb{E} \left[ \|X_{n,t}^*\|^2 \right] \right)^{1/2} \leq K M_n^{1/2} r_n^{1/2}, \tag{SB.8}
\end{aligned}$$

where the first inequality is due to the Cauchy–Schwarz inequality and the fact that  $\|\alpha_l\| \leq 1$ , the second inequality is by Jensen’s inequality and the fact that  $(X_{n,t}^* \mathbf{1}_{\{t > \tau_n \vee k_n^*\}})_{t \geq 1}$  are martingale differences, and the remaining inequalities are obtained similarly as (SB.6) while using the fact that  $\|X_{n,t}^*\|_2 \leq K M_n^{1/2}$ .

Combining (SB.7) and (SB.8), we see that for some finite constant  $C_2 > 0$ ,

$$\mathbb{E} \left[ \max_l \left| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* \mathbf{1}_{\{t > \tau_n \vee k_n^*\}} \right| \right] \leq C_2 \min \left\{ L_n^{1/q} \max_{l,t} \|\alpha_l^\top X_{n,t}^*\|_q, M_n^{1/2} \right\} r_n^{1/2}.$$

Applying (SB.4) with

$$a_n = 2\varepsilon^{-1} C_2 \min \left\{ L_n^{1/q} \max_{l,t} \|\alpha_l^\top X_{n,t}^*\|_q, M_n^{1/2} \right\} r_n^{1/2},$$

and then use Markov’s inequality and (SB.3), we deduce

$$\limsup_n \mathbb{P} \left( \max_l \left| n^{-1/2} \sum_{t=1}^n \alpha_l^\top X_{n,t}^* \mathbf{1}_{\{t > \tau_n\}} \right| > a_n \right) < \varepsilon.$$

This finishes the proof of (SB.1).

Step 3. Fix some constant  $\varepsilon > 0$  and consider the same  $k_n^*$  as in (SB.3). Note that  $\Xi_n$  is  $\mathcal{H}_{n,n}$ -measurable, and conditional on  $\mathcal{H}_{n,n}$ ,  $k_n^{-1/2} \sum_{t=n+1}^{n+k_n} \alpha_l^\top (\Xi_n^{1/2} \eta_{n,t})$  is centered Gaussian with variance  $\alpha_l^\top \Xi_n \alpha_l$ . By the maximum inequality under Orlicz norm (see, e.g., Lemma 2.2.2 in van der Vaart and Wellner (1996) with  $\psi(x) = \exp(x^2) - 1$ ),

$$\mathbb{E} \left[ \max_l \left| k_n^{-1/2} \sum_{t=n+1}^{n+k_n} \alpha_l^\top (\Xi_n^{1/2} \eta_{n,t}) \right| \middle| \mathcal{H}_{n,n} \right] \leq K (\log L_n)^{1/2} \max_l (\alpha_l^\top \Xi_n \alpha_l)^{1/2}. \tag{SB.9}$$



Recall that  $\Xi_n = \Sigma_n^* - \sum_{t=1}^{\tau_n} V_{n,t}^*$ . By the triangle inequality and the Cauchy–Schwarz inequality, we have in restriction to the event  $\{\tau_n \geq k_n^*\}$ ,

$$\begin{aligned}
& \max_l \alpha_l^\top \Xi_n \alpha_l \\
& \leq \max_l \alpha_l^\top \left( \Sigma_n^* - \sum_{t=1}^n V_{n,t}^* \right) \alpha_l + \max_l \alpha_l^\top \left( \sum_{t=\tau_n+1}^n V_{n,t}^* \right) \alpha_l \\
& \leq \left\| \Sigma_n^* - \sum_{t=1}^n V_{n,t}^* \right\|_S + \max_l \alpha_l^\top \left( \sum_{t=k_n^*+1}^n V_{n,t}^* \right) \alpha_l \\
& \leq 2 \left\| \sum_{t=1}^n (V_{n,t}^* - \mathbb{E}[V_{n,t}^*]) \right\|_S + \left\| \sum_{t=1}^{k_n^*} (V_{n,t}^* - \mathbb{E}[V_{n,t}^*]) \right\|_S + \max_l \sum_{t=k_n^*+1}^n \alpha_l^\top \mathbb{E}[V_{n,t}^*] \alpha_l \\
& \equiv R_n
\end{aligned} \tag{SB.10}$$

Since  $\mathbb{E}[V_{n,t}^*] = n^{-1} \mathbb{E}[X_{n,t}^* X_{n,t}^{*\top}]$  and  $\mathbb{E}[X_{n,t}^* X_{n,t}^{*\top}]$  has bounded eigenvalues under Assumption 4(ii), we have uniformly in  $l$ ,

$$\sum_{t=k_n^*+1}^n \alpha_l^\top \mathbb{E}[V_{n,t}^*] \alpha_l \leq K(n - k_n^*)/n \leq Kr_n.$$

In addition, since  $k_n^*/n \rightarrow 1$ , Assumption 4(iii) implies that

$$2 \left\| \sum_{t=1}^n (V_{n,t}^* - \mathbb{E}[V_{n,t}^*]) \right\|_S + \left\| \sum_{t=1}^{k_n^*} (V_{n,t}^* - \mathbb{E}[V_{n,t}^*]) \right\|_S = O_p(r_n).$$

The  $R_n$  sequence defined in (SB.10) is thus  $O_p(r_n)$ . Therefore, there exists some finite constant  $C_3 > 0$  such that  $\mathbb{P}(R_n > C_3 r_n) < \varepsilon/2$ . Hence,

$$\begin{aligned}
& \limsup_n \mathbb{P} \left( \max_l \alpha_l^\top \Xi_n \alpha_l > C_3 r_n \right) \\
& \leq \limsup_n \mathbb{P}(R_n > C_3 r_n, \tau_n \geq k_n^*) + \limsup_n \mathbb{P}(\tau_n < k_n^*) < \varepsilon,
\end{aligned}$$

which shows that  $\max_l (\alpha_l^\top \Xi_n \alpha_l)^{1/2} = O_p(r_n^{1/2})$ . From this estimate and (SB.9), we deduce (SB.2) as claimed.  $\square$

Next, we prove Lemma A2 in the appendix of the main text. We recall that  $(\zeta_{n,t})_{1 \leq t \leq n+k_n}$  are independent  $M_n$ -dimensional standard normal vectors and  $\tilde{\zeta}_{n,t} \equiv (V_{n,t}^+)^{1/2} \zeta_{n,t}$ . We need an additional lemma.

**Lemma SB1.** *Under Assumption 4, we have*

- (a)  $\sum_{t=n+1}^{n+k_n} \mathbb{E}[\max_l |\alpha_l^\top Z_{n,t}|^3] \leq K(\log L_n)^{3/2} k_n^{-1/2}$ ;
- (b)  $\sum_{t=1}^{n+k_n} \mathbb{E}[\max_l |\alpha_l^\top \tilde{\zeta}_{n,t}|^3] \leq K(\log L_n)^{3/2} \sum_{t=1}^n \mathbb{E}[\max_l (\alpha_l^\top V_{n,t}^* \alpha_l)^{3/2}] + K(\log L_n)^{3/2} k_n^{-1/2}$ .

PROOF. (a) By definition, for any  $t \in \{n+1, \dots, n+k_n\}$ ,

$$\max_l \left| \alpha_l^\top Z_{n,t} \right|^3 = k_n^{-3/2} \max_l \left| \alpha_l^\top (\Xi_n^{1/2} \eta_{n,t}) \right|^3. \quad (\text{SB.11})$$

Note that  $\Xi_n$  is  $\mathcal{H}_{n,n}$ -measurable and  $k_n^{-1/2} \sum_{t=n+1}^{n+k_n} \alpha_l^\top (\Xi_n^{1/2} \eta_{n,t})$  is  $\mathcal{H}_{n,n}$ -conditionally centered Gaussian with variance  $\alpha_l^\top \Xi_n \alpha_l$ . By the maximum inequality under Orlicz norm (see, e.g., Lemma 2.2.2 in van der Vaart and Wellner (1996) with  $\psi(x) = \exp(x^{2/3}) - 1$ ),

$$\mathbb{E} \left[ \max_l \left| \alpha_l^\top (\Xi_n^{1/2} \eta_{n,t}) \right|^3 \middle| \mathcal{H}_{n,n} \right] \leq K (\log L_n)^{3/2} \max_l (\alpha_l^\top \Xi_n \alpha_l)^{3/2}. \quad (\text{SB.12})$$

Since  $\alpha_l^\top \Xi_n \alpha_l \leq \alpha_l^\top \Sigma_n^* \alpha_l$  by the definition of  $\Xi_n$ , we further have

$$\max_l (\alpha_l^\top \Xi_n \alpha_l)^{3/2} \leq \max_l (\alpha_l^\top \Sigma_n^* \alpha_l)^{3/2} \leq K, \quad (\text{SB.13})$$

where the second inequality holds because  $\|\alpha_l\|$  and the eigenvalues of  $\Sigma_n^*$  are bounded. The assertion of part (a) is then obtained by combining the estimates in (SB.11), (SB.12), and (SB.13).

(b) By the definition of  $\tilde{\zeta}_{n,t}$ , we can write

$$\begin{aligned} \sum_{t=1}^{n+k_n} \max_l \left| \alpha_l^\top \tilde{\zeta}_{n,t} \right|^3 &= \sum_{t=1}^n \max_l \left| \alpha_l^\top (V_{n,t}^*)^{1/2} \zeta_{n,t} 1_{\{t \leq \tau_n\}} \right|^3 \\ &\quad + k_n^{-3/2} \sum_{t=n+1}^{n+k_n} \max_l \left| \alpha_l^\top \Xi_n^{1/2} \zeta_{n,t} \right|^3. \end{aligned} \quad (\text{SB.14})$$

Using the same argument in part (a), we can show that

$$k_n^{-3/2} \sum_{t=n+1}^{n+k_n} \mathbb{E} \left[ \max_l \left| \alpha_l^\top \Xi_n^{1/2} \zeta_{n,t} \right|^3 \right] \leq K (\log L_n)^{3/2} k_n^{-1/2}. \quad (\text{SB.15})$$

For any  $t \leq n$ ,  $\alpha_l^\top (V_{n,t}^*)^{1/2} \zeta_{n,t}$  is  $\mathcal{H}_{n,t-1}$ -conditionally centered Gaussian with conditional variance  $\alpha_l^\top V_{n,t}^* \alpha_l$ . Therefore,

$$\mathbb{E} \left[ \max_l \left| \alpha_l^\top (V_{n,t}^*)^{1/2} \zeta_{n,t} \right|^3 \middle| \mathcal{H}_{n,t-1} \right] \leq K (\log L_n)^{3/2} \max_l (\alpha_l^\top V_{n,t}^* \alpha_l)^{3/2} \quad (\text{SB.16})$$

which implies that

$$\mathbb{E} \left[ \max_l \left| \alpha_l^\top (V_{n,t}^*)^{1/2} \zeta_{n,t} 1_{\{t \leq \tau_n\}} \right|^3 \right] \leq K (\log L_n)^{3/2} \mathbb{E} \left[ \max_l (\alpha_l^\top V_{n,t}^* \alpha_l)^{3/2} \right]. \quad (\text{SB.17})$$

The assertion of part (b) then follows from (SB.14), (SB.15), and (SB.17). *Q.E.D.*

PROOF OF LEMMA A2. Step 1. We outline the proof in this step. For any  $x = (x_1, \dots, x_{L_n}) \in \mathbb{R}^{L_n}$  and  $\sigma > 0$ , we define a function

$$F_\sigma(x) = \sigma \log \left( \sum_{l=1}^{L_n} \exp(\sigma^{-1} x_l) \right).$$

By equation (17) in Chernozhukov, Chetverikov, and Kato (2014),

$$\max_l x_l \leq F_\sigma(x) \leq \max_l x_l + \sigma \log L_n.$$

For any given sequence  $(\alpha_l)_{1 \leq l \leq L_n}$ , we denote  $\alpha = (\alpha_1, \dots, \alpha_{L_n})$ , which is an  $M_n \times L_n$  matrix. We can then write

$$\left(\alpha_1^\top S_n^+, \dots, \alpha_{L_n}^\top S_n^+\right)^\top = \alpha^\top S_n^+, \quad \left(\alpha_1^\top \tilde{S}_n^*, \dots, \alpha_{L_n}^\top \tilde{S}_n^*\right)^\top = \alpha^\top \tilde{S}_n^*.$$

With this notation, we see that

$$\left| \max_l \alpha_l^\top S_n^+ - F_\sigma \left( \alpha^\top S_n^+ \right) \right| \leq \sigma \log L_n, \quad \left| \max_l \alpha_l^\top \tilde{S}_n^* - F_\sigma \left( \alpha^\top \tilde{S}_n^* \right) \right| \leq \sigma \log L_n. \quad (\text{SB.18})$$

Let  $\sigma_n \equiv B_{1,n}^{1/3} (\log L_n)^{-1/3}$ . With any positive constant  $C > 1$ , we associate the following real sequences

$$\delta_n(C) \equiv C \sigma_n \log L_n, \quad \psi_n(C) \equiv (C \log L_n) \sqrt{\exp \left( 1 - (C \log L_n)^2 \right)}.$$

Note that as  $C \rightarrow \infty$ ,

$$\sup_{n \geq 1} \psi_n(C) \rightarrow 0. \quad (\text{SB.19})$$

By the first estimate in (SB.18), we have for any Borel set  $A \subseteq \mathbb{R}$ ,

$$\mathbb{P} \left( \max_l \alpha_l^\top S_n^+ \in A \right) \leq \mathbb{P} \left( F_{\sigma_n} \left( \alpha^\top S_n^+ \right) \in A^{\delta_n(C)} \right). \quad (\text{SB.20})$$

For ease of notation, below, we shall write  $\delta_n$  and  $\psi_n$  in place of  $\delta_n(C)$  and  $\psi_n(C)$ , respectively.

Given  $\sigma_n$  and  $\delta_n$ , we define a function  $f(\cdot)$  as follows. First define

$$g(x) \equiv \max \left\{ 0, 1 - \frac{d(x, A^{2\delta_n})}{\delta_n} \right\},$$

where  $d(x, A^{2\delta_n})$  denotes the distance between the point  $x$  and the set  $A^{2\delta_n}$ . We then set for any  $x \in \mathbb{R}$ ,

$$f(x) \equiv \mathbb{E} [g(x + \sigma_n \mathcal{N})],$$

where  $\mathcal{N}$  is a generic standard normal random variable. By Lemma 4.2 of Chernozhukov, Chetverikov, and Kato (2014), we have for all  $x \in \mathbb{R}$ ,

$$|\partial f(x)| \leq \delta_n^{-1}, \quad |\partial^2 f(x)| \leq C_f \sigma_n^{-1} \delta_n^{-1}, \quad |\partial^3 f(x)| \leq C_f \sigma_n^{-2} \delta_n^{-1}, \quad (\text{SB.21})$$

where  $C_f$  is a fixed finite constant and  $\partial^j f(x)$  denotes the  $j$ th derivative of  $f(x)$ ; moreover,

$$(1 - \psi_n) 1_{\{x \in A^{\delta_n}\}} \leq f(x) \leq \psi_n + (1 - \psi_n) 1_{\{x \in A^{4\delta_n}\}}. \quad (\text{SB.22})$$

We further define  $\Delta_{\sigma_n}(x) = f \circ F_{\sigma_n}(x)$ . The first inequality in (SB.22) implies

$$(1 - \psi_n)1_{\{F_{\sigma_n}(\alpha^\top S_n^+) \in A^{\delta_n}\}} \leq \Delta_{\sigma_n}(\alpha^\top S_n^+). \quad (\text{SB.23})$$

By (SB.20) and (SB.23),

$$\mathbb{P}\left(\max_l \alpha_l^\top S_n^+ \in A\right) \leq \frac{\mathbb{E}[\Delta_{\sigma_n}(\alpha^\top S_n^+)]}{1 - \psi_n}. \quad (\text{SB.24})$$

In addition, by first using (SB.22) and then (SB.18), we can deduce that

$$\begin{aligned} \mathbb{E}[\Delta_{\sigma_n}(\alpha^\top \tilde{S}_n^*)] &\leq \psi_n + (1 - \psi_n)\mathbb{P}\left(F_{\sigma_n}(\alpha^\top \tilde{S}_n^*) \in A^{4\delta_n}\right) \\ &\leq \psi_n + (1 - \psi_n)\mathbb{P}\left(\max_l \alpha_l^\top \tilde{S}_n^* \in A^{5\delta_n}\right). \end{aligned} \quad (\text{SB.25})$$

Combining (SB.24) and (SB.25), we obtain

$$\begin{aligned} \mathbb{P}\left(\max_l \alpha_l^\top S_n^+ \in A\right) &\leq \mathbb{P}\left(\max_l \alpha_l^\top \tilde{S}_n^* \in A^{5\delta_n}\right) \\ &\quad + \frac{\psi_n}{1 - \psi_n} + \frac{\mathbb{E}[\Delta_{\sigma_n}(\alpha^\top S_n^+)] - \mathbb{E}[\Delta_{\sigma_n}(\alpha^\top \tilde{S}_n^*)]}{1 - \psi_n}. \end{aligned} \quad (\text{SB.26})$$

In step 2, below, we show that, for some constant  $K$  that does not depend on  $C$  or  $n$ ,

$$\left|\mathbb{E}[\Delta_{\sigma_n}(\alpha^\top S_n^+)] - \mathbb{E}[\Delta_{\sigma_n}(\alpha^\top \tilde{S}_n^*)]\right| \leq K \frac{B_{1,n}}{\sigma_n^2 \delta_n}. \quad (\text{SB.27})$$

By construction,  $\sigma_n^2 \delta_n = CB_{1,n}$ . Hence, (SB.26) and (SB.27) imply that

$$\mathbb{P}\left(\max_l \alpha_l^\top S_n^+ \in A\right) \leq \mathbb{P}\left(\max_l \alpha_l^\top \tilde{S}_n^* \in A^{5\delta_n}\right) + \frac{\psi_n(C) + K/C}{1 - \psi_n(C)}. \quad (\text{SB.28})$$

Recall that  $\delta_n = CB_{1,n}^{1/3}(\log L_n)^{2/3}$ . Define  $\epsilon_n(\cdot)$  such that

$$\epsilon_n(5C) \equiv \frac{\psi_n(C) + K/C}{1 - \psi_n(C)}.$$

Then the assertion of the lemma readily follows from (SB.19) and (SB.28).

Step 2. In this step, we finish the proof of Lemma A2 by establishing (SB.27). We set for each  $t \in \{1, \dots, n + k_n\}$ ,

$$D_{n,t} \equiv \sum_{s \in [1,t]} Z_{n,s} + \sum_{s \in (t, n+k_n]} \tilde{\zeta}_{n,s}.$$

By this definition, we have  $S_n^+ = D_{n, n+k_n} + Z_{n, n+k_n}$ ,  $\tilde{S}_n^* = D_{n,1} + \tilde{\zeta}_{n,1}$ , and

$$D_{n,t} + Z_{n,t} = D_{n,t+1} + \tilde{\zeta}_{n,t+1} \text{ for any } t = 1, \dots, n + k_n - 1.$$

Hence,

$$\begin{aligned} & \mathbb{E}[\Delta_{\sigma_n}(\alpha^\top S_n^+)] - \mathbb{E}[\Delta_{\sigma_n}(\alpha^\top \tilde{S}_n^*)] \\ &= \sum_{t=1}^{n+k_n} \left( \mathbb{E}[\Delta_{\sigma_n}(\alpha^\top D_{n,t} + \alpha^\top Z_{n,t})] - \mathbb{E}[\Delta_{\sigma_n}(\alpha^\top D_{n,t} + \alpha^\top \tilde{\zeta}_{n,t})] \right). \end{aligned} \quad (\text{SB.29})$$

Using a third-order mean-value expansion, we deduce

$$\begin{aligned} & \left| \mathbb{E} \left[ \Delta_{\sigma_n}(\alpha^\top D_{n,t} + \alpha^\top Z_{n,t}) \right] - \mathbb{E} \left[ \Delta_{\sigma_n}(\alpha^\top D_{n,t}) \right] - \mathbb{E}[\partial \Delta_{\sigma_n}(\alpha^\top D_{n,t})^\top \alpha^\top Z_{n,t}] \right. \\ & \quad \left. - \frac{1}{2} \mathbb{E} \left[ \text{Tr}(\partial^2 \Delta_{\sigma_n}(\alpha^\top D_{n,t}) \alpha^\top Z_{n,t} Z_{n,t}^\top \alpha) \right] \right| \\ & \leq \frac{1}{6} \sum_{l_1, l_2, l_3=1}^{L_n} \mathbb{E} \left[ \left| (\alpha_{l_1}^\top Z_{n,t}) (\alpha_{l_2}^\top Z_{n,t}) (\alpha_{l_3}^\top Z_{n,t}) \partial_{l_1} \partial_{l_2} \partial_{l_3} \Delta_{\sigma_n}(\bar{\chi}_{n,t}) \right| \right], \end{aligned} \quad (\text{SB.30})$$

where  $\bar{\chi}_{n,t}$  is some mean-value between  $\alpha^\top D_{n,t}$  and  $\alpha^\top D_{n,t} + \alpha^\top Z_{n,t}$ . By (SB.21) and Lemma 4.3 in Chernozhukov, Chetverikov, and Kato (2014),

$$\begin{aligned} & \sum_{l_1, l_2, l_3=1}^{L_n} \mathbb{E} \left[ \left| (\alpha_{l_1}^\top Z_{n,t}) (\alpha_{l_2}^\top Z_{n,t}) (\alpha_{l_3}^\top Z_{n,t}) \partial_{l_1} \partial_{l_2} \partial_{l_3} \Delta_{\sigma_n}(\bar{\chi}_{n,t}) \right| \right] \\ & \leq \mathbb{E} \left[ \max_{1 \leq l_1, l_2, l_3 \leq L_n} \left| (\alpha_{l_1}^\top Z_{n,t}) (\alpha_{l_2}^\top Z_{n,t}) (\alpha_{l_3}^\top Z_{n,t}) \right| \sum_{l_1, l_2, l_3=1}^{L_n} \left| \partial_{l_1} \partial_{l_2} \partial_{l_3} \Delta_{\sigma_n}(\bar{\chi}_{n,t}) \right| \right] \\ & \leq \frac{13C_f}{\sigma_n^2 \delta_n} \mathbb{E} \left[ \max_l \left| \alpha_l^\top Z_{n,t} \right|^3 \right], \end{aligned} \quad (\text{SB.31})$$

which together with (SB.30) implies that

$$\begin{aligned} & \left| \mathbb{E} \left[ \Delta_{\sigma_n}(\alpha^\top D_{n,t} + \alpha^\top Z_{n,t}) \right] - \mathbb{E} \left[ \Delta_{\sigma_n}(\alpha^\top D_{n,t}) \right] - \mathbb{E}[\partial \Delta_{\sigma_n}(\alpha^\top D_{n,t})^\top \alpha^\top Z_{n,t}] \right. \\ & \quad \left. - \frac{1}{2} \mathbb{E} \left[ \text{Tr}(\partial^2 \Delta_{\sigma_n}(\alpha^\top D_{n,t}) \alpha^\top Z_{n,t} Z_{n,t}^\top \alpha) \right] \right| \\ & \leq \frac{13C_f}{6\sigma_n^2 \delta_n} \mathbb{E} \left[ \max_l \left| \alpha_l^\top Z_{n,t} \right|^3 \right]. \end{aligned} \quad (\text{SB.32})$$

Similarly, we can show that

$$\begin{aligned} & \left| \mathbb{E} \left[ \Delta_{\sigma_n}(\alpha^\top D_{n,t} + \alpha^\top \tilde{\zeta}_{n,t}) \right] - \mathbb{E} \left[ \Delta_{\sigma_n}(\alpha^\top D_{n,t}) \right] - \mathbb{E}[\partial \Delta_{\sigma_n}(\alpha^\top D_{n,t})^\top \alpha^\top \tilde{\zeta}_{n,t}] \right. \\ & \quad \left. - \frac{1}{2} \mathbb{E} \left[ \text{Tr}(\partial^2 \Delta_{\sigma_n}(\alpha^\top D_{n,t}) \alpha^\top \tilde{\zeta}_{n,t} \tilde{\zeta}_{n,t}^\top \alpha) \right] \right| \\ & \leq \frac{13C_f}{6\sigma_n^2 \delta_n} \mathbb{E} \left[ \max_l \left| \alpha_l^\top \tilde{\zeta}_{n,t} \right|^3 \right]. \end{aligned} \quad (\text{SB.33})$$

Since  $\tilde{\zeta}_{n,t} = (V_{n,t}^+)^{1/2} \zeta_{n,t}$  and  $\zeta_{n,t}$  is a standard normal random vector independent of  $D_{n,t}$  and  $V_{n,t}^+$ , we have

$$\begin{cases} \mathbb{E}[\partial \Delta_{\sigma_n}(\alpha^\top D_{n,t})^\top \alpha^\top \tilde{\zeta}_{n,t}] = 0, \\ \mathbb{E} \left[ \text{Tr}(\partial^2 \Delta_{\sigma_n}(\alpha^\top D_{n,t}) \alpha^\top \tilde{\zeta}_{n,t} \tilde{\zeta}_{n,t}^\top \alpha) \right] = \mathbb{E}[\text{Tr}(\partial^2 \Delta_{\sigma_n}(\alpha^\top D_{n,t}) \alpha^\top V_{n,t}^+ \alpha)]. \end{cases} \quad (\text{SB.34})$$

Let

$$\tilde{\Sigma}_{n,t}^+ \equiv \sum_{s \in (t, n+k_n]} V_{n,s}^+ = \Sigma_n^* - \sum_{s \leq t} V_{n,s}^+.$$

Note that  $\tilde{\Sigma}_{n,t}^+$  is  $\mathcal{H}_{n,t-1}$ -measurable because  $\Sigma_n^*$  is deterministic and  $\sum_{s \leq t} V_{n,s}^+$  is  $\mathcal{H}_{n,t-1}$ -measurable. By construction,  $\alpha^\top \sum_{s \in (t, n+k_n]} \tilde{\zeta}_{n,s}$  is  $\mathcal{H}_{n,n}$ -conditionally centered Gaussian with conditional variance  $\alpha^\top \tilde{\Sigma}_{n,t}^+ \alpha$ . Hence, the normalized variable

$$\frac{\alpha^\top \sum_{s \in (t, n+k_n]} \tilde{\zeta}_{n,s}}{\sqrt{\alpha^\top \tilde{\Sigma}_{n,t}^+ \alpha}} \quad (\text{SB.35})$$

is independent of  $\mathcal{H}_{n,n}$ . Let  $\mathcal{H}'_{n,t}$  be the  $\sigma$ -field generated by  $\mathcal{H}_{n,t}$  and the variable defined in (SB.35). Due to the aforementioned independence property, we see that  $Z_{n,t}$  is also a martingale difference array with respect to  $\mathcal{H}'_{n,t}$ . In addition, we can rewrite

$$\alpha^\top D_{n,t} = \alpha^\top \sum_{s \in [1, t)} Z_{n,s} + \sqrt{\alpha^\top \tilde{\Sigma}_{n,t}^+ \alpha} \frac{\alpha^\top \sum_{s \in (t, n+k_n]} \tilde{\zeta}_{n,s}}{\sqrt{\alpha^\top \tilde{\Sigma}_{n,t}^+ \alpha}},$$

which implies that  $\alpha^\top D_{n,t}$  is  $\mathcal{H}'_{n,t-1}$ -measurable. Hence,

$$\mathbb{E}[\partial \Delta_{\sigma_n} (\alpha^\top D_{n,t})^\top \alpha^\top Z_{n,t}] = \mathbb{E} \left[ \partial \Delta_{\sigma_n} (\alpha^\top D_{n,t})^\top \alpha^\top \mathbb{E} [Z_{n,t} | \mathcal{H}'_{n,t-1}] \right] = 0, \quad (\text{SB.36})$$

and

$$\begin{aligned} & \mathbb{E} \left[ \text{Tr}(\partial^2 \Delta_{\sigma_n} (\alpha^\top D_{n,t}) \alpha^\top Z_{n,t} Z_{n,t}^\top \alpha) \right] \\ &= \mathbb{E} \left[ \text{Tr}(\partial^2 \Delta_{\sigma_n} (\alpha^\top D_{n,t}) \alpha^\top \mathbb{E}[Z_{n,t} Z_{n,t}^\top | \mathcal{H}'_{n,t-1}] \alpha) \right] \\ &= \mathbb{E} \left[ \text{Tr}(\partial^2 \Delta_{\sigma_n} (\alpha^\top D_{n,t}) \alpha^\top \mathbb{E}[Z_{n,t} Z_{n,t}^\top | \mathcal{H}_{n,t-1}] \alpha) \right] \\ &= \mathbb{E} \left[ \text{Tr}(\partial^2 \Delta_{\sigma_n} (\alpha^\top D_{n,t}) \alpha^\top V_{n,t}^+ \alpha) \right]. \end{aligned} \quad (\text{SB.37})$$

Summarizing the results in (SB.34), (SB.36), and (SB.37), we have

$$\begin{aligned} \mathbb{E}[\partial \Delta_{\sigma_n} (\alpha^\top D_{n,t})^\top \alpha^\top Z_{n,t}] &= \mathbb{E}[\partial \Delta_{\sigma_n} (\alpha^\top D_{n,t})^\top \alpha^\top \tilde{\zeta}_{n,t}], \\ \mathbb{E} \left[ \text{Tr}(\partial^2 \Delta_{\sigma_n} (\alpha^\top D_{n,t}) \alpha^\top Z_{n,t} Z_{n,t}^\top \alpha) \right] &= \mathbb{E} \left[ \text{Tr}(\partial^2 \Delta_{\sigma_n} (\alpha^\top D_{n,t}) \alpha^\top \tilde{\zeta}_{n,t} \tilde{\zeta}_{n,t}^\top \alpha) \right]. \end{aligned}$$

Then, from (SB.29), (SB.32), and (SB.33), we deduce

$$\begin{aligned} & \left| \mathbb{E}[\Delta_{\sigma_n} (\alpha^\top S_n^+)] - \mathbb{E}[\Delta_{\sigma_n} (\alpha^\top \tilde{S}_n^*)] \right| \\ & \leq \frac{5C_f}{\sigma_n^2 \delta_n} \sum_{t=1}^{n+k_n} \left( \mathbb{E} \left[ \max_l |\alpha_l^\top Z_{n,t}|^3 \right] + \mathbb{E} \left[ \max_l |\alpha_l^\top \tilde{\zeta}_{n,t}|^3 \right] \right). \end{aligned} \quad (\text{SB.38})$$

This estimate and Lemma SB1 imply that

$$\begin{aligned} \left| \mathbb{E}[\Delta_{\sigma_n}(\alpha^\top S_n^+)] - \mathbb{E}[\Delta_{\sigma_n}(\alpha^\top \tilde{S}_n^*)] \right| &\leq \frac{10C_f K}{\sigma_n^2 \delta_n} \left( B_{1,n} + (\log L_n)^{3/2} k_n^{-1/2} \right) \\ &\leq K \frac{B_{1,n}}{\sigma_n^2 \delta_n}, \end{aligned} \quad (\text{SB.39})$$

where the second line follows from the condition that  $k_n \geq (\log L_n)^3 B_{1,n}^{-2}$ . Note that the constant  $K$  does not depend on  $n$  or  $C$ . This finishes the proof of (SB.27).  $\square$

## SC Pre-whitened HAC estimator with growing dimensions

In this appendix, we provide details about the pre-whitened HAC estimator. Section SC.1 describes the estimator and its asymptotic properties. The proofs are in Section SC.2.

### SC.1 The estimator and its asymptotic properties

To implement the CSPA test, we need an estimator  $\hat{A}_n$  that satisfies Assumption 1 in the present context with growing dimensions. While Newey–West type estimators are theoretically valid (see Li and Liao (2019)), they may lead to size distortions in finite samples when the data is highly serially dependent as shown in our Monte Carlo experiments. This concern motivates us to analyze a more general class of HAC estimators allowing for pre-whitening as proposed by Andrews and Monahan (1992). Since this HAC estimation result is useful not only for the CSPA test, but also for other types of time series inference problems involving growing dimensions, we develop the theory in a general setting for a  $M_n$ -dimensional mixingale array  $(e_{n,t})_{1 \leq t \leq n, n \geq 1}$  with respect to a filtration  $(\mathcal{F}_{n,t})_{1 \leq t \leq n, n \geq 1}$ . If our goal is to verify Assumption 1 in the main text, we can set  $e_{n,t} = u_t \otimes P(X_t)$  and  $M_n = Jm_n$ . Some regularity conditions are needed.

**Assumption SC1.** *We have the following conditions: (i) for some  $q \geq 3$ ,  $(e_{n,t})$  is a  $M_n$ -dimensional  $L_q$ -mixingale array with respect to a filtration  $(\mathcal{F}_{n,t})$ , that is, for each  $1 \leq j \leq M_n$  and  $h \geq 0$ ,*

$$\|\mathbb{E}[e_{j,n,t} | \mathcal{F}_{n,t-h}]\|_q \leq \bar{c}_n \psi_h, \quad \|e_{j,n,t} - \mathbb{E}[e_{j,n,t} | \mathcal{F}_{n,t+h}]\|_q \leq \bar{c}_n \psi_{h+1},$$

*for a positive sequence  $\bar{c}_n$  and some finite constants  $(\psi_k)_{k \geq 0}$ ; (ii)  $\sum_{h \geq 0} \psi_h < \infty$  and  $\sup_{h \geq 0} h \psi_h^2 < \infty$ ; (iii)  $\mathbb{E}[e_{n,t}] = 0$  and  $\mathbb{E}[e_{n,t} e_{n,t+h}^\top]$  does not depend on  $t$ ; (iv) for all  $h \geq 0$  and  $s \geq 0$ ,*

$$\sup_t \max_{1 \leq l, k \leq M_n} \|\mathbb{E}[e_{l,n,t} e_{k,n,t+h} | \mathcal{F}_{n,t-s}] - \mathbb{E}[e_{l,n,t} e_{k,n,t+h}]\|_2 \leq \bar{c}_n^2 \psi_s;$$

*(v)  $\sup_t \max_{1 \leq l, k \leq M_n} \|e_{l,n,t} e_{k,n,t+h}\|_2 \leq \bar{c}_n^2$  for all  $h \geq 0$ ; (vi) the largest eigenvalues of  $\mathbb{E}[e_{n,t} e_{n,t}^\top]$  are uniformly bounded.*

Our goal is to construct pre-whitened HAC estimators for the long-run variance-covariance matrix

$$A_n \equiv \mathbb{E} \left[ \frac{1}{n} \left( \sum_{t=1}^n e_{n,t} \right) \left( \sum_{t=1}^n e_{n,t} \right)^\top \right].$$

To pre-whiten the data, we consider a linear filter  $B_n(\mathbb{L})$  of the form

$$B_n(\mathbb{L}) \equiv I_{M_n} - \sum_{s=1}^p B_{n,s} \mathbb{L}^s,$$

where  $\mathbb{L}$  denotes the lag operator,  $I_{M_n}$  is the  $M_n$ -dimensional identity matrix, and for each  $s \in \{1, \dots, p\}$ ,  $B_{n,s}$  is a  $M_n \times M_n$  nonrandom matrix. The associated pre-whitened array and its long-run variance-covariance matrix are then defined as

$$\begin{aligned} e_{n,t}^* &\equiv B_n(\mathbb{L})e_{n,t}, \quad \text{for } t \in \{p+1, \dots, n\}, \\ A_n^* &\equiv \mathbb{E} \left[ \frac{1}{n-p} \left( \sum_{t=p+1}^n e_{n,t}^* \right) \left( \sum_{t=p+1}^n e_{n,t}^* \right)^\top \right]. \end{aligned}$$

We impose the following conditions on the filter and the pre-whitened data.

**Assumption SC2.** (i)  $\max_{1 \leq s \leq p} \|B_{n,s}\|_S \leq C$  for some finite constant  $C > 0$  and the eigenvalues of  $B_n(1)$  are bounded away from zero; (ii)  $\sum_{h=1}^{\infty} h^{r_2^*} \|\Gamma_n^*(h)\|_S \leq K \bar{c}_n^2 M_n$  for some positive constants  $r_2^*$  and  $K$ , where  $\Gamma_n^*(h) \equiv \mathbb{E}[e_{n,t}^* e_{n,t+h}^{*\top}]$  is the  $h$ th autocovariance matrix of  $e_{n,t}^*$ .

Pre-whitening is motivated by the fact that  $A_n^*$  tends to be easier to estimate nonparametrically than  $A_n$  because the autocovariances of the pre-whitened data generally decay faster. This effect is captured by the  $r_2^*$  constant in Assumption SC2 (i.e., higher  $r_2^*$  corresponds to faster decay). For example, if the pre-whitened array  $e_{n,t}^*$  is a white noise,  $r_2^*$  can be taken to be an arbitrarily large number, which results in a faster rate of convergence in the HAC estimation.

In typical applications,  $e_{n,t}$  is not directly observed, but relies on some preliminary estimator. We formalize the setup with generated variables as follows. Suppose that

$$e_{n,t} = g(Z_{n,t}, \theta_0),$$

where  $Z_{n,t}$  is an observed time series and  $g(z, \theta)$  is a measurable function known up to a parameter  $\theta_0$ . We note that the unknown parameter  $\theta_0$  may be infinite dimensional. This is indeed the case for the CSPA application, where  $\theta_0$  represents the conditional mean functions  $(h_j(\cdot))_{1 \leq j \leq J}$  in our CSPA test. We suppose that a preliminary estimator  $\hat{\theta}_n$  for  $\theta_0$  is available and use

$$\hat{e}_{n,t} = g(Z_{n,t}, \hat{\theta}_n)$$



as a proxy for  $e_{n,t}$ . Similarly, the filter  $B_n(\mathbb{L})$  may also rely on parameters that need estimation. We consider preliminary estimators  $(\widehat{B}_{n,s})_{1 \leq s \leq p}$  and set

$$\widehat{B}_n(\mathbb{L}) \equiv I_{M_n} - \sum_{s=1}^p \widehat{B}_{n,s} \mathbb{L}^s.$$

The following high-level assumptions on the preliminary estimators' convergence rates will be used in our analysis.

**Assumption SC3.** (i)  $n^{-1} \sum_{t=1}^n \|g(Z_{n,t}, \hat{\theta}_n) - g(Z_{n,t}, \theta_0)\|^2 = O_p(\delta_{\theta,n}^2)$  for some positive sequence  $\delta_{\theta,n} = o(1)$ ; (ii)  $\max_{1 \leq s \leq p} \|\widehat{B}_{n,s} - B_{n,s}\|_S = O_p(\delta_{B,n})$  for some positive sequence  $\delta_{B,n} = o(1)$ .

Assumption SC3(i) mainly concerns the convergence rate of the preliminary estimator  $\hat{\theta}_n$ . Quite commonly,  $g(\cdot)$  is stochastically Lipschitz in  $\theta$  and  $\delta_{\theta,n}$  is determined by the convergence rate of  $\hat{\theta}_n$ , for which there are well-known results in the literature. Condition (ii) specifies the convergence rate of  $(\widehat{B}_{n,s})_{1 \leq s \leq p}$ . These conditions can be easily verified under more primitive conditions.

We are now ready to describe the pre-whitened HAC estimator for  $A_n$ . The feasible proxy of the pre-whitened series  $e_{n,t}^*$  is given by

$$\hat{e}_{n,t}^* = \widehat{B}_n(\mathbb{L}) \hat{e}_{n,t}, \quad p+1 \leq t \leq n.$$

We then estimate  $\Gamma_n^*(h) \equiv \mathbb{E}[e_{n,t}^* e_{n,t+h}^{*\top}]$  using

$$\widehat{\Gamma}_n^*(h) \equiv \frac{1}{n-p} \sum_{t=p+1}^{n-h} \hat{e}_{n,t}^* \hat{e}_{n,t+h}^{*\top}, \quad \widehat{\Gamma}_n^*(-h) \equiv \widehat{\Gamma}_n^*(h)^\top, \quad \text{for } 0 \leq h \leq n-p-1.$$

Consider a kernel function  $\mathcal{K}(\cdot)$  that is bounded, Lebesgue-integrable, symmetric, continuous at zero with  $\mathcal{K}(0) = 1$  such that, for some constants  $C \in \mathbb{R}$  and  $r_1 \in (0, \infty]$ ,

$$\lim_{x \rightarrow 0} \frac{1 - \mathcal{K}(x)}{|x|^{r_1}} = C; \tag{SC.1}$$

see many examples in Andrews (1991). In addition, we choose a bandwidth sequence  $\kappa_n$  that goes to infinity with the sample size  $n$ . The HAC estimator for the pre-whitened data is then given by

$$\widehat{A}_n^* \equiv \sum_{h=-(n-p-1)}^{n-p-1} \mathcal{K}\left(\frac{h}{\kappa_n}\right) \widehat{\Gamma}_n^*(h).$$

Finally, with  $\widehat{H}_n \equiv (I_{M_n} - \sum_{s=1}^p \widehat{B}_{n,s})^{-1}$ , we define the pre-whitened HAC estimator for  $A_n$  as

$$\widehat{A}_n \equiv \widehat{H}_n \widehat{A}_n^* \widehat{H}_n^\top.$$

Theorem SC1, below, describes the asymptotic behavior of this estimator.

**Theorem SC1.** *Suppose that Assumptions SC1, SC2, and SC3 hold, and*

$$\bar{c}_n^2 M_n (\kappa_n n^{-1/2} + \kappa_n^{-r_1 \wedge r_2^*}) + \kappa_n M_n^{1/2} \delta_{\theta, n} + \kappa_n \delta_{B, n} M_n = o(1). \quad (\text{SC.2})$$

Then,

$$\left\| \widehat{A}_n - A_n \right\|_S = O_p(\bar{c}_n^2 M_n (\kappa_n n^{-1/2} + \kappa_n^{-r_1 \wedge r_2^*})) + O_p(\kappa_n M_n^{1/2} \delta_{\theta, n}) + O_p(\kappa_n M_n \delta_{B, n}). \quad (\text{SC.3})$$

Theorem SC1 quantifies the magnitudes of three sources of estimation errors in the pre-whitened HAC estimator  $\widehat{A}_n$ . The first  $O_p(\bar{c}_n^2 M_n (\kappa_n n^{-1/2} + \kappa_n^{-r_1 \wedge r_2^*}))$  component is from the infeasible pre-whitened HAC estimator constructed with known  $\theta_0$  and  $(B_{n,s})_{1 \leq s \leq p}$ . Other things equal, this term tends to be smaller when the autocovariance of  $e_{n,t}^*$  decays faster (i.e., larger  $r_2^*$ ). This provides a theoretical reason for the better finite-sample performance of the pre-whitened estimator. The other two error terms are resulted from the estimation error in  $\hat{\theta}_n$  and  $(\widehat{B}_{n,s})_{1 \leq s \leq p}$ , respectively. Under quite general settings,  $\|\widehat{A}_n - A_n\|_S$  converges to zero in probability at polynomial rate as required in Assumption 1.

## SC.2 Proof of Theorem SC1

In this subsection, we prove Theorem SC1. We first prove two preliminary lemmas concerning the infeasible estimator constructed using known  $\theta_0$  and  $B_n(\mathbb{L})$ .

**Lemma SC1.** *Let  $H_n \equiv B_n(1)^{-1}$ . Under Assumptions SC1 and SC2,*

$$\left\| A_n - H_n A_n^* H_n^\top \right\|_S = O(\bar{c}_n^2 M_n n^{-1}). \quad (\text{SC.4})$$

PROOF. Denote  $n_p = n - p$ . It is easy to see that

$$A_n = \sum_{h=-(n-1)}^{n-1} \frac{n - |h|}{n} \Gamma_n(h), \quad A_n^* = \sum_{h=-(n_p-1)}^{n_p-1} \frac{n_p - |h|}{n_p} \Gamma_n^*(h),$$

where  $\Gamma_n(h) = \mathbb{E}[e_{n,t} e_{n,t+h}^\top]$  and  $\Gamma_n^*(h) = \mathbb{E}[e_{n,t}^* e_{n,t+h}^{*\top}]$ . For notational simplicity, we further denote

$$\beta_{n,0} = I_{M_n}, \quad \beta_{n,s} = -B_{n,s}, \quad \text{for } 1 \leq s \leq p. \quad (\text{SC.5})$$

We can then rewrite  $e_{n,t}^* = \sum_{s=0}^p \beta_{n,s} e_{n,t-s}$ , and subsequently,

$$\Gamma_n^*(h) = \sum_{s,u=0}^p \beta_{n,s} \Gamma_n(h + s - u) \beta_{n,u}^\top = \sum_{l=-p}^p \sum_{s=0 \vee l}^{p \wedge (l+p)} \beta_{n,s} \Gamma_n(h + l) \beta_{n,s-l}^\top.$$

Using this representation, we can rewrite  $A_n^*$  as

$$\begin{aligned} A_n^* &= \sum_{h=-(n_p-1)}^{n_p-1} \frac{n_p - |h|}{n_p} \left( \sum_{l=-p}^p \sum_{s=0 \vee l}^{p \wedge (l+p)} \beta_{n,s} \Gamma_n(h + l) \beta_{n,s-l}^\top \right) \\ &= \sum_{h=-(n-1)}^{n-1} \sum_{l=-p}^p \sum_{s=0 \vee l}^{p \wedge (l+p)} \mathbf{1}_{\{h-n_p+1 \leq l \leq h+n_p-1\}} \frac{n_p - |h-l|}{n_p} \beta_{n,s} \Gamma_n(h) \beta_{n,s-l}^\top. \end{aligned} \quad (\text{SC.6})$$

Since  $B_n(1) = \sum_{s=0}^p \beta_{n,s}$ , we also have

$$\begin{aligned} B_n(1) A_n B_n(1)^\top &= \sum_{h=-(n-1)}^{n-1} \sum_{s=0}^p \sum_{u=0}^p \frac{n-|h|}{n} \beta_{n,s} \Gamma_n(h) \beta_{n,u}^\top \\ &= \sum_{h=-(n-1)}^{n-1} \sum_{l=-p}^p \sum_{u=0 \vee l}^{p \wedge (l+p)} \frac{n-|h|}{n} \beta_{n,s} \Gamma_n(h) \beta_{n,s-l}^\top. \end{aligned} \quad (\text{SC.7})$$

By (SC.6) and (SC.7),

$$\begin{aligned} &\left\| A_n^* - B_n(1) A_n B_n(1)^\top \right\|_S \\ &\leq \sum_{h=-(n-1)}^{n-1} \sum_{l=-p}^p \sum_{s=0 \vee l}^{p \wedge (l+p)} \left( \mathbb{1}_{\{h-n_p+1 \leq l \leq h+n_p-1\}} \frac{n_p - |h-l|}{n_p} - \frac{n-|h|}{n} \right) \left\| \beta_{n,s} \Gamma_n(h) \beta_{n,s-l}^\top \right\|_S \\ &\leq K n^{-1} \sum_{h=-(n-1)}^{n-1} \left\| \Gamma_n(h) \right\|_S, \end{aligned} \quad (\text{SC.8})$$

where the first inequality is by the triangle inequality, and the second inequality holds because  $\|B_{n,s}\|_S$  is bounded and for  $|l| \leq p$ ,

$$\left| \mathbb{1}_{\{h-n_p+1 \leq l \leq h+n_p-1\}} \frac{n_p - |h-l|}{n_p} - \frac{n-|h|}{n} \right| \leq K n^{-1}.$$

Let  $\Gamma_{k,l,n}(h) = \mathbb{E}[e_{k,n,t} e_{l,n,t}]$  be the  $(k, l)$  element of  $\Gamma_n(h)$ . Note that

$$\begin{aligned} |\Gamma_{k,l,n}(h)| &\leq \left| \mathbb{E}[(e_{k,n,t} - \mathbb{E}[e_{k,n,t} | \mathcal{F}_{n,t+\lfloor h/2 \rfloor}]) e_{l,n,t+h}] \right| \\ &\quad + \left| \mathbb{E}[\mathbb{E}[e_{k,n,t} | \mathcal{F}_{n,t+\lfloor h/2 \rfloor}] \mathbb{E}[e_{l,n,t+h} | \mathcal{F}_{n,t+\lfloor h/2 \rfloor}]] \right| \\ &\leq K \bar{c}_n^2 \psi_0 \psi_{\lfloor h/2 \rfloor}, \end{aligned}$$

where the first inequality is by the triangle inequality, and the second inequality follows from the Cauchy–Schwarz inequality and mixingale properties. Hence,

$$\sum_{h=-(n-1)}^{n-1} \left\| \Gamma_n(h) \right\|_S \leq \sum_{h=-\infty}^{\infty} \left( \sum_{k=1}^{M_n} \sum_{l=1}^{M_n} \Gamma_{k,l,n}(h)^2 \right)^{1/2} \leq K \bar{c}_n^2 M_n \sum_{h=0}^{\infty} \psi_{\lfloor h/2 \rfloor}.$$

Since  $\sum_{h \geq 0} \psi_h < \infty$  by assumption, we further deduce that  $\sum_{h=-(n-1)}^{n-1} \left\| \Gamma_n(h) \right\|_S = O(\bar{c}_n^2 M_n)$ . In view of (SC.8), we then have

$$\left\| A_n^* - B_n(1) A_n B_n(1)^\top \right\|_S = O(\bar{c}_n^2 M_n n^{-1}).$$

Since the eigenvalues of  $B_n(1)$  are bounded away from zero, the assertion of the lemma readily follows from this estimate.  $\square$

**Lemma SC2.** *Under Assumptions SC1 and SC2,*

$$\left\| \tilde{A}_n^* - A_n^* \right\|_S = O_p(\tilde{c}_n^2 M_n (\kappa_n n^{-1/2} + \kappa_n^{-r_1 \wedge r_2^*})),$$

where  $\tilde{A}_n^* \equiv \sum_{h=-(n-p-1)}^{n-p-1} \mathcal{K}(h/\kappa_n) \tilde{\Gamma}_n^*(h)$  and  $\tilde{\Gamma}_n^*(h) \equiv (n-p)^{-1} \sum_{t=p+1}^{n-h} e_{n,t}^* e_{n,t+h}^{*\top}$ .

PROOF. For ease of notation, we denote  $n_p = n - p$  and write  $\sum_h$  in place of  $\sum_{h=-(n-p-1)}^{n-p-1}$ . By definition, we can decompose

$$\tilde{A}_n^* - A_n^* = \sum_h \mathcal{K}\left(\frac{h}{\kappa_n}\right) \left( \tilde{\Gamma}_n^*(h) - \mathbb{E}[\tilde{\Gamma}_n^*(h)] \right) + \sum_h \left( \mathcal{K}\left(\frac{h}{\kappa_n}\right) - 1 \right) \frac{n_p - |h|}{n_p} \Gamma_n^*(h). \quad (\text{SC.9})$$

Under Assumption SC1, we can follow the same proof as Lemma A6 of Li and Liao (2019) to show that

$$\mathbb{E} \left[ \left\| \tilde{\Gamma}_n^*(h) - \mathbb{E}[\tilde{\Gamma}_n^*(h)] \right\|_S \right] \leq K \tilde{c}_n^2 M_n n^{-1/2}.$$

By the triangle inequality, we further deduce

$$\begin{aligned} & \mathbb{E} \left[ \left\| \sum_h \mathcal{K}\left(\frac{h}{\kappa_n}\right) \left( \tilde{\Gamma}_n^*(h) - \mathbb{E}[\tilde{\Gamma}_n^*(h)] \right) \right\|_S \right] \\ & \leq K \tilde{c}_n^2 M_n n^{-1/2} \sum_h |\mathcal{K}(h/\kappa_n)| \leq K \tilde{c}_n^2 M_n \kappa_n n^{-1/2}. \end{aligned}$$

Hence, the first term on the right-hand side of (SC.9) satisfies

$$\left\| \sum_h \mathcal{K}\left(\frac{h}{\kappa_n}\right) \left( \tilde{\Gamma}_n^*(h) - \mathbb{E}[\tilde{\Gamma}_n^*(h)] \right) \right\|_S = O_p(\tilde{c}_n^2 M_n \kappa_n n^{-1/2}). \quad (\text{SC.10})$$

Let  $r = r_1 \wedge r_2^*$ . By the properties of the kernel function  $\mathcal{K}(\cdot)$ , we can find some (small) constant  $\varepsilon \in (0, 1)$  such that

$$\frac{|1 - \mathcal{K}(x)|}{|x|^r} \leq \frac{|1 - \mathcal{K}(x)|}{|x|^{r_1}} \leq K \text{ for } x \in [-\varepsilon, \varepsilon].$$

Hence,  $|\mathcal{K}(h/\kappa_n) - 1| \leq K |h|^r \kappa_n^{-r}$  when  $|h| \leq \varepsilon \kappa_n$ . Therefore,

$$\sum_{|h| \leq \varepsilon \kappa_n} \left| \left( \mathcal{K}\left(\frac{h}{\kappa_n}\right) - 1 \right) \frac{n_p - |h|}{n_p} \right| \|\Gamma_n^*(h)\|_S \leq K \kappa_n^{-r} \sum_{|h| \leq \varepsilon \kappa_n} |h|^r \|\Gamma_n^*(h)\|_S.$$

Since  $\mathcal{K}(\cdot)$  is bounded,

$$\begin{aligned} & \sum_{|h| > \varepsilon \kappa_n} \left| \left( \mathcal{K}\left(\frac{h}{\kappa_n}\right) - 1 \right) \frac{n_p - |h|}{n_p} \right| \|\Gamma_n^*(h)\|_S \\ & \leq K \sum_{|h| > \varepsilon \kappa_n} \left( \frac{|h|}{\varepsilon \kappa_n} \right)^r \|\Gamma_n^*(h)\|_S \leq K \kappa_n^{-r} \sum_{|h| > \varepsilon \kappa_n} |h|^r \|\Gamma_n^*(h)\|_S. \end{aligned}$$

Combining the two displayed estimates above yields

$$\sum_h \left| \left( \mathcal{K} \left( \frac{h}{\kappa_n} \right) - 1 \right) \frac{n_p - |h|}{n_p} \right| \|\Gamma_n^*(h)\|_S \leq K \kappa_n^{-r} \sum_h |h|^r \|\Gamma_n^*(h)\|_S.$$

By Assumption SC2,  $\sum_{h=1}^{\infty} h^{r_2^*} \|\Gamma_n^*(h)\|_S \leq K \bar{c}_n^2 M_n$ . Hence, the second term on the right-hand side of (SC.9) satisfies

$$\left\| \sum_h \left( \mathcal{K} \left( \frac{h}{\kappa_n} \right) - 1 \right) \frac{n_p - |h|}{n_p} \Gamma_n^*(h) \right\|_S \leq K \bar{c}_n^2 M_n \kappa_n^{-r}. \quad (\text{SC.11})$$

The assertion of the lemma then follows from (SC.9), (SC.10), and (SC.11).  $\square$

We are now ready to prove Theorem SC1.

PROOF OF THEOREM SC1. By Lemma SC1, Lemma SC2, and (SC.2),

$$\left\| A_n - H_n \tilde{A}_n^* H_n^\top \right\|_S = O_p(\bar{c}_n^2 M_n (\kappa_n n^{-1/2} + \kappa_n^{-r_1 \wedge r_2^*})) = o_p(1), \quad (\text{SC.12})$$

where  $H_n$  and  $\tilde{A}_n^*$  are defined in those lemmas. Since  $A_n$  has bounded eigenvalues under Assumption SC1,  $\tilde{A}_n^*$ 's largest eigenvalue is also  $O_p(1)$ .

We complement the definitions in (SC.5) by setting

$$\hat{\beta}_{n,0} = I_{M_n}, \quad \hat{\beta}_{n,s} = -\hat{B}_{n,s}, \quad \text{for } 1 \leq s \leq p.$$

Under Assumption SC2 and Assumption SC3,  $\|\hat{\beta}_{n,s}\|_S = O_p(1)$  for each  $s$ . Note that

$$\hat{e}_{n,t}^* = \sum_{s=0}^p \hat{\beta}_{n,s} \hat{e}_{n,t-s}, \quad e_{n,t}^* = \sum_{s=0}^p \beta_{n,s} e_{n,t-s}.$$

We then observe that, for each  $h$ ,

$$\begin{aligned} & \sum_{t=p+1}^{n-h} \|\hat{e}_{n,t}^* - e_{n,t}^*\|^2 \\ &= \sum_{t=p+1}^{n-h} \left\| \sum_{s=0}^p \hat{\beta}_{n,s} \hat{e}_{n,t-s} - \sum_{s=0}^p \beta_{n,s} e_{n,t-s} \right\|^2 \\ &\leq K \sum_{t=p+1}^{n-h} \left\| \sum_{s=0}^p \hat{\beta}_{n,s} (\hat{e}_{n,t-s} - e_{n,t-s}) \right\|^2 + K \sum_{t=p+1}^{n-h} \left\| \sum_{s=0}^p (\hat{\beta}_{n,s} - \beta_{n,s}) e_{n,t-s} \right\|^2 \\ &\leq K \sum_{t=p+1}^{n-h} \sum_{s=0}^p \|\hat{\beta}_{n,s}\|_S^2 \|\hat{e}_{n,t-s} - e_{n,t-s}\|^2 + K \sum_{t=p+1}^{n-h} \sum_{s=0}^p \|\hat{\beta}_{n,s} - \beta_{n,s}\|_S^2 \|e_{n,t-s}\|^2 \\ &\leq K \left( \sum_{s=0}^p \|\hat{\beta}_{n,s}\|_S^2 \right) \left( \sum_{t=1}^n \|\hat{e}_{n,t} - e_{n,t}\|^2 \right) + K \left( \sum_{s=0}^p \|\hat{\beta}_{n,s} - \beta_{n,s}\|_S^2 \right) \sum_{t=1}^n \|e_{n,t}\|^2 \end{aligned}$$

where the first inequality is by the  $C_r$ -inequality, the second inequality is by the Cauchy–Schwarz inequality, and the third inequality is obvious. Since the eigenvalues of  $\mathbb{E}[e_{n,t}e_{n,t}^\top]$  are bounded by Assumption SC1,  $\sum_{t=1}^n \|e_{n,t}\|^2 = O_p(nM_n)$ . By Assumption SC3, we can further deduce that, uniformly in  $h$ ,

$$\frac{1}{n-p} \sum_{t=p+1}^{n-h} \|\hat{e}_{n,t}^* - e_{n,t}^*\|^2 = O_p(\delta_{\theta,n}^2 + \delta_{B,n}^2 M_n). \quad (\text{SC.13})$$

We further note that, uniformly in  $h$ ,

$$\begin{aligned} & \left\| \widehat{\Gamma}_n^*(h) - \widetilde{\Gamma}_n^*(h) \right\| \\ &= \left\| \frac{1}{n-p} \sum_{t=p+1}^{n-h} \left( \hat{e}_{n,t}^* \hat{e}_{n,t+h}^{*\top} - e_{n,t}^* e_{n,t+h}^{*\top} \right) \right\| \\ &\leq \frac{1}{n-p} \left\| \sum_{t=p+1}^{n-h} (\hat{e}_{n,t}^* - e_{n,t}^*) e_{n,t+h}^{*\top} \right\| + \frac{1}{n-p} \left\| \sum_{t=p+1}^{n-h} e_{n,t}^* (\hat{e}_{n,t+h}^* - e_{n,t+h}^*)^\top \right\| \\ &\quad + \frac{1}{n-p} \left\| \sum_{t=p+1}^{n-h} (\hat{e}_{n,t}^* - e_{n,t}^*) (\hat{e}_{n,t+h}^* - e_{n,t+h}^*)^\top \right\| \\ &\leq \frac{2}{n-p} \left( \sum_{t=1}^n \|\hat{e}_{n,t}^* - e_{n,t}^*\|^2 \right)^{1/2} \left( \sum_{t=1}^n \|e_{n,t}^*\|^2 \right)^{1/2} + \frac{1}{n-p} \sum_{t=1}^n \|\hat{e}_{n,t}^* - e_{n,t}^*\|^2 \\ &= O_p\left(\delta_{\theta,n} M_n^{1/2} + \delta_{B,n} M_n\right), \end{aligned}$$

where the first equality is by definition, the first inequality is by the triangle inequality, the second inequality is by the Cauchy–Schwarz inequality, and the last line follows from (SC.13). This estimate further implies that

$$\left\| \widehat{A}_n^* - \widetilde{A}_n^* \right\|_S = O_p\left(\kappa_n \left(\delta_{\theta,n} M_n^{1/2} + \delta_{B,n} M_n\right)\right). \quad (\text{SC.14})$$

Noting that  $\|\widehat{H}_n^{-1}\|_S = O_p(1)$ , we also have

$$\left\| \widehat{H}_n \widehat{A}_n^* \widehat{H}_n^\top - \widehat{H}_n \widetilde{A}_n^* \widehat{H}_n^\top \right\|_S = O_p\left(\kappa_n \left(\delta_{\theta,n} M_n^{1/2} + \delta_{B,n} M_n\right)\right). \quad (\text{SC.15})$$

Since  $\|\widehat{H}_n - H_n\|_S = O_p(\delta_{B,n})$  and  $\|\widetilde{A}_n^*\|_S = O_p(1)$ ,

$$\left\| \widehat{H}_n \widetilde{A}_n^* \widehat{H}_n^\top - H_n \widetilde{A}_n^* H_n^\top \right\|_S = O_p(\delta_{B,n}). \quad (\text{SC.16})$$

The assertion of the theorem then follows from (SC.12), (SC.15), and (SC.16).  $\square$

## References

- ANDREWS, D. W. K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59(3), 817–858.
- ANDREWS, D. W. K., AND J. C. MONAHAN (1992): “An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator,” *Econometrica*, 60(4), 953–966.
- CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2014): “Gaussian Approximation of Suprema of Empirical Processes,” *Annals of Statistics*, 42(4), 1564–1597.
- CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): “Intersection Bounds: Estimation and Inference,” *Econometrica*, 81(2), 667–737.
- LI, J., AND Z. LIAO (2019): “Uniform Nonparametric Inference for Time Series,” *Journal of Econometrics*, Forthcoming.
- RIO, E. (2009): “Moment Inequalities for Sums of Dependent Random Variables under Projective Conditions,” *Journal of Theoretical Probability*, 22(1), 146–163.
- VAN DER VAART, A., AND J. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag.