Journal of the Royal Statistical Society Series B: Statistical Methodology, 2023, **00**, 1–27
https://doi.org/10.1093/jrsssb/qkad108

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society **Series B** Statistical Methodology B

Original Article

Another look at bandwidth-free inference: a sample splitting approach

Yi Zhang and Xiaofeng Shao

Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

Address for correspondence: Xiaofeng Shao, Department of Statistics, University of Illinois at Urbana-Champaign, 605 East Springfield Avenue, 135 Computer Application Building, Champaign, IL 61820, USA. Email: xshao@illinois.edu

Abstract

The bandwidth-free tests for a multi-dimensional parameter have attracted considerable attention in econometrics and statistics literature. These tests can be conveniently implemented due to their tuning-parameter free nature and possess more accurate size as compared to the traditional heteroskedasticity and autocorrelation consistent-based approaches. However, when sample size is small/medium, these bandwidth-free tests exhibit large size distortion when both the dimension of the parameter and the magnitude of temporal dependence are moderate, making them unreliable to use in practice. In this paper, we propose a sample splitting-based approach to reduce the dimension of the parameter to one for the subsequent bandwidth-free inference. Our SS–SN (sample splitting plus self-normalisation) idea is broadly applicable to many testing problems for time series, including mean testing, testing for zero autocorrelation, and testing for a change point in multivariate mean, among others. Specifically, we propose two types of SS–SN test statistics and derive their limiting distributions under both the null and alternatives and show their effectiveness in alleviating size distortion via simulations. In addition, we obtain the limiting distributions for both SS–SN test statistics in the multivariate mean testing problem when the dimension is allowed to diverge.

Keywords: fixed-*b* asymptotics, hypothesis testing, long run variance, self-normalisation, time series

1 Introduction

Hypothesis testing for a multi-dimensional parameter is often encountered in the analysis of economic time series. Classical approaches involve conducting consistent estimation of the variance–covariance matrix of the parameter estimate non-parametrically using spectral methods [e.g. heteroskedasticity and autocorrelation consistent (HAC) estimators] and constructing standard tests based on the asymptotic normality of the parameter estimate and consistency of HAC estimator. The use of HAC estimator has been extensively analysed in econometrics literature; see Andrews (1991), Andrews and Monahan (1992), Gallant (2009), Hansen (1992), Newey and West (1987), and Robinson (1991, 1998) for important contributions. It has become a long tradition in time series analysis and econometrics to use HAC estimator, and it has been implemented in many statistical and econometrics softwares.

Since the pioneering work of Kiefer et al. (2000) (KVB thereafter), bandwidth-free inference has become an important alternative, due to the difficulty of choosing the optimal bandwidth in the use of HAC estimator, and good statistical property of the KVB test. Specifically, the KVB test was developed for linear regression model with dynamic regressors and heteroscedastic and serially correlated errors. Their test statistics have non-standard asymptotic distributions that only depend on the number of restrictions being tested, and critical values are easy to simulate using standard techniques. The main advantage of the KVB approach compared to standard HAC-based counterpart is that estimates of the variance–covariance matrix are not explicitly required so the sensitivity of HAC estimator with respect to the choice of bandwidth (or truncation lag) is avoided, as no bandwidth is involved in the KVB test.

Received: February 21, 2023. **Revised:** September 1, 2023. **Accepted:** September 9, 2023
© The Royal Statistical Society 2023. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

In statistics literature, Lobato (2001) proposed a bandwidth-free test for the uncorrelation at top K lags of a time series and described the principle of bandwidth-free inference using the simple mean testing example, which can be viewed as a special case of KVB test. Inspired by Lobato (2001) and Kiefer et al. (2000), Shao (2010) proposed the self-normalisation (SN, hereafter) technique for the inference (testing and confidence region construction) of a general parameter, including marginal means, quantiles, and autocorrelation at specific lags, in the setting of stationary time series. Shao and Zhang (2010) further extended self-normalisation to testing for a change point in a parameter associated with a weakly dependent time series and modified the self-normaliser to adapt to the change-point testing problem. For many follow-up work on self-normalisation for time series, we refer the reader to the review by Shao (2015). A major message from this line of literature is that no tuning parameter (i.e. bandwidth or truncation lag) is needed in conducting hypothesis testing or confidence interval construction as we can use an inconsistent estimator of asymptotic variance–covariance matrix (or long run covariance matrix) and the resulting studentised statistic is asymptotically pivotal. Both theoretical and empirical research suggest that the size associated with bandwidth-free test is typically more accurate as compared to the classical HAC-based method with some degree of power loss (Jansson, 2004; Kiefer & Vogelsang, 2005; Sun et al., 2008; Zhang & Shao, 2013).

Despite the implementational convenience and size accuracy of bandwidth-free tests, it has been empirically observed that the size can still be quite distorted when the dimension of the parameter is moderate and the temporal dependence is moderate/strong; see Figure 1 for an illustration in the mean testing context. This phenomenon is not superising in view of the theoretical work by Sun (2014c), where the impact of dimensionality and serial dependence on the size distortion was carefully investigated via edgeworth expansion for a class of F-test statistics under both small-b and fixed-b asymptotics (Kiefer & Vogelsang, 2005). Note that in the mean testing problem, the SN test statistic corresponds to fixed-b asymptotics with $b = 1$ and the use of Bartlett kernel (Kiefer & Vogelsang, 2002).

Moderate dimensional time series with moderate/strong temporal dependence are prevalent in practice. Therefore, there is a strong need to develop new testing methods that can control the size when the dimension of the parameter is moderate and the temporal dependence is moderate/strong. When the time series is very strongly autocorrelated, Müller (2014) and Sun (2014a) proposed methods to control the size in a near unit root model and focused on the univariate setting. In contrast, the temporal dependence in our framework is relatively weak compared to those examined in their work as the focus is more on reducing the size distortion due to the moderate dimensionality.

In this article, we develop a sample splitting-based approach (called SS–SN, sample splitting plus self-normalisation) to reduce the size distortion associated with bandwidth-free inference. The

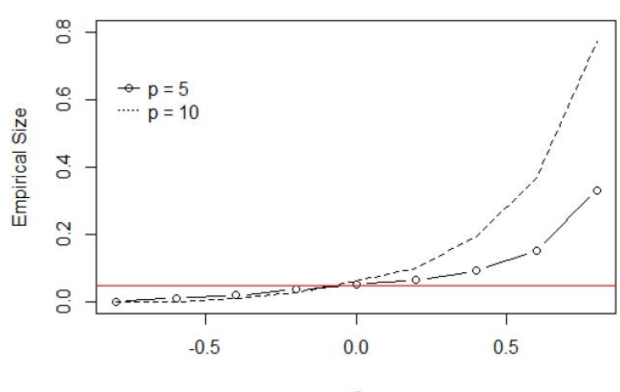

Figure 1. Empirical size for traditional SN test on multivariate mean.

basic idea is to split the full sample into two parts, with one part used to reduce the dimension of the parameter to one, and the other part used to perform bandwidth-free testing for the dimension-reduced (i.e. one-dimensional) time series. We show that this SS–SN approach is generally applicable to testing for the multivariate mean, uncorrelation at a finite number of lags, regression coefficients in linear regression models with time series regressor/error and a change point in multivariate mean, among others. Using the orthogonal increment property of Brownian motion and a novel conditioning argument, we obtain the limiting null distributions of our L_∞-type and L_2-type SS–SN test statistics when the dimension is fixed, which is pivotal and is independent of the sample splitting proportion. We also derive the asymptotic power function for our proposed test and compare to its bandwidth-free counterpart. By using a recent result on sequential Gaussian approximation for time series in a growing-dimensional environment (Mies & Steland, 2023), we show the asymptotic validity of our SS–SN test statistics in a multivariate mean testing problem when the dimension diverges as sample size grows to infinity. Under the same setting, we further obtain the asymptotic independence of L_∞-type and L_2-type SS–SN test statistics under the null, which justifies the Bonferroni test that combines the L_∞-type and L_2-type tests in achieving all-round power against dense and sparse alternatives. The theoretical tools we develop for the growing-dimensional setting are of independent interest.

The idea of sample splitting-based inference is not new, and there is a large literature in statistics and machine learning; see Shafer and Vovk (2008), Wasserman and Roeder (2009), Rinaldo et al. (2019), Wasserman et al. (2020), and Du et al. (2023) among others. However, it seems that sample splitting is mostly used for the inference of independent data. In the context of time series, sample splitting was used for the post-selection inference in Lunde (2019), for the identification testing for structural VAR models in Maciejowska (2022) and for unit root testing in Chang et al. (2022). These are the only references we are aware of. The scope and property of our proposed SS–SN inference are substantially different from these papers and have no overlap with the existing literature.

The rest of this paper is organised as follows. Section 2.1 describes the SN method in a multi-dimensional mean testing problem and illustrates its large size distortion due to moderate dimension and temporal dependence. Then, we propose our SS–SN test statistics and investigate their asymptotic properties under the null and local alternatives in Sections 2.1 and 2.2. In Sections 2.3–2.5, we present the asymptotic theories for the two SS–SN test statistics when the dimension is allowed to diverge. In Section 3, we present several extensions, including testing for zero autocorrelation in a time series, linear hypothesis testing in a regression model and testing for a change point in multivariate mean. Simulation results are provided in Section 4 and Section 5 concludes. Proofs for main results and auxiliary lemmas are gathered in the online supplemental material, which also contains some variants of SS–SN test statistics based on different rescaling methods, corresponding simulation results, and a real data illustration.

2 Methodology and theory

In this section, we introduce our L_∞-type SS–SN test statistic in the case of testing the mean of a multivariate stationary time series in Section 2.1, and we develop an L_2-type SS–SN statistic which targets the dense alternative in Section 2.2. We present the asymptotic theories for the two SS–SN test statistics in the growing-dimensional setting in Sections 2.3–2.5, respectively.

2.1 Hypothesis testing on multi-dimensional mean

Let $\mathbf{X}_t = (X_t^1, X_t^2, \ldots, X_t^p)^\top$ be a p-dimensional stationary time series with mean $E(\mathbf{X}_t) = \boldsymbol{\mu} = (\mu^1, \mu^2, \ldots, \mu^p)^\top \in \mathbb{R}^p$. We want to test the null hypothesis $H_0{:}\boldsymbol{\mu} = \boldsymbol{\mu}_0 = (\mu_0^1, \mu_0^2, \ldots, \mu_0^p)^\top$ against $H_A{:}\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. Denote $\boldsymbol{S}_{a,b} = \sum_{t=a}^{b} \mathbf{X}_t$, $S_{a,b}^j = \sum_{t=a}^{b} X_t^j$, the autocovariance matrix $\boldsymbol{\Gamma}(k) = E[(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_{t+k} - \boldsymbol{\mu})^\top]$ and let $\boldsymbol{\Gamma} = \sum_{k=-\infty}^{\infty} \boldsymbol{\Gamma}(k)$ be the long run covariance matrix with the (i, j) element being Γ_{ij}. Also denote the ith row of $\boldsymbol{\Gamma}^{1/2}$ as $\boldsymbol{\Gamma}_i^\top$, so we have $\Gamma_{ij} = \boldsymbol{\Gamma}_i^\top \boldsymbol{\Gamma}_j$. The following functional central limit theorem (FCLT) is needed in deriving the asymptotic properties. Here, we let $D^d[0, 1]$ (when $d = 1$, we omit the superscript and just use $D[0, 1]$) denote the space of $\mathbb{R}^d$ valued functions on $[0, 1]$ which are right continuous and have left limit, endowed with the topology induced by the multi-dimensional Skorokhod metric (Billingsley, 2013).

Assumption 1 (Functional central limit theorem). We assume that

$$n^{-1/2}(\boldsymbol{S}_{1,\lfloor nr\rfloor} - \lfloor nr\rfloor\boldsymbol{\mu}) \Rightarrow \boldsymbol{\Gamma}^{1/2}\mathbf{B}_p(r) \quad \text{on } D^p[0,1], \tag{1}$$

where $\mathbf{B}_p(r){:}[0,1] \to \mathbb{R}^p$ is a p-dimensional vector of independent Brownian motions (we omit the subscript and use $B(r)$ when $p = 1$), and '$\Rightarrow$' signifies weak convergence in $D^p[0,1]$ (Billingsley, 2013).

Throughout, p is held fixed except for Sections 2.3–2.5, where $p = p_n$ is diverging as $n \to \infty$. When p is fixed, the above FCLT holds under mild moment and weak dependence assumptions; see Lobato (2001) for discussion on the primitive assumptions for FCLT. The SN (self-normalised) test statistic is

$$T_n^p = n^{-1}(\boldsymbol{S}_{1,n} - n\boldsymbol{\mu}_0)^\top(\boldsymbol{V}_n^p)^{-1}(\boldsymbol{S}_{1,n} - n\boldsymbol{\mu}_0),$$

where $\boldsymbol{V}_n^p = n^{-2}\sum_{t=1}^n\{\boldsymbol{S}_{1,t} - (t/n)\boldsymbol{S}_{1,n}\}\{\boldsymbol{S}_{1,t} - (t/n)\boldsymbol{S}_{1,n}\}^\top$. Under the null, T_n^p converges in distribution to $U_p = \boldsymbol{B}_p(1)^\top\boldsymbol{V}_p^{-1}\boldsymbol{B}_p(1)$, where $\boldsymbol{V}_p = \int_0^1[\boldsymbol{B}_p(r) - r\boldsymbol{B}_p(1)][\boldsymbol{B}_p(r) - r\boldsymbol{B}_p(1)]^\top \mathrm{d}r$. Since the distribution of U_p is pivotal and its upper critical values have been tabulated in Lobato (2001), we reject the null hypothesis at level ζ if T_n^p is larger than the $100(1-\zeta)\%$ upper critical value of U_p, denoted as $U_{p,\zeta}$.

One major drawback of this SN testing procedure is that there is large size distortion under the null when n is small/moderate, and when p is moderately large or the autocorrelation is moderate/strong. To show numerically how large the size distortion is, we test the null hypothesis $\boldsymbol{\mu} = \mathbf{0}$, where $\mathbf{0}$ is a vector in $\mathbb{R}^p$ with all elements being 0, and simulate the data from the VAR(1) process $\mathbf{X}_t = \rho\mathbf{I}_p\mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t$, where $\mathbf{I}_p$ is the p-dimensional identity matrix and $\boldsymbol{\epsilon}_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_p)$. We set the nominal level at 5% and repeat the mean test 5,000 times with the length of time series $n = 100$ and $p \in \{5, 10\}$. As shown in Figure 1, the size distortion for the above SN test when $p = 10$ is much larger than when $p = 5$ and the test is severely oversized when ρ is close to 1 and severely undersized when ρ is close to -1.

Next, we introduce an SS–SN test statistic to reduce the size distortion. The SS–SN procedure consists of two steps: (a) we split the sample into two parts: $\mathcal{P}_1 := \{\mathbf{X}_1, \ldots, \mathbf{X}_{\lfloor n\alpha\rfloor}\}$ and $\mathcal{P}_2 := \{\mathbf{X}_{\lfloor n\alpha\rfloor+1}, \ldots, \mathbf{X}_n\}$, where $\alpha \in (0,1)$ is the splitting ratio. For $i = 1, 2, \ldots, p$, denote $\sigma_i^2 = Var(X_1^i)$, $\hat{\sigma}_i^2 = \frac{1}{\lfloor n\alpha\rfloor}\sum_{t=1}^{\lfloor n\alpha\rfloor}(X_t^i - S_{1,\lfloor n\alpha\rfloor}^i/\lfloor n\alpha\rfloor)^2$ and based on the first part $\mathcal{P}_1$, define

$$\hat{j} = \underset{j=1,2,\ldots,p}{\operatorname{argmax}} \frac{n^{-1}(S_{1,\lfloor n\alpha\rfloor}^j - \lfloor n\alpha\rfloor\mu_0^j)^2}{\hat{\sigma}_j^2}, \tag{2}$$

which represents the coordinate that corresponds to the largest deviation from the null. Note that $\hat{j}$ bears the signal and is solely determined by the difference between sample mean and true mean, rescaled by the sample variance of each component time series in $\mathcal{P}_1$. Under the alternative, $\hat{j}$ estimates the coordinate with the strongest deviation from the null as scaled by its corresponding marginal variance, see Theorem 1 below. Note that there are other sensible ways of rescaling in determining $\hat{j}$ in equation (2). We refer the reader to Remark 4 in Section 2.2 and online supplementary Appendix A. (b) Then we construct a SN test statistic based on the $\hat{j}$th dimension/component of the second part $\mathcal{P}_2$, or the projected sample $\{\mathbf{e}_{\hat{j}}^\top\mathbf{X}_{\lfloor n\alpha\rfloor+1}, \ldots, \mathbf{e}_{\hat{j}}^\top\mathbf{X}_n\}$, where $\mathbf{e}_j$ is a vector in $\mathbb{R}^p$ with jth element being 1 and all other elements being 0. So the SS–SN$_1$ statistic is defined as

$$T_n^{(\mathrm{M})}(\alpha, \hat{j}) = \frac{(n - \lfloor n\alpha\rfloor)^{-1}(S_{\lfloor n\alpha\rfloor+1,n}^{\hat{j}} - (n - \lfloor n\alpha\rfloor)\mu_0^{\hat{j}})^2}{V_n^{(\mathrm{M})}(\hat{j})} \tag{3}$$

where $V_n^{(M)}(j) = (n - \lfloor n\alpha \rfloor)^{-2} \sum_{k=\lfloor n\alpha\rfloor+1}^{n} (S^j_{\lfloor n\alpha\rfloor+1,k} - \frac{k-\lfloor n\alpha\rfloor}{n-\lfloor n\alpha\rfloor} S^j_{\lfloor n\alpha\rfloor+1,n})^2$. To derive the limiting distributions of $T_n^{(M)}(\alpha, \hat{j})$ under both null and alternative, we introduce another assumption on σ_i^2 and $\hat{\sigma}_i^2$.

Assumption 2 $\sigma_i^2 > 0$ and $\hat{\sigma}_i^2 \xrightarrow{p} \sigma_i^2$ for $i = 1, 2, \ldots, p$.

Assumption 2 is mild and can be verified by imposing suitable moment and weak dependence assumptions on $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$. As shown below, the limiting null distribution of $T_n^{(M)}(\alpha, \hat{j})$ is U_1, so the level ζ test is $\mathbf{1}(T_n^{(M)}(\alpha, \hat{j}) > U_{1,\zeta})$. The following theorem shows the asymptotic properties of $T_n^{(M)}(\alpha, \hat{j})$ under the null and alternatives.

Theorem 1 Suppose Assumptions 1 and 2 hold. Then (i) under H_0, we have

$$T_n^{(M)}(\alpha, \hat{j}) \xrightarrow{\mathcal{D}} U_1, \tag{4}$$

where '$\xrightarrow{\mathcal{D}}$' signifies convergence in distribution. (ii) Under H_A, let the true mean be $\boldsymbol{\mu} = \boldsymbol{\mu}_n$ and denote $\|\boldsymbol{\mu}_n - \boldsymbol{\mu}_0\|_\infty = \max_{j=1,2,\ldots,p} |\mu_n^j - \mu_0^j|$.

1. If $\sqrt{n}\|\boldsymbol{\mu}_n - \boldsymbol{\mu}_0\|_\infty \to \infty$, then $T_n^{(M)}(\alpha, \hat{j}) \xrightarrow{p} \infty$, thus the limiting power is 1.
2. If $\sqrt{n}(\boldsymbol{\mu}_n - \boldsymbol{\mu}_0) \to \mathbf{c} := (c^1, c^2, \ldots, c^p)$ and $\|\mathbf{c}\|_\infty \neq 0$, then we have

$$\hat{j} \xrightarrow{\mathcal{D}} \operatorname*{argmax}_{j=1,2,\ldots,p} \frac{\left\{B^{(j)}(\alpha) + \alpha c^j\right\}^2}{\sigma_j^2} \stackrel{d}{=} j^*,$$

$$T_n^{(M)}(\alpha, \hat{j}) \xrightarrow{\mathcal{D}} U^*,$$

where $B^{(j)}(r) = \boldsymbol{\Gamma}_j^\top \mathbf{B}_p(r)$ is mean zero Brownian motion with covariance $Cov(B^{(i)}(u), B^{(j)}(v)) = \min\{u, v\}\Gamma_{ij}$ and the conditional distribution of U^* given $j^* = j$ is

$$U^*|_{j^*=j} \stackrel{d}{=} \frac{\left\{B(1) + \sqrt{\frac{1-\alpha}{\Gamma_{jj}}} c^j\right\}^2}{\int_0^1 \left\{B(r) - rB(1)\right\}^2 \mathrm{d}r}.$$

Since the non-central chi-square distribution is statistically larger than chi-square distribution and $\{B(r) - rB(1)\}_{r\in[0,1]}$ is independent of $B(1)$, our test has non-trivial power asymptotically.

3. If $\sqrt{n}\|\boldsymbol{\mu}_n - \boldsymbol{\mu}_0\|_\infty \to 0$, then $T_n^{(M)}(\alpha, \hat{j}) \xrightarrow{\mathcal{D}} U_1$, so our test has trivial power asymptotically.

The limiting null distribution U_1 is the same as the one in Lobato (2001) when $p = 1$ and the critical values are already tabulated there. Also it is interesting to note that the limiting null distribution does not depend on the sample splitting proportion $\alpha \in (0, 1)$. We shall study the impact of α on size accuracy and power later.

2.2 L_2-type SS–SN statistic

The SS–SN$_1$ test statistic is expected to have good power when the alternative is sparse and strong, as only the $\hat{j}$th component time series is used in the testing after dimension reduction. As will be shown in Figure 2a later, SS–SN$_1$ test has more power loss under the dense alternative (i.e. a substantial portion of coordinates of $\boldsymbol{\mu} - \boldsymbol{\mu}_0$ is non-zero) than under the sparse alternative (i.e. a small portion of coordinates of $\boldsymbol{\mu} - \boldsymbol{\mu}_0$ is non-zero), as compared to the traditional SN test. This

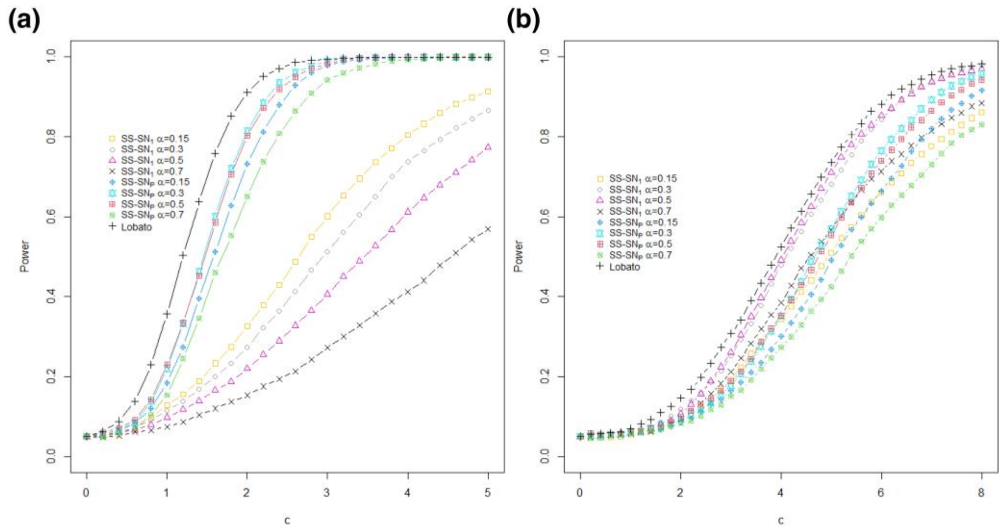

Figure 2. Asymptotic power under the dense (a) and sparse (b) alternatives when testing hypothesis on multivariate mean.

motivates us to propose another SS–SN statistic which can preserve the power under the dense alternative. To be specific, on $\mathcal{P}_1$, define

$$\hat{\boldsymbol{P}} = \operatorname{diag}\left\{\frac{1}{\sqrt{\hat{\sigma}_1^2}}, \ldots, \frac{1}{\sqrt{\hat{\sigma}_p^2}}\right\}\frac{1}{\sqrt{n}}(\boldsymbol{S}_{1,\lfloor n\alpha\rfloor} - \lfloor n\alpha\rfloor\boldsymbol{\mu}_0). \tag{5}$$

Note that under the alternative, $\hat{\boldsymbol{P}}$ estimates the direction with the strongest deviation from the null as measured by the squared L_2 norm of the rescaled signal $(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^\top\operatorname{diag}\{\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_p^2}\}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)$. Then the L_2-type SS–SN statistic (i.e. SS–SN$_P$) is defined as

$$Q_n^{(M)}(\alpha) = \frac{(n-\lfloor n\alpha\rfloor)^{-1}\left\{\hat{\boldsymbol{P}}^\top\left[\mathbf{S}_{\lfloor n\alpha\rfloor+1,n}-(n-\lfloor n\alpha\rfloor)\boldsymbol{\mu}_0\right]\right\}^2}{V_n(\alpha)}, \tag{6}$$

where $V_n(\alpha) = (n-\lfloor n\alpha\rfloor)^{-2}\sum_{k=\lfloor n\alpha\rfloor+1}^{n}\{\hat{\boldsymbol{P}}^\top[\mathbf{S}_{\lfloor n\alpha\rfloor+1,k} - \frac{k-\lfloor n\alpha\rfloor}{n-\lfloor n\alpha\rfloor}\mathbf{S}_{\lfloor n\alpha\rfloor+1,n}]\}^2$. Instead of constructing the test statistic using the $\hat{j}$th coordinate of the second part $\mathcal{P}_2$ as done for SS–SN$_1$, we construct the SN statistic based on the projected sample $\{\hat{\boldsymbol{P}}^\top\mathbf{X}_{\lfloor n\alpha\rfloor+1}, \ldots, \hat{\boldsymbol{P}}^\top\mathbf{X}_n\}$. The following theorem shows the asymptotic properties of $Q_n^{(M)}(\alpha)$ under the null and alternatives.

Theorem 2 Suppose Assumptions 1 and 2 hold. Then (i) under H_0, we have

$$Q_n^{(M)}(\alpha)\xrightarrow{\mathcal{D}}U_1. \tag{7}$$

(ii) Under H_A, let the true mean be $\boldsymbol{\mu}=\boldsymbol{\mu}_n$ and denote $\|\boldsymbol{\mu}_n-\boldsymbol{\mu}_0\| = \sqrt{\sum_{i=1}^{p}(\mu_n^i-\mu_0^i)^2}$.

1. If $\sqrt{n}\|\boldsymbol{\mu}_n-\boldsymbol{\mu}_0\| \to \infty$, then $Q_n^{(M)}(\alpha)\xrightarrow{p}\infty$, thus the limiting power is 1.

2. If $\sqrt{n}(\boldsymbol{\mu}_n - \boldsymbol{\mu}_0) \to \mathbf{c} := (c^1, c^2, \ldots, c^p)$ and $\|\mathbf{c}\| \neq 0$, then we have

$$\hat{\boldsymbol{P}} \xrightarrow{\mathcal{D}} \mathrm{diag}\left\{\frac{1}{\sqrt{\sigma_1^2}}, \ldots, \frac{1}{\sqrt{\sigma_p^2}}\right\}\left[\boldsymbol{\Gamma}^{1/2}\mathbf{B}_p(\alpha) + \alpha\mathbf{c}\right] \stackrel{d}{=} \boldsymbol{P}^*,$$

$$Q_n^{(M)}(\alpha) \xrightarrow{\mathcal{D}} U^{**},$$

and the conditional distribution of U^{**} given $\boldsymbol{P}^* = \boldsymbol{P}$ is

$$U^{**}\big|_{\boldsymbol{P}^*=\boldsymbol{P}} \stackrel{d}{=} \frac{\left\{B(1) + \sqrt{\frac{1-\alpha}{\boldsymbol{P}^\top \boldsymbol{\Gamma} \boldsymbol{P}}}\boldsymbol{P}^\top \mathbf{c}\right\}^2}{\int_0^1 \left\{B(r) - rB(1)\right\}^2 \mathrm{d}r}.$$

In this case, our test has non-trivial power asymptotically.

3. If $\sqrt{n}\|\boldsymbol{\mu}_n - \boldsymbol{\mu}_0\| \to 0$, then $Q_n^{(M)}(\alpha) \xrightarrow{\mathcal{D}} U_1$, so our test has trivial power asymptotically.

Remark 1 For the traditional SN statistic, the computational cost is of order $O(p^2 n + p^3)$, which scales quadratically in p and is $O(p^2 n)$ if $p \ll n$. In contrast, the computational cost for both our SS–SN statistics are of order $O(pn)$, which is linear in p. This could result in substantial saving in computation when p is moderate.

Remark 2 To understand how SS–SN$_P$ statistic can reduce power loss incurred by SS–SN$_1$ under dense alternative, we shall focus on the local alternative as in part (ii).2 of Theorem 2 with $\mathbf{c} = c\mathbf{1}$ and $\boldsymbol{\Gamma} = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_p^2\} = \mathbf{I}_p$, where $\mathbf{1}$ is a vector in $\mathbb{R}^p$ with all elements being 1. According to Theorem 6 in Magnus (1986), $E\frac{\boldsymbol{P}^\top \mathbf{c}\mathbf{c}^\top \boldsymbol{P}}{\boldsymbol{P}^\top \boldsymbol{P}} = pc^2$, so on average, the non-central constant for the numerator of U^{**} is $(1-\alpha)pc^2$, which is p times the non-central constant for the numerator of U^*. Hence, SS–SN$_P$ statistic is expected to outperform SS–SN$_1$ statistic in power under dense alternative when the same α is used.

Remark 3 Under the null, the limiting distribution U_1 is pivotal and does not depend on the splitting ratio α. Under the local alternative $\sqrt{n}(\boldsymbol{\mu}_n - \boldsymbol{\mu}_0) \to \mathbf{c}$, the limiting distributions of our SS–SN$_1$ and SS–SN$_P$ test statistics depend on α, $\mathbf{c}$, $\{\sigma_j^2\}_{j=1}^p$, and $\boldsymbol{\Gamma}$. According to Lemma 4 in Lobato (2001), the limiting distribution of the traditional SN test statistic is $[\boldsymbol{B}_p(1) + \boldsymbol{\Gamma}^{-1/2}\mathbf{c}]^\top \boldsymbol{V}_p^{-1}[\boldsymbol{B}_p(1) + \boldsymbol{\Gamma}^{-1/2}\mathbf{c}]$, which depends on $\mathbf{c}$ and $\boldsymbol{\Gamma}$.

To understand the power behaviour of SS–SN$_1$ and SS–SN$_P$ statistics, as compared to the traditional SN test, we set $p = 10$ and calculate the asymptotic power $P(U^* > U_{1,0.05})$, $P(U^{**} > U_{1,0.05})$, and $P([\boldsymbol{B}_{10}(1) + \boldsymbol{\Gamma}^{-1/2}\mathbf{c}]^\top \boldsymbol{V}_{10}^{-1}[\boldsymbol{B}_{10}(1) + \boldsymbol{\Gamma}^{-1/2}\mathbf{c}] > U_{10,0.05})$ under the sparse alternative $\mathbf{c} = c\mathbf{e}_1$ and dense alternative $\mathbf{c} = c\mathbf{1}$. Here, $\boldsymbol{\Gamma} = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_p^2\} = \mathbf{I}_{10}$ and $U_{1,0.05}$, $U_{10,0.05}$ are the 95th upper percentile of U_1 and U_{10}, respectively. We plot the asymptotic power as a function of c. Here, we approximate the asymptotic power by approximating the p-dimensional Brownian motion with standardised partial sum of 5,000 iid $N(0, I_p)$ random vectors and setting the number of replications as 3,000.

As shown in Figure 2a and b, both SS–SN$_1$ and SS–SN$_P$ statistics have power loss compared with the traditional SN test by Lobato, which is expected since only the second part of data (i.e. $\mathcal{P}_2$) is directly used in constructing the SN statistics. This is a price we have to pay to achieve more accurate size and lower computational cost. For dense alternative, the power loss of SS–SN$_1$ statistic appears substantially larger than that of SS–SN$_P$ statistic for all α, which is consistent with the theoretical finding (cf. Remark 2). It appears that when $\alpha = 0.3$, 0.5, the SS–SN$_P$ statistic achieves the best power as compared to other SS–SN counterparts, and the power loss relative to Lobato's test is moderate. In contrast, the optimal α for SS–SN$_1$ is 0.15, suggesting that the optimal α in general depends on which SS–SN statistic we use. For sparse alternative, the optimal power corresponds to SS–SN$_1$ with $\alpha = 0.3$, 0.5, which outperforms other SS–SN counterparts and the power loss is very small compared with traditional SN statistic. It is worth noting that there is no advantage to set $\alpha > 0.5$, as that is always dominated by $\alpha = 0.5$ in power.

As shown in Figure 2, SS–SN$_P$ outperforms SS–SN$_1$ under dense alternative and SS–SN$_1$ outperforms SS–SN$_P$ under sparse alternative. In practice, if the practitioner has the prior knowledge about the type of alternative, then he/she is recommended to choose the one of SS–SN test statistics accordingly. In the absence of such knowledge, we recommend to combine the two SS–SN test statistics via a simple Bonferroni procedure. Since when $\alpha = 0.5$, SS–SN$_1$ have almost best power against sparse alternative and SS–SN$_P$ have almost best power against dense alternative, we combine SS–SN$_1$ and SS–SN$_P$ with $\alpha = 0.5$ and name it SS–SN$_b$. To be specific, the test using SS–SN$_b$ rejects the null at 5% level if either the test using SS–SN$_1$ with $\alpha = 0.5$ or the test using SS–SN$_P$ with $\alpha = 0.5$ rejects the null at 2.5% level. In Section 4, we show through simulation that the power for SS–SN$_b$ is close to the best of two SS–SN statistics with overall good performance under both sparse and dense alternatives.

Remark 4 As pointed out by one of the reviewers, the projection $\hat{\mathbf{P}}$ defined in equation (5) is not necessarily the optimal direction of projection in terms of power maximisation. To see that, assume $\boldsymbol{\mu}_0 = \mathbf{0}$ and the true mean is $\boldsymbol{\mu}_n \neq \mathbf{0}$ under the alternative. For any fixed $\mathbf{P} \in \mathbb{R}^p$, we project the data in the second subsample along the direction $\mathbf{P}$ and construct a one-dimensional SN statistic. Then similar to part (ii).2 of Theorem 2, the statistic approximately follows the same distribution as

$$U_n = \frac{\left\{B(1) + \sqrt{\frac{1-\alpha}{\boldsymbol{P}^\top \boldsymbol{\Gamma} \boldsymbol{P}}}\boldsymbol{P}^\top \sqrt{n}\boldsymbol{\mu}_n\right\}^2}{\int_0^1 \left\{B(r) - rB(1)\right\}^2 \mathrm{d}r}$$

for large enough n. Note that the numerator and denominator of U_n are independent and conditioning on $\int_0^1 \{B(r) - rB(1)\}^2 \mathrm{d}r$, $\{B(1) + \sqrt{\frac{1-\alpha}{\boldsymbol{P}^\top \boldsymbol{\Gamma} \boldsymbol{P}}}\boldsymbol{P}^\top \sqrt{n}\boldsymbol{\mu}_n\}^2$ follows non-central chi-square distribution with one degree of freedom and non-central constant $n(1-\alpha)\frac{\boldsymbol{P}^\top \boldsymbol{\mu}_n^\top \boldsymbol{\mu}_n \boldsymbol{P}}{\boldsymbol{P}^\top \boldsymbol{\Gamma} \boldsymbol{P}}$. Following similar argument as in Theorem 3.4.1 of Huang (2015), we can show the optimal direction of projection which maximise $P(U_n \geq t)$ for all $t > 0$ is the one that maximise the non-central constant. According to A.4.11 in Seber and Lee (2003), it is proportional to $\mathbf{P}_n^* = \boldsymbol{\Gamma}^{-1/2}\boldsymbol{\mu}_n$. The projection $\hat{\mathbf{P}}$ defined in Section 2.2 of the paper is an estimator of $\tilde{\mathbf{P}}_n = \mathrm{diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2\}^{-1/2}\boldsymbol{\mu}_n$, which is not the optimal direction in theory.

To pursue the optimal projection, we need to provide a consistent long run covariance matrix estimator, which is hard for moderate dimensional time series when the sample size is small/medium. This point was also expressed in Korkas and Fryzlewicz (2017) for a one-dimensional change-point detection problem, where the authors state that estimating long run variance is a difficult problem in time series analysis and the estimation error would likely not make it worthwhile and they opt to rescale using marginal sample variance. This might suggest we estimate $\boldsymbol{\Gamma}(0)^{-1/2}\boldsymbol{\mu}_n$ instead of $\boldsymbol{\Gamma}^{-1/2}\boldsymbol{\mu}_n$, where $\boldsymbol{\Gamma}(0)$ is the marginal covariance matrix. However, the moderate and possibly growing dimensionality (see Sections 2.3–2.5) further adds complication to the estimation of $\boldsymbol{\Gamma}(0)^{-1/2}$, which is well recognised in high-dimensional mean testing for iid data; see Srivastava and Du (2008) and Srivastava et al. (2013). The latter authors propose to use the diagonal elements of the covariance matrix to replace $\boldsymbol{\Gamma}(0)$ in their testing to avoid the error accumulation due to high dimensionality. Based on these considerations, we opt to use an estimator of the simple projection vector $\tilde{\mathbf{P}}_n$, which does not involve any bandwidth parameter and appears to work well in finite sample.

In Appendix A of the online supplementary material, we compare different SS–SN statistics where the projection vectors are defined using different rescaling methods, and we show that for the SS–SN statistics rescaled by the long run covariance matrix estimator (SSSN-L_1 and SSSN-L_P), the power loss are larger compared with most marginally rescaled SS–SN statistics when there is no or weak cross-sectional dependence (see online supplementary Figures S1a, b and S3a, b), which confirmed the claim made in Korkas and Fryzlewicz (2017) (see the discussion before Section 4.1 therein).

2.3 Asymptotic theory for SS–SN$_1$ when the dimension is diverging

In this subsection, we justify the asymptotic validity of the SS–SN$_1$ statistic in the multivariate mean testing problem, when the dimension is diverging as sample size grows to infinity. This is consistent with the main theme of this work, that is, to address the large size distortion due to moderate dimensionality of the parameter we test. We shall use a set of notations with their dependence on n being explicit. Specifically, for $t = 1, 2, \ldots, n$, let $\mathbf{X}_{nt} = (X_{nt}^1, X_{nt}^2, \ldots, X_{nt}^{p_n})^\top$ be a stationary time series with mean $E(\mathbf{X}_{nt}) = \boldsymbol{\mu}_n = (\mu_n^1, \mu_n^2, \ldots, \mu_n^{p_n})^\top \in \mathbb{R}^{p_n}$ and with long run covariance matrix $\Gamma_n = \sum_{h=-\infty}^{\infty} Cov(\mathbf{X}_{nt}, \mathbf{X}_{n(t+|h|)}) = (\gamma_{nij})_{i,j=1}^{p_n}$. For any $i = 1, 2, \ldots, p_n$, let σ_{ni}^2 and $\hat{\sigma}_{ni}^2$ be the variance of X_{n1}^i and its sample version calculated on $\{X_{n1}^i, \ldots, X_{n\lfloor n\alpha\rfloor}^i\}$ and let $\gamma_{ni} = \gamma_{nii}$ be the ith diagonal element of Γ_n. For two functions $p(x)$ and $q(x)$ we write $p \lesssim q$ if there exist constant $C > 0$ such that $\lim\sup_{x\to\infty} |\frac{p(x)}{q(x)}| \le C$ and we write $p \asymp q$ if $p \lesssim q$ and $q \lesssim p$. In this section, we allow the dimension p_n to grow with n and we want to test the sequence of null hypotheses $H_{n0}{:}\boldsymbol{\mu}_n = \mathbf{0}$ against $H_{nA}{:}\boldsymbol{\mu}_n \neq \mathbf{0}$.

We use the physical dependence measure of Wu (2005) to describe the dependence structure of $\mathbf{X}_{nt}$. Let $\epsilon_i, \tilde{\epsilon}_i, i \in \mathbb{Z}$ be iid $U[0, 1]$ random variables and denote $\boldsymbol{\epsilon}_t = (\epsilon_t, \epsilon_{t-1}, \ldots) \in \mathbb{R}^\infty$, $\tilde{\boldsymbol{\epsilon}}_{t,j} = (\epsilon_t, \ldots, \epsilon_{j+1}, \tilde{\epsilon}_j, \epsilon_{j-1}, \ldots) \in \mathbb{R}^\infty$ and $\bar{\boldsymbol{\epsilon}}_{t,j} = (\epsilon_t, \ldots, \epsilon_{j+1}, \tilde{\epsilon}_j, \tilde{\epsilon}_{j-1}, \ldots) \in \mathbb{R}^\infty$. For some measurable function $\mathbf{G}_n = (G_n^1, \ldots, G_n^{p_n})^\top{:}\mathbb{R}^\infty \to \mathbb{R}^{p_n}$, define

$$\theta_{n,j,q} = (E\|\mathbf{G}_n(\boldsymbol{\epsilon}_0) - \mathbf{G}_n(\tilde{\boldsymbol{\epsilon}}_{0,-j})\|^q)^{\frac{1}{q}}, \quad j = 0, 1, 2, \ldots,$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^{p_n}$. The following assumption is needed to derive a strong approximation result for the partial sum process of $\mathbf{X}_{nt}$.

Assumption 3 Assume that $\mathbf{X}_{nt} - \boldsymbol{\mu}_n = \mathbf{G}_n(\epsilon_t)$, and for some constant $q > 4, \beta > 2$ and $\Theta_n > 0$, we have

$$\theta_{n,j,q} \le \Theta_n \frac{1}{(j \vee 1)^\beta}, \quad j = 0, 1, 2, \ldots \tag{8}$$

$$(E\|\mathbf{X}_{n1} - \boldsymbol{\mu}_n\|^q)^{\frac{1}{q}} \le \Theta_n. \tag{9}$$

Also, assume that there exist $0 < \gamma_{\min} < \gamma_{\max} < \infty$, $0 < \sigma_{\min}^2 < \sigma_{\max}^2 < \infty$ such that $\gamma_{\min} \le \gamma_{ni} \le \gamma_{\max}$, $\sigma_{\min}^2 \le \sigma_{ni}^2 \le \sigma_{\max}^2$ for any $n = 1, 2, \ldots$ and $i = 1, 2, \ldots, p_n$.

Define

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

Mies and Steland (2023) provided a sequential Gaussian approximation result for non-stationary time series in high dimension. Here, we provide a slightly refined version of Theorem 3.1 in Mies and Steland (2023) for stationary time series in the following proposition.

Proposition 1 Suppose Assumption 3 holds, then on a potentially different probability space, there exist random vectors $\{\mathbf{X}'_{nt}\}_{t=1}^{n} \overset{d}{=} \{\mathbf{X}_{nt}\}_{t=1}^{n}$ and a standard p_n-dimensional Brownian motion $B_{p_n}(r)$ such that for any small $\epsilon > 0$,

$$E \sup_{r\in[0,1]} \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} (\mathbf{X}'_{nt} - \boldsymbol{\mu}_n) - \mathbf{W}_n(r) \right\|^2 \leq C\left\{ \Theta_n^2 \log(n) \left(\frac{p_n}{n}\right)^{2\xi} + \frac{p_n \Theta_n^2}{n^{1-\epsilon}} \right\}, \quad (10)$$

where $\mathbf{W}_n(r) = (W_{n1}(r), W_{n2}(r), \ldots, W_{np_n}(r))^\top = \boldsymbol{\Gamma}_n^{1/2} \mathbf{B}_{p_n}(r)$ and $C > 0$ is a generic constant.

Note that in equation (10), the term $\Theta_n^2 \log(n)(\frac{p_n}{n})^{2\xi}$ comes from Theorem 3.1 in Mies and Steland (2023) and the term $\frac{p_n\Theta_n^2}{n^{1-\epsilon}}$ quantifies the difference between $W_n(\frac{\lfloor nr \rfloor}{n})$ and $W_n(r)$. The right-hand side of equation (10) converges to 0 if $\Theta_n = O(\sqrt{p_n})$ and $p_n \asymp n^{\psi}$ for some $0 < \psi < \frac{\xi}{\xi + \frac{1}{2}}$ and under these two conditions the first term $\Theta_n^2 \log(n)(\frac{p_n}{n})^{2\xi}$ dominates.

As in Section 2.1, we define the test statistic as

$$T_n^{(D)}(\alpha, \hat{j}_n) = \frac{(n - \lfloor n\alpha \rfloor)^{-1} [S_{\lfloor n\alpha \rfloor + 1, n}^{n\hat{j}_n}]^2}{V_n^{(D)}(\hat{j}_n)}, \quad (11)$$

where

$$\hat{j}_n = \operatorname*{argmax}_{j=1,2,\ldots,p_n} \frac{n^{-1}[S_{1,\lfloor n\alpha \rfloor}^{nj}]^2}{\hat{\sigma}_{nj}^2}, \quad (12)$$

$S_{a,b}^{nj} = \sum_{t=a}^{b} X_{nt}^{j}$ and $V_n^{(D)}(j) = (n - \lfloor n\alpha \rfloor)^{-2} \sum_{k=\lfloor n\alpha \rfloor + 1}^{n} (S_{\lfloor n\alpha \rfloor + 1, k}^{nj} - \frac{k - \lfloor n\alpha \rfloor}{n - \lfloor n\alpha \rfloor} S_{\lfloor n\alpha \rfloor + 1, n}^{nj})^2$. We derive the asymptotic properties of $T_n^{(D)}(\alpha, \hat{j}_n)$ under two sets of assumptions on the matrix Γ_n and dimensionality.

Assumption 4
(a) $\gamma_{nij} = \rho_{nij}\sqrt{\gamma_{ni}\gamma_{nj}}$ with $|\rho_{nij}| < \bar{\rho} \in (0, 1)$ for any $i, j = 1, 2, \ldots, p_n$ and $n = 1, 2, \ldots$.
(b) $\Theta_n = O(\sqrt{p_n})$.
(c) $p_n \asymp n^{\psi}$ for some $0 < \psi < \frac{\xi}{\xi + \frac{9}{2}}$.

Assumption 5
(a) Γ_n is diagonal.
(b) $\Theta_n = O(\sqrt{p_n})$.
(c) $p_n \asymp n^{\psi}$ for some $0 < \psi < \frac{\xi}{\xi + \frac{1}{2}}$.

Note that by Minkowski inequality, a sufficient condition for (b) in previous two assumptions is that $E|X_{n1}^j - \mu_n^j|^q$ is uniformly bounded for $n = 1, 2, \ldots$ and $j = 1, 2, \ldots, p_n$. The following theorem shows the asymptotic distribution of $T_n^{(D)}(\alpha, \hat{j}_n)$ under the null.

Theorem 3 Suppose either Assumptions 3, 4 or Assumptions 3, 5 hold. Then under H_{n0}, we have

$$T_n^{(D)}(\alpha, \hat{j}_n) \xrightarrow{\mathcal{D}} U_1. \tag{13}$$

The following theorem shows the consistency of our test under alternatives.

Theorem 4 Under H_{nA} and Assumption 3, denote $\|\boldsymbol{\mu}_n\|_\infty = \max_{j=1,2,\ldots,p_n} |\mu_n^j|$. Assume that $p_n \asymp n^\psi$ for some $0 < \psi < \frac{\xi}{\xi + \frac{1}{2}}$ and there exists $\kappa > 0$ such that $\frac{\sqrt{n}\|\boldsymbol{\mu}_n\|_\infty}{p_n^\kappa} \to \infty$. Then we have $T_n^{(D)}(\alpha, \hat{j}_n) \xrightarrow{p} \infty$.

2.4 Asymptotic theory for SS–SN$_P$ when the dimension is diverging

In this subsection, we justify the asymptotic validity of the SS–SN$_P$ statistic in the multivariate mean testing problem, when the dimension is diverging as sample size grows to infinity. As in Section 2.2, we define the test statistic as

$$Q_n^{(D)}(\alpha) = \frac{(n - \lfloor n\alpha \rfloor)^{-1}\left\{\hat{\boldsymbol{P}}_n^\top \mathbf{S}_{\lfloor n\alpha \rfloor+1,n}\right\}^2}{V_n^{(2)}(\alpha)}, \tag{14}$$

where

$$\hat{\boldsymbol{P}}_n = \left(\frac{S_{1,\lfloor n\alpha \rfloor}^{n1}}{\sqrt{n}\hat{\sigma}_{n1}}, \ldots, \frac{S_{1,\lfloor n\alpha \rfloor}^{np_n}}{\sqrt{n}\hat{\sigma}_{np_n}}\right)^\top, \tag{15}$$

$V_n^{(2)}(\alpha) = (n - \lfloor n\alpha \rfloor)^{-2} \sum_{k=\lfloor n\alpha \rfloor+1}^n \{\hat{\boldsymbol{P}}_n^\top [\mathbf{S}_{\lfloor n\alpha \rfloor+1,k} - \frac{k - \lfloor n\alpha \rfloor}{n - \lfloor n\alpha \rfloor} \mathbf{S}_{\lfloor n\alpha \rfloor+1,n}]\}^2$ and $\mathbf{S}_{a,b} = (S_{a,b}^{n1}, \ldots, S_{a,b}^{np_n})^\top = (\sum_{t=a}^b X_{nt}^1, \ldots, \sum_{t=a}^b X_{nt}^{p_n})^\top$.

The following theorem shows the asymptotic distribution of $Q_n^{(D)}(\alpha)$ under the null.

Theorem 5 Suppose Assumptions 3 and 5(b,c) hold. Then under H_{n0}, we have

$$Q_n^{(D)}(\alpha) \xrightarrow{\mathcal{D}} U_1. \tag{16}$$

Note that no restrictions on the correlation between different coordinates of $\mathbf{X}_{nt}$ are imposed, so both weak and strong cross-sectional dependence are allowed. The following theorem shows the consistency of our test under alternatives.

Theorem 6 Under H_{nA} and Assumption 3, denote $\|\boldsymbol{\mu}_n\| = \sqrt{\sum_{j=1}^{p_n} (\mu_n^j)^2}$. Assume that $p_n \asymp n^\psi$ for some $0 < \psi < \frac{\xi}{\xi + \frac{1}{2}}$ and there exists $\kappa > 0$ such that $\frac{\sqrt{n}\|\boldsymbol{\mu}_n\|}{p_n^{1/2+\kappa}} \to \infty$. Then we have $Q_n^{(D)}(\alpha) \xrightarrow{p} \infty$.

The asymptotic validity for the original SN test statistic T_n^p has only been provided for the fixed-p case, and whether it works under the diverging p setting is unknown. Therefore, we view the justification for SS–SN test statistics in the growing p setting an interesting theoretical contribution to the literature.

With that being said, it is worth noting that self-normalisation has been extended to do inference for high-dimensional time series via the use of U-statistics; see Wang and Shao (2020) and Wang et al. (2022). In particular, Wang and Shao (2020) adopted a trimmed U-statistic and developed a new SN test statistic to test for the mean of high-dimensional stationary time series. The restriction on the growth rate of the dimensionality in their work is minimal and they require $p \to \infty$ but allow $p \gg$, and in some special cases, p can grow exponentially. In contrast, we are focusing on the testing problem where the dimension of parameter is moderate, and the regime corresponds to either p is fixed or growing p with $p \ll n$. So the applicability of the tests developed in Wang and Shao (2020) and ours are fairly different. The test in Wang and Shao (2020) targets the dense alternative and requires weak cross-sectional dependence, whereas our two SS–SN test statistics together can capture both dense and sparse alternatives, and can accommodate both weak and strong cross-sectional dependence. The technical tools involved are also very different. Here, we rely on the strong approximation for partial sum process and a careful analysis of the maximum spacing for an independent but not identically distributed chi-square random variables, whereas the theory in Wang and Shao (2020) is built on the weak convergence of sequential U-statistic of high-dimensional dependent observations.

2.5 Asymptotic independence of SS–SN$_1$ and SS–SN$_P$

In the literature, there has been a sizeable amount of work on the asymptotic independence between the sum and maximum of a weakly dependent sequence (Hsing, 1995; Peng & Nadarajah, 2003) and between the sum-of-squares type test statistic and the maximum-type test statistic in high-dimensional testing problems; see Li and Xue (2015), Xu et al. (2016), and He et al. (2021), among others. It is natural to ask whether our L_2-type and L_∞-type SS–SN statistics are asymptotically independent in the growing-dimensional setting. We shall provide an affirmative answer to this question below.

As in the proof of Theorem 5, denote $\mathbf{D}_\text{g} = Diag\{\frac{1}{\sigma_{n1}}, \dots, \frac{1}{\sigma_{np_n}}\}$ and $\boldsymbol{\Lambda}_n = \boldsymbol{\Gamma}_n^{1/2}\mathbf{D}_\text{g}\boldsymbol{\Gamma}_n^{1/2}\boldsymbol{\Gamma}_n^{1/2}\mathbf{D}_\text{g}\boldsymbol{\Gamma}_n^{1/2}$ with eigenvalues $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{p_n}$. The following assumption is needed to prove the asymptotic independence.

Assumption 6 There exist $\epsilon > 0$ such that $\lambda_{p_n}/p_n^{1-\epsilon} \to 0$ as $p_n \to \infty$.

Let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra on $\mathbb{R}$, we now state asymptotic independence of $T_n^{(D)}(\alpha, \hat{j}_n)$ and $Q_n^{(D)}(\alpha)$, the proof of which is deferred to Appendix C.10 in the online supplementary material.

Theorem 7 Under H_{n0}, suppose Assumptions 3, 6 and either Assumption 4 or Assumtion 5 hold, then $T_n^{(D)}(\alpha, \hat{j}_n)$ and $Q_n^{(D)}(\alpha)$ are asymptotically independent in the sense that

$$\left| P\big(T_n^{(D)}(\alpha, \hat{j}_n) \in A; Q_n^{(D)}(\alpha) \in B\big) - P\big(T_n^{(D)}(\alpha, \hat{j}_n) \in A\big)P\big(Q_n^{(D)}(\alpha) \in B\big) \right| \to 0$$

for any $A, B \in \mathcal{B}(\mathbb{R})$.

A few remarks are in order. Note that the largest eigenvalue of $\boldsymbol{\Gamma}_n^{1/2}\mathbf{D}_\text{g}\boldsymbol{\Gamma}_n^{1/2}$ is $\lambda_{p_n}^{1/2}$. Denote the largest eigenvalue of $\boldsymbol{\Gamma}_n$ as $\tilde{\gamma}_n$, then according to online supplementary Lemma 2, if Assumption 3 holds, Assumption 6 is equivalent to $\tilde{\gamma}_n/p_n^{(1-\epsilon)/2} \to 0$ as $p_n \to \infty$. According to the Gershgorin Circle Theorem (see Bell, 1965), $\tilde{\gamma}_n$ is upper bounded by the largest absolute row sum of $\boldsymbol{\Gamma}_n$, so Assumption 6 holds if, under Assumption 3, $\boldsymbol{\Gamma}_n$ is diagonal or $\gamma_{nij} = c^{|i-j|}$ for some $c \in (-1, 1)$ [i.e. AR(1) type correlation]. This suggests that when the p components are independent or weakly correlated in the long run, Assumption 6 is satisfied and asymptotic independence between our L_2-type and L_∞-type SS–SN statistics holds.

On the other hand, if $\gamma_{nij} = c \in (0, \gamma_{\min})$ for all $i \ne j$, then we have $\boldsymbol{\Gamma}_n = Diag\{\gamma_{n1} - c, \dots, \gamma_{np_n} - c\} + c\mathbf{1}_n\mathbf{1}_n^\top$ where $\mathbf{1}_n$ is a vector in $\mathbb{R}^{p_n}$ with all elements being 1. Under

Assumption 3 all eigenvalues of $Diag\{\gamma_{n1}-c, \ldots, \gamma_{np_n}-c\}$ are non-negative and the largest eigenvalue of $c\mathbf{1}_n\mathbf{1}_n^\top$ is cp_n, so we have $\tilde{\gamma}_n \geq cp_n$ and Assumption 6 does not hold. This corresponds to the case with strong correlation among the p_n components in the long run. We conjecture that the asymptotic independence between our L_2-type and L_∞-type SS–SN statistics does not hold for this case, which is confirmed in our unreported simulations.

3 Extensions to other testing problems

In this section, we generalise the SS–SN approach to testing for zero autocorrelation in a time series in Section 3.1, linear hypothesis testing in a regression model in Section 3.2 and testing for a change point in multivariate mean in Section 3.3. For simplicity, we only consider the SS–SN$_1$ statistic and similar results for SS–SN$_P$ statistic can be obtained in an analogous fashion.

3.1 Testing for zero autocorrelation

Let $\{X_t\}$ be a univariate stationary time series with mean $E(X_t) = \mu$. Testing for white noise is an important problem in time series analysis and there is a rich literature; see Horowitz et al. (2006), Shao (2011a, 2011b), and Liu et al. (2022) and cited references therein. To be specific, we shall test the null hypothesis $H_0{:}r_1 = r_2 = \cdots = r_p = 0$ against $H_A{:}r_i \neq 0$ for some $i = 1, 2, \ldots, p$, where p is a positive integer and $r_i = E[(X_t - \mu)(X_{t+i} - \mu)]$ is the autocovariance at lag i.

We now apply the SS–SN approach to test H_0. Define $\mathbf{Z}_t = (Z_t^1, Z_t^2, \ldots, Z_t^p)^\top$ and $\hat{\mathbf{Z}}_t = (\hat{Z}_t^1, \hat{Z}_t^2, \ldots, \hat{Z}_t^p)^\top$ where $Z_t^i = (X_t - \mu)(X_{t+i} - \mu)$, $\hat{Z}_t^i = (X_t - \bar{X}_n)(X_{t+i} - \bar{X}_n)$ and $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. The null hypothesis is equivalent to the hypothesis that $\{\mathbf{Z}_t\}$ is a p-dimensional mean zero stationary time series. We prove the following proposition about the FCLT for $\{\hat{\mathbf{Z}}_t\}$.

Proposition 2 If Assumptions 1 and 2 hold for $\{X_t\}$ and $\{\mathbf{Z}_t\}$, then (i) Assumption 1 also holds for $\{\hat{\mathbf{Z}}_t\}$ and (ii) for $i = 1, 2, \ldots, p$, the sample variance of $\{\hat{Z}_t^i\}$ converges in probability to the variance of Z_t^i.

By Proposition 2, we can use $\{\hat{\mathbf{Z}}_t\}$ to construct a similar test statistic as in equation (3) to test the zero autocorrelation hypothesis. The asymptotic property of this statistic is stated in the following proposition.

Proposition 3 Suppose Assumption 1 holds for $\{X_t\}$ and $\{\mathbf{Z}_t\}$ and the δth moment of $|X_t|$ is finite for some $\delta > 2$. Define the test statistic $T_n^{(A)}(\alpha, \hat{\jmath}) = T_{n'}^{(M)}(\alpha, \hat{\jmath})$ according to equations (3) and (2), with $\{\mathbf{X}_t\}$ replaced by $\{\hat{\mathbf{Z}}_t\}$, $\boldsymbol{\mu}_0$ replaced by $\mathbf{0}$ and n replaced by $n' = n - p$, then we have under H_0

$$T_n^{(A)}(\alpha, \hat{\jmath}) \xrightarrow{\mathcal{D}} U_1. \tag{17}$$

At level ζ, we reject H_0 if $T_n^{(A)}(\alpha, \hat{\jmath}) > U_{1,\zeta}$. In Section 4.2, we show that our test has accurate size even when the white noise process is not independent over time (i.e. contains higher order dependence) and when the sample size is small.

3.2 Testing linear hypotheses in a regression model

Kiefer et al. (2000) pioneered the bandwidth-free test for general linear hypotheses of the parameters in a time series regression model. We now show that SS–SN method is also applicable to their setting. Consider the regression model

$$y_t = \mathbf{X}_t^\top \boldsymbol{\beta} + \epsilon_t, \quad t = 1, 2, \ldots, n, \tag{18}$$

where $\boldsymbol{\beta}$ is a p-dimensional parameter, $\mathbf{X}_t$ is a p-dimensional regressor and ϵ_t is a mean zero (conditional on $\mathbf{X}_t$) random process. Let $\mathbf{v}_t = \mathbf{X}_t\epsilon_t$ and $\boldsymbol{\Omega} = \sum_{k=-\infty}^{\infty} E(\mathbf{v}_t\mathbf{v}_{t+k}^\top)$, we assume that the following condition from Kiefer et al. (2000) holds.

Assumption 7 (i) $n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor}\mathbf{v}_t \Rightarrow \mathbf{\Omega}^{1/2}\mathbf{B}_p(r)$ on $D^p[0,1]$.

(ii) $1/n\sum_{t=1}^{\lfloor nr\rfloor}\mathbf{X}_t\mathbf{X}_t^\top \xrightarrow{p} r\mathbf{Q}$ for all $r\in[0,1]$ and $\mathbf{Q}^{-1}$ exists.

Suppose we are interested in testing H_0:$\mathbf{R}\boldsymbol{\beta}=\mathbf{s}$ against H_A:$\mathbf{R}\boldsymbol{\beta}\neq\mathbf{s}$, where $\mathbf{s}\in\mathbb{R}^d$ and $\mathbf{R}$ is a $(d\times p)$ matrix of rank d. The test statistic proposed by Kiefer et al. (2000) is $A_n=n(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{s})^\top\mathbf{V}_n^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}-\mathbf{s})/d$, where

$$\mathbf{V}_n=\mathbf{R}\left(\frac{1}{n}\sum_{t=1}^{n}\mathbf{X}_t\mathbf{X}_t^\top\right)^{-1}\left[\frac{1}{n^2}\sum_{k=1}^{n}\left(\sum_{t=1}^{k}\mathbf{X}_t\hat{\epsilon}_t\right)\left(\sum_{t=1}^{k}\mathbf{X}_t\hat{\epsilon}_t\right)^\top\right]\left(\frac{1}{n}\sum_{t=1}^{n}\mathbf{X}_t\mathbf{X}_t^\top\right)^{-1}\mathbf{R}^\top,$$

$\hat{\boldsymbol{\beta}}$ is the OLS (Ordinary Least Squares) estimator of $\boldsymbol{\beta}$ and $\hat{\epsilon}_t$ is the residual. The limiting null distribution of A_n is U_d/d. As shown in Kiefer et al. (2000), the size distortion for their test $\mathbf{1}(A_n>U_{d,\zeta}/d)$ increases as d increases, and our SS–SN$_1$ test tackles this problem by focusing on the single hypothesis among d hypotheses which deviates most from the null. Let $\hat{\boldsymbol{\beta}}_{a:b}$ be the OLS estimator of $\boldsymbol{\beta}$ based on $(y_t,\mathbf{X}_t^\top)$ for $t=a,a+1,\ldots,b$ and define

$$\hat{j}^{(R)}=\underset{j=1,2,\ldots,d}{\operatorname{argmax}}\frac{n[\mathbf{e}_j^\top(\mathbf{R}\hat{\boldsymbol{\beta}}_{1:\lfloor n\alpha\rfloor}-\mathbf{s})]^2}{\lfloor n\alpha\rfloor^{-1}\sum_{t=1}^{\lfloor n\alpha\rfloor}(g_t^j-\frac{1}{\lfloor n\alpha\rfloor}\sum_{k=1}^{\lfloor n\alpha\rfloor}g_k^j)^2},$$

where $(g_t^1,\ldots,g_t^d)^\top=\mathbf{R}(\frac{1}{\lfloor n\alpha\rfloor}\sum_{t=1}^{\lfloor n\alpha\rfloor}\mathbf{X}_t\mathbf{X}_t^\top)^{-1}\mathbf{X}_t(y_t-\mathbf{X}_t^\top\hat{\boldsymbol{\beta}}_{1:\lfloor n\alpha\rfloor})$ for $t=1,2,\ldots,\lfloor n\alpha\rfloor$. So $\hat{j}^{(R)}$ represents the coordinate of $\mathbf{R}\boldsymbol{\beta}-\mathbf{s}$ that deviates most from $\mathbf{0}$ at the sample level. The following assumption on $(g_t^1,\ldots,g_t^d)^\top$ is needed to derive the asymptotic properties of our test statistic.

Assumption 8 For $j=1,2,\ldots,d$, $\lfloor n\alpha\rfloor^{-1}\sum_{t=1}^{\lfloor n\alpha\rfloor}(g_t^j-\frac{1}{\lfloor n\alpha\rfloor}\sum_{k=1}^{\lfloor n\alpha\rfloor}g_k^j)^2\xrightarrow{p}\Upsilon_j>0$.

We then find the OLS $\hat{\boldsymbol{\beta}}_{\lfloor n\alpha\rfloor+1:n}$ on the second part of the data and the residual is $\hat{\mu}_t=y_t-\mathbf{X}_t^\top\hat{\boldsymbol{\beta}}_{\lfloor n\alpha\rfloor+1:n}$. Define

$$\hat{\mathbf{s}}_t=(\hat{s}_t^1,\hat{s}_t^2,\ldots,\hat{s}_t^d)=\mathbf{R}\left(\frac{1}{n-\lfloor n\alpha\rfloor}\sum_{t=\lfloor n\alpha\rfloor+1}^{n}\mathbf{X}_t\mathbf{X}_t^\top\right)^{-1}\mathbf{X}_t\hat{\mu}_t,$$

and $\tilde{S}_{a,b}^j=\sum_{t=a}^{b}\hat{s}_t^j$ for $\lfloor n\alpha\rfloor+1\le a\le b\le n$, $j=1,\ldots,d$. Our test statistic can be defined as

$$T_n^{(R)}(\alpha,\hat{j}^{(R)})=\frac{(n-\lfloor n\alpha\rfloor)\mathbf{e}_{\hat{j}^{(R)}}^\top(\mathbf{R}\hat{\boldsymbol{\beta}}_{\lfloor n\alpha\rfloor+1:n}-\mathbf{s})(\mathbf{R}\hat{\boldsymbol{\beta}}_{\lfloor n\alpha\rfloor+1:n}-\mathbf{s})^\top\mathbf{e}_{\hat{j}^{(R)}}}{V_n^{(R)}(\hat{j}^{(R)})},\tag{19}$$

where $V_n^{(R)}(j)=(n-\lfloor n\alpha\rfloor)^{-2}\sum_{k=\lfloor n\alpha\rfloor+1}^{n}(\tilde{S}_{\lfloor n\alpha\rfloor+1,k}^j-\frac{k-\lfloor n\alpha\rfloor}{n-\lfloor n\alpha\rfloor}\tilde{S}_{\lfloor n\alpha\rfloor+1,n}^j)^2$. The following proposition shows the asymptotic property of $T_n^{(R)}(\alpha,\hat{j}^{(R)})$.

Proposition 4 Suppose Assumption 7 holds, then (i) under H_0, we have

$$T_n^{(R)}(\alpha,\hat{j}^{(R)})\xrightarrow{\mathcal{D}}U_1.\tag{20}$$

(ii) Under H_A, denote $\|\mathbf{R}\boldsymbol{\beta}-\mathbf{s}\|_\infty=\max_{j=1,2,\ldots,d}|\mathbf{R}_j^\top\boldsymbol{\beta}-s_j|$ where $\mathbf{R}_j^\top$ is the jth row of $\mathbf{R}$. Let the jth row of $\mathbf{R}\mathbf{Q}^{-1}\mathbf{\Omega}^{1/2}$ be $\mathbf{h}_j^\top\in\mathbb{R}^p$, then we have

1. If $\sqrt{n}\|\mathbf{R}\boldsymbol{\beta}-\mathbf{s}\|_\infty\to\infty$, then $T_n^{(R)}(\alpha,\hat{j}^{(R)})\xrightarrow{p}\infty$, thus the limiting power is 1.

2. If $\sqrt{n}(\mathbf{R}\boldsymbol{\beta}-\mathbf{s}) \to \mathbf{c}=(c^1, c^2, \ldots, c^d)^\top$ and $\|\mathbf{c}\|_\infty \neq 0$, then we have

$$\hat{j}^{(R)} \xrightarrow{\mathcal{D}} \underset{j=1,2,\ldots,d}{\operatorname{argmax}} \frac{\{\tilde{B}^{(j)}(\alpha)+\alpha c^j\}^2}{\mathrm{Y}_j} \overset{d}{=} j^{(R)} \quad \text{and} \quad T_n^{(R)}(\alpha, \hat{j}^{(R)}) \xrightarrow{\mathcal{D}} U^{(R)},$$

where $\tilde{B}^{(j)}(r)=\mathbf{h}_j^\top \mathbf{B}_p(r)$ are mean zero Brownian motions with covariance $Cov(\tilde{B}^{(i)}(u), \tilde{B}^{(j)}(v))=\min\{u, v\}\mathbf{h}_i^\top \mathbf{h}_j$ and the conditional distribution of $U^{(R)}$ given $j^{(R)}=j$ is

$$U^{(R)}\Big|_{j^{(R)}=j} \overset{d}{=} \frac{\left\{B(1)+\sqrt{\dfrac{1-\alpha}{\mathbf{h}_j^\top \mathbf{h}_j}}c^j\right\}^2}{\int_0^1 \left\{B(r)-rB(1)\right\}^2 \mathrm{d}r}.$$

3. If $\sqrt{n}\|\mathbf{R}\boldsymbol{\beta}-\mathbf{s}\|_\infty \to 0$, then $T_n^{(R)}(\alpha, \hat{j}^{(R)}) \xrightarrow{\mathcal{D}} U_1$, so our test has trivial power asymptotically.

In Section 4.3, we show that our test has less size distortion, at the cost of a small loss of power, compared with the test used in Kiefer et al. (2000) when the number of restrictions under the null is moderate and strong autocorrelation is present in the data. We do not provide proofs for Propositions 3 and 4 since they are trivial in view of the proofs we provided for Theorems 1 and 2.

3.3 Testing for a change point in multivariate mean

Let $\mathbf{X}_t=(X_t^1, X_t^2, \ldots, X_t^p)^\top$ be a p-dimensional time series and let $E(\mathbf{X}_t)=\boldsymbol{\mu}_t:=(\mu_t^1, \mu_t^2, \ldots, \mu_t^p)^\top \in \mathbb{R}^p$. Suppose we want to test the null hypothesis H_0:$\boldsymbol{\mu}_1=\boldsymbol{\mu}_2=\cdots=\boldsymbol{\mu}_n$ against H_A:$\boldsymbol{\mu}_1=\cdots=\boldsymbol{\mu}_{k^*} \neq \boldsymbol{\mu}_{k^*+1}=\cdots=\boldsymbol{\mu}_n$ where $k^*=\lfloor nr_0 \rfloor$ for some unknown $r_0 \in (0, 1)$. As in Section 2.1, define the autocovariance matrix as $\boldsymbol{\Gamma}(k)=E[(\mathbf{X}_t-\boldsymbol{\mu}_t)(\mathbf{X}_{t+k}-\boldsymbol{\mu}_{t+k})^\top]$, and let $\boldsymbol{\Gamma}=\sum_{k=-\infty}^{\infty}\boldsymbol{\Gamma}(k)$ with (i, j) element being Γ_{ij}. The following assumption is needed in deriving the asymptotic distribution of our test statistic.

Assumption 9 Assume that (a)

$$n^{-1/2}\sum_{t=1}^{\lfloor nr \rfloor}(\mathbf{X}_t-\boldsymbol{\mu}_t) \Rightarrow \boldsymbol{\Gamma}^{1/2}\mathbf{B}_p(r) \text{ on } D^p[0, 1]. \tag{21}$$

(b) $\lfloor nb \rfloor+1 \leq k^* \leq n-\lfloor nb \rfloor-1$ for some $b \in (0, 0.5)$

Under this assumption, the change point cannot lie in the first and last $\lfloor nb \rfloor$ sample points. Here, b is usually called a trimming parameter, see Andrews (1993). For $k=1, 2, \ldots, \lfloor nb \rfloor$, define $\boldsymbol{W}_{1,k}=(W_{1,k}^1, \ldots, W_{1,k}^p)^\top=\mathbf{X}_k-\mathbf{X}_{n-\lfloor nb \rfloor+k}$, $\mathbf{M}_{1,k}=(M_{1,k}^1, \ldots, M_{1,k}^p)^\top=\sum_{t=1}^{k}\mathbf{W}_{1,t}$ and $\vartheta_j^2=Var(W_{1,1}^j)$, $\hat{\vartheta}_j^2=\lfloor nb \rfloor^{-1}\sum_{t=1}^{\lfloor nb \rfloor}(W_{1,t}^j-\frac{1}{\lfloor nb \rfloor}\sum_{k=1}^{\lfloor nb \rfloor}W_{1,k}^j)^2$ for $j=1, 2, \ldots, p$. We use the difference between the first and last $\lfloor nb \rfloor$ points of the data to find the coordinate of $\mathbf{X}_t$ that has the strongest signal of a mean change, then we apply the SN test statistic used in Shao and Zhang (2010). To be specific, define

$$\hat{j}=\underset{j\in\{1,2,\ldots,p\}}{\operatorname{argmax}} \frac{n^{-1}[\mathbf{e}_j^\top \boldsymbol{M}_{1,\lfloor nb \rfloor}]^2}{\hat{\vartheta}_j^2}.$$

Then we apply the SN-based test statistic used in Shao and Zhang (2010) on $\{X_{t,\hat{j}}\}_{t=\lfloor nb\rfloor+1}^{n-\lfloor nb\rfloor}$. Define $S_{a,b}^{j}=\sum_{t=a}^{b}X_t^j$. Let

$$G_n=\sup_{k=\lfloor nb\rfloor+1,\lfloor nb\rfloor+2,\dots,n-\lfloor nb\rfloor-1}\frac{T_n(\hat{j},k)^2}{V_n(\hat{j},k)},$$

where

$$T_n(j,k)=\frac{1}{\sqrt{n-2\lfloor nb\rfloor}}\sum_{t=\lfloor nb\rfloor+1}^{k}\left(X_t^j-\frac{S_{\lfloor nb\rfloor+1,n-\lfloor nb\rfloor}^{j}}{n-2\lfloor nb\rfloor}\right),$$

$$V_n(j,k)=\frac{1}{(n-2\lfloor nb\rfloor)^2}\left[\sum_{t=\lfloor nb\rfloor+1}^{k}\left\{S_{\lfloor nb\rfloor+1,t}^{j}-\frac{t-\lfloor nb\rfloor}{k-\lfloor nb\rfloor}S_{\lfloor nb\rfloor+1,k}^{j}\right\}^2\right.$$
$$\left.+\sum_{t=k+1}^{n-\lfloor nb\rfloor}\left\{S_{t,n-\lfloor nb\rfloor}^{j}-\frac{n-\lfloor nb\rfloor-t+1}{n-\lfloor nb\rfloor-k}S_{k+1,n-\lfloor nb\rfloor}^{j}\right\}^2\right].$$

The following theorem shows the asymptotic properties of G_n under the null and alternative.

Theorem 8 Suppose Assumption 9 holds and $\hat{\vartheta}_j^2\xrightarrow{p}\vartheta_j^2>0$ for $j=1,2,\dots,p$, then (i) under H_0, we have

$$G_n\xrightarrow{\mathcal{D}}G\overset{d}{=}\sup_{r\in[0,1]}\frac{\left\{B(r)-rB(1)\right\}^2}{\int_0^r\left\{B(s)-\frac{s}{r}B(r)\right\}^2ds+\int_r^1\left\{B(1)-B(s)-\frac{1-s}{1-r}(B(1)-B(r))\right\}^2ds},\tag{22}$$

(ii) under H_A, denote $\mathbf{\Delta}_n=(\Delta_n^1,\Delta_n^2,\dots,\Delta_n^p)^\top=E(\mathbf{X}_{k^*+1})-E(\mathbf{X}_{k^*})$ and $\|\mathbf{\Delta}_n\|_\infty=\max_{j=1,2,\dots,p}|\Delta_n^p|$, we have

1. If $\sqrt{n}\|\mathbf{\Delta}_n\|_\infty\to\infty$, then $G_n\xrightarrow{p}\infty$, thus the limiting power of the level ζ test $\mathbf{1}(G_n>G_\zeta)$ for $\zeta\in(0,1)$ is 1, where G_ζ is the $100(1-\zeta)$th upper percentile of G.
2. If $\sqrt{n}\mathbf{\Delta}_n\to\mathbf{c}:=(c^1,c^2,\dots,c^p)\in\mathbb{R}^p$ and $\|\mathbf{c}\|_\infty\neq0$, then we have

$$\hat{j}\xrightarrow{\mathcal{D}}\operatorname*{argmax}_{j\in\{1,2,\dots,p\}}\frac{\left[B^{(j)}(b)-\left(B^{(j)}(1)-B^{(j)}(1-b)\right)-bc^j\right]^2}{\vartheta_j^2}\overset{d}{=}j^*,$$
$$G_n\xrightarrow{\mathcal{D}}G^*,$$

where $B^{(j)}(r)$ are mean zero Brownian motions with covariance $Cov(B^{(i)}(u),B^{(j)}(v))=2\min\{u,v\}\Gamma_{ij}$. The conditional distribution of G^* given $j^*=j$ is

$$G^*|_{j^*=j}\overset{d}{=}\sup_{r\in[0,1]}\frac{\left\{B'(r)-rB'(1)\right\}^2}{\int_0^r\left\{B'(s)-\frac{s}{r}B'(r)\right\}^2ds+\int_r^1\left\{B'(1)-B'(s)-\frac{1-s}{1-r}(B'(1)-B'(r))\right\}^2ds},\tag{23}$$

where the process $B'(r)$ is defined as $B'(r) = B(r) + \frac{1}{\sqrt{1-2b}} H_j((1-2b)r + b)$ with $H_j = \frac{1}{\sqrt{\Gamma_{jj}}} \mathbf{e}_j^\top \mathbf{H}(r)$ and $\mathbf{H}(r) = (r - r_0)\mathbf{c}\mathbf{1}_{r \geq r_0}$ where $\mathbf{1}_{r \geq r_0} = 1$ if $r \geq r_0$ and 0 otherwise.

3. If $\sqrt{n}\|\boldsymbol{\Delta}_n\|_\infty \to 0$, then

$$G_n \xrightarrow{\mathcal{D}} G,$$

so our test has trivial power asymptotically.

Interestingly, the limiting null distribution G is pivotal and is identical to the one for the SN test in Shao and Zhang (2010), who has tabulated its critical values. In Section 4.4, we show that our test has substantially smaller size distortion than the test used in Shao and Zhang (2010) when the dimension of $\mathbf{X}_t$ is moderate and there is strong autocorrelation in the data.

4 Simulation studies

In this section, we examine the size and power properties of our SS–SN test statistics in finite sample. Specifically in Section 4.1, we examine the empirical size and power of our test statistics in testing hypotheses on multivariate mean and compare with the traditional SN statistic proposed in Lobato (2001). In Section 4.2, we show the favourable size performance of our test statistics when testing for uncorrelation in a univariate time series. In Sections 4.3 and 4.4, we present the size and size-adjusted power of our test statistics for testing linear hypotheses in a regression model and the existence of a change point in multivariate mean, respectively.

4.1 Finite sample size and power for multivariate mean tests

In this subsection, we examine the empirical size and power of our test statistics in testing hypotheses on multivariate mean. Under the null, we assume the data comes from the following VAR(1) model: $\mathbf{X}_t = \rho \mathbf{I}_p \mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t$, where $\boldsymbol{\epsilon}_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_p)$. We set the nominal level at 5%. The experiment is repeated 5,000 times with the length of time series $n \in \{100, 300\}$, $\rho \in \{-0.7, -0.5, 0.2, 0.5, 0.7\}$ and $p \in \{5, 10\}$. We compare the empirical sizes for $T_n^{(M)}(\alpha, \hat{j})$ (denoted as SS–SN$_1$), $Q_n^{(M)}(\alpha)$ (denoted as SS–SN$_P$), their Bonferroni combination when $\alpha = 0.5$ (denoted as SS–SN$_b$), and the test statistic used in Lobato (2001) (denoted as Lobato) for different combinations of n, p, and ρ. As Table 1 shows, SS–SN$_1$, SS–SN$_P$, and SS–SN$_b$ have more accurate size than Lobato when $|\rho|$ is close to 1 and the sizes for SS–SN$_1$ and SS–SN$_p$ are very similar, while the size of SS–SN$_b$ is often slightly more distorted compared with these two. The distortion for Lobato gets more severe when p increases from 5 to 10, whereas for our tests the impact of the dimension on the size is minimal. When we increase the sample size from $n = 100$ to 300, we see noticeable improvements in size distortion for all tests. For both SS–SN$_1$ and SS–SN$_P$, the choice of α seems to have little impact on the size distortion and no particular value of α dominates others in size accuracy. Furthermore, our SS–SN tests exhibit more size stability across the range of ρs as compared to Lobato, especially at $n = 300$. This stability, which is achieved by dimension reduction step involved in the SS–SN procedure, is attractive since in practice the amount of temporal dependence is usually unknown.

For the size-adjusted power, we generate the data from the process: $\mathbf{X}_t - \mu\mathbf{e}_1 = \rho\mathbf{I}_p(\mathbf{X}_{t-1} - \mu\mathbf{e}_1) + \boldsymbol{\epsilon}_t$ under the sparse alternative and from the process: $\mathbf{X}_t - \mu\mathbf{1} = \rho\mathbf{I}_p(\mathbf{X}_{t-1} - \mu\mathbf{1}) + \boldsymbol{\epsilon}_t$ under the dense alternative, where $\boldsymbol{\epsilon}_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_p)$. We set $n = 300$, $\rho = 0.2$, $p = 10$ and the experiment is repeated 2,000 times at nominal level 5%. As Figure 3a and b shows, SS–SN$_1$ has relatively larger power loss than SS–SN$_P$, as compared with Lobato under the dense alternative. The power loss is relatively smaller under sparse alternative and in this case SS–SN_1 outperforms SS–SN$_P$. Note that under both dense and sparse alternatives, the power curve of SS–SN$_b$ is close to that of the SS–SN statistic which performs better. Hence, the SS–SN$_b$ can have good all-round power against both types of alternatives. The power loss of SS–SN$_b$ relative to Lobato is moderate, but its gain in size stability and accuracy can be substantial, especially when p is moderate and temporal

Table 1. Empirical rejection rate (in percentage) under the null when testing hypothesis on multivariate mean

n	p	ρ	SS–SN$_1$			SS–SN$_P$			SS–SN$_b$	Lobato
			$\alpha=0.15$	$\alpha=0.3$	$\alpha=0.5$	$\alpha=0.15$	$\alpha=0.3$	$\alpha=0.5$		
100	5	−0.7	3.22	2.94	2.20	2.96	2.72	2.40	1.52	0.62
		−0.5	3.88	3.84	3.38	4.02	3.80	3.48	2.60	1.88
		0.2	6.00	5.72	5.72	5.40	5.90	5.38	5.14	7.08
		0.5	6.56	6.94	7.14	6.28	6.78	7.12	6.54	11.90
		0.7	7.92	8.62	9.40	7.46	8.24	9.68	10.12	21.04
	10	−0.7	3.10	2.92	2.10	2.80	2.92	2.28	1.44	0.06
		−0.5	3.80	3.88	2.94	3.84	3.73	3.42	2.42	0.64
		0.2	5.20	5.70	5.52	5.18	5.50	5.48	5.16	10.42
		0.5	6.02	6.92	7.44	6.18	7.08	7.34	7.42	26.00
		0.7	7.38	8.42	10.48	7.16	8.78	9.52	10.62	52.78
300	5	−0.7	4.34	3.80	3.74	4.04	3.62	3.84	2.84	2.26
		−0.5	4.50	4.10	4.48	4.34	4.18	4.30	3.64	3.28
		0.2	5.12	4.52	5.32	4.98	4.96	5.18	4.98	6.18
		0.5	5.22	4.74	5.56	5.44	5.20	5.92	5.54	7.18
		0.7	5.82	5.54	6.32	5.82	5.96	6.72	6.38	9.42
	10	−0.7	4.64	3.92	4.06	3.84	3.86	3.92	3.36	0.56
		−0.5	4.92	4.42	4.60	4.22	4.36	4.16	4.08	2.18
		0.2	5.28	5.40	5.06	4.98	5.00	5.12	4.88	6.42
		0.5	5.62	5.74	5.52	5.46	5.50	5.60	5.54	10.68
		0.7	5.96	6.22	6.46	5.90	6.04	6.38	6.28	18.34

dependence is strong. We also tried other settings (e.g. $n=300, \rho=0.7, p=10$) for the size-adjusted power and the results are quantitatively similar so are skipped.

4.2 Finite sample size for testing zero autocorrelation

In this subsection, we present the empirical size of $T_n^{(A)}(\alpha, \hat{j})$ statistic (denoted as SS–SN$_1$), its L_2-type counterpart (denoted as SS–SN$_P$), and their Bonferroni combination (denoted as SS–SN$_b$) in testing zero autocorrelation at nominal level 5%. Under the null hypothesis, we assume the data comes from the same models used in Lobato (2001). Let $u_t \overset{i.i.d}{\sim} N(0,1)$ and the eight models are (a) $i.i.dN(0,1)$; (b) $t(6)$; (c) demeaned standard log normal; (d) 1-dependent process $X_t = u_t u_{t-1}$; (e) the heteroscedastic process $X_t = s_t u_t u_{t-1}$, where s_t is the infinite repetition of the sequence $\{1, 1, 1, 2, 3, 1, 1, 1, 1, 2, 4, 6\}$; (f) the uncorrelated non-martingale difference process $X_t = u_{t-2}u_{t-1}(u_{t-2} + u_t + 1)$; (g) the GARCH(1,1) model $X_t = \delta_t u_t$, where $\delta_t^2 = 0.001 + 0.02X_{t-1}^2 + 0.8\delta_{t-1}^2$; (h) the bilinear model $X_t = u_t = 0.5u_{t-1}X_{t-2}$. The experiment is repeated 5,000 times with $n \in \{100, 500\}$ and the results are shown in Table 2. The size for our tests and the Ljung-Box test are close to the nominal level for N(0,1), t(6) and GARCH(1) models, while Lobato test is severely undersized when $n=100$ or $p=20$. For the bilinear model, our test statistics have accurate size, while Ljung-Box test is oversized. For LogNormal model, SS–SN$_1$ and Ljung-Box have slightly more accurate size than SS–SN$_P$ and the Lobato test is noticeably undersized for all cases. For the RT, Hetero, and No-MDS models, Lobato and Ljung-Box test both have severe size distortion, while our tests are mildly undersized. In addition, for different splitting ratio α and p, the size for our two SS–SN tests does not change much. Overall it is fair to say that our SS–SN tests have the most accurate and stable sizes across all DGPs. Note that the size for SS–SN$_b$ is generally slightly more distorted than SS–SN$_1$ and SS–SN$_P$ but the difference is small.

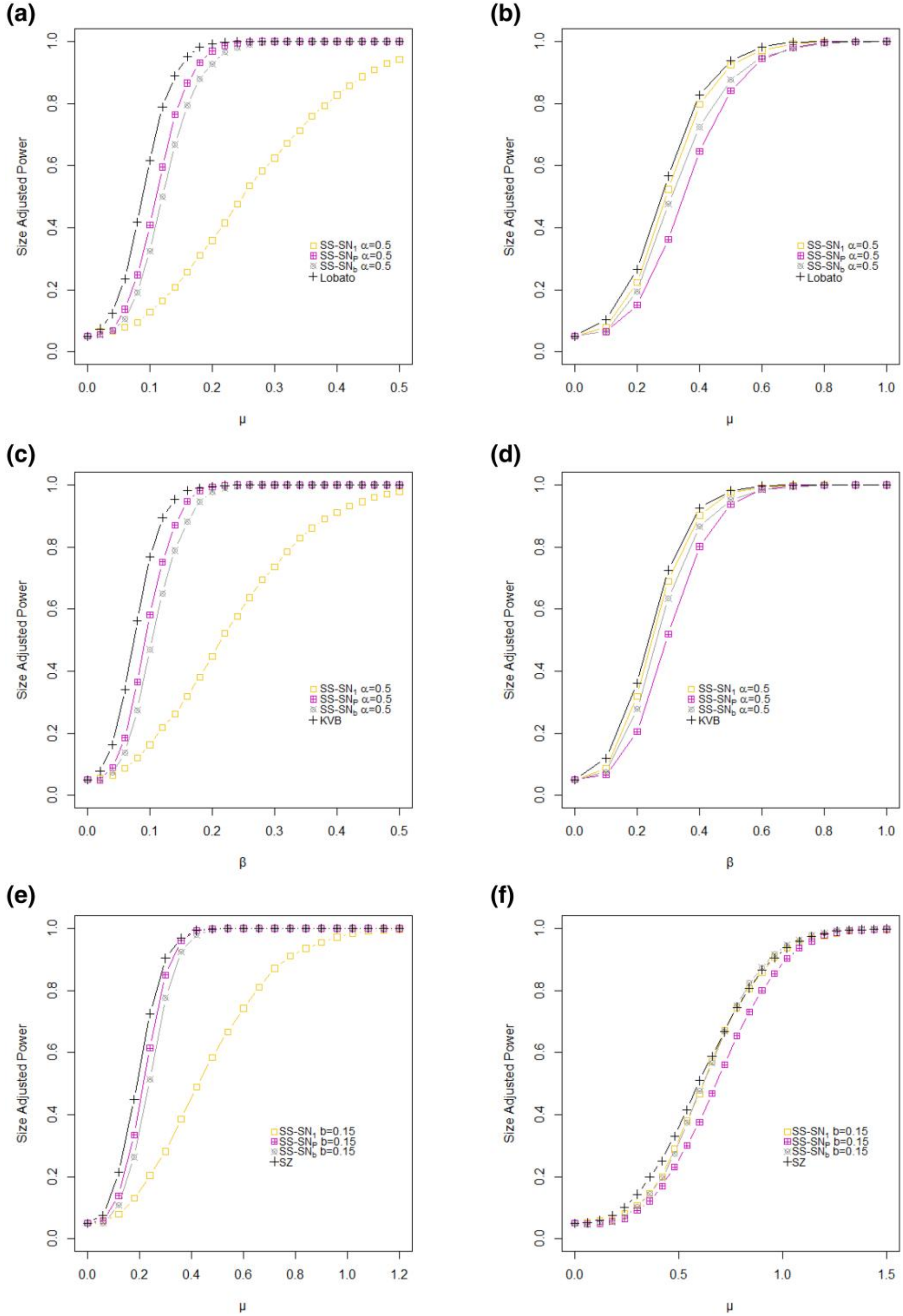

Figure 3. Size-adjusted power for testing hypothesis on multivariate mean (first row), in a regression model (second row) and the existence of a change point (last row) under the dense (left column) and sparse (right column) alternatives.

Table 2. Empirical rejection rate (in percentage) under the null with different models when testing zero autocorrelation

Model	n	p	SS–SN$_1$			SS–SN$_P$			SS–SN$_b$	Lobato	Ljung-Box
			$\alpha = 0.15$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.15$	$\alpha = 0.3$	$\alpha = 0.5$			
$N(0,1)$	100	5	4.90	4.60	4.36	4.46	4.60	4.58	3.30	4.16	5.48
		10	4.40	4.42	4.26	4.70	4.66	3.70	3.52	1.46	5.24
		20	4.63	4.68	4.50	4.28	4.32	4.46	4.22	0.66	4.76
	500	5	4.72	5.40	5.16	4.72	4.50	4.68	4.56	5.20	4.96
		10	4.96	5.32	4.78	4.62	4.60	4.82	4.58	5.50	5.12
		20	5.18	5.30	4.48	4.66	4.90	4.96	4.58	2.98	4.52
$t(6)$	100	5	4.52	3.88	3.80	4.42	3.90	3.74	2.90	2.44	4.76
		10	4.58	3.74	4.36	4.38	4.20	3.72	3.62	0.80	4.44
		20	4.70	4.40	4.42	4.28	4.16	4.92	4.28	0.48	4.58
	500	5	4.78	4.80	4.90	4.52	4.52	4.68	3.84	5.04	4.24
		10	4.80	4.84	4.86	4.94	4.98	5.12	4.36	4.24	5.56
		20	4.32	4.42	4.66	4.18	4.26	4.28	4.56	2.12	4.98
LogNormal	100	5	3.78	3.36	3.14	2.98	2.88	2.62	2.46	0.62	3.14
		10	3.76	3.96	3.82	3.22	3.00	2.92	3.16	0.24	3.58
		20	4.16	4.40	4.16	3.30	3.68	3.74	3.34	0.12	3.58
	500	5	4.68	5.16	4.74	3.98	4.22	4.16	3.42	3.06	4.10
		10	4.32	5.38	4.28	3.76	4.20	3.70	3.58	1.54	4.08
		20	4.70	4.88	4.50	3.72	3.76	3.62	3.64	0.26	4.12
RT	100	5	3.40	3.30	2.64	3.04	3.44	2.52	1.50	0.86	21.34
		10	2.80	2.94	2.58	2.88	2.92	2.64	1.88	0.12	21.90
		20	3.28	3.30	2.94	3.46	3.60	3.44	2.62	0.02	21.94
	500	5	4.62	4.46	4.44	4.70	4.40	4.08	3.58	3.86	24.10
		10	4.62	4.18	4.10	4.84	4.22	4.40	3.54	1.60	24.58
		20	4.60	4.28	4.00	4.66	4.62	4.34	3.62	0.38	24.08

(*continued*)

Table 2. Continued

Model	n	p	SS–SN$_1$			SS–SN$_p$			SS–SN$_b$	Lobato	Ljung-Box
			$\alpha = 0.15$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.15$	$\alpha = 0.3$	$\alpha = 0.5$			
Hetero	100	5	2.80	2.58	1.92	2.56	2.48	1.66	1.12	0.48	26.06
		10	3.00	2.58	2.72	2.54	2.28	2.26	1.88	0.16	25.94
		20	3.10	3.20	2.96	3.32	3.06	3.86	2.74	0.10	25.48
	500	5	4.06	3.64	3.46	4.46	4.16	3.14	2.56	2.80	32.48
		10	4.06	4.14	3.64	4.26	4.06	4.12	2.82	1.08	32.52
		20	3.78	3.92	3.30	4.28	3.68	3.54	2.98	0.24	31.92
No-MDS	100	5	2.02	2.34	1.54	1.90	1.98	1.80	1.08	0.30	27.26
		10	1.86	1.82	2.38	2.10	2.04	1.76	1.44	0.02	28.70
		20	1.96	2.12	2.48	2.26	2.54	2.50	2.04	0.06	28.28
	500	5	3.56	3.54	3.18	3.86	3.58	3.20	2.32	2.32	38.42
		10	3.22	3.68	3.22	3.82	3.32	3.30	2.48	0.48	39.22
		20	3.20	3.40	3.32	3.54	3.20	3.08	2.58	0.06	39.60
GARCH(1)	100	5	4.52	4.70	4.08	5.06	4.56	4.28	3.66	3.68	5.66
		10	4.46	4.22	3.96	4.30	4.66	4.08	3.52	1.44	5.34
		20	4.78	4.66	4.32	4.76	4.88	4.32	4.24	0.60	5.06
	500	5	4.72	4.94	4.70	4.38	5.06	4.86	4.02	5.02	4.98
		10	5.22	5.00	4.90	4.72	5.00	5.00	4.76	4.96	5.62
		20	5.30	4.84	4.54	4.40	4.90	4.66	4.56	2.66	5.58
Bilinear	100	5	5.32	4.44	4.08	4.94	3.88	3.70	2.88	2.90	12.66
		10	5.18	4.58	3.84	4.08	3.68	3.86	3.24	1.00	12.16
		20	4.44	4.50	3.62	4.42	4.38	3.88	3.40	0.30	13.10
	500	5	5.16	4.34	4.86	4.90	5.44	4.64	4.02	6.18	14.32
		10	4.70	4.70	5.26	5.10	4.80	4.84	4.22	4.24	14.22
		20	4.90	4.96	4.94	4.62	3.96	4.50	4.36	2.20	14.54

4.3 Finite sample size and power for testing linear hypotheses in a regression model

In this subsection, we report the result of a simulation experiment to compare the finite sample size and power of the statistic $T_n^{(R)}(\alpha, \hat{j}^{(R)})$ (denoted as SS–SN$_1$) defined in equation (19), its L_2-type counterpart (denoted as SS–SN$_P$), their Bonferroni combination (denoted as SS–SN$_b$), and the test statistic used in Kiefer et al. (2000) (denoted as KVB). For $p \in \{5, 10, 20\}$, $n \in \{300, 600\}$ and $\rho \in \{-0.7, -0.5, 0.2, 0.5, 0.7\}$, we assume the data is generated from the following model

$$y_t = \sum_{i=1}^{p} X_t^i \beta_i + \epsilon_t, \quad t = 1, 2, \ldots, n,$$

where $\{X_t^i\}$ and $\{\epsilon_t\}$ come from $(p+1)$ independent AR(1) processes $\eta_t = \rho\eta_t + e_t$ with $e_t \overset{i.i.d}{\sim} N(0, 1-\rho^2)$ so that the marginal distribution of η_t is $N(0, 1)$. The null hypothesis is $H_0{:}\beta_1 = \beta_2 = \cdots = \beta_p = 0$. The empirical rejection rate based on 5,000 Monte Carlo replications under H_0 is shown in Table 3. In general SS–SN$_1$, SS–SN$_P$, SS–SN$_b$, and KVB have relatively accurate size when $p = 5$ and $\rho = 0.2$, i.e. when the dimension and temporal dependence is small, and are oversized for other parameter combinations. When $n = 300$, as p and $|\rho|$ increases, the size distortion for KVB increases drastically, while the size distortion for SS–SN$_1$ and SS–SN$_P$ are small when $|\rho| < 0.7$ for all values of p. When $n = 600$, the size for all SS–SN statistics are less than 10% for all but one parameter combination, while KVB still have large size distortion when $p = 10, 20$ and $|\rho| = 0.7$. Overall, the improvement of size stability and accuracy across the dimension and range of ρs from KVB to SS–SN is apparent, and this is mainly due to the dimension reduction step in our SS–SN procedure.

Next, we examine the power of our SS–SN statistics under two alternative hypotheses. For the sparse alternative, we assume $\boldsymbol{\beta} = \beta \mathbf{e}_1^\top$ for some $\beta > 0$, so only the first component of $\boldsymbol{\beta}$ deviates from H_0. For the dense alternative, we assume $\boldsymbol{\beta} = \beta \mathbf{1}^\top$. We assume the same model as in Section 4.3 with $n = 300$, $p = 10$ and $\rho = 0.2$. We repeat the experiment 2,000 times and the curve for size-adjusted power against β for the dense and sparse alternatives is shown in Figure 3c and d. The findings here are qualitatively similar to those reported in Figure 3a and b. Under dense alternative, the power loss of SS–SN$_P$ compared with KVB is significantly smaller than that of SS–SN$_1$. Under sparse alternative, the power curve of SS–SN$_1$ is very close to that of KVB. Overall, we recommend the user to employ SS–SN$_b$, which achieves good all-round power and exhibits moderate power loss as compared to KVB under both alternatives.

4.4 Finite sample size and power for testing a change point in multivariate mean

In this subsection, we calculate the empirical size of our proposed tests in testing the existence of a change point in the mean of a VAR(1) process. As in previous simulations, the L_∞-type, L_2-type, and the Bonferroni combination are denoted as SS–SN$_1$, SS–SN$_P$, and SS–SN$_b$. Under the null hypothesis, we assume the data comes from the VAR(1) process $\mathbf{X}_t = \rho \mathbf{I}_p \mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t$, where $\boldsymbol{\epsilon}_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_p)$ and we set the trimming constant b in Assumption 9 to be 0.15, following the convention (Andrews, 1993). The experiment, with nominal level at 5%, is repeated 5,000 times with the length of time series $n \in \{300, 600\}$, $\rho \in \{-0.7, -0.5, 0.2, 0.5, 0.7\}$ and $p \in \{5, 10, 20\}$. We also calculate the empirical size for the test used in Shao and Zhang (2010) (denoted as SZ) and compare them under different combinations of n, p, and ρ.

As shown in Table 4, when $n = 300$, our tests are slightly undersized when $\rho = -0.7$ and oversized when $\rho = 0.7$, but the size distortion does not get worse as p increases, which is not the case for SZ. When $n = 600$, the sizes for our tests are more accurate than that for $n = 300$ and close to the nominal level uniformly over p and ρ. For SZ, the size also gets more accurate, but there is still large size distortion when p is large and $|\rho|$ is close to 1. Again SS–SN improves the size stability and accuracy across the dimension and ρs.

To examine the size-adjusted power, assume data $\mathbf{Y}_t$ comes from the following model:

$$\mathbf{Y}_t = \begin{cases} \mathbf{X}_t, & 1 \le t \le k_0 = \lfloor n r_0 \rfloor \\ \mathbf{X}_t + \mu & k_0 < t \le n, \end{cases}$$

Table 3. Empirical rejection rate (in percentage) under the null at level 5% in a regression model

n	p	ρ	SS–SN$_1$			SS–SN$_P$			SS–SN$_b$	KVB
			$\alpha = 0.15$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.15$	$\alpha = 0.3$	$\alpha = 0.5$		
300	5	−0.7	6.80	7.58	8.62	7.26	6.96	8.08	8.34	10.62
		−0.5	6.14	6.30	6.72	5.78	5.88	7.04	6.32	7.46
		0.2	5.50	5.34	6.00	5.54	5.38	5.90	5.84	5.60
		0.5	6.24	6.48	6.42	6.10	6.24	6.52	6.30	7.52
		0.7	6.80	8.04	8.58	6.76	7.60	8.40	7.90	10.28
	10	−0.7	7.58	8.80	9.58	7.90	8.62	10.50	11.00	20.18
		−0.5	6.60	7.32	7.68	6.72	6.96	7.78	7.74	11.56
		0.2	5.48	5.38	5.76	5.74	5.78	5.92	5.46	7.36
		0.5	6.34	6.30	7.28	6.54	6.52	7.44	7.22	11.72
		0.7	7.66	7.78	10.66	7.74	8.28	10.58	10.68	20.08
	20	−0.7	9.64	10.74	12.52	9.16	11.16	12.14	14.00	49.20
		−0.5	7.58	7.88	8.76	7.38	8.16	8.58	8.98	23.70
		0.2	6.38	7.00	6.90	6.00	6.00	6.88	7.34	11.58
		0.5	7.64	8.46	9.34	7.48	8.08	8.48	9.96	23.86
		0.7	10.40	10.72	12.76	9.98	11.30	13.08	14.12	50.32
600	5	−0.7	6.24	5.56	6.60	6.88	6.26	6.66	6.26	7.96
		−0.5	5.62	5.64	5.78	5.88	6.14	5.94	5.10	6.64
		0.2	5.06	5.64	5.10	5.30	5.02	5.04	4.66	5.52
		0.5	6.04	5.86	5.54	5.32	5.92	5.60	5.12	6.26
		0.7	5.80	6.36	6.28	5.54	6.48	5.54	5.56	7.70
	10	−0.7	6.92	6.80	7.62	6.40	6.50	7.28	7.70	12.66
		−0.5	6.14	6.28	6.48	5.96	5.96	6.48	6.62	8.64
		0.2	5.42	5.70	5.72	5.22	5.16	5.46	5.34	5.82
		0.5	5.72	5.48	5.40	5.50	6.18	6.22	5.90	7.76
		0.7	6.26	6.80	6.84	7.02	7.12	7.22	7.66	12.02
	20	−0.7	7.46	7.80	9.18	7.84	8.60	8.86	9.86	27.26
		−0.5	6.72	6.48	7.12	6.50	6.60	7.18	6.92	13.96
		0.2	5.24	5.58	6.18	5.28	5.90	6.40	5.78	8.48
		0.5	5.68	6.76	7.72	6.00	6.34	7.34	7.56	13.74
		0.7	6.42	8.50	9.10	8.02	7.36	8.96	10.10	26.74

where $\mathbf{X}_t$ is generated from the model in the null hypothesis. For the sparse alternative, we let $\boldsymbol{\mu} = \mu e_1$ and for the dense alternative, we let $\boldsymbol{\mu} = \mu\mathbf{1}$. We set $n = 300, \rho = 0.2, p = 10, r_0 = 1/2$ and the experiment is repeated 2,000 times. The results, as shown in Figure 3e and f, are qualitatively similar to those reported in Figure 3a–d. Note that under sparse alternative, the power curve of SS–SN$_1$ and SS–SN$_b$ are very close to that of SZ and even slightly outperforms SZ when μ is large. The use of SS–SN$_b$ is again recommended due to its overall good performance under both alternatives.

Based on the simulation results reported in Sections 4.1–4.4, we conclude that both SS–SN test statistics offer very stable and relatively accurate size across a wide range of data generating processes for most combinations of (n, p, ρ) we examined, as compared to the traditional bandwidth-free tests. The latter often yield very large size distortion in the case of small sample size and/or large dimension when the magnitude of temporal dependence is moderate. The size stability and accuracy with respect to the dimension and magnitude of dependence is a major gain of the SS–SN procedures. As a consequence of the usual size-power trade-off, there is a power loss for

Table 4. Empirical rejection rate (in percentage) under the null at level 5% when testing for change point in multivariate mean

p	ρ	$n=300$				$n=600$			
		SS–SN$_1$	SS–SN$_P$	SS–SN$_b$	SZ	SS–SN$_1$	SS–SN$_P$	SS–SN$_b$	SZ
5	−0.7	2.94	3.30	2.50	1.10	3.38	3.70	2.98	2.80
	−0.5	3.98	4.54	3.38	2.20	3.84	4.36	3.34	3.12
	0.2	5.06	6.06	4.88	5.10	4.88	4.72	4.02	5.02
	0.5	6.04	6.96	5.96	7.86	5.20	5.38	4.40	6.84
	0.7	7.10	7.94	7.44	12.88	5.82	6.18	5.00	8.38
10	−0.7	3.10	2.90	2.50	0.22	4.28	4.06	3.80	1.26
	−0.5	4.22	3.96	3.66	1.22	5.26	4.78	4.58	2.32
	0.2	5.48	5.20	5.24	7.70	5.92	5.30	5.52	6.28
	0.5	6.62	6.28	6.24	13.46	6.02	5.64	5.86	9.46
	0.7	7.36	7.50	7.66	28.82	6.50	6.22	6.70	14.24
20	−0.7	2.82	3.12	2.56	0.00	4.52	4.30	4.00	0.18
	−0.5	3.62	3.90	3.26	0.34	5.12	4.82	5.08	1.02
	0.2	4.92	5.14	4.90	11.80	5.58	5.44	5.56	7.92
	0.5	5.82	5.94	6.04	32.84	5.82	5.90	6.10	16.80
	0.7	7.20	6.88	7.60	73.20	6.36	6.32	7.04	35.82

SS–SN tests due to the use of sample splitting. However, when comparing the optimal SS–SN test to the traditional bandwidth-free counterparts (e.g. SS–SN$_P$ for dense alternative, and SS–SN$_1$ for sparse alternative), the power loss is mild. In a sense, this is similar to the 'more accurate size but less power' phenomenon when comparing tests based on fixed-b asymptotics versus small-b asymptotics (Kiefer & Vogelsang, 2005). In practice, when there is no prior knowledge about the type of alternative, we recommend the user to employ the Bonferonni test, i.e. SS–SN$_b$ by setting $\alpha = 0.5$. The asymptotic independence between L_2-type and L_∞-type SS–SN test statistics as stated in Theorem 7 further lends theoretical support to the Bonferroni test as it is expected to be non-conservative when the dimension of the parameter is moderate and sample size is large.

5 Conclusion

In this article, we propose a class of new tests for hypotheses on a multi-dimensional parameter based on SN and sample splitting. Our two SS–SN statistics do not involve any bandwidth parameter and the asymptotic null distribution is pivotal and is independent of the sample splitting proportion α. The construction of both SS–SN statistics are rather straightforward and the test statistics applied to the second part of sample $\mathcal{P}_2$ after dimension reduction based on the first part $\mathcal{P}_1$ are effectively targeting at parameter of dimension one. This sample splitting approach is broadly applicable to many time series testing problems, and we only cover testing hypotheses on marginal means, autocorrelations, regression parameter and a change point in multivariate mean to illustrate its usefulness. Overall, the SS–SN methodology provides an important addition to the existing SN toolbox owing to its superior ability of dealing with moderate dimensional parameter in the inference of low or moderate dimensional time series.

Below we shall highlight several appealing features of our test statistics. (a) For a moderate dimensional parameter, the size of our test statistics is considerably more accurate than traditional SN statistic, especially when temporal dependence is strong. As a price to pay, the SS–SN test loses some power. However, the power loss is moderate as seen from both theoretical power analysis and simulation studies. In practice, we recommend the practitioner to set $\alpha = 0.5$, and use the Bonferroni test that combines the two SS–SN test statistics so the power is adaptive to both sparse and dense alternatives. Simulation results show that the Bonferroni test exhibits accurate size and all-round good power in all settings. (b) We managed to show the asymptotic validity of SS–SN_1 and SS–SN_P test statistics and their asymptotic independence under the null in multivariate mean testing problem in a growing-dimensional setting, which is an interesting theoretical contribution to the SN literature. The theory is consistent with the empirical observation that the size is robust for SS–SN test statistics for a broad range of dimensions. (c) As a by-product of dimension reduction involved in SS–SN, there is substantial saving in computational cost as compared to traditional SN test statistics. In the mean testing problem, the cost of our SS–SN test statistics scales linearly in p, which is superior to that for the traditional SN statistic.

To conclude, we mention some possible extensions. The scope of this paper can be considerably expanded by using the GMM (Generalized Method of Moment) framework of Kiefer and Vogelsang (2005). Also one can regard KVB's test as a special case of the so-called fixed-b asymptotics (Kiefer & Vogelsang, 2005) with $b = 1$ and the use of Bartlett kernel. It is expected that the fixed-b-based tests and also other fixed-smoothing-based tests as advocated in Sun (2014b), Hwang and Sun (2017), and Wang and Sun (2020) will encounter the same size distortion problem when the dimension is moderate and temporal dependence is moderate/strong. Hence, it would be interesting to extend the SS–SN idea to fixed-smoothing methods and to GMM settings. These topics are left for future research.

Acknowledgments

We would like to thank the Associate Editor and two reviewers, whose thoughtful comments led to a substantial improvement of the paper.

Conflict of interest: None declared.

Funding

The research project is partially supported by U.S. National Science Foundation (DMS-2210002).

Data availability

The data used in the empirical example in online supplementary Appendix B is contained in the dataset *NelPlo* in the R package *tseries* (https://cran.r-project.org/web/packages/tseries/index.html).

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series B*

References

Andrews D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*(3), 817–858. https://doi.org/10.2307/2938229

Andrews D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*(4), 821–856. https://doi.org/10.2307/2951764

Andrews D. W., & Monahan J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, *60*(4), 953–966. https://doi.org/10.2307/2951574

Bell H. E. (1965). Gershgorin's theorem and the zeros of polynomials. *The American Mathematical Monthly*, *72*(3), 292–295. https://doi.org/10.2307/2313703

Billingsley P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Chang J., Cheng G., & Yao Q. (2022). Testing for unit roots based on sample autocovariances. *Biometrika*, *109*(2), 543–550. https://doi.org/10.1093/biomet/asab034

Du J.-H., Patil P., Roeder K., & Kuchibhotla A. K. (2023). 'Extrapolated cross-validation for randomized ensembles', arXiv, arXiv:2302.13511, preprint: not peer reviewed.

Gallant A. R. (2009). *Nonlinear statistical models*. John Wiley & Sons.

Hansen B. E. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica*, *60*(4), 967–972. https://doi.org/10.2307/2951575

He Y., Xu G., Wu C., & Pan W. (2021). Asymptotically independent u-statistics in high-dimensional testing. *Annals of Statistics*, *49*(1), 154–181. https://doi.org/10.1214/20-AOS1951

Horowitz J. L., Lobato I., Nankervis J. C., & Savin N. (2006). Bootstrapping the Box–Pierce Q test: A robust test of uncorrelatedness. *Journal of Econometrics*, *133*(2), 841–862. https://doi.org/10.1016/j.jeconom.2005.06.014

Hsing T. (1995). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *Annals of Probability*, *23*(2), 938–947. https://doi.org/10.1214/aop/1176988296

Huang Y. (2015). *Projection test for high-dimensional mean vectors with optimal direction* [Ph.D. dissertation]. Department of Statistics, The Pennsylvania State University at University Park.

Hwang J., & Sun Y. (2017). Asymptotic F and t tests in an efficient GMM setting. *Journal of Econometrics*, *198*(2), 277–295. https://doi.org/10.1016/j.jeconom.2017.02.003

Jansson M. (2004). The error in rejection probability of simple autocorrelation robust tests. *Econometrica*, *72*(3), 937–946. https://doi.org/10.1111/ecta.2004.72.issue-3

Kiefer N. M., & Vogelsang T. J. (2002). Heteroskedasticity-autocorrelation robust standard errors using the bartlett kernel without truncation. *Econometrica*, *70*(5), 2093–2095. https://doi.org/10.1111/ecta.2002.70.issue-5

Kiefer N. M., & Vogelsang T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, *21*(6), 1130–1164. https://doi.org/10.1017/S0266466605050565

Kiefer N. M., Vogelsang T. J., & Bunzel H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, *68*(3), 695–714. https://doi.org/10.1111/ecta.2000.68.issue-3

Korkas K. K., & Fryzlewicz, P. (2017). Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica Sinica*, *27*(1), 287–311. http://dx.doi.org/10.5705/ss.202015.0262

Li D., & Xue L. (2015). 'Joint limiting laws for high-dimensional independence tests', arXiv, arXiv:1512.08819, preprint: not peer reviewed.

Liu M., Zhu F., & Zhu K. (2022). Multifrequency-band tests for white noise under heteroscedasticity. *Journal of Business and Economic Statistics*, *40*(2), 799–814. https://doi.org/10.1080/07350015.2020.1870478

Lobato I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, *96*(455), 1066–1076. https://doi.org/10.1198/016214501753208726

Lunde R. (2019). 'Sample splitting and weak assumption inference for time series', arXiv, arXiv:1902.07425, preprint: not peer reviewed.

Maciejowska K. (2022). 'Identification testing via sample splitting–an application to structural var models', arXiv, arXiv:2206.03831, preprint: not peer reviewed.

Magnus J. R. (1986). The exact moments of a ratio of quadratic forms in normal variables. *Annales d'Economie et de Statistique* (4), 95–109. https://doi.org/10.2307/20075629

Mies F., & Steland A. (2023). Sequential gaussian approximation for nonstationary time series in high dimensions. *Bernoulli*, *29*(4), 3114–3140. https://doi.org/10.3150/22-BEJ1577

Müller U. K. (2014). Hac corrections for strongly autocorrelated time series. *Journal of Business & Economic Statistics*, *32*(3), 311–322. https://doi.org/10.1080/07350015.2014.931238

Newey W. K., & West K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708. https://doi.org/10.2307/1913610

Peng Z., & Nadarajah S. (2003). On the joint limiting distribution of sums and maxima of stationary normal sequence. *Theory Probability and its Applications*, 47(4), 706–709. https://doi.org/10.1137/S0040585X97980142

Rinaldo A., Wasserman L., & G'Sell M. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *The Annals of Statistics*, *47*(6), 3438–3469. https://doi.org/10.1214/18-AOS1784

Robinson P. M. (1991). Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica*, *59*(5), 1329–1363. https://doi.org/10.2307/2938370

Robinson P. M. (1998). Inference-without-smoothing in the presence of nonparametric autocorrelation. *Econometrica*, *66*(5), 1163–1182. https://doi.org/10.2307/2999633

Seber G. A., & Lee A. J. (2003). *Linear regression analysis* (Vol. *330*). John Wiley & Sons.

Shafer G., & Vovk V. (2008). A tutorial on conformal prediction. *Journal of Machine Learning Research*, *9*(3), 371–421.

Shao X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(3), 343–366. https://doi.org/10.1111/j.1467-9868.2009.00737.x

Shao X. (2011a). A bootstrap-assisted spectral test of white noise under unknown dependence. *Journal of Econometrics*, *162*(2), 213–224. https://doi.org/10.1016/j.jeconom.2011.01.001

Shao X. (2011b). Testing for white noise under unknown dependence and its applications to goodness-of-fit for time series models. *Econometric Theory*, *27*(2), 312–343. https://doi.org/10.1017/S0266466610000253

Shao X. (2015). Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association*, *110*(512), 1797–1817. https://doi.org/10.1080/01621459.2015.1050493

Shao X., & Zhang X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, *105*(491), 1228–1240. https://doi.org/10.1198/jasa.2010.tm10103

Srivastava M. S., & Du M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, *99*(3), 386–402. https://doi.org/10.1016/j.jmva.2006.11.002

Srivastava M. S., Katayama S., & Kano Y. (2013). A two sample test in high dimensional data. *Journal of Multivariate Analysis*, *114*(1), 349–358.

Sun Y. (2014a). Comment on "HAC corrections for strongly autocorrelated time series". *Journal of Business and Economic Statistics*, *32*(3), 330–334. https://doi.org/10.1080/07350015.2014.926817

Sun Y. (2014b). Fixed-smoothing asymptotics in a two-step GMM framework. *Econometrica*, *82*(6), 2327–2370. https://doi.org/10.3982/ECTA11684

Sun Y. (2014c). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroscedasticity and autocorrelation robust inference. *Journal of Econometrics*, *178*(3), 659–677. https://doi.org/10.1016/j.jeconom.2013.10.001

Sun Y., Phillips P. C., & Jin S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. *Econometrica*, *76*(1), 175–194. https://doi.org/10.1111/ecta.2008.76.issue-1

Wang R., & Shao X. (2020). Hypothesis testing for high-dimensional time series via self-normalization. *The Annals of Statistics*, *48*(5), 2728–2758. http://dx.doi.org/10.1214/19-AOS1904

Wang R., Zhu C., Volgushev S., & Shao X. (2022). Inference for change points in high dimensional data via self-normalization. *The Annals of Statistics*, *50*(2), 781–806. https://doi.org/10.1214/21-AOS2127

Wang X., & Sun Y. (2020). An asymptotic f test for uncorrelatedness in the presence of time series dependence. *Journal of Time Series Analysis*, *41*(4), 536–550. https://doi.org/10.1111/jtsa.v41.4

Wasserman L., Ramdas A., & Balakrishnan S. (2020). Universal inference. *Proceedings of the National Academy of Sciences*, *117*(29), 16880–16890. https://doi.org/10.1073/pnas.1922664117

Wasserman L., & Roeder K. (2009). High dimensional variable selection. *The Annals of Statistics*, *37*(5A), 2178–2201. https://doi.org/10.1214/08-AOS646

Wu W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, *102*(40), 14150–14154. https://doi.org/10.1073/pnas.0506715102

Xu G., Lin L., Wei P., & Pan W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika*, *103*(3), 609–624. https://doi.org/10.1093/biomet/asw029

Zhang X., & Shao X. (2013). Fixed-smoothing asymptotics for time series. *The Annals of Statistics*, *41*(3), 1329–1349. https://doi.org/10.1214/13-AOS1113