FIXED-SMOOTHING ASYMPTOTICS FOR TIME SERIES

BY XIANYANG ZHANG AND XIAOFENG SHAO

University of Missouri-Columbia and University of Illinois
at Urbana-Champaign

In this paper, we derive higher order Edgeworth expansions for the finite sample distributions of the subsampling-based $t$-statistic and the Wald statistic in the Gaussian location model under the so-called fixed-smoothing paradigm. In particular, we show that the error of asymptotic approximation is at the order of the reciprocal of the sample size and obtain explicit forms for the leading error terms in the expansions. The results are used to justify the second-order correctness of a new bootstrap method, the Gaussian dependent bootstrap, in the context of Gaussian location model.

1. Introduction. Many economic and financial applications involve time series data with autocorrelation and heteroskedasticity properties. Often the unknown dependence structure is not the chief object of interest but the inference on the parameter of interest involves the estimation of unknown dependence. In stationary time series models estimated by generalized method of moments (GMM), robust inference is typically accomplished by consistently estimating the asymptotic covariance matrix, which is proportional to the long run variance (LRV) matrix of the estimating equations or moment conditions defining the estimator, using a kernel smoothing method. In the econometrics and statistics literature, the bandwidth parameter/truncation lag involved in the kernel smoothing method is assumed to grow slowly with sample size in order to achieve consistency. The inference is conducted by plugging in a covariance matrix estimator that is consistent under heteroskedasticity and autocorrelation. This approach dates back to Newey and West [25] and Andrews [1]. Recently, Kiefer and Vogelsang [13] (KV, hereafter) developed an alternative first-order asymptotic theory for the HAC
(heteroskedasticity and autocorrelation consistent) based robust inference, where the proportion of the bandwidth involved in the HAC estimator to the sample size $T$, denoted as $b$, is held fixed in the asymptotics. Under the fixed-$b$ asymptotics, the HAC estimator converges to a nondegenerate yet nonstandard limiting distribution. The tests based on the fixed-$b$ asymptotic approximation were shown to enjoy better finite sample properties than the tests based on the small-$b$ asymptotic theory under which the HAC estimator is consistent, and the limiting distribution of the studentized statistic admits a standard form, such as standard normal or $\chi^2$ distribution. Using the higher order Edgeworth expansions, Jansson [12], Sun et al. [31] and Sun [28] rigorously proved that the fixed-$b$ asymptotics provides a high-order refinement over the traditional small-$b$ asymptotics in the Gaussian location model. Sun et al. [31] also provided an interesting decision theoretical justification for the use of fixed-$b$ rules in econometric testing. For non-Gaussian linear processes, Gonçalves and Vogelsang [6] obtained an upper bound on the convergence rate of the error in the fixed-$b$ approximation and showed that it can be smaller than the error of the normal approximation under suitable assumptions.

Since the seminal contribution by KV, there has been a growing body of work in econometrics and statistics to extend and expand the fixed-$b$ idea in the inference for time series data. For example, Sun [30] developed a procedure for hypothesis testing in time series models by using the non-parametric series method. The basic idea is to project the time series onto a space spanned by a set of fourier basis functions (see Phillips [26] and Müller [24] for early developments) and construct the covariance matrix estimator based on the projection vectors with the number of basis functions held fixed. Also see Sun [29] for the use of a similar idea in the inference of the trend regression models. Ibragimov and Müller [10] proposed a subsampling based $t$-statistic for robust inference where the unknown dependence structure can be in the temporal, spatial or other forms. In their paper, the number of non-overlapping blocks is held fixed. The $t$-statistic-based approach was extended by Bester et al. [3] to the inference of spatial and panel data with group structure. In the context of misspecification testing, Chen and Qu [5] proposed a modified $M$ test of Kuan and Lee [15] which involves dividing the full sample into several recursive subsamples and constructing a normalization matrix based on them. In the statistical literature, Shao [27] developed the self-normalized approach to inference for time series data that uses an inconsistent LRV estimator based on recursive subsample estimates. The self-normalized method is an extension of Lobato [21] from the sample autocovariances to more general approximately linear statistics, and it coincides with KVs fixed-$b$ approach in the inference of the mean of a stationary time series by using the Bartlett kernel and letting $b = 1$. Although the above inference procedures are proposed in different settings
and for different problems and data structures, they share a common feature in the sense that the underlying smoothing parameters in the asymptotic covariance matrix estimators such as the number of basis functions, the number of cluster groups and the number of recursive subsamples, play a similar role as the bandwidth in the HAC estimator. Throughout the paper, we shall call these asymptotics, where the smoothing parameter (or function of smoothing parameter) is held fixed, the fixed-smoothing asymptotics. In contrast, when the smoothing parameter grows with respect to sample size, we use the term increasing-domain asymptotics. At some places the terms fixed-\( K \) (or fixed-\( b \)) and increasing-\( K \) (or small-\( b \)) asymptotics are used to follow the convention in the literature.

In this article, we derive higher order expansions of the finite sample distributions of the subsampling-based \( t \)-statistic and the Wald statistic with HAC covariance estimator when the underlying smoothing parameters are held fixed, under the framework of the Gaussian location model. Specifically, we show that the error in the rejection probability (ERP, hereafter) is of order \( O(1/T) \) under the fixed-smoothing asymptotics. Under the assumption that the eigenfunctions of the kernel in the HAC estimator have zero mean and other mild assumptions, we derive the leading error term of order \( O(1/T) \) under the fixed-smoothing framework. These results are similar to those obtained under the fixed-\( b \) asymptotics (see Sun et al. [31]), but are stronger in the sense that we are able to derive the exact form of the leading error term with order \( O(1/T) \). The explicit form of the leading error term in the approximation provides a clear theoretical explanation for the empirical findings in the literature regarding the direction and magnitude of size distortion for time series with various degrees of dependence. To the best of our knowledge, this is the first time that the leading error terms are made explicit through the higher order Edgeworth expansion under the fixed-smoothing asymptotics. It is also worth noting that our nonstandard argument differs from that in Jansson [12] and Sun et al. [31], and it may be of independent theoretical interest and be useful for future follow-up work.

Second, we propose a novel bootstrap method for time series, the Gaussian dependent bootstrap, which is able to mimic the second-order properties of the original time series and produces a Gaussian bootstrap sample. For the Gaussian location model, we show that the inference based on the Gaussian dependent bootstrap is more accurate than the first-order approximation under the fixed-smoothing asymptotics. This seems to be the first time a bootstrap method is shown to be second-order correct under the fixed-smoothing asymptotics; see Gonçalves and Vogelsang [6] for a recent attempt for the moving block bootstrap in the non-Gaussian setting.

We now introduce some notation. For a vector \( x = (x_1, x_2, \ldots, x_q) \in \mathbb{R}^q \), we let \( \|x\| = (\sum_{i=1}^{q} x_i^2)^{1/2} \) be the Euclidean norm. For a matrix \( A = \)
\((a_{ij})^{q_0}_{i,j=1} \in \mathbb{R}^{q_0 \times q_0}\), denote by \(\|A\|_2 = \sup_{\|x\|=1}\|Ax\|\) the spectral norm and \(\|A\|_\infty = \max_{1 \leq i,j \leq q_0}|a_{ij}|\) the max norm. Denote by \([a]\) the integer part of a real number \(a\). Let \(L^2[0,1]\) be the space of square integrable functions on \([0,1]\). Denote by \(D[0,1]\) the space of functions on \([0,1]\) which are right continuous and have left limits, endowed with the Skorokhod topology; see Billingsley [4]. Denote by \(\Rightarrow\) weak convergence in the \(\mathbb{R}^{q_0}\)-valued function space \(D^{q_0}[0,1]\), where \(q_0 \in \mathbb{N}\). Denote by \(\rightarrow_d\) and \(\rightarrow_p\) convergence in distribution and convergence in probability, respectively. The notation \(N(\mu, \Sigma)\) is used to denote the multivariate normal distribution with mean \(\mu\) and covariance \(\Sigma\). Let \(\chi_k^2\) be a random variable following \(\chi^2\) distribution with \(k\) degrees of freedom and \(G_k\) be the corresponding distribution function.

The layout of the paper is as follows. Section 2 contains the higher order expansions of the finite sample distributions of the subsampling \(t\)-statistic and the Wald statistic with HAC estimator. We introduce the Gaussian dependent bootstrap and the results about its second-order accuracy in Section 3. Section 4 concludes. Technical details and simulation results are gathered in the supplementary material [34].

2. Higher order expansions. This paper is partially motivated by recent studies on the ERP for the Gaussian location model by Jansson [12] and Sun et al. [31], who showed that the ERP is of order \(O(1/T)\) under the fixed-\(b\) asymptotics, which is smaller than the ERP under the small-\(b\) asymptotics. A natural question is to what extent the ERP result can be extended to the recently proposed fixed-smoothing based inference methods under the fixed-smoothing asymptotics. Following Jansson [12] and Sun et al. [31], we focus on the inference of the mean of a univariate stationary Gaussian time series or equivalently, a Gaussian location model. We conjecture that the higher order terms in the asymptotic expansion under the Gaussian assumption will also show up in the general expansion without the Gaussian assumption.

2.1. Expansion for the finite sample distribution of subsampling-based \(t\)-statistic. We first investigate the Edgeworth expansion of the finite sample distribution of subsampling-based \(t\)-statistic (Ibragimov and Müller [10]). Here we treat the subsampling-based \(t\)-statistic and other cases separately, because the \(t\)-statistic corresponds to a different choice of normalization factor (compare with the Wald statistic in Section 2.2). Given the observations \(\{X_1, X_2, \ldots, X_T\}\) from a Gaussian stationary time series, we divide the sample into \(K\) approximately equal sized groups of consecutive observations. The observation \(X_i\) is in the \(j\)th group if and only if \(i \in M_j = \{s \in \mathbb{Z}: (j-1)T/K < s \leq jT/K\}, j = 1, 2, \ldots, K\). Define the sample mean of the \(k\)th group as

\[
\hat{\mu}_k = \frac{1}{|M_k|} \sum_{i \in M_k} X_i, \quad k = 1, 2, \ldots, K.
\]
where \( | \cdot | \) denotes the cardinality of a finite set. Let \( \hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K)' \), \( \bar{\mu}_n = \frac{1}{K} \sum_{i=1}^{K} \hat{\mu}_i \) and \( S_n^2 = \frac{1}{K-1} \sum_{i=1}^{K} (\hat{\mu}_i - \bar{\mu}_n)^2 \). Then the subsampling-based \( t \)-statistic for testing the null hypothesis \( H_0: \mu = \mu_0 \) versus the alternative \( H_a: \mu \neq \mu_0 \) is given by

\[
T_K = \frac{\sqrt{K}(\hat{\mu}_n - \mu_0)}{S_n} = \frac{\sqrt{K}(\bar{\mu}_n - \mu_0)}{\left(\frac{\sum_{i=1}^{K} (\hat{\mu}_i - \bar{\mu}_n)^2}{K-1}\right)^{1/2}}.
\]

Our goal here is to develop an Edgeworth expansion of \( P(|T_K| \leq x) \) when \( K \) is fixed and sample size \( T \to \infty \). It is not hard to see that the distribution of \( T_K \) is symmetric, so it is sufficient to consider \( P(|T_K| \leq x) \) since \( P(T_K \leq x) = \frac{1 + P(|T_K| \leq x)}{2} \) for any \( x \geq 0 \). Denote by \( t_k \) a random variable following \( t \) distribution with \( k \) degrees of freedom. The following theorem gives the higher order expansion under the Gaussian assumption.

**Theorem 2.1.** Assume that \( \{X_i\} \) is a stationary Gaussian time series satisfying that \( \sum_{h=0}^{\infty} \gamma_X(h) > 0 \) and \( \sum_{h=0}^{\infty} h^2 |\gamma_X(h)| < \infty \). Further suppose that \( |M_1| = |M_2| = \cdots = |M_K| \) and \( K \) is fixed. Then under \( H_0 \), we have

\[
\sup_{x \in [0, +\infty)} |P(|T_K| \leq x) - \Psi(x; K)| = O(1/T^2),
\]

where \( \Psi(x; K) = P(|t_{K-1}| \leq x) - \frac{B}{\sigma^2 T} \gamma(x; K) \) with

\[
\gamma(x; K) = -K^2 P(|t_{K-1}| \leq x) + (K + 1)E\left[\chi_{K-1}^2 G_1\left(\frac{\chi_{K-1}^2 x^2}{K-1}\right)\right] - E\left[\chi_{1}^2 G_{K-1}\left(\frac{(K-1)\chi_{1}^2}{x^2}\right)\right] + 1
\]

and \( B = \sum_{h=-\infty}^{+\infty} |h| \gamma_X(h) \).

We present the proof of Theorem 2.1 in Section 5, which requires some nonstandard arguments. From the above expression, we see that the leading error term is of order \( O(1/T) \), and the magnitude and direction of the error depend upon \( B/\sigma^2 \), which is related to the second-order properties of time series, and \( \gamma(x; K) \), which is independent of the dependence structure of \( \{X_i\} \) and can be approximated numerically for given \( x \) and \( K \). Figure 1 plots the approximated values of \( \gamma(t_{K-1}(1-\alpha); K)/K \) for different \( K \) and \( \alpha \), where \( t_{K-1}(1-\alpha) \) denotes the \( 100(1-\alpha)\% \) quantile of the \( t \) distribution with \( K-1 \) degrees of freedom. It can be seen from Figure 1 that \( \gamma(t_{K-1}(1-\alpha); K)/K \) increases rapidly for \( K < 10 \), and it becomes stable for relatively large \( K \). For each \( K \geq 2 \), \( \gamma(t_{K-1}(1-\alpha); K)/K \) is an increasing function of \( \alpha \). In the simulation work of Ibragimov and Müller [10] (see Figure 2 therein),
they found that the size of the subsampling-based $t$-test is relatively robust to the correlations if $K$ is small (say $K = 4$ in their simulation). This finding is in fact supported by our theory. For $K \leq 4$, the magnitude of $\Upsilon(x; K)$ is rather small, so the leading error term is small across a range of correlations. As $K$ increases, the first-order approximation deteriorates, which is reflected in the increasing magnitude of $\Upsilon(t_{K-1}(1-\alpha); K)$ with respect to $K$.

Notice that $\Upsilon(t_{K-1}(1-\alpha); K)$ is always positive and $\sigma^2 > 0$ by assumption, so the sign of the leading error term, that is, $-\frac{B}{2\sigma^2 T} \Upsilon(x; K)$, is determined by $B$. When $B > 0$ [e.g., AR(1) process with positive coefficient], the first-order based inference tends to be oversized, and conversely it tends to be undersized when $B < 0$ [e.g., MA(1) process with negative coefficient]. Some simulations for AR(1) and MA(1) models in the Gaussian location model support these theoretical findings. We decide not to report these results to conserve space. Given the sample size $T$, the size distortion for the first-order based inference may be severe if the ratio $B/\sigma^2$ is large. For example, this is the case for AR(1) model, $X_t = \rho X_{t-1} + \varepsilon_t$, as the correlation $\rho$ gets closer to 1. As indicated by Figure 1, we show in the following proposition that $\Upsilon(t_{K-1}(1-\alpha); K)/K$ converges as $K \to +\infty$.

**Proposition 2.1.** As $K \to +\infty$, we have $\Upsilon(x; K)/K = 2x^2 G_1'(x^2) + O(1/K)$, for any fixed $x \in \mathbb{R}$. 

**Fig. 1.** Simulated values of $\Upsilon(t_{K-1}(1-\alpha); K)/K$ based on 500,000 replications.
Under the local alternative $H'_a : \mu = \mu_0 + (\delta \sigma) / \sqrt{T}$ with $\delta \neq 0$, we can derive a similar expansion for $T_K$ with $K$ fixed. Formally let $Z$ be a random variable following the standard normal distribution and $S_{K-1} = \sqrt{\chi^2_{K-1} / (K-1)}$ with the $\chi^2_{K-1}$ distribution being independent with $Z$. Then the quantity $t_{K-1,\delta} = (Z + \delta) / S_{K-1}$ follows a noncentral $t$ distribution with noncentral parameter $\delta$. Define $e_1(x) = E[I\{|t_{K-1,\delta}| > x\}Z^2]$ and $e_2(x) = E[I\{|t_{K-1,\delta}| > x\}\chi^2_{K-1}]$. Then under the local alternative, we have

$$P(|T_K| \leq x) = P(|t_{K-1,\delta}| \leq x) - \frac{B}{2\sigma^2 T} \Upsilon(x; K) + O(1/T^2),$$

where $\Upsilon(x; K) = K^2 P(|t_{K-1,\delta}| > x) - e_1(x) - (K + 1)e_2(x)$. For fixed $\delta$, $P(|t_{K-1,\delta}| > t_{K-1}(1 - \alpha))$ is a monotonic increasing functions of $K$. An unreported numerical study shows that $\Upsilon(x; K)$ is roughly monotonic with respect to $K$ for $\delta \in (0, 1]$, which suggests that larger $K$ tends to deliver more power when $B > 0$. Combined with the previous discussion, we see that the choice of $K$ leads to a trade-off between the size distortion and power loss.

**Remark 2.1.** Theorem 2.1 gives the ERP and the exact form of the leading error term under the fixed-$K$ asymptotics. The higher order expansion derived here is based on an expansion of the density function of $(\hat{\mu}_1, \ldots, \hat{\mu}_K)$ which is made possible by the Gaussian assumption. Extension to the general GMM setting without the Gaussian assumption may require a different strategy in the proof. Expansion for a distribution function or equivalently characteristic function has been used in the higher order expansion of the finite sample distribution under the Gaussian assumption (see, e.g., Velasco and Robinson [32] and Sun et al. [31]). With $K$ fixed in the asymptotics, the leading term of the variance of the LRV estimator is captured by the first order fixed-$K$ limiting distribution and the leading term of the bias of the LRV estimator is reflected in the leading error term $-\frac{B}{2\sigma^2 T} T(x, K)$. Specifically, let $\Sigma_T = (\sigma_{ij})_{i,j=1}^K$ with $\sigma_{ij} = q \text{Cov}(\hat{\mu}_i, \hat{\mu}_j)$. Then the leading error term captures the difference between $\Sigma_T$ and $\sigma^2 I_K$, and the effect of the off-diagonal elements $\sigma_{ij}$ with $|i - j| > 1$ is of order $O(1/T^2)$ and thus is not reflected in the leading term.

**Remark 2.2.** When the number of groups $K$ grows slowly with the sample size $T$, the Edgeworth expansion for $T_K$ was developed for $P(T_K \leq x)$ in Lahiri [18, 19] under the general non-Gaussian setup. The expansion given here is different from the usual Edgeworth expansion under the increasing-domain asymptotics in terms of the form and the convergence rate. Using the same argument, we can show that under the fixed-$K$ asymptotics, the leading error term in the expansion of $P(T_K \leq x)$ is of order $O(1/T)$ under
the Gaussian assumption. In the non-Gaussian case, we conjecture that the order of the leading error term is $O(1/\sqrt{T})$, which is due to the effect of the third and fourth-order cumulants.

The higher order Edgeworth expansion results in Sun et al. [31] suggest that the fixed-$b$ based approximation is a refinement of the approximation provided by the limiting distribution derived under the small-$b$ asymptotics. In a similar spirit, it is natural to ask if the fixed-$K$ based approximation refines the first-order approximation under the increasing-$K$ asymptotics. To address this question, we consider the expansion under the increasing-domain asymptotics, where $K$ grows slowly with the sample size $T$.

**Proposition 2.2.** Under the same conditions in Theorem 2.1 but with $\lim_{T \to \infty} (1/K + K/T) = 0$, we have

$$P(|T_K| \leq x) = G_1(x^2) + \frac{1}{K-1}x^4G''_1(x^2) - \frac{BK}{T\sigma^2}x^2G'_1(x^2) + O(1/T).$$

**Remark 2.3.** Since $P(|t_{K-1}| \leq x) = G_1(x^2) + \frac{1}{K-1}x^4G''_1(x^2) + O(1/K^2)$ (see, e.g., Sun [30]), we know that the fixed-$K$ based approximation captures the first two terms in (3), whereas the increasing-$K$ based approximation (i.e., $\chi^2_1$) only captures the first term. In view of Proposition 2.1, it is not hard to see that

$$\Psi(x; K) = G_1(x^2) + \frac{1}{K-1}x^4G''_1(x^2) - \frac{BK}{T\sigma^2}x^2G'_1(x^2) + O(1/K^2) + O(1/T),$$

which implies that the fixed-$K$ based expansion is able to capture all the three terms in (3) as the smoothing parameter $K \to \infty$ with $T^{1/3} = o(K)$. Loosely speaking, this suggests that the fixed-$K$ based expansion holds for a broad range of $K$, and it gets close to the corresponding increasing-$K$ based expansion when $K$ is large.

**2.2. Fixed-$b$ expansion.** Consider a semi-positive definite bivariate kernel $G(\cdot, \cdot)$ which satisfies the spectral decomposition

$$G(r, t) = \sum_{j=1}^{+\infty} \lambda_j \phi_j(r) \phi_j(t), \quad 0 \leq r, t \leq 1,$$

where $\{\phi_j\}$ are the eigenfunctions, and $\{\lambda_j\}$ are the eigenvalues which are in a descending order, that is, $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. Suppose we have the observations $\{X_1, X_2, \ldots, X_T\}$ from a stationary Gaussian time series with mean
Let $\mu$ and autocovariance function $\gamma_X(i-j) = E[(X_i - \mu)(X_j - \mu)]$. The LRV estimator based on the kernel $G(\cdot, \cdot)$ and bandwidth $bT$ with $b \in (0, 1]$ is given by

$$
\hat{D}_{T,b} = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} G\left(\frac{i}{bT}, \frac{j}{bT}\right) (X_i - \bar{X}_T)(X_j - \bar{X}_T),
$$

where $\bar{X}_T = \frac{1}{T} \sum_{i=1}^{T} X_i$ is the sample mean. For the convenience of presentation, we set $b = 1$. To illustrate the idea, we define the projection vectors $\xi_j = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \phi_0(i/T) X_i$ with $\phi_0(t) = \phi_j(t) - \frac{1}{T} \sum_{i=1}^{T} \phi_j(i/T)$ for $j = 1, 2, \ldots$. Here the dependence of $\xi_j$ on $T$ is suppressed to simplify the notation. Following Sun [30], we limit our attention to the case $\int_0^1 \phi_j(t) dt = 0$ (e.g., Fourier basis and Haar wavelet basis). For any semi-positive definite kernel $G(\cdot, \cdot)$, we can define the demeaned kernel,

$$
\tilde{G}(r,t) = G(r,t) - \int_0^1 G(s,t) \, ds - \int_0^1 G(r,p) \, dp + \int_0^1 \int_0^1 G(s,p) \, ds \, dp.
$$

Suppose $\tilde{G}(\cdot, \cdot)$ admits the spectral decomposition $\tilde{G}(r,t) = \sum_{i=1}^{+\infty} \tilde{\lambda}_i \tilde{\phi}_i(r) \tilde{\phi}_i(t)$ with $\{\tilde{\phi}_i\}$ and $\{\tilde{\lambda}_i\}$ being the eigenfunctions and eigenvalues, respectively. Notice that

$$
\int_0^1 \int_0^1 \tilde{G}(r,t) \, dr \, dt = \sum_{i=1}^{+\infty} \tilde{\lambda}_i \left( \int_0^1 \tilde{\phi}_i(t) \, dt \right)^2 = 0,
$$

which implies $\int_0^1 \tilde{\phi}_i(t) \, dt = 0$ whenever $\lambda_i > 0$, that is, the eigenfunctions of the demeaned kernel $\tilde{G}(\cdot, \cdot)$ are all mean zero. Based on the spectral decomposition (4) of $G(\cdot, \cdot)$, the LRV estimator with $b = 1$ can be rewritten as

$$
\hat{D}_{T,1} = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} G\left(\frac{i}{T}, \frac{j}{T}\right) (X_i - \bar{X}_T)(X_j - \bar{X}_T) = \sum_{i=1}^{+\infty} \lambda_i \xi_i^2.
$$

We focus on testing the null hypothesis $H_0 : \mu = \mu_0$ versus the alternative $H_a : \mu \neq \mu_0$. Define a sequence of random variables

$$
F_T(K) = \frac{\xi_0^2}{\sum_{j=1}^{K} \lambda_j \xi_j^2}, \quad K = 1, \ldots, \infty,
$$

with $\xi_0 = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} (X_i - \mu_0)$. The Wald test statistic with HAC estimator is given by $F_T(\infty) = \xi_0^2 / \hat{D}_{T,1}$. Let $\{v_i\}_{i=0}^{+\infty}$ be a sequence of independent and identically distributed (i.i.d.) standard normal random variables. Further
define $\mathcal{F}(K) := \mathcal{F}(v; K) = \frac{v_0^2}{\sum_{j=1}^K \lambda_j v_j^2}$ and

$$\mathcal{N}_T(x; K) = \frac{1}{2\sigma^2} \sum_{i=0}^K (\text{var}(\xi_i) - \sigma^2) E[(v_i^2 - 1)I\{\mathcal{F}(v; K) \leq x\}],$$

(5) \hspace{1cm} K = 1, \ldots, \infty,

with $\sigma^2 = \sum_{h=-\infty}^{+\infty} \gamma_X(h)$ being the LRV. The following theorem establishes the asymptotic expansion of the finite sample distribution of $F_T(K)$ with $1 \leq K \leq \infty$.

**Theorem 2.2.** Assume the kernel $G(\cdot, \cdot)$ satisfies the following conditions:

1. The second derivatives of the eigenfunctions $\{\phi_i^{(2)}(\cdot)\}_{i=1}^{+\infty}$ exist. Further assume that the eigenfunctions are mean zero and satisfy that

$$\sup_{1 \leq i \leq J} \|\phi_i^{(j)}(t)\| < CJ^j$$

for $j = 0, 1, 2, J \in \mathbb{N}$, and some constant $C$ which does not depend on $j$ and $J$.

2. The eigenvalues $\lambda_n = O(1/n^a)$, for some $a > 19$.

Under the assumption that $\{X_i\}$ is a stationary Gaussian time series with $\sigma^2 = \sum_{h=-\infty}^{+\infty} \gamma_X(h) > 0$ and $\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty$, and the null hypothesis $H_0$, we have

$$\sup_{x \in [0, +\infty)} |P(F_T(K) \leq x) - P(\mathcal{F}(K) \leq x) - \mathcal{N}_T(x; K)| = o(1/T)$$

(6) \hspace{1cm} for any $1 \leq K \leq \infty$.

The proof of Theorem 2.2 is based on the arguments of the proof of Theorem 2.1 given in Section 5 and the truncation argument. The technical details are provided in Zhang and Shao [34]. For $K < \infty$, Theorem 2.2 shows that the $O(1/T)$ ERP rate can be extended to the Wald statistic with series variance estimator (Sun [30]). When $K = \infty$, Theorem 2.2 gives the asymptotic expansion of the Wald test statistic $F_T(\infty)$ which is of particular interest. The leading error term $\mathcal{N}_T(x; \infty)$ reflects the departure of $\{v_j\}_{j=0}^{+\infty}$ from the i.i.d. standard normal random variables $\{v_j\}_{j=0}^{+\infty}$. Specifically, the form of $\mathcal{N}_T(x; \infty)$ suggests that the leading error term captures the difference between the LRV and the variances of $\xi_i$’s which are not exactly the same across $i = 0, 1, 2, \ldots$. By the orthogonality assumption, the covariance between $\xi_i$ and $\xi_j$ with $i \neq j$ is of smaller order and hence is not reflected in the leading term. Assume $\int_0^1 G(r, r) dr = \sum_{j=1}^{+\infty} \lambda_j = 1$. As seen from Theorem 2.2, the bias of the LRV estimator [i.e., $\sum_{i=1}^{+\infty} \lambda_i (\text{var}(\xi_i) - \sigma^2)$] is reflected in the leading error term $\mathcal{N}_T(x; \infty)$, which is a weighted sum of the relative
difference of \( \text{var}(\xi) \) and \( \sigma^2 \). Note that the difference \( \text{var}(\xi) - \sigma^2 \) relies on the second-order properties of the time series and the eigenfunctions of \( G(\cdot, \cdot) \), and the weight \( E[(\alpha^2 - 1)\mathcal{F}(\infty) \leq x] \) which depends on the eigenvalues of \( G(\cdot, \cdot) \) is of order \( O(\lambda_i) \), as seen from the arguments used in the proof of Theorem 2.2.

In the econometrics and statistics literature, the bivariate kernel \( G(\cdot, \cdot) \) is usually defined through a semi-positive definite univariate kernel \( K(\cdot) \), that is, \( G(r, t) = K(r - t) \). In what follows, we make several remarks regarding this special case.

**Remark 2.4.** For \( 0 < b \leq 1 \), we define \( G_b(\cdot, \cdot) = G(\cdot/b, \cdot/b) \). If \( G(\cdot, \cdot) \) is semi-positive definite on \( [0, 1/b]^2 \), then \( G_b(\cdot, \cdot) \) satisfies the spectral decomposition \( G_b(r, t) = \sum_{j=1}^{\infty} \lambda_{j,b} \phi_{j,b}(r) \phi_{j,b}(t) \) with \( 0 \leq r, t \leq 1 \). The eigencoeficients of \( G_b(r, t) \) can be obtained by solving a homogenous Fredholm integral equation of the second kind, where the solutions can be approximated numerically when analytical solutions are unavailable. When \( G(r, t) = K(r - t) \), it was shown in Knessl and Keller [14] that under suitable assumptions on \( K(\cdot) \), \( \lambda_{j,b} = b \int_{-\infty}^{+\infty} K(r) \, dr - (\pi^2 b^2 / 2) \int_{-\infty}^{+\infty} r^2 K(r) \, dr + o(b^3) \) and \( \phi_{j,b} \approx \sqrt{2} \sin(\pi j x) \) for \( x \) bounded away from 0 and 1 as \( b \to 0 \), which implies that \( \lambda_{M,b}/\lambda_{1,b} \to 1 \) for any fixed \( M \in \mathbb{N} \) and \( b \to 0 \). Our result can be extended to the case where \( b < 1 \) if the assumptions in Theorem 2.2 hold for \( \{\lambda_{j,b}\} \) and \( \{\phi_{j,b}\} \). It is also worth noting that our result is established under different assumptions as compared to Theorem 6 in Sun et al. [31], where the bivariate kernel is defined as \( G(r, t) = K(r - t) \) and the technical assumption \( b < 1/(16 \int_{-\infty}^{+\infty} |K(r)| \, dr) \) is required, which rules out the case \( b = 1 \) for most kernels. Here we provide an alternative way of proving the \( O(1/T) \) ERP when the eigenfunctions are mean zero. Furthermore, we provide the exact form of the leading error term which has not been obtained in the literature.

**Remark 2.5.** The assumption on the eigenvalues is satisfied by the bivariate kernel defined through the QS kernel and the Daniel kernel with \( 0 < b \leq 1 \), and the Tukey–Hanning kernel with \( b = 1 \) because these kernels are analytical on the corresponding regions, and their eigenvalues decay exponentially fast; see Little and Reade [20]. However, the assumption does not hold for the Bartlett kernel because the decay rate of its eigenvalues is of order \( O(1/n^2) \). For the demeaned Tukey–Hanning kernel with \( b = 1 \), we have that the eigenfunctions \( \phi_1(t) = \sqrt{2} \cos \pi t \) and \( \phi_2(t) = \frac{\sin \pi t - 2/\pi}{\sqrt{1/2 - 4/\pi^2}} \) with eigenvalues \( \lambda_1 = 0.25 \), \( \lambda_2 = 0.0474 \) and \( \lambda_j = 0 \) for \( j \geq 3 \). It is not hard to construct a kernel that satisfies the conditions in Theorem 2.2. For example, one can consider the kernel \( K(r - t) = \sum_{j=1}^{+\infty} \lambda_j \{ \cos(2\pi jr) \cos(2\pi jt) + \sin(2\pi jr) \sin(2\pi jt) \} = \sum_{j=1}^{+\infty} \lambda_j \cos(2\pi j(r - t)) \) with \( \sum_{j=1}^{+\infty} \lambda_j = 1 \) and \( \lambda_j = \)
that the higher order expan-
bisian location model based on both fixed-
approximations for the distribution of studentized sample mean in the Gaus-
Small-
defined of $F$
For the Bartlett kernel $b$
the second-order asymptotic expansions under the fixed-
and $\mathbb{P}$
it is worth noting that $P$
the Wald statistic based on the difference kernel $G$
above. The formulas for the second-order approximation und er the small-
small-
$X. ZHANG AND X. SHAO
Table 1
Asymptotic comparison between the first and second-order approximations based on fixed-$b$ and small-$b$ asymptotics

<table>
<thead>
<tr>
<th>Asymptotics</th>
<th>First order</th>
<th>Second order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed-$b$</td>
<td>$P(F_b(\infty) \leq x)$</td>
<td>$P(F_b(\infty) \leq x) + \mathcal{R}_{T,b}(x; \infty)$</td>
</tr>
<tr>
<td>Small-$b$</td>
<td>$G_1(x)$</td>
<td>$G_1(x) + (c_2 G_1''(x)x^2 - c_1 G_1'(x)b) + \sum_{b \to 0} \frac{1}{s^2(\mathcal{G}_b)} G_1'(x)x$</td>
</tr>
</tbody>
</table>

$O(1/j^{19+\epsilon})$ for some $\epsilon > 0$. Then the asymptotic expansion (6) holds for 
the Wald statistic based on the difference kernel $G(r,t) = \mathcal{K}(r-t)$.

Define the Parzen characteristic exponent

$$q = \max \left\{ q_0 : q_0 \in \mathbb{Z}^+, g_{q_0} = \lim_{x \to 0} \frac{1 - \mathcal{K}(x)}{|x|^{q_0}} < \infty \right\}.$$ 

For the Bartlett kernel $q$ is 1; For the Parzen and QS kernels, $q$ is equal 
to 2. Let $c_1 = \int_{-\infty}^{+\infty} \mathcal{K}(x) \, dx$ and $c_2 = \int_{-\infty}^{+\infty} \mathcal{K}^2(x) \, dx$. Further define $F_b(\infty)$ and $\mathcal{R}_{T,b}(x; \infty)$ with $\phi_j$ and $\lambda_j$ being replaced with $\phi_{j,b}$ and $\lambda_{j,b}$ in the definition of $F(\infty)$ and $\mathcal{R}_T(x; \infty)$. We summarize the first and second-order approximations for the distribution of studentized sample mean in the Gaussian location model based on both fixed-$b$ and small-$b$ asymptotics in Table 1 above. The formulas for the second-order approximation under the small-$b$ asymptotics is from Velasco and Robinson [32].

**Remark 2.6.** A few remarks are in order regarding Table 1. First of all, it is worth noting that $P(F_b(\infty) \leq x) = G_1(x) + (c_2 G_1''(x)x^2 - c_1 G_1'(x)b) + O(b^2)$ as $b \to 0$ in Sun et al. [31], which suggests that the fixed-$b$ limiting distribution captures the first two terms in the higher order asymptotic expansion under the small-$b$ asymptotics and thus provides a better approximation than the $\chi_1^2$ approximation. Second, it is interesting to compare 
the second-order asymptotic expansions under the fixed-$b$ asymptotics and small-$b$ asymptotics. We show in Proportion 2.3 that the higher order expansion under fixed-$b$ asymptotics is consistent with the corresponding higher order expansion under small-$b$ asymptotics as $b$ approaches zero.

Because our fixed-$b$ expansion is established under the assumption that 
the eigenfunctions have mean zero, we shall consider the Wald statistic $F_T(\infty)$ based on the demeaned kernel $\tilde{G}_b(r,t) = \mathcal{K}_b(r-t) - \int_0^1 \mathcal{K}_b(s-t) \, ds - \int_0^1 \mathcal{K}_b(r-p) \, dp + \int_0^1 \int_0^1 \mathcal{K}_b(s-p) \, ds \, dp$ with $\mathcal{K}_b(\cdot) = \mathcal{K}(\cdot/b)$ and $b \in (0,1]$. Let $\{\tilde{\phi}_{j,b}\}$ and $\{\tilde{\lambda}_{j,b}\}$ be the corresponding eigenfunctions and eigenvalues of $\tilde{G}_b(\cdot,\cdot)$.
Proposition 2.3. Suppose $K(\cdot): \mathbb{R} \to [0, 1]$ is symmetric, semi-positive definite, piecewise smooth with $K(0) = 1$ and $\int_0^{+\infty} xK(x)\,dx < \infty$. The Parzen characteristic exponent of $K$ is no less than one. Further assume that

$$
\sup_{k \in \mathbb{N}} \left| \sum_{i=1}^k \tilde{\lambda}_{i,b}(\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) \right| = O \left( \int_{-\infty}^{+\infty} \frac{h^q}{\sigma^2(bT)^q} |\gamma_X(h)| \,dh \right) \quad (7)
$$

as $b + 1/(bT) \to 0$, where $\tilde{\xi}_{i,b}$ is defined by replacing $\phi_j$ with $\tilde{\phi}_{j,b}$ in the definition of $\xi_i$. Then under the assumption that $\sigma^2 > 0$ and $\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty$, we have

$$
\mathbb{N}_{T,b}(x; \infty) = \frac{g_q \sum_{h=-\infty}^{+\infty} |h|^q \gamma_X(h)}{\sigma^2(bT)^q} G_1'(x) x(1 + o(1)) + O(1/T)
$$

for fixed $x \in \mathbb{R}$, as $b \to 0$ and $bT \to +\infty$.

In Proposition 2.3, condition (7) is not primitive, and it requires that the bias for the LRV estimators based on the kernel $\tilde{G}_{k,b}(r,t) = \sum_{i=1}^k \tilde{\lambda}_{j,b} \tilde{\phi}_{j,b}(r) \tilde{\phi}_{j,b}(t)$ is at the same or smaller order of the bias for the LRV estimator based on $\tilde{G}_b(r,t)$. This condition simplifies our technical arguments and it can be verified through a case-by-case study. As shown in Proposition 2.3, the fixed-$b$ expansion is consistent with the small-$b$ expansion as $b$ approaches zero, and it is expected to be more accurate in terms of approximating the finite sample distribution when $b$ is relatively large. Overall speaking, the above result suggests that the fixed-$b$ expansion provides a good approximation to the finite sample distribution which holds for a broad range of $b$.

3. Gaussian dependent bootstrap. Given the higher order expansions presented in Section 2, it seems natural to investigate if bootstrapping can help to improve the first-order approximation. Though the higher order corrected critical values can also be obtained by direct estimation of the leading error term, it involves estimation of the eigencomponents of the kernel function and a choice of truncation number for the leading error term $\mathbb{N}_T(x; \infty)$ [see (5)] besides estimating the second-order properties of the time series. Therefore it is rather inconvenient to implement this analytical approach because numerical or analytical calculation of eigencomponents can be quite involved, the truncation number and the bandwidth parameter used in estimating second-order properties are both user-chosen numbers, and it seems difficult to come up with good rules about their (optimal) choice in the current context. By contrast, the bootstrap procedure proposed below, which involves only one user-chosen number, aims to estimate the leading error term in an automatic fashion and the computational cost is moderate given current high computing power.
To present the idea, we again limit our attention to the univariate Gaussian location model. Consider a consistent estimate of the covariance matrix of \{X_i\}_{i=1}^T which takes the form \( \hat{\Xi}(\omega; l) \in \mathbb{R}^{T \times T} \) with the \((i, j)\)th element given by \( \omega_l(i-j)\hat{\gamma}_X(|i-j|) \) for \( i, j = 1, 2, \ldots, T \), where \( \omega \) is a kernel function with \( \omega_l(\cdot) = \omega(\cdot/l) \) and \( \hat{\gamma}_X(h) = \frac{1}{T} \sum_{i=1}^{T-h} (X_i - \bar{X}_T)(X_{i+h} - \bar{X}_T) \) for \( h = 0, 1, 2, \ldots, T - 1 \). Estimating the covariance matrix of a stationary time series has been investigated by a few researchers. See Wu and Pourahmadi \cite{33} for the use of a banded sample covariance matrix and McMurry and Politis \cite{23} for a tapered version of the sample covariance matrix. In what follows, we shall consider the Bartlett kernel, that is, \( \hat{\Xi}(\omega; l) \geq 0 \).

We now introduce a simple bootstrap procedure which can be shown to be second-order correct. Suppose \( X_1^*, \ldots, X_T^* \) is the bootstrap sample generated from \( N(0, \hat{\Xi}(\omega, l)) \). It is easy to see that \( X_1^* \)’s are stationary and Gaussian conditional on the data. This is why we name this bootstrap method “Gaussian dependent bootstrap.” There is a large literature on bootstrap for time series; see Lahiri \cite{17} for a review. However, most of the existing bootstrap methods do not deliver a conditionally normally distributed bootstrap sample. Since our higher order results are obtained under the Gaussian assumption, we need to generate Gaussian bootstrap samples in order for our expansion results to be useful.

Denote by \( T_K^* \) the bootstrapped subsampling \( t \)-statistic obtained by replacing \( (X_1 - \mu_0, X_2 - \mu_0, \ldots, X_T - \mu_0) \) with \( (X_1^*, X_2^*, \ldots, X_T^*) \). Define the bootstrapped projection vectors \( \xi_0^* = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} X_j^* \) and \( \xi_j^* = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \phi_j^0(i/T) X_i^* \) for \( j = 1, \ldots \). Let \( P^* \) be the bootstrap probability measure conditional on the data. The following theorems state the second-order accuracy of the Gaussian dependent bootstrap in the univariate Gaussian location model.

**Theorem 3.1.** For the Gaussian location model, under the same conditions in Theorem 2.1 and \( 1/l + l^3/T \to 0 \), we have

\[
\sup_{x \in [0, +\infty)} |P(|T_K| \leq x) - P^*(|T_K^*| \leq x)| = o_p(1/T).
\]

**Theorem 3.2.** For the Gaussian location model, under the assumptions in Theorem 2.2 and that \( 1/l + l^3/T \to 0 \), we have

\[
\sup_{x \in [0, +\infty)} |P(F_T(\infty) \leq x) - P^*(F_T^*(\infty) \leq x)| = o_p(1/T),
\]

where \( F_T^*(\infty) = \frac{(\xi_0^*)^2}{\sum_{j=1}^{\infty} \lambda_j \xi_j^2} \) with \( \{\lambda_j\}_{j=1}^{+\infty} \) given in (4). Note that \( F_T^*(\infty) = (\xi_0^*)^2/\hat{D}_{T,1}^* \), where \( \hat{D}_{T,1}^* = T^{-1} \sum_{i,j=1}^{T} G(i/T, j/T)(X_i^* - \bar{X}_T^*)(X_j^* - \bar{X}_T^*) \) and \( \bar{X}_T^* \) is the bootstrap sample mean.
Remark 3.1. The higher order terms in the small-\(b\) expansion and the increasing-\(K\) expansion (see Table 1 and Proposition 2.2) depend on the second-order properties only through the quantities \(\sum_{h=-\infty}^{+\infty} |h|^k \gamma_X(h)\) for \(k = 0, 1, \ldots, q\). It suggests that the Gaussian dependent bootstrap also preserves the second-order accuracy under the increasing-domain asymptotics provided that

\[\sum_{h=-\infty}^{+\infty} |h|^{q+1} \gamma_X(h) < \infty.\]

A rigorous proof is omitted due to space limitation.

The bootstrap-based autocorrelation robust testing procedures have been well studied in both econometrics and statistics literature under the increasing-domain asymptotics. In the statistical literature, Lahiri [16] showed that for the studentized \(M\)-estimator, the ERP of the moving block bootstrap (MBB)-based one-sided testing procedure is of order \(o_p(T^{-1/2})\) which provides an asymptotic refinement to the normal approximation. Under the framework of the smooth function model, Götze and Küchler [7] showed that the ERP for the MBB-based one-sided test is of order \(O_p(T^{-3/4+\epsilon})\) for any \(\epsilon > 0\) when the HAC estimator is constructed using the truncated kernel. Note that in the latter paper, the HAC estimator used in the studentized bootstrap statistic needs to take a different form from the original HAC estimator to achieve the higher order accuracy. Also see Lahiri [18] for a recent contribution. In the econometric literature, the Edgeworth analysis for the block bootstrap has been conducted by Hall and Horowitz [8], Andrews [2] and Inoue and Shintani [11], among others, in the GMM framework. Within the increasing-domain asymptotic framework, it is still unknown whether the bootstrap can achieve an ERP of \(o_p(1/T)\) when a HAC covariance matrix estimator is used for studentization; see Härdle, Horowitz and Kreiss [9]. Note that Hall and Horowitz [8] and Andrews [2] obtained the \(o_p(1/T)\) results for symmetrical tests but they assumed the uncorrelatedness of the moment conditions after finite lags. Note that all the above results were obtained under the non-Gaussian assumption.

Within the fixed-smoothing asymptotic framework, Jansson [12] established that the error of the fixed-\(b\) approximation to the distribution of two-sided test statistic is of order \(O(\log(T)/T)\) for the Gaussian location model and the case \(b = 1\), which was further refined by Sun et al. [31] by dropping the \(\log(T)\) term. In the non-Gaussian setting, Gonçalves and Vogelsang [6] showed that the fixed-\(b\) approximation to the distribution of one-sided test statistic has an ERP of order \(o(T^{-1/2+\epsilon})\) for any \(\epsilon > 0\) when all moments exist. The latter authors further showed that the MBB (with i.i.d. bootstrap as a special case) is able to replicate the fixed-\(b\) limiting distribution.
and thus provides more accurate approximation than the normal approximation. However, because the exact form of the leading error term was not obtained in their studies, their results seem not directly applicable to show the higher order accuracy of bootstrap under the fixed-$b$ asymptotics. Using the asymptotic expansion results developed in Section 2, we show that the Gaussian dependent bootstrap can achieve an ERP of order $o_p(1/T)$ under the Gaussian assumption. This appears to be the first result that shows the higher order accuracy of bootstrap under the fixed-smoothing asymptotics.

Our result also provides a positive answer to the open question mentioned in Härdle, Horowitz and Kreiss [9] that whether the bootstrap can achieve an ERP of $o_p(1/T)$ in the dependence case when a HAC covariance matrix estimator is used for studentization. It is worth noting that our result is established for the symmetrical distribution functions under the fixed-smoothing asymptotics and the Gaussian assumption. It seems that in general the ERP of order $o_p(1/T)$ cannot be achieved under the increasing-domain asymptotics or for the non-Gaussian case. In the supplementary material [34], we provide some simulation results which demonstrate the effectiveness of the proposed Gaussian dependent bootstrap in both Gaussian and non-Gaussian settings. The MBB is expected to be second-order accurate, as seen from its empirical performance, but a rigorous theoretical justification seems very difficult. Finally, we mention that it is an important problem to choose $l$. For a given criterion, the optimal $l$ presumably depends on the second-order property of the time series in a sophisticated fashion. Some of the rules proposed for block-based bootstrap (see Lahiri [17], Chapter 7) may still work, but a serious investigation is beyond the scope of this article.

4. Conclusion. In this paper, we derive the Edgeworth expansions of the subsampling-based $t$-statistic and the Wald statistic with HAC estimator in the Gaussian location model. Our work differs from the existing ones in two important aspects: (i) the expansion is derived under the fixed-smoothing asymptotics and the ERP of order $O(1/T)$ is shown for a broad class of fixed-smoothing inference procedures; (ii) we obtain an explicit form for the leading error term, which is unavailable in the literature. An in-depth analysis of the behavior of the leading error term when the smoothing parameter grows with sample size (i.e., $K \to \infty$ in the subsampling $t$-statistic or $b \to 0$ in the Wald statistic with the HAC estimator) shows the consistency of our results with the expansion results under the increasing-domain asymptotics. Building on these expansions, we further propose a new bootstrap method, the Gaussian dependent bootstrap, which provides a higher order correction than the first-order fixed-smoothing approximation.

We mention a few directions that are worthy of future research. First, it would be interesting to relax the Gaussian assumption in all the expansions we obtained in the paper. For non-Gaussian time series, Edgeworth
expansions have been obtained by Götze and Kunsch [7], Lahiri [18, 19], among others, for studentized statistics of a smooth function model under weak dependence assumption, but their results were derived under the increasing-smoothing asymptotics. For the location model and studentized sample mean, the extension to the non-Gaussian case may require an expansion of the corresponding characteristic function, which involves calculation of the high-order cumulants under the fixed-smoothing asymptotics. The detailed calculation of the high-order terms can be quite involved and challenging. We conjecture that under the fixed-smoothing asymptotics, the leading error term in the expansion of its distribution function involves the third and fourth-order cumulants, which reflects the non-Gaussianness, and the order of the leading error term is $O(T^{-1/2})$ instead of $O(T^{-1})$. Second, we expect that our expansion results will be useful in the optimal choice of the smoothing parameter, the kernel and its corresponding eigenvalues and eigenfunctions, for a given loss function. The optimal choice of the smoothing parameter has been addressed in Sun et al. [31] using the expansion derived under the increasing-smoothing asymptotics. As the finite sample distribution is better approximated by the corresponding fixed-smoothing based approximations at either first or second order than its increasing-smoothing counterparts, the fixed-smoothing asymptotic theory proves to be more relevant in terms of explaining the finite sample results; see Gonçalves and Vogelsang [6]. Therefore, it might be worth reconsidering the choice of the optimal smoothing parameter under the fixed-smoothing asymptotics. Third, we restrict our attention to the Gaussian location model when deriving the higher order expansions. It would be interesting to extend the results to the general GMM setting. A recent attempt by Sun [28] for the HAC-based inference seems to suggest this is feasible. Finally, under the fixed-smoothing asymptotics, the second correctness of the moving block bootstrap for studentized sample mean, although suggested by the simulation results, is still an open but challenging topic for future research.

5. Proof of Theorem 2.1. Consider the $K + 1$-dimensional multivariate normal density function which takes the form

$$f(y, \Sigma) = (2\pi)^{-(K+1)/2} |\Sigma|^{-1/2} \exp(-\frac{1}{2} y' \Sigma^{-1} y).$$

We assume the $(i, j)$th element and the $(j, i)$th element of $\Sigma$ are functionally unrelated. The results can be extended to the case where symmetric matrix elements are considered functionally equal; see, for example, McCulloch [22]. In the following, we use $\otimes$ to denote the Kronecker product in matrix algebra and use vec to denote the operator that transforms a matrix into a column vector by stacking the columns of the matrix one underneath the other. For a vector $y \in \mathbb{R}^{l \times 1}$ whose elements are differential functions of a vector
$x \in \mathbb{R}^{k \times 1}$, we define $\frac{\partial u}{\partial x}$ to be a $k \times l$ matrix with the $(i, j)$th element being $\frac{\partial y_i}{\partial x_j}$. The notation $u \propto v$ represents $u = O(v)$ and $v = O(u)$. We first present the following lemmas whose proofs are given in the online supplement [34].

**Lemma 5.1.**
\[
\frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \Sigma) = \frac{f(y, \Sigma)}{2} \{(\Sigma^{-1} y) \otimes (\Sigma^{-1} y) - \text{vec}(\Sigma^{-1})\}.
\]

**Lemma 5.2.**
\[
\frac{\partial^2 f}{\partial \text{vec}(\Sigma) \, \text{vec}(\Sigma)}(y, \Sigma)
\]
\[
= \frac{1}{4}\{(\Sigma^{-1} y) \otimes (\Sigma^{-1} y) - \text{vec}(\Sigma^{-1})\}\{(\Sigma^{-1} y) \otimes (\Sigma^{-1} y) - \text{vec}(\Sigma^{-1})\}'
\]
\[
- \frac{1}{2}\{(\Sigma^{-1} yy'\Sigma^{-1}) \otimes \Sigma^{-1} + \Sigma^{-1} \otimes (\Sigma^{-1} yy'\Sigma^{-1}) - \Sigma^{-1} \otimes \Sigma^{-1}\}
\]
\[
\times f(y, \Sigma).
\]

**Lemma 5.3.** Let $\{\Sigma_T\} \subset \mathbb{R}^{(K+1)\times(K+1)}$ be a sequence of positive definite matrices with $K + 1 \leq T$. If $K$ is fixed with respect to $T$ and $\|\Sigma_T - \Sigma\|_2 = O(1/T)$ for a positive definite matrix $\Sigma$, then we have
\[
\|\Sigma_T^{-1} - \Sigma^{-1}\|_2 = O(1/T).
\]

**Lemma 5.4.** Let $\hat{\Sigma}_T(y)$ be a $(K + 1) \times (K + 1)$ positive symmetric matrix which depends on $y \in \mathbb{R}^{K+1}$. Assume that $\sup_{y \in \mathbb{R}^{K+1}} \|\hat{\Sigma}_T(y) - \Sigma\|_2 \leq \|\Sigma_T - \Sigma\|_2 = O(1/T)$ for a positive definite matrix $\Sigma$. Let $R_T = \Sigma_T - \Sigma$. If $K$ is fixed with respect to $T$, we have
\[
\int_{y \in \mathbb{R}^{K+1}} \left| \text{vec}(R_T)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \, \text{vec}(\Sigma)}(y, \hat{\Sigma}_T(y)) \text{vec}(R_T) \right| dy = O(1/T^2).
\]

**Proof of Theorem 2.1.** For the convenience of our presentation, we ignore the functional symmetry of the covariance matrix in the proof. With some proper modifications, we can extend the results to the case where the functional symmetry is taken into consideration. Let $|\mathcal{M}_1| = |\mathcal{M}_2| = \cdots = |\mathcal{M}_K| = q$. Define $Y_i = \sqrt{q}(\bar{\mu}_i - \mu_0)$, and $\bar{Y} = \frac{1}{K} \sum_{i=1}^{K} Y_i$ and $S_{\bar{Y}}^2 = \frac{1}{K-1} \sum_{i=1}^{K} (Y_i - \bar{Y})^2$ as the sample mean and sample variance of $\{Y_i\}_{i=1}^{K}$, respectively. Note that $T_K(Y) = \sqrt{K} \bar{Y} / S_Y$, where $Y = (Y_1, Y_2, \ldots, Y_K)'$. Simple algebra yields that
\[
\sigma_{ij} := \text{Cov}(Y_i, Y_j) = \sum_{h=1-q}^{q-1} \left( \frac{q - |h|}{q} \right) \gamma_X(h - (j-i)q).
\]
Notice that $Y$ follows a normal distribution with mean zero and covariance matrix $\Sigma_T$, where $\Sigma_T = (\sigma_{ij})_{i,j=1}^K$. The density function of $Y$ is given by

$$f(y, \Sigma_T) = (2\pi)^{-K/2}|\Sigma_T|^{-1/2}\exp(-\frac{1}{2}y'\Sigma_T^{-1}y).$$

Under the assumption $\sum_{h=-\infty}^{\infty}h^2|\gamma_X(h)| < \infty$, it is straightforward to see that $\|\Sigma_T - \sigma^2 I_K\|_2 = O(1/T)$. Taking a Taylor expansion of $f(y, \Sigma_T)$ around elements of the matrix $\sigma^2 I_K$, we have

$$f(y, \Sigma_T) = f(y, \sigma^2 I_K) + \left\{\frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \sigma^2 I_K)\right\}' \text{vec}(\Sigma_T - \sigma^2 I_K)$$

$$+ \text{vec}(\Sigma_T - \sigma^2 I_K)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma)\partial \text{vec}(\Sigma)}(y, \tilde{\Sigma}_T(y)) \text{vec}(\Sigma_T - \sigma^2 I_K),$$

where $\sup_{y \in \mathbb{R}^K} \|\tilde{\Sigma}_T(y) - \sigma^2 I_K\|_2 \leq \|\Sigma_T - \sigma^2 I_K\|_2 = O(1/T)$. By Lemmas 5.1 and 5.4, we get

$$\frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \sigma^2 I_K) = f(y, \sigma^2 I_K)\left\{-\frac{1}{2\sigma^2} \text{vec}(I_K) + \frac{1}{2\sigma^4} y \otimes y\right\}$$

and

$$\int_{y \in \mathbb{R}^K} \left|\text{vec}(\Sigma_T - \sigma^2 I_K)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma)\partial \text{vec}(\Sigma)}(y, \tilde{\Sigma}_T(y)) \text{vec}(\Sigma_T - \sigma^2 I_K)\right| dy = O\left(\frac{1}{T^2}\right),$$

which imply that

$$f(y, \Sigma_T) = f(y, \sigma^2 I_K)\left\{1 - \frac{1}{2\sigma^2} \sum_{i=1}^K (\sigma_{ii} - \sigma^2)\right\}$$

$$+ \frac{1}{2\sigma^4} f(y, \sigma^2 I_K) \sum_{i=1}^K \sum_{j=1}^K (\sigma_{ij} - \sigma^2 \delta_{ij})y_iy_j + R(y)$$

$$= g(y, \sigma^2 I_K) + R(y),$$

where $g$ denotes the major term, $R(y)$ is the remainder term and $\delta_{ij} = I\{i = j\}$ is the Kronecker’s delta. Define $\tilde{\Psi}(x; K) = \int_{\{|T_K(y)| > x\}} g(y, \sigma^2 I_K) dy$. By (10), we see that

$$\sup_{x \in \mathbb{R}} \left|\int_{\{|T_K(y)| > x\}} f(y, \Sigma_T) dy - \tilde{\Psi}(x; K)\right| \leq \int_{\mathbb{R}^K} |R(y)| dy = O(1/T^2).$$
It follows from some simple calculation that
\[
\tilde{\Psi}(x; K) = \left\{ 1 - \frac{1}{2\sigma^2} \sum_{i=1}^{K} (\sigma_{ii} - \sigma^2) \right\} P(|t_{K-1}| > x) + \frac{1}{2\sigma^2} (J_1 + J_2),
\]
where
\[
J_1 = \sum_{i=1}^{K} (\sigma_{ii} - \sigma^2) E[I\{|\tilde{T}_K(v)| > x\}v_i^2], \quad J_2 = \sum_{i \neq j} \sigma_{ij} E[I\{|\tilde{T}_K(v)| > x\}v_i v_j].
\]
Here \(\{v_i\}_{i=1}^{K}\) are i.i.d. standard normal random variables and \(\tilde{T}_K(v) = \sqrt{K}\bar{v}/S_v\) is the \(t\) statistic based on \(\{v_i\}\) with \(\bar{v} = \frac{1}{K} \sum_{i=1}^{K} v_i\) and \(S_v^2 = \frac{1}{K-1} \sum_{i=1}^{K} (v_i - \bar{v})^2\). Let \(U = K\bar{v}^2\) and \(D = (K-1)S_v^2\). Then \(U \sim \chi^2_1\), \(D \sim \chi^2_{K-1}\) and \(U\) and \(D\) are independent. We define that
\[
E[I\{|\tilde{T}_K(v)| > x\}v_i^2] = \frac{1}{K} E\left[I\{|\tilde{T}_K(v)| > x\} \sum_{i=1}^{K} v_i^2\right]
\]
and
\[
E[I\{|\tilde{T}_K(v)| > x\}v_i v_j] = \frac{1}{K(K-1)} E\left[I\{|\tilde{T}_K(v)| > x\} \sum_{i \neq j} v_i v_j\right]
\]
We then have
\[
P(|T_K| > x) = \tilde{\Psi}(x; K) + O(1/T^2)
\]
(11)
\[
= \{1 - \alpha\} P(|t_{K-1}| > x) + \beta E\left[UG_{K-1}\left(\frac{(K-1)U}{x^2}\right)\right] + \tau \left\{ K - 1 - E\left[DG_1\left(\frac{Dx^2}{K-1}\right)\right] \right\} + O(1/T^2),
\]
uniformly for $x \in \mathbb{R}$, where the coefficients are given by

$$
\alpha = \frac{1}{2\sigma^2} \sum_{i=1}^{K} (\sigma_{ii} - \sigma^2) = -\frac{K^2 B}{2\sigma^2 T} + O(1/T^2),
$$

$$
\beta = \frac{1}{2K\sigma^2} \sum_{i=1}^{K} \sum_{j=1}^{K} (\sigma_{ij} - \delta_{ij} \sigma^2) = -\frac{B}{2\sigma^2 T} + O(1/T^2)
$$

and

$$
\tau = \frac{1}{2K\sigma^2} \sum_{i=1}^{K} (\sigma_{ii} - \sigma^2) - \frac{1}{2K(K-1)\sigma^2} \sum_{i\neq j} \sigma_{ij} = -\frac{(K+1)B}{2\sigma^2 T} + O(1/T^2).
$$

The conclusion thus follows from equation (11). □

**Acknowledgments.** The authors would like to thank the Associate Editor and the reviewers for their constructive comments, which substantially improve the paper.

**SUPPLEMENTARY MATERIAL**

**Proofs of the other results in Sections 2–3 and simulation results.** (DOI: 10.1214/13-AOS1113SUPP; .pdf). This supplement contains proofs of the other main results in Sections 2–3 and some simulation results.

**REFERENCES**


1. Proofs of the main results.

1.1. Proof of the main results in Section 2.1. Consider the \( K + 1 \) dimensional multivariate normal density function which takes the form

\[
f(y, \Sigma) = (2\pi)^{-\frac{K+1}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y' \Sigma^{-1} y \right).
\]

We assume the \((i, j)\)th element and the \((j, i)\)th element of \( \Sigma \) are functionally unrelated. The results can be extended to the case where symmetric matrix elements are considered functionally equal (see e.g., McCulloch [5]). In the following, we use \( \otimes \) to denote the Kronecker product in matrix algebra and use \( \text{vec} \) to denote the operator that transforms a matrix into a column vector by stacking the columns of the matrix one underneath the other. For a vector \( y \in \mathbb{R}^{l \times 1} \) whose elements are differential functions of a vector \( x \in \mathbb{R}^{k \times 1} \), we define \( \frac{\partial y}{\partial x} \) to be a \( k \times l \) matrix with the \((i, j)\)th element being \( \frac{\partial y_j}{\partial x_i} \). The notation \( u \asymp v \) represents \( u = O(v) \) and \( v = O(u) \). Lemma 1.1 and Lemma 1.2 below are straightforward consequences of matrix calculus (see e.g., Vetter [9], Brewer [1] and Turkington [8]).

**Lemma 1.1.**

\[
\frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \Sigma) = \frac{f(y, \Sigma)}{2} \{(\Sigma^{-1} y) \otimes (\Sigma^{-1} y) - \text{vec}(\Sigma^{-1})\}.
\]

**Proof.** By matrix calculus, we get

\[
\frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \Sigma) = (2\pi)^{-\frac{K+1}{2}} \left\{ \exp \left( -\frac{1}{2} y' \Sigma^{-1} y \right) \frac{\partial |\Sigma|^{-\frac{1}{2}}}{\partial \text{vec}(\Sigma)} + |\Sigma|^{-\frac{1}{2}} \frac{\partial}{\partial \text{vec}(\Sigma)} \exp \left( -\frac{1}{2} y' \Sigma^{-1} y \right) \right\}
\]

\[
= (2\pi)^{-\frac{K+1}{2}} \left\{ -\frac{1}{2} |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y' \Sigma^{-1} y \right) \text{vec}(\Sigma^{-1}) + \frac{1}{2} |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y' \Sigma^{-1} y \right) (\Sigma^{-1} y) \otimes (\Sigma^{-1} y) \right\}
\]

\[
= \frac{f(y, \Sigma)}{2} \{(\Sigma^{-1} y) \otimes (\Sigma^{-1} y) - \text{vec}(\Sigma^{-1})\},
\]

*Supported in part by NSF Grant DMS08-04937 and DMS11-04545.*
where we have used the formulas \( \frac{\partial X^{-1}}{\partial \text{vec}(X)} = -X^{-1}b \otimes (X^{-1})'a \) and \( \frac{\partial |X|^m}{\partial \text{vec}(X)} = m|X|^{m-1} \) (see Theorem 4.3 and Theorem 4.19 in Turkington [8]).

Lemma 1.2.

\[
\frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \Sigma) = \frac{1}{4}((\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})) \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\}'f(y, \Sigma) - \frac{1}{2}((\Sigma^{-1}yy'\Sigma^{-1}) \otimes \Sigma^{-1} + \Sigma^{-1} \otimes (\Sigma^{-1}yy'\Sigma^{-1}) - \Sigma^{-1} \otimes \Sigma^{-1})f(y, \Sigma).
\]

Proof. From Lemma 1.1, we have

\[
\frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \Sigma) = \frac{\partial}{\partial \text{vec}(\Sigma)} \left( \frac{f(y, \Sigma)}{2} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\} \right) = \left( \frac{\partial}{\partial \text{vec}(\Sigma)} \frac{f(y, \Sigma)}{2} \right) \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\}' + \frac{f(y, \Sigma)}{2} \frac{\partial}{\partial \text{vec}(\Sigma)} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\} = I_1 + I_2.
\]

Again from Lemma 1.1, it is not hard to see that

\[
I_1 = \frac{f(y, \Sigma)}{4} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1})\}'.
\]

In view of Lemma 4.3 in Turkington [8], we have

\[
\frac{\partial \text{vec}(\Sigma^{-1}yy'\Sigma^{-1})}{\partial \text{vec}(\Sigma)} = \frac{\partial \text{vec}(\Sigma^{-1}y)}{\partial \text{vec}(\Sigma)} (y'y^{-1} \otimes I_{K+1}) + \frac{\partial \text{vec}(y'y')}{\partial \text{vec}(\Sigma)} (I_{K+1} \otimes y'y^{-1}).
\]

Also by Theorem 4.3 in Turkington [8], we get

\[
\frac{\partial \text{vec}(\Sigma^{-1}y)}{\partial \text{vec}(\Sigma)} = -\Sigma^{-1}y \otimes \Sigma^{-1}; \quad \frac{\partial \text{vec}(y'y')}{\partial \text{vec}(\Sigma)} = -\Sigma^{-1} \otimes \Sigma^{-1}y,
\]

which implies that

\[
\frac{\partial \text{vec}(\Sigma^{-1}yy'\Sigma^{-1})}{\partial \text{vec}(\Sigma)} = -(\Sigma^{-1}yy'\Sigma^{-1}) \otimes \Sigma^{-1} - \Sigma^{-1} \otimes (\Sigma^{-1}yy'\Sigma^{-1}).
\]

Further by Theorem 4.2 in Turkington [8], we obtain \( \frac{\partial \text{vec}(\Sigma^{-1})}{\partial \text{vec}(\Sigma)} = -\Sigma^{-1} \otimes \Sigma^{-1} \). The conclusion thus follows directly from the above derivation.

Lemma 1.3. Let \( \{\Sigma_T\} \subset \mathbb{R}^{(K+1) \times (K+1)} \) be a sequence of positive definite matrices with \( K + 1 \leq T \). If \( K \) is fixed with respect to \( T \) and \( ||\Sigma_T - \Sigma||_2 = O(1/T) \) for a positive definite matrix \( \Sigma \), then we have

\[
||\Sigma_T^{-1} - \Sigma^{-1}||_2 = O(1/T).
\]

\[\Box\]
PROOF. Let $\Sigma_T = \Sigma + R_T$ with $\|R_T\|_2 = O(1/T)$. For sufficiently large $T$, we have $\|\Sigma^{-1}R_T\|_2 \leq \|\Sigma^{-1}\|_2\|R_T\|_2 < 1$. By the last equation at p. 355 of Horn and Johnson [4], we have

$$\|\Sigma_T^{-1} - \Sigma^{-1}\|_2 \leq \frac{\|\Sigma^{-1}\|^2\|R_T\|_2}{1 - \|\Sigma^{-1}R_T\|_2} = O(1/T).$$

\[\diamondsuit\]

**Lemma 1.4.** Let $\tilde{\Sigma}_T(y)$ be a $(K + 1) \times (K + 1)$ positive symmetric matrix which depends on $y \in \mathbb{R}^{K+1}$. Assume that $\sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T(y) - \Sigma\|_2 \leq \|\Sigma_T - \Sigma\|_2 = O(1/T)$ for a positive definite matrix $\Sigma$. Let $R_T = \Sigma_T - \Sigma$. If $K$ is fixed with respect to $T$, we have

$$\int_{y \in \mathbb{R}^{K+1}} \left| \text{vec}(R_T)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \cdot \text{vec}(\Sigma)}(y, \tilde{\Sigma}_T(y)) \text{vec}(R_T) \right| dy = O(1/T^2).$$

**Proof.** Let $\tilde{R}_T(y) = \tilde{\Sigma}_T(y) - \Sigma$. Note that

$$\sup_{y \in \mathbb{R}^{K+1}} \|\Sigma^{-1}\tilde{R}_T(y)\|_2 \leq \|\Sigma^{-1}\|_2 \sup_{y \in \mathbb{R}^{K+1}} \|\tilde{R}_T(y)\|_2 \leq \|\Sigma^{-1}\|_2\|\Sigma_T - \Sigma\|_2 < 1,$$

for large enough $T$. By the same arguments in Lemma 1.3, we have $\sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1}\|_2 = O(1/T)$. Therefore, when $T$ is sufficiently large, we have $y'(\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1}/2)y = y'(\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1})y + y'y^{-1}/2 \geq (\lambda_{\min}(\Sigma^{-1})/2 - \|\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1}\|_2)\|y\|^2 \geq 0$ for all $y$, where $\lambda_{\min}(\Sigma^{-1})$ denotes the smallest eigenvalue of $\Sigma^{-1}$. On the other hand, for sufficiently large $T$, we have $\sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T^{-1}(y)\|_2^{K+1} \leq (\|\Sigma^{-1}\|_2 + \sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1}\|_2)^{K+1} \leq C\|\Sigma^{-1}\|^k$ with $C > 0$. Combining the above arguments, we get $f(y, \tilde{\Sigma}_T(y)) \leq C\|\Sigma^{-1}\|^{1/2} \exp(-y'\Sigma^{-1}y/4) \leq C f(y, 2\Sigma)$ for all $y$. When $K$ is fixed, $\|\cdot\|_2$ and $\|\cdot\|_{\infty}$ are equivalent, which implies $\sup_{y \in \mathbb{R}^{K+1}} \|\Sigma_T(y)^{-1} - \Sigma^{-1}\|_{\infty} = O(1/T)$. Since the elements of $\tilde{\Sigma}_T^{-1}(y)$ are uniformly bounded for all $y$, in view of Lemma 1.2, it is straightforward to see

$$\left| \text{vec}(R_T)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \cdot \text{vec}(\Sigma)}(y, \tilde{\Sigma}_T(y)) \text{vec}(R_T) \right| \leq Cp(y) f(y, 2\Sigma)/T^2,$$

where $p(y)$ is a polynomial of degree 4. The conclusion follows by noting that $\int p(y) f(y, 2\Sigma) dy < \infty$. \[\diamondsuit\]

**Proof of Proposition 2.1.** Note first that

$$\Upsilon(x; K)K = -KP(|t_{K-1}| \leq x) + \frac{K + 1}{K} E \left[ \frac{2}{\chi_{K-1}^2} G_1 \left( \frac{\chi_{K-1}}{K - 1} x^2 \right) \right] + O(1/K).$$
Using the fact that \( P(|t_{K-1}| \leq x) = G_1(x^2) + \frac{1}{K-1} x^4 G_1''(x^2) + O(1/K^2) \), we get

\[
\begin{align*}
\Upsilon(x; K) / K & = -K G_1(x^2) - \frac{K}{K-1} x^4 G_1''(x^2) + \frac{K+1}{K} E \left[ \frac{\chi_{K-1}^2}{K-1} \right] G_1(x^2) \\
& \quad \left( \frac{\chi_{K-1}^2}{K-1} - 1 \right) x^2 G_1'(x^2) + \frac{1}{2} \left( \frac{\chi_{K-1}^2}{K-1} - 1 \right)^2 x^4 G_1''(x^2) \right] + O(1/K) \\
& = 2x^2 G_1'(x^2) + O(1/K).
\end{align*}
\]

\[\Box\]

**Proof of Proposition 2.2.** Recall that \( q = T/K \) is assumed to be an integer. Using the notation in the proof of Theorem 2.1, let \( S_Y^2 = \frac{1}{K-1} \sum_{i=1}^K Y_i - \bar{Y} \). Notice that

\[
cov(Y) = \begin{pmatrix}
\sigma^2 - B/q & B/(2q) & 0 & \ldots & 0 \\
B/(2q) & \sigma^2 - B/q & B/(2q) & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & B/(2q) & \sigma^2 - B/q
\end{pmatrix}
K \times K
\]

\[+ O(1/q^2) l_K l'_K,\]

\[= \sigma^2 I_K + \frac{B}{2q} \begin{pmatrix}
-2 & 1 & 0 & \ldots & 0 \\
1 & -2 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1 & -2
\end{pmatrix} K \times K
\]

\[= \sigma^2 I_K + \frac{B}{2q} M + O(1/q^2) l_K l'_K,
\]

where \( l'_K = (1, 1, \ldots, 1)_{1 \times K} \) and the summation of all the \( O(1/q^2) \) is of order \( O(K/q^2) \). Because

\[
E[Y_i^2] = \sum_{h=1-q}^{q-1} \frac{q - |h|}{q} \gamma_X(h) = \sigma^2 - B/q + O(1/q^2),
\]

and

\[
E[Y_i Y_j] = \frac{1}{K^2} \sum_{i,j=1}^K E[Y_i Y_j] = \frac{1}{K^2} \{ K \sigma^2 - B/q + O(K/q^2) \}
\]

\[= \sigma^2/K + O(1/(K^2q)) + O(1/(Kq^2)),
\]

we obtain

\[
E[S_Y^2] - \sigma^2 = \frac{K}{K-1} \{ \sigma^2 - B/q - \sigma^2/K + o(1/T) \} - \sigma^2 = -B/q + O(1/T).
\]
Consider the covariance matrix of $\hat{Y}' = (Y_1 - \bar{Y}, Y_2 - \bar{Y}, \ldots, Y_K - \bar{Y})$. It is easy to see that $\hat{Y} = (I_K - l_k l_k' / K) Y = H_K Y$, where $H_K = I_K - l_k l_k' / K$ is an idempotent matrix. Ignoring the $O(1/q^2)$ order term in $\text{cov}(Y)$, we have

$$(c_{ij})_{i,j=1}^K := \text{cov}(\hat{Y}) = H_K \text{cov}(Y) H_K \approx H_K \{ \sigma^2 I_K + BM/(2q) \} H_K$$

where

$$A = \begin{pmatrix} -2 & -1 & -1 & \ldots & -2 \\ -1 & 1 & 0 & \ldots & 0 \\ -1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -2 & -1 & -1 & \ldots & -2 \end{pmatrix}_{K \times K}$$

Since $\hat{Y}$ is Gaussian, we get

$$E[S_Y^4] = \frac{1}{(K-1)^2} \sum_{i,j=1}^K E[(Y_i - \bar{Y})(Y_j - \bar{Y})^2] = \frac{1}{(K-1)^2} \sum_{i,j=1}^K \{ c_{ii} c_{jj} + 2c_{ij} \},$$

where $c_{ii} = (1 - 1/K) \sigma^2 - B/q + O(1/T)$ and $c_{ij} = -1/K \sigma^2 + B/(2q) I(|i - j| = 1) + O(1/T)$, for $i \neq j$. It implies that

$$\sum_{i,j=1}^K c_{ij}^2 = \sum_{i=1}^K c_{ii}^2 + \sum_{|i-j|=1} c_{ij}^2 + \sum_{|i-j|>1} c_{ij}^2 = K \left( 1 - \frac{1}{K} \right)^2 \sigma^4 + \frac{KB^2}{q^2} - 2(K-1)B \sigma^2$$

$$\quad \quad \quad + 2(K-1) \left( \frac{\sigma^4}{K^2} + \frac{B^2}{4q^2} - \frac{\sigma^2 B}{Kq} \right) + \frac{(K-1)(K-2)}{K^2} \sigma^4 + O(1/q)$$

$$= (K-1)\sigma^4 + O(K/q),$$

and

$$\sum_{i,j=1}^K c_{ii} c_{jj} = K^2 c_{11}^2 + O(K/q) = (K-1)^2 \sigma^4 - \frac{2BK(K-1)\sigma^2}{q} + O(K/q).$$

Therefore we get

$$E[S_Y^4] = \frac{K+1}{K-1} \sigma^4 - \frac{2BK\sigma^2}{(K-1)q} + O(1/T),$$

which implies

$$\text{var}(S_Y^2) = \frac{K+1}{K-1} \sigma^4 - \frac{2BK\sigma^2}{(K-1)q} - (\sigma^2 - B/q)^2 + O(1/T) = \frac{2\sigma^4}{K-1} + O(1/T).$$

Let $X = (X_1, X_2, \ldots, X_T)'$, $\tilde{\mu}_{GLS} = (l_T' \text{cov}(X)^{-1} l_T)^{-1} l_T' \text{cov}(X)^{-1} X$ and $\sigma^2_{GLS} = T \text{var}(\tilde{\mu}_{GLS}) = T (l_T' \text{cov}(X)^{-1} l_T)^{-1}$. Note that $\tilde{\mu}_{GLS} - \mu_0$ is independent of $S_Y$ and
Let \( \sigma_{GLS}^2 = \sigma^2 + O(1/T) \) (see Grenander and Rosenblatt [3]). Using similar arguments in Lemma 1 of Sun [6], we have

\[
P(\{T_k \leq x\} \leq x) = P \left( \frac{T(\hat{\mu}_{GLS} - \mu_0)^2/\sigma_{GLS}^2}{S_T^2/\sigma_{GLS}^2} \leq x^2 \right) + O(1/T)
\]

\[
= E[G_1(S_T^2/x^2) + O(1/T)
\]

\[
= G_1(x^2) + \frac{x^2}{\sigma^2} G_1'(x^2) E[S_T^2 - \sigma^2] + \frac{x^4}{2\sigma^4} G_1''(x^2) E[(S_T^2 - \sigma^2)^2] + O(1/T)
\]

\[
= G_1(x^2) - \frac{B K}{T \sigma^2} x^2 G_1'(x^2) + \frac{1}{K-1} x^4 G_1''(x^2) + O(1/T).
\]

\[\diamondsuit\]

1.2. Proof of the main results in Section 2.2. We first establish a high order expansion for Wald statistic based on the kernel \( G_{k,1}(r, t) = \sum_{j=1}^{K} \lambda_j \phi_j(r) \tilde{\phi}_j(t) \) in Lemma 1.6 below. Let \( \xi = (\xi_0, \xi_1, \ldots, \xi_K) \) with \( \xi_0 = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} (X_i - \mu_0) \) and \( \xi_j = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \phi_j(i/T) X_i \) for \( j = 1, 2, \ldots, K \), and \( \Sigma_\xi \) be the covariance matrix of \( \xi \). Define \( Q_{ij}(x) = P(\mathcal{F}(J) \leq x) \) for \( 1 \leq J \leq \infty \). We present the following lemma regarding the convergence rate of \( \Sigma_\xi \) for the basis functions \( \{\phi_j(t)\}_{j=1}^{K} \) without the mean zero and orthogonality assumption. Define \( R = (R_{ij})_{i,j=1}^{K} \) with \( R_{ij} = \int_{0}^{1} \phi_i(t) \tilde{\phi}_j(t) dt \), where \( \tilde{\phi}_j(t) = \phi_j(t) - \int_{0}^{1} \phi_j(t) dt \), and \( \bar{R} = \text{diag}(1, R) = (\bar{R}_{i,j})^{K}_{i,j=0} \).

**Lemma 1.5.** Assume the basis functions \( \{\phi_j(t)\}_{j=1}^{K} \) are bounded with finite discontinuous points and satisfy \( \sup_{\alpha \in (0,1)} \left\{ \frac{1}{\alpha} \int_{0}^{1} \phi_s(x) (\tilde{\phi}_r(x + \alpha) - \tilde{\phi}_r(x)) dx \right\} < \infty \), for \( 1 \leq s, r \leq K \). If \( \sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty \) and \( K \) is fixed, then we have \( \|\Sigma_\xi - \sigma^2 R\|_\infty = O(1/T) \).

**Proof of Lemma 1.5.** For \( s = 1, 2, \ldots, K \), we have

\[
\text{cov}(\xi_0, \xi_s) = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} \gamma_X(j - i) \phi_s^0 \left( \frac{j}{T} \right) = \frac{1}{T} \sum_{h=1-T}^{T-1} \gamma_X(h) \sum_{1 \leq i, h+i \leq T} \phi_s^0 \left( \frac{h+i}{T} \right).
\]

Simple algebra gives us

\[
\frac{1}{T} \sum_{1 \leq i, h+i \leq T} \phi_s^0 \left( \frac{h+i}{T} \right) = \left\{ \frac{h}{|h|} \sum_{i=1}^{T} \phi_s(i/T) - \frac{h}{|h|} \sum_{i=1}^{T} \phi_s(i/T) - \frac{h}{|h|} \sum_{i=1}^{T-|h|+1} \phi_s(i/T), \quad h > 0; \right.
\]

\[
\left. \frac{h}{|h|} \sum_{i=1}^{T} \phi_s(i/T) - \frac{h}{|h|} \sum_{i=1}^{T} \phi_s(i/T) - \frac{h}{|h|} \sum_{i=1}^{T-|h|+1} \phi_s(i/T), \quad h < 0. \right\}
\]

It implies that

\[
\text{cov}(\xi_0, \xi_s) = \frac{1}{T} \int_{0}^{1} \phi_s(t) dt \sum_{h=-\infty}^{+\infty} |h| \gamma_X(h)
\]

\[
- \frac{1}{T} \sum_{0 < h < T} \gamma_X(h) \left( \sum_{i=1}^{h} \phi_s(i/T) + \sum_{i=T-h+1}^{T} \phi_s(i/T) \right) + O(1/T^2).
\]
Note that the second term on the right hand side of (1) is of order $O(1/T)$ because the basis functions $\{\phi_s(t)\}$ are bounded. Consider the covariance between $\xi_s$ and $\xi_r$ with $1 \leq s, r \leq K$. Straightforward calculation yields

$$\text{cov}(\xi_s, \xi_r) = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} \phi^0_s \left( \frac{i}{T} \right) \phi^0_r \left( \frac{j}{T} \right) \gamma_X(i-j)$$

$$= \frac{1}{T} \sum_{h=1}^{T-1} \sum_{1 \leq j, j+h \leq T} \phi^0_s \left( \frac{j+h}{T} \right) \phi^0_r \left( \frac{j}{T} \right) \gamma_X(h)$$

$$+ \frac{1}{T} \sum_{h=1}^{T-1} \sum_{1 \leq j, j+h \leq T} \phi^0_s \left( \frac{j+h}{T} \right) \phi^0_r \left( \frac{j}{T} \right) \gamma_X(h)$$

$$+ \gamma_X(0) \frac{1}{T} \sum_{j=1}^{T} \phi^0_s \left( \frac{j}{T} \right) \phi^0_r \left( \frac{j}{T} \right) = I_1 + I_2 + I_3.$$

Notice that

$$\frac{1}{T} \sum_{1 \leq j \leq T} \phi^0_s \left( \frac{j}{T} \right) \phi^0_r \left( \frac{j}{T} \right) = \int_0^1 \tilde{\phi}_s(t) \tilde{\phi}_r(t) dt + C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t)) = R_{sr} + C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t)),$$

where $C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t))$ is of order $O(1/T)$. It is not hard to see that

$$I_1 = \frac{1}{T} \sum_{h=1}^{T-1} \frac{1}{T} \gamma_X(h) \left[ \sum_{j=1}^{T-h} \phi^0_r \left( \frac{j}{T} \right) \left\{ \phi^0_s \left( \frac{j+h}{T} \right) - \phi^0_s \left( \frac{j}{T} \right) \right\} \right], \quad \text{say } J_{1,T}$$

$$+ \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left\{ \sum_{j=1}^{h} \phi^0_r \left( \frac{j}{T} \right) \phi^0_s \left( \frac{j}{T} \right) - \sum_{j=h+1}^{T} \phi^0_r \left( \frac{j}{T} \right) \phi^0_s \left( \frac{j}{T} \right) \right\},$$

and

$$I_2 = \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left[ \sum_{j=1+|h|}^{T} \phi^0_r \left( \frac{j}{T} \right) \left\{ \phi^0_s \left( \frac{j+h}{T} \right) - \phi^0_s \left( \frac{j}{T} \right) \right\} \right], \quad \text{say } J_{2,T}$$

$$+ \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left\{ \sum_{j=1}^{T-h} \phi^0_r \left( \frac{j}{T} \right) \phi^0_s \left( \frac{j}{T} \right) - \sum_{j=1}^{h} \phi^0_r \left( \frac{j}{T} \right) \phi^0_s \left( \frac{j}{T} \right) \right\}.$$
Using (2), we have

\[
\text{cov}(\xi_s, \xi_r) = \left\{ \sigma^2 - \sum_{|h| \geq T} \gamma_X(h) \right\} \left\{ R_{sr} + C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t)) \right\} - \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left\{ \sum_{j=1}^{h} \phi_r^0 \left( \frac{j}{T} \right) \phi_s^0 \left( \frac{j}{T} \right) \right\} + \sum_{j=T-h+1}^{T} \phi_r^0 \left( \frac{j}{T} \right) \phi_s^0 \left( \frac{j}{T} \right) + J_{1,T} + J_{2,T}.
\]

(3)

Under the assumption that \(\sup_{\alpha \in [0,1]} \left| \frac{1}{T} \int_0^{1-\alpha} \tilde{\phi}_r(x)(\tilde{\phi}_s(x + \alpha) - \tilde{\phi}_s(x))dx \right| < \infty\), it is straightforward to see that

\[
|J_{1,T}| \leq \frac{1}{T} \sum_{h=1}^{T-1} |h\gamma_X(h)| \sup_{1 \leq k \leq T} \left| \frac{1}{h} \sum_{j=1}^{h} \phi_r^0 \left( \frac{j}{T} \right) \left\{ \phi_s^0 \left( \frac{j + h}{T} \right) - \phi_s^0 \left( \frac{j}{T} \right) \right\} \right| \leq C \frac{T}{T} \sum_{h=1}^{T-1} |h\gamma_X(h)| \left\{ \sup_{\alpha \in [0,1]} \left| \frac{1}{\alpha} \int_0^{1-\alpha} \tilde{\phi}_r(x)(\tilde{\phi}_s(x + \alpha) - \tilde{\phi}_s(x))dx \right| \right\},
\]

which implies that \(J_{1,T} = O(1/T)\). The same argument applies to \(J_{2,T}\). The proof is then complete.

The assumption regarding the basis functions in Lemma 1.5 is mild. If \(\{\phi_j(t)\}_{j=1}^K\) are lipschitz continuous of order one, then the assumption is satisfied.

**Lemma 1.6.** Suppose \(\sigma^2 > 0\) and the basis functions \(\{\phi_j(t)\}_{j=1}^K\) are mean zero and orthogonal. Under the assumptions in Lemma 1.5 and \(H_0\), we have \(\sup_{x \in [0, +\infty]} |\mathcal{N}_T(x; K)| = O(1/T)\) and

\[
(4) \quad \sup_{x \in [0, +\infty]} |P(F_T(K) \leq x) - Q_K(x) - \mathcal{N}_T(x; K)| = O(1/T^2),
\]

with \(K\) fixed and \(T \to \infty\).

**Proof of Lemma 1.6.** Note that when \(\{\phi_j(t)\}_{j=1}^K\) are mean zero and orthogonal, we have \(R = I_{K+1}\). It follows directly from Lemma 1.5 that \(\sup_{x \in [0, +\infty]} |\mathcal{N}_T(x; K)| = O(1/T)\). To show the second part, we first note that under the Gaussian assumption, the density function of \(\xi\) is given by \(f(u, \Sigma_\xi) = (2\pi)^{-(K+1)/2} |\Sigma_\xi|^{-1/2} \exp \left( -\frac{1}{2} u \Sigma_\xi^{-1} u \right) \).

Taking a Taylor expansion of the density function \(f(u, \Sigma_\xi)\) around the covariance matrix \(\sigma^2 I_{K+1}\), we get

\[
f(u, \Sigma_\xi) = f(u, \sigma^2 I_{K+1}) + \frac{\partial f}{\partial \text{vec}(\Sigma)}(u, \sigma^2 I_{K+1}) \text{vec}(\Sigma_\xi - \sigma^2 I_{K+1}) + R_T(u).
\]
By Lemma 1.4, the remainder term $R_T(u)$ satisfies that $\int_{\mathbb{R}^{K+1}} |R_T(u)| \, dv = O(1/T^2)$. Following Lemma 1.1, we have $\frac{\partial f}{\partial \text{vec}(\Sigma)}(u, \sigma^2 I_{K+1}) = f(u, \sigma^2 I_{K+1}) \left\{ \frac{1}{2\sigma^2} u \otimes u - \frac{1}{2\sigma^2} \text{vec}(I_{K+1}) \right\}$, which implies that

$$P(F_T(K) \leq x) = Q_K(x) \left\{ 1 - \frac{1}{2\sigma^2} \sum_{i=0}^{K} (\text{var}(\xi_i) - \sigma^2) \right\} + \zeta_T(x),$$

where $\zeta_T(x) = \frac{1}{2\sigma^2} \int_{\mathcal{F}(u; K) \leq x} f(u, \sigma^2 I_{K+1}) u^t \otimes u' \text{vec}(\Sigma_\xi - \sigma^2 I_{K+1}) \, du + \int_{\mathcal{F}(u; K) \leq x} R_T(u) \, du$

and $\mathcal{F}(u; K) = \frac{u_0}{\sum_{j=1}^{K} \lambda_j u_j^2}$. By letting $v = u/\sigma$ and noting that $E[\mathbf{1}[\mathcal{F}(v; K) \leq x] v, v_r] = 0$ for $s \neq r$, we obtain

$$\zeta_T(x) = \frac{1}{2\sigma^2} E[\mathbf{1}[\mathcal{F}(v; K) \leq x] (v \otimes v) \text{vec}(\Sigma_\xi - \sigma^2 I_{K+1})] + \int_{\mathcal{F}(u; K) \leq x} R_T(u) \, du$$

where $v = (v_0, v_1, \ldots, v_K)$ is a $(K+1)$-dimensional vector of i.i.d. standard normal random variables. Therefore, we get

$$\sup_{x \in [0, +\infty)} |P(F_T(K) \leq x) - Q_K(x)| = \sup_{x \in [0, +\infty)} \left| \int_{\mathcal{F}(u; K) \leq x} R_T(u) \, du \right|$$

$$\leq \int_{\mathbb{R}^{K+1}} |R_T(u)| \, du = O(1/T^2),$$

which completes the proof. \hfill \Box

**Lemma 1.7.** Let $\{\Sigma_{T,J+1}\} \subset \mathbb{R}^{(J+1)\times(J+1)}$ be an array of positive definite matrices with $J + 1 \leq T$. Assume that $\|\Sigma_{T,J+1} - \Sigma_{J+1}\|_\infty = O(J/T)$ for a sequence of positive definite matrices $\{\Sigma_{J}\}_{J=1}^\infty$ with $\sup_j \|\Sigma_{j}\|_2 < \infty$. If $J$ satisfies that $1/J + J^2/T \to 0$, then we have $\|\Sigma_{T,J+1} - \Sigma_{J+1}\|_\infty = O(J^2/T)$. \hfill \Box

**Proof.** Let $\Sigma_{T,J+1} = \Sigma_{J+1} + R_T,J+1$. For sufficiently large $T$, we have $\|\Sigma_{J+1}^{-1} R_T,J+1\|_2 \leq (J + 1)\|\Sigma_{J+1}^{-1}\|_2 \|R_T,J+1\|_\infty < 1$, where we are using the fact that $\|R_T,J+1\|_2 \leq (J + 1)\|R_T,J+1\|_\infty$. It follows that

$$\|\Sigma_{T,J+1}^{-1} - \Sigma_{J+1}^{-1}\|_\infty \leq \|\Sigma_{T,J+1}^{-1} - \Sigma_{J+1}^{-1}\|_2 \leq (J + 1)\|\Sigma_{J+1}^{-1}\|_2 \|R_T,J+1\|_\infty = O(J^2/T).$$

\hfill \Box
Choose $y$. The above arguments imply that when $T$ is sufficiently large. On the other hand, we have

$$\int_{y \in \mathbb{R}^{J^2 + 1}} \left| \text{vec}(RT_{J+1})' \frac{\partial^2 f}{\partial \text{vec}(\Sigma)^2} (y, \tilde{\Sigma}_{T,J+1}(y)) \text{vec}(RT_{J+1}) \right| dy = o(1/T).$$

**Proof.** Let $\tilde{R}_{T,J+1}(y) = \tilde{\Sigma}_{T,J+1}(y) - \sigma^2 I_{J+1}$. Note first that $\sup_{y \in \mathbb{R}^{J^2 + 1}} \tilde{R}_{T,J+1}(y)/\sigma^2 \leq (J+1)\sup_{y \in \mathbb{R}^{J^2 + 1}} |\tilde{R}_{T,J+1}(y)|/\sigma^2 \leq (J+1)|\Sigma_{T,J+1} - \sigma^2 I_{J+1}|/\sigma^2 < 1$, for large enough $T$. Following the arguments in Lemma 1.7, we know that

$$\sup_{y \in \mathbb{R}^{J^2 + 1}} |\tilde{\Sigma}_{T,J+1}^{-1}(y) - \sigma^{-2} I_{J+1}|_2 \leq \frac{C(J+1)}{1 - (J+1)|\Sigma_{T,J+1} - \sigma^2 I_{J+1}|/\sigma^2} = O(J^2/T).$$

Choose $r = J^3/T$. Then we have

$$y' \left( \tilde{\Sigma}_{T,J+1}^{-1}(y) - \frac{1}{(1+r)\sigma^2} I_{J+1} \right) y = y' \left( \tilde{\Sigma}_{T,J+1}^{-1}(y) - \frac{1}{\sigma^2} I_{J+1} \right) y + \frac{r}{(r+1)\sigma^2} |y|^2 \geq \left( \frac{r}{(r+1)\sigma^2} - |\tilde{\Sigma}_{T,J+1}^{-1}(y) - I_{J+1}/\sigma^2|_2 \right) |y|^2 \geq 0,$$

when $T$ is sufficiently large. On the other hand, we have

$$\sup_{y \in \mathbb{R}^{J^2 + 1}} |\tilde{\Sigma}_{T,J+1}^{-1}(y)| \leq \sup_{y \in \mathbb{R}^{J^2 + 1}} |\tilde{\Sigma}_{T,J+1}^{-1}(y)|_2^{J+1} \leq \left( \frac{1}{\sigma^2} + \frac{CJ^2}{T} \right)^{J+1} \left( \frac{1}{(r+1)\sigma^2} I_{J+1} \right) \left( 1 + \frac{C(r+1)J^2}{T} \right)^{J+1} \leq \frac{1}{(r+1)\sigma^2 I_{J+1}} (1 + C_T)^{(1/r)(J+1)r} \leq C \frac{1}{(r+1)\sigma^2 I_{J+1}}.$$

The above arguments imply that $f(y, \tilde{\Sigma}_{T,J+1}(y)) \leq C f(y, (1+r)\sigma^2 I_{J+1})$ for all $y$. Therefore we get

$$\int_{y \in \mathbb{R}^{J^2 + 1}} \left| \text{vec}(RT_{J+1})' \frac{\partial^2 f}{\partial \text{vec}(\Sigma)^2} (y, \tilde{\Sigma}_{J+1}(y)) \text{vec}(RT_{J+1}) \right| dy \leq C \int_{y \in \mathbb{R}^{J^2 + 1}} \left| \text{vec}(RT_{J+1})' \left( (\tilde{\Sigma}_{J+1}(y)y) \otimes (\tilde{\Sigma}_{J+1}(y)y) - \text{vec}(\tilde{\Sigma}_{J+1}(y)y) \right) \right| \text{vec}(RT_{J+1}) \left| f(y, (1+r)\sigma^2 I_{J+1}) \right| dy$$

$$+ C \int_{y \in \mathbb{R}^{J^2 + 1}} \left| \text{vec}(RT_{J+1})' \left( \tilde{\Sigma}_{J+1}(y)y y' \tilde{\Sigma}_{J+1}(y) \right) \otimes \tilde{\Sigma}_{J+1}(y) + \tilde{\Sigma}_{J+1}(y) \otimes \tilde{\Sigma}_{J+1}(y) y y' \tilde{\Sigma}_{J+1}(y) \right| \left| f(y, (1+r)\sigma^2 I_{J+1}) \right| dy \leq C J^6/T^2 = o(1/T),$$
where the first inequality in the last row comes from the fact that $\sup_{y \in \mathbb{R}^{J+1}} |\Sigma_{J+1}^{-1}(y) - \sigma^{-2} I_{J+1}| |\leq \sup_{y \in \mathbb{R}^{J+1}} |\Sigma_{J+1}^{-1}(y) - \sigma^{-2} I_{J+1}|_2 = O(J^2/T)$.

**Lemma 1.9.** Recall that $Q_J(x) = P(\mathcal{F}(J) \leq x)$ for $1 \leq J \leq \infty$. We have

\begin{equation}
\sup_{x \in [0, +\infty)} |Q_J(x) - Q_\infty(x)| = O \left( \sum_{j=J+1}^{\infty} \lambda_j \right).
\end{equation}

**Proof.** Let $\{v_j\}_{j=0}^{\infty}$ be a sequence of i.i.d. standard normal random variables. Define $U(J) = \sum_{j=1}^{J} 2 \lambda_j v_j^2$, $V(J) = \sum_{j=J+1}^{\infty} \lambda_j v_j^2$ and $Q_J = v_0^2/U(J)$ for $2 \leq J \leq \infty$. For any $x \in [0, +\infty)$ and large enough $J$ with $J \geq 3$, we have,

\begin{align*}
|Q_J(x) - Q_\infty(x)| &\leq |E[E[I\{Q_J \leq x\}|U(J)] - E[E[I\{Q_\infty \leq x\}|U(\infty)]|| \\
&\leq |E[G_1(xU(J)) - E[G_1(xU(\infty))]|| \\
&\leq |E[xV(J)G_1'(xU(J))]| = \left| E \left[ \frac{V(J)}{U(J)} xU(J)G_1'(xU(J)) \right] \right| \\
&\leq CE \left[ \frac{V(J)}{U(J)} \right] \leq CE[V(J)]E \left[ \frac{1}{U(J)} \right] \leq C \sum_{j=J+1}^{\infty} \lambda_j,
\end{align*}

where $U(J) \leq U(J) \leq U(J) + V(J)$ and $C$ does not depend on $x$. Note that we are using the mean value theorem, and the facts that $E[1/U(J)] \leq E \left[ 1/(\lambda_1^2) \right] < \infty$ and $\sup_{x \in [0, +\infty)} |xG_1'(x)| < \infty$.

**Lemma 1.10.** Let $V_T(J) = \sum_{j=J+1}^{\infty} 2 \lambda_j^2$. Assume that $\sup_{1 \leq i \leq \infty} \sup_{t \in [0,1]} \phi_i(t) < \infty$ and $\{X_i\}$ is a stationary Gaussian time series. Then we have $E[V_T^2(J) = O((\sum_{j=J+1}^{\infty} \lambda_j)^2)$.

**Proof.** Let $\sigma_{ij} = \gamma_X(i-j)$. For $i, j \geq J+1$, we have

\begin{align*}
E[\xi_i^2 \xi_j^2] &= \frac{1}{T^2} \sum_{i_1, i_2=1}^{T} \sum_{j_1, j_2=1}^{T} \phi_{i_1}^0(i_1/T)\phi_{i_2}^0(i_2/T)\phi_{j_1}^0(j_1/T)\phi_{j_2}^0(j_2/T)E[(X_{i_1} - \mu)(X_{i_2} - \mu)(X_{j_1} - \mu)(X_{j_2} - \mu)] \\
&= \frac{1}{T^2} \sum_{i_1, i_2=1}^{T} \sum_{j_1, j_2=1}^{T} \phi_{i_1}^0(i_1/T)\phi_{i_2}^0(i_2/T)\phi_{j_1}^0(j_1/T)\phi_{j_2}^0(j_2/T) \sigma_{i_1j_1} \sigma_{i_2j_2} + \sigma_{i_1j_2} \sigma_{i_2j_1} + \sigma_{i_1j_2} \sigma_{i_2j_1} \\
&= I_{1,T} + I_{2,T} + I_{3,T}.
\end{align*}

For the first term, we have

\begin{align*}
I_{1,T} &= \left( \frac{1}{T} \sum_{i_1, i_2=1}^{T} \phi_{i_1}^0(i_1/T)\phi_{i_2}^0(i_2/T) \sigma_{i_1i_2} \right) \left( \frac{1}{T} \sum_{j_1, j_2=1}^{T} \phi_{j_1}^0(j_1/T)\phi_{j_2}^0(j_2/T) \sigma_{j_1j_2} \right) = L_{1,T}L_{2,T}.
\end{align*}
Note that
\[
|L_{1,T}| = \left| \frac{1}{T} \sum_{h=1-T}^{T-1} \sum_{1 \leq i_1, i_1 + h \leq T} \phi_{10}(i_1/T) \phi_{0}(h/T) \gamma_X(h) \right | \\
\leq C \sum_{h=-\infty}^{+\infty} |\gamma_X(h)| \frac{1}{T} \sum_{1 \leq i_1 \leq T} |\phi_{0}(i_1/T)| \leq C \sum_{h=-\infty}^{+\infty} |\gamma_X(h)|,
\]
which implies that \(|I_{1,T}| \leq C(\sum_{h=-\infty}^{+\infty} |\gamma_X(h)|)^2\). Similar arguments apply to the other terms \(I_{2,T}\) and \(I_{3,T}\). We then have \(\sup_{1 \leq i,j \leq \infty} E(\xi_1^2 \xi_j^2) < C\). Therefore, we obtain \(E[V_T(J)^2] = \sum_{i,j=J}^{\infty} \sum_{j=J+1}^{\infty} \lambda_i \lambda_j E[\xi_i^2 \xi_j^2] \leq C \left( \sum_{i,J=J+1}^{\infty} \lambda_i \right)^2\).

\(\diamond\)

**Lemma 1.11.** Assume the eigenfunctions are continuously differentiable, mean zero and uniformly bounded, and \(\sum_{j=1}^{\infty} \lambda_j < \infty\). Suppose that \(\{X_i\}\) is a stationary Gaussian time series with \(\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty\). When \(1/J + J/T \to 0\), we have
\[
\sup_{x \in [0, +\infty)} \left| P(F_T(J) \leq x) - P(F_T(\infty) \leq x) \right| = O \left( \left( \sum_{j=J+1}^{\infty} \lambda_j \right)^{1/3} \right).
\]
Recall that \(F_T(J) = \frac{\xi_J^2}{\sum_{j=1}^{J} \lambda_j \xi_j^2}\) for \(J = 1, 2, \ldots, \infty\).

**Proof.** Let \(R_T(J) = F_T(J) - F_T(\infty) = \frac{\xi_J^2 V_T(J)}{(\sum_{j=1}^{\infty} \lambda_j \xi_j^2) (\sum_{j=1}^{J} \lambda_j \xi_j^2)}\). For any \(\delta > 0\), we have
\[
P(F_T(\infty) \leq x - \delta) - P(|R_T(J)| \geq \delta) \leq P(F_T(J) \leq x) \leq P(F_T(\infty) \leq x + \delta) + P(|R_T(J)| \geq \delta).
\]
Observe that
\[
P(|R_T(J)| \geq \delta) \leq \frac{E[|R_T(J)|]}{\delta} \leq \left( \frac{E[V_T(J)]}{\delta} \right)^{1/2} \left( E \left[ \frac{\xi_J^4}{(\sum_{j=1}^{J} \lambda_j \xi_j^2)^4} \right] \right)^{1/2}.
\]
Choose a fixed \(J_0 \geq 9\), denote by \(\Sigma_{T,J_0+1}\) the covariance matrix of \((\xi_0, \ldots, \xi_{J_0})\). By Lemma 1.5, we know that \(|\Sigma_{T,J_0+1} - \sigma^2 I_{J_0+1}|_2 \leq (J_0 + 1)|\Sigma_{T,J_0+1} - \sigma^2 I_{J_0+1}|_\infty = O(1/T)\). For large enough \(T\), we have \(|\Sigma_{T,J_0+1}|_2 \leq 2\sigma^2\). Let \(\lambda = \min(1, \frac{1}{2\sigma^2}) > 0\), we know that \(\Sigma_{T,J_0+1}^{-1} - \lambda I_{J_0+1}\) is semi-positive definite, i.e., for any \(x \in \mathbb{R}^{J_0+1}, x^T \Sigma_{T,J_0+1}^{-1} x \geq \lambda x^T x\). Using similar arguments in Lemma 1.3, we know that \(|\Sigma_{T,J_0+1}|^{-1} \leq \lambda I_{J_0+1}\).
In what follows, we show that \( \sup_x \{ R_T(J) \} \geq \delta \leq C \left( \sum_{j=J+1}^{\infty} \lambda_j / \delta \right) \).

In what follows, we show that \( \sup_{x \in [0, \infty)} |P(F_T(\infty) \leq x - \delta) - P(F_T(\infty) \leq x| \leq C \sqrt{\delta} \) for any \( \delta > 0 \). Let \( X = (X_1, X_2, \ldots, X_T)' \), \( l_T = (1, 1, \ldots, 1)' \), \( X^* = X - l_T \mu_0 \) and \( \Omega_T = \text{cov}(X) \). Then the GLS estimate of \( \mu \) is given by \( \hat{\mu}_{\text{GLS}} = (l_T' \Omega_T^{-1} l_T)^{-1} l_T' \Omega_T^{-1} X \) and \( \hat{\mu}_{\text{OLS}} - \mu_0 = \hat{\mu}_{\text{GLS}} - \mu_0 + \frac{1}{l_T} l_T' \hat{X} \), where \( \hat{X} = (I_T - l_T(l_T' \Omega_T^{-1} l_T)^{-1} l_T' \Omega_T^{-1}) X^* \).

The following facts which can be found in Sun et al. [7] play an important role in the proof presented below: (1) \( \hat{\mu}_{\text{GLS}} - \mu_0 \) is independent of \( \hat{X} \); (2) \( \hat{\mu}_{\text{GLS}} - \mu_0 \) is independent of \( X - l_T \hat{\mu}_{\text{OLS}} \). Notice that \( \hat{D}_T = \sum_{j=1}^{\infty} \lambda_j \xi_j^2 = \frac{1}{T} (X - l_T \hat{\mu}_{\text{OLS}})' g_T (X - l_T \hat{\mu}_{\text{OLS}}) \) with \( g_T = (g(i/T, j/T))_{i,j=1}^{\infty} \). Then \( \hat{\mu}_{\text{GLS}} - \mu_0 \) is also independent of \( \hat{D}_T \). Define \( \sigma_{\text{GLS}}^2 = T \text{var}(\hat{\mu}_{\text{GLS}}) = T(l_T' \Omega_T^{-1} l_T)^{-1} \). Denote by \( \Phi_{\text{norm}} \) and \( \phi_{\text{norm}} \) the cumulative distribution function and density function of the standard normal distribution. Therefore, we get

\[
P(F_T(\infty) \leq x) = 2P(\frac{\sqrt{T} (\hat{\mu}_{\text{OLS}} - \mu_0)}{\sqrt{\hat{D}_T}} \leq x) - 1 = 2P(\sqrt{T} (\hat{\mu}_{\text{OLS}} - \mu_0) \leq \sqrt{x \hat{D}_T}) - 1
\]

\[
= 2P(\sqrt{T} (\hat{\mu}_{\text{GLS}} - \mu_0)/\sigma_{\text{GLS}} \leq \sqrt{x \hat{D}_T}/\sigma_{\text{GLS}}) - 1
\]

\[
= 2E \left[ \Phi_{\text{norm}} \left( \sqrt{x \hat{D}_T}/\sigma_{\text{GLS}} - l_T' \hat{X}/(\sqrt{T} \sigma_{\text{GLS}}) \right) \right] - 1,
\]

which implies that for \( x, \delta \geq 0 \) with \( x - \delta \geq 0 \),

\[
|P(F_T(\infty) \leq x + \delta) - P(F_T(\infty) \leq x)| \leq 2E \left[ \Phi_{\text{norm}} \left( (x + \delta) \hat{D}_T/\sigma_{\text{GLS}} - l_T' \hat{X}/(\sqrt{T} \sigma_{\text{GLS}}) \right) \right] - 2E \left[ \Phi_{\text{norm}} \left( x \hat{D}_T/\sigma_{\text{GLS}} - l_T' \hat{X}/(\sqrt{T} \sigma_{\text{GLS}}) \right) \right]
\]

\[
\leq C \sqrt{\delta} E[\sqrt{\hat{D}_T/\sigma_{\text{GLS}}}] < C \sqrt{\delta} (E[\hat{D}_T])^{1/2}/\sigma_{\text{GLS}} < C \sqrt{\delta},
\]
with \( \sqrt{xD_T/\sigma_{GLS}} \leq a^* \leq \sqrt{(x + \delta)D_T/\sigma_{GLS}} \) or \( \sqrt{(x - \delta)D_T/\sigma_{GLS}} \leq a^* \leq \sqrt{xD_T/\sigma_{GLS}} \). Here we are using the fact that \( \sigma_{GLS}^2 = \sigma^2 + O(1/T) \) and \( E[D_T] \) is uniformly bounded for all \( T \). Choosing \( \delta = (\sum_{j=J+1}^{\infty} \lambda_j)^{2/3} \), the conclusion follows in view of (6), (7) and (8).

\[ \square \]

**Lemma 1.12.** Under the assumptions in Theorem 2.2, we have \( ||\Sigma_{\xi,J+1} - \sigma^2 I_{J+1}||_\infty = O(J/T) \) with \( J \leq T \), where \( \Sigma_{\xi,J+1} \) denotes the covariance matrix of \( (\xi_0, \xi_1, \ldots, \xi_J) \).

**Proof of Lemma 1.12.** Using the arguments in Lemma 1.5, we have for any \( 1 \leq s \leq J \),

\[
|\text{cov}(\xi_s, \xi_r)| \leq C \left[ \frac{1}{T^2} \sum_{i=1}^{T} \phi_s(i/T) \right] + \frac{1}{T} \sum_{0 < h < T} \left| \gamma_X(h) \left\{ \sum_{i=1}^{h} \phi_s(i/T) + \sum_{i=T-h+1}^{T} \phi_s(i/T) \right\} \right| \leq C/T,
\]

where \( C \) is a generic constant which does not depend on \( s \). Again by the arguments in Lemma 1.5, we have

\[
|\text{cov}(\xi_s, \xi_r) - \sigma^2 \delta_{sr}| \leq \sum_{h=1-T}^{T-1} |\gamma_X(h)C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t))| + \sum_{|h| \geq T} |\gamma_X(h)| \delta_{sr}
\]

\[
+ \frac{1}{T} \sum_{h=1}^{T-1} \left| \gamma_X(h) \left\{ \sum_{j=1}^{h} \phi_r^0 \left( \frac{j}{T} \right) \phi_s^0 \left( \frac{j}{T} \right) + \sum_{j=T-h+1}^{T} \phi_r^0 \left( \frac{j}{T} \right) \phi_s^0 \left( \frac{j}{T} \right) \right\} \right|
\]

\[
+ |J_{1,T}| + |J_{2,T}|, \quad 1 \leq s, r \leq J,
\]

where \( J_{1,T}, J_{2,T} \) and \( C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t)) \) are defined in the proof of Lemma 1.5. By the Trapezoidal rule and the assumption that \( \sup_{t \in [0,1]} |\phi_r''(t)| < CJ^2 \), we have

\[ (9) \quad |C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t))| \leq C(J^2/T^2 + 1/T), \]

which implies that \( |\text{cov}(\xi_s, \xi_r) - \sigma^2 \delta_{sr}| \leq CJ/T + |J_{1,T}| + |J_{2,T}| \) for \( J \leq T \). By the mean value theorem and the assumption that \( \sup_{1 \leq i \leq J, \sup_{t \in [0,1]} |\phi_r'(t)| < CJ \), we get

\[ (10) \quad |J_{1,T}| \leq \frac{1}{T} \sum_{h=1}^{T-1} |\gamma_X(h)| \left| \sum_{j=1}^{T-h} \phi_r^0 \left( \frac{j}{T} \right) \left\{ \phi_s^0 \left( \frac{j+h}{T} \right) - \phi_s^0 \left( \frac{j}{T} \right) \right\} \right|
\]

\[
\leq \frac{CJ}{T} \sum_{h=1}^{T-1} |h\gamma_X(h)| \frac{1}{T} \sum_{j=1}^{T-h} \left| \phi_r^0 \left( \frac{j}{T} \right) \right| \leq \frac{CJ}{T}.
\]

Using the same argument for \( J_{2,T} \), we get \( |\text{cov}(\xi_s, \xi_r) - \sigma^2 \delta_{sr}| \leq CJ/T \), which completes the proof.

\[ \square \]
Proof of Theorem 2.2. Suppose $J = o(T^{1/6})$. By Lemma 1.12, we know $\|\Sigma_{\xi J} - J^2 I_{J+1}\|_\infty = O(J/T)$. Using Lemma 1.8 and similar arguments in the proof of Lemma 1.6, we can show that

$$\sup_{x \in \mathbb{R}} |P(F_T(J) \leq x) - Q_J(x) - N_T(x; J)| = o(1/T),$$

where $N_T(x; J) = \frac{1}{2\sigma^2} \sum_{i=0}^J (\text{var}(\xi_i) - \sigma^2) E[(v_i^2 - 1)I\{F(v; J) \leq x\}]$ with $v = (v_0, v_1, \ldots, v_J) \sim N(0, I_{J+1})$. Next, we show that $N_T(x; J)$ converges uniformly as $J \to +\infty$. Note first that

$$\sup_{x \in (0, +\infty)} |N_T(x; J + p) - N_T(x; J)| \leq \sup_{x \in (0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=J+1}^{J+p} (\text{var}(\xi_i) - \sigma^2) E[(v_i^2 - 1)I\{F(v; J + p) \leq x\}] \right|$$

$$+ \sup_{x \in (0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=1}^J (\text{var}(\xi_0) - \sigma^2) E[(v_i^2 - 1)(I\{F(v; J + p) \leq x\} - I\{F(v; J) \leq x\})] \right|$$

$$+ \sup_{x \in (0, +\infty)} \left| \frac{1}{2\sigma^2} (\text{var}(\xi_0) - \sigma^2) E[(v_0^2 - 1)(I\{F(v; J + p) \leq x\} - I\{F(v; J) \leq x\})] \right| = I_1 + I_2 + I_3,$$

for any $J, p \in \mathbb{Z}^+$. In view of (9) and (10), we have

$$|\text{var}(\xi_i) - \sigma^2| < C(i/T + i^2/T^2),$$

for $1 \leq i < \infty$. Hence we get, for sufficiently large $J$,

$$I_1 \leq \frac{1}{2\sigma^2} \sup_{x \in (0, +\infty)} \sum_{i=J+1}^{J+p} (\text{var}(\xi_i) - \sigma^2) E\left[ (v_i^2 - 1)G_1 \left( x \sum_{j=1}^{J+p} \lambda_j v_j^2 \right) \right]$$

$$\leq \frac{C}{T} \sup_{x \in (0, +\infty)} \sum_{i=J+1}^{J+p} (i + i^2/T) E\left[ (v_i^2 - 1)G_1 \left( x \sum_{j=1}^{J+p} \lambda_j v_j^2 \right) \right]$$

$$\leq \frac{C}{T} \sup_{x \in (0, +\infty)} \sum_{i=J+1}^{J+p} \sum_{j=1}^{J+1} (i + i^2/T) \lambda_i E\left[ v_i^2 (v_i^2 + 1) x G_1'(y_i) \right]$$

$$\leq \frac{C}{T} \sum_{i=J+1}^{J+p} (i + i^2/T) \lambda_i E\left[ v_i^2 (v_i^2 + 1) \right]$$

$$\leq \frac{C}{T} \sum_{i=J+1}^{J+p} (i + i^2/T) \lambda_i E\left[ v_i^2 (v_i^2 + 1) \right]$$

$$\leq \frac{C}{T} \sum_{i=J+1}^{J+p} \left( i \lambda_i + \frac{1}{T} \sum_{j=J+1}^{+\infty} i^2 \lambda_j \right) = O\left( \frac{J - a^2 + 2}{T} \right),$$
where \( y_i = x(\sum_{j \neq i} \lambda_j v_j^2 + \alpha_i \lambda_i^2 v_i^2) \) for some \( 0 \leq \alpha_i \leq 1 \). On the other hand, we get
\[
I_2 \leq \frac{C J}{T} \sup_{x \in [0, +\infty)} \left\{ \left( \frac{\sum_{i=1}^{J} \left( v_i^2 - 1 \right)}{\sum_{j=J+1}^{J+p} \lambda_j v_j^2} \right)^{1/2} \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{j=J+1}^{J+p} \lambda_j v_j^2 \right)^2 \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{j=J+1}^{J+p} \lambda_j v_j^2 \right)^2 \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{j=J+1}^{J+p} \lambda_j v_j^2 \right)^2 \right] \right\} \\
\leq \frac{C J}{T} \sup_{x \in [0, +\infty)} \left\{ \left( \frac{\sum_{i=1}^{J} \left( v_i^2 - 1 \right)}{\sum_{j=J+1}^{J+p} \lambda_j v_j^2} \right)^{1/2} \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{j=J+1}^{J+p} \lambda_j v_j^2 \right)^2 \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{j=J+1}^{J+p} \lambda_j v_j^2 \right)^2 \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{j=J+1}^{J+p} \lambda_j v_j^2 \right)^2 \right] \right\} \\
\leq \frac{C J^2}{T} \left( \sum_{j=J+1}^{J+p} \lambda_j \right) \leq \frac{C J^2}{T} \left( \sum_{j=J+1}^{J+p} \lambda_j \right) \cdot \frac{O \left( \frac{J^{-a+3}}{T} \right)}.
\]

Finally using the Cauchy-Schwarz inequality and similar arguments in Lemma 1.9, we know
\[
I_3 \leq \frac{C}{T} \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \\
\leq \frac{C}{T} \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \\
\leq \frac{C}{T} \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \cdot \sup_{x \in [0, +\infty)} \left\{ E \left[ \left( \sum_{i=1}^{J} \left( v_i^2 - 1 \right)^2 \right)^{1/2} \right] \right\} \\
\leq \frac{O \left( \frac{J^{-a+1/2}}{T} \right)}.
\]

Therefore, it is straightforward to see that \( \sup_{x \in [0, +\infty)} |N_T(x; J) - N_T(x; \infty)| = O \left( \frac{J^{-a+1/2}}{T} \right) \) and \( \sup_{x \in [0, +\infty)} |N_T(x; \infty)| = O(1/T) \), which imply that
\[
\sup_{x \in [0, +\infty)} |P(F_T(J) \leq x) - Q_J(x) - N_T(x; \infty)| = o(1/T),
\]
for \( J = o(T^{1/6}) \). Let \( J = T^{1/6}/\log(T) \) and note that \( (\sum_{j=J+1}^{J+p} \lambda_j)^{1/3} = o(1/T) \). The proof is completed in view of Lemma 1.9 and Lemma 1.11.

**Proof of Proposition 2.3.** Under the assumption that \( \sup_{x \in \mathbb{R}} |K(x)|dx < \infty \), we have
\[
\sum_{j=1}^{+\infty} (\tilde{\lambda}_{j,b})^2 = \int_0^1 \int_0^1 \tilde{G}_b^2(r,t)drdt \leq \sup_{t \in [0,1]} \int_0^1 \tilde{G}_b^2(t,r)dr \leq 4 \sup_{t \in [0,1]} \int_0^1 |\tilde{G}_b(r,t)|dr \leq 16 \sup_{t \in [0,1]} \int_{-t}^{+t} |\tilde{K}_b(r)|dr \leq Cb,
\]
and \( \tilde{\lambda}_{1,b} \leq (\int_0^1 \int_0^1 \tilde{G}_b^2(r,t)drdt)^{1/2} \leq C\sqrt{b} \). Suppose \( \{\tilde{a}_i\} \) is a sequence of random variables such that \( 0 \leq \tilde{a}_i \leq 1 \). Using the fact that \( \sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} = \int_0^1 \tilde{G}_b(r,t)dr = 1 + O(b) \), we get
\[
\left( \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 + \tilde{a}_i \tilde{\lambda}_{i,b} v_i^2 - 1 \right)^2 = \sup_i \left\{ \sum_{j \neq i}^{+\infty} (\tilde{\lambda}_{j,b})^2 + (\tilde{\lambda}_{i,b})^2 E(\tilde{a}_i v_i^2 - 1)^2 \right\} + O(b) \leq Cb.
\]
By the Taylor expansion, we have

\[
\begin{align*}
N_{T,b}(x; \infty) &= \frac{1}{2\sigma^2} \sum_{i=1}^{+\infty} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E\left( (v_i^2 - 1)I\{F(v; \infty) \leq x \} \right) + O(1/T) \\
&= \frac{1}{2\sigma^2} \sum_{i=1}^{+\infty} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E\left( (v_i^2 - 1)G_1 \left( x \sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 \right) \right) + O(1/T) \\
&= \frac{x}{\sigma^2} \sum_{i=1}^{+\infty} \tilde{\lambda}_{i,b} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E\left( G_1 \left( x \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 \right) \right) \\
&\quad + \frac{x^2}{4\sigma^4} \sum_{i=1}^{+\infty} (\tilde{\lambda}_{i,b})^2 (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E\left( v_i^4 (v_i^2 - 1)G''_1 \left( x \left( \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 + a_i \tilde{\lambda}_{i,b} v_i^2 \right) \right) \right) + O(1/T) \\
&= I_{1T,b} + I_{2T,b} + O(1/T),
\end{align*}
\]

where 0 ≤ a_i ≤ 1. Let \( A_{i,b} = E \left[ G_1' \left( x \left( \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 \right) \right) \right] \), \( B_{i,b} = \tilde{\lambda}_{i,b} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) \), \( C_{i,b} = \sum_{j=1}^{i} B_{j,b} \) and \( S_{N,b} = \sum_{i=1}^{N} A_{i,b} B_{i,b} \). Using summation by parts, we have \( S_{N,b} = A_{N,b} C_{N,b} - \sum_{i=1}^{N-1} (A_{i+1,b} - A_{i,b}) C_{i,b} \). Note that \( \{A_{i,b}\}_{i=1}^{+\infty} \) is a nonincreasing sequence and \( \lim_{b \to 0} \sup \{A_{i,b}\} = G_1'(x) \) as seen from (13). Let \( D_{T,b} \) be defined by replacing \( \phi_j \) and \( \lambda_j \) with \( \phi_{j,b} \) and \( \tilde{\lambda}_{j,b} \) in the definition of \( D_T \). It is not hard to see that as \( b + 1/(bT) \to 0 \),

\[
\lim_{N \to +\infty} A_{N,b} C_{N,b} = \sigma^2 G'(x) \left( E[D_{T,b}]/\sigma^2 - \sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} \right) \left( 1 + o(1) \right)
\]

\[
= - \frac{G'(x) g_4 \sum_{i=-\infty}^{+\infty} |h|^q \gamma_X(h)}{(bT)^q} \left( 1 + o(1) \right) + O(1/T),
\]

where we have used the fact \( E[D_{T,b}]/\sigma^2 - \sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} = - \frac{g_4 \sum_{i=-\infty}^{+\infty} |h|^q \gamma_X(h)}{\sigma^2 (bT)^q} \left( 1 + o(1) \right) + O(1/T) \), which can be proved by using similar arguments in the proof of Lemma 2 in Sun et al. [7]. On the other hand, observe that \( |\sum_{i=1}^{N-1} (A_{i+1,b} - A_{i,b}) C_{i,b}| \leq \sup_{i \in \mathbb{N}} |C_{i,b}| \sum_{i=1}^{N-1} |A_{i,b} - A_{i+1,b}| \leq \sup_{i \in \mathbb{N}} |C_{i,b}| (A_{1,b} - \lim_{N \to +\infty} A_{N,b}) = o(\lim_{N \to +\infty} A_{N,b}) \) as \( b + 1/(bT) \to 0 \), for all \( N \). Hence we get

\[
I_{1T,b} = - \frac{x G'(x) g_4 \sum_{h=-\infty}^{+\infty} |h|^q \gamma_X(h)}{\sigma^2 (bT)^q} \left( 1 + o(1) \right) + O(1/T),
\]

Define \( H_{i,b} = \tilde{\lambda}_{i,b} E \left[ v_i^4 (v_i^2 - 1)G''_1 \left( x \left( \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 + a_i \tilde{\lambda}_{i,b} v_i^2 \right) \right) \right] \) and \( S_{N,b} = \sum_{i=1}^{+\infty} H_{i,b} B_{i,b} \). Again using summation by parts, we obtain \( S_{N,b} = H_{N,b} C_{N,b} - \sum_{i=1}^{N-1} (H_{i+1,b} - H_{i,b}) C_{i,b} \). By (13), we can show that \( \sup_{i} |H_{i,b}/\tilde{\lambda}_{i,b} - 12G''_1(x)| = O(\sqrt{b}) \). Therefore,
we get \( \lim_{N \to +\infty} C_{N,b} H_{N,b} = o(\lim_{N \to +\infty} C_{N,b}) \) and
\[
\left| \sum_{i=1}^{N-1} (H_{i+1,b} - H_{i,b})C_{i,b} \right| \leq \sup_{i \in \mathbb{N}} |C_{i,b}| \left\{ \sum_{i=1}^{N-1} \left[ |H_{i+1,b} - 12\tilde{\lambda}_{i+1,b}G_1''(x)| + 12G_2''(x)(\tilde{\lambda}_{i,b} - \tilde{\lambda}_{i+1,b}) \right. \right. \\
\left. \left. + |12\tilde{\lambda}_{i,b}G_1''(x) - H_{i,b}| \right] \right\} = O\left( \sqrt{b} \lim_{N \to +\infty} C_{N,b} \right).
\]
The conclusion follows from the above arguments by noting that \( I_{2T,b} = o(I_{1T,b}) \).

\[ \Diamond \]

1.3. Proof of the main results in Section 3.

**Lemma 1.13.** Let \( \omega_l(x) = (1-\lfloor x/l \rfloor)\mathbb{I}\{ \lfloor x/l \rfloor < 1 \} \). Suppose that \( m^3/l^2 + (ml)^3/T + 1/m \to 0 \) and \( \sum_{h=-\infty}^{+\infty} h^2|\gamma_X(h)| < \infty \). Then under the Gaussian assumption, we have
\[
\sup_{0 \leq k \leq m} \left| \sum_{h=1-l}^{T-1} g_k,T(h)\omega_l(h)\gamma_X(h) - \sum_{h=1-T}^{T-1} g_k,T(h)\gamma_X(h) \right| = O_p(\sqrt{m^3/l^2 + (ml)^3/T}),
\]
where \( |g_k,T(h)| = C(|h| + |h| + 1) \) for \( 0 \leq k \leq m \) and \( |h| \leq T \), and the constant \( C \) does not depend on \( k \) and \( h \).

**Proof of Lemma 1.13.** Note first that for any \( \epsilon > 0 \),
\[
P\left( \sup_{0 \leq k \leq m} \left| \sum_{h=1-l}^{T-1} g_k,T(h)\omega_l(h)\gamma_X(h) - \sum_{h=1-T}^{T-1} g_k,T(h)\gamma_X(h) \right| > \epsilon \right) \\
\leq \sum_{k=0}^{m} \left( \frac{1}{C^2} \sum_{k=0}^{m} E \left| \sum_{h=1-l}^{T-1} g_k,T(h)\omega_l(h)\gamma_X(h) - \sum_{h=1-T}^{T-1} g_k,T(h)\gamma_X(h) \right|^2 \right) \\
\leq \frac{2}{C^2} \sum_{k=0}^{m} E \left( \sum_{h=1-l}^{T-1} g_k,T(h)\omega_l(h)\gamma_X(h) - \gamma_X(h) \right)^2 + \frac{Cm^3}{T^2C^2}.
\]
Let \( z_i = X_i - E[X_i] \) and \( w_i|h| = z_i z_i + |h| - \gamma_X(h) \). Simple calculation yields that
\[
\left| \sum_{h=1-l}^{T-1} g_k,T(h)\omega_l(h)\gamma_X(h) - \gamma_X(h) \right| = \left| \sum_{h=1-l}^{T-1} g_k,T(h)\omega_l(h)\{ \gamma_X(h) - \gamma_X(h) \} \right| + C(k+1)/l \\
\leq \sum_{h=1-l}^{T-1} g_k,T(h)\omega_l(h) \left\{ \frac{1}{T} \sum_{i=1}^{T-|h|} w_i|h| \right\} + \sum_{h=1-l}^{T-1} g_k,T(h)\omega_l(h) \left\{ \frac{T - |h|}{T} \right\} \\
+ \sum_{h=1-l}^{T-1} g_k,T(h)\omega_l(h) \left\{ \frac{z_i + z_i + |h|}{T} \right\} + C(k+1)/l := I_{1T} + I_{2T} + I_{3T} + C(k+1)/l, \]
where
which implies that $E \left[ \sum_{h=1-l}^{l-1} g_{k,T}(h)\{\omega(h)\hat{\gamma}_X(h) - \gamma_X(h)\} \right]^2 \leq C(EI_{1T}^2 + EI_{2T}^2 + EI_{3T}^2 + (k + 1)^2/l^2)$. We proceed to derive the order of $EI_{1T}^2$. Notice that

$$
EI_{1T}^2 = \frac{1}{T^2} \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \sum_{i_3=1}^{h_1} \sum_{i_4=1}^{h_2} \text{cov}(w_{i_1|h_1}, w_{i_2|h_2}) g_{k,T}(h_1)g_{k,T}(h_2)\omega_i(h_1)\omega_i(h_2)
$$

$$
\leq \frac{C(k+1)^2}{T^2} \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \sum_{i_3=1}^{h_1} \sum_{i_4=1}^{h_2} (|h_1| + 1)(|h_2| + 1)|\gamma_X(i_1 - i_2)\gamma_X(i_1 - i_2 + |h_1| - |h_2|)|
$$

$$
+ \frac{C(k+1)^2}{T^2} \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \sum_{i_3=1}^{h_1} \sum_{i_4=1}^{h_2} (|h_1| + 1)(|h_2| + 1)|\gamma_X(i_1 - i_2 - |h_2|)\gamma_X(i_1 - i_2 + |h_1|)|
$$

$$
\leq \frac{C(k+1)^2}{T^2} \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \sum_{i_3=1}^{h_1} \sum_{i_4=1}^{h_2} (T - |s|)(|h_1| + 1)(|h_2| + 1)|\gamma_X(s)\gamma_X(s + |h_1| - |h_2|)|
$$

$$
+ \frac{C(k+1)^2}{T^2} \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \sum_{i_3=1}^{h_1} \sum_{i_4=1}^{h_2} (T - |s|)(|h_1| + 1)(|h_2| + 1)|\gamma_X(s - |h_2|)\gamma_X(s + |h_1|)| := J_{1T} + J_{2T}.
$$

Then we get

$$
J_{1T} \leq \frac{C(k+1)^2}{T} \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \sum_{i_3=1}^{h_1} \sum_{i_4=1}^{h_2} |\gamma_X(s)|\gamma_X(s + |h_1| - |h_2|)|
$$

$$
\leq \frac{C(k+1)^2}{T} \sum_{s=0}^{+\infty} |\gamma_X(s)| \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} |\gamma_X(s + |h_1| - |h_2|)| \leq \frac{C(k+1)^2}{T} \frac{1}{l^3},
$$

and

$$
J_{2T} \leq \frac{C(k+1)^2}{T} \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \sum_{i_3=1}^{h_1} \sum_{i_4=1}^{h_2} |\gamma_X(s)|\gamma_X(s + |h_1| + |h_2|)|
$$

$$
\leq \frac{C(k+1)^2}{T} \sum_{s=0}^{+\infty} |\gamma_X(s)| \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} |\gamma_X(s + |h_1| + |h_2|)|
$$

$$
\leq \frac{C(k+1)^2}{T} \sum_{s=0}^{+\infty} |\gamma_X(s)| \sum_{v=1}^{2l-1} v^2|\gamma_X(s + v)| \leq \frac{C(k+1)^2}{T} \frac{l}{l^3}.
$$

It implies that $EI_{1T}^2 \leq \frac{C(k+1)^2}{T} \frac{1}{l^3}$. Applying similar arguments to $J_{2T}$ and $I_{3T}$, we get $EI_{2T}^2 \leq C(k+1)^2l^4/T^2$ and $EI_{3T}^2 \leq C(k+1)^2l^4/T^2$. Note the constant $C$ above does not depend on $m$ by the assumption. We then have

$$
P \left( \sup_{0 \leq k \leq m} \left| \sum_{h=1-l}^{l-1} g_{k,T}(h)\omega_i(h)\hat{\gamma}_X(h) - \sum_{h=1-T}^{T-1} g_{k,T}(h)\gamma_X(h) \right| > \epsilon \right) \leq \frac{C}{l^2} (m^3/l^2 + (ml)^3/T) \rightarrow 0.
$$

\end{document}
Proof of Theorem 3.2. We choose \( m \) so that \( m^3/l^2 + (ml)^3/T + 1/m \to 0 \) (e.g., \( l \asymp T^{1/5} \) and \( m \asymp T^{2/15-\epsilon} \) for some \( \epsilon > 0 \)). From equation (3) in Lemma 1.5, we know that

\[
\var(\xi_i) - \sigma^2 - (\var^*(\xi^*_i) - \sigma^2) = \frac{1}{T} \left\{ \sum_{h=1-T}^{T-1} \left( \sum_{h=1-h}^{T-1} g_i(t)\gamma_i(h) \right) - \sum_{h=1-T}^{l-1} g_i(t)\omega_i(h)\gamma_i(h) \right\} - \sum_{|h| \geq T} \gamma_i(h),
\]

where \( \hat{\sigma}^2 = \sum_{h=1-l}^{l-1} \omega_i(h)\hat{\gamma}_i(h) \) and \( g_{0,i}(h) = -|h| \),

\[
g_{i,T}(h) = TCT(\phi_i(s), \phi_i(t)) - \left[ \sum_{j=1}^{h} \left\{ \phi_i^0 \left( \frac{j}{T} \right) \right\}^2 + \sum_{j=T-h+1}^{T} \left\{ \phi_i^0 \left( \frac{j}{T} \right) \right\}^2 \right] \mathbb{I}\{h \geq 1\} + \left[ \sum_{j=1}^{T-h} \phi_i^0 \left( \frac{j}{T} \right) \left\{ \phi_i^0 \left( \frac{j + h}{T} \right) - \phi_i^0 \left( \frac{j}{T} \right) \right\} \right] \mathbb{I}\{h \geq 1\} + \left[ \sum_{j=1+h}^{T} \phi_i^0 \left( \frac{j}{T} \right) \left\{ \phi_i^0 \left( \frac{j + h}{T} \right) - \phi_i^0 \left( \frac{j}{T} \right) \right\} \right] \mathbb{I}\{h \leq -1\},
\]

for \( 1 \leq i \leq m \). Note that \( \sup_{1 \leq i \leq m} |TC_T(\phi_i(s), \phi_i(t))| \leq C \). It is not hard to see that \( |g_{i,T}(h)| \leq C(|ih| + |h| + 1) \) for \( 0 \leq i \leq m \). By Lemma 1.13, we know

\[
\sup_{0 \leq i \leq m} \var(\xi_i) - \sigma^2 - \var^*(\xi^*_i) + \hat{\sigma}^2 = O_p \left( \frac{\sqrt{m^3/l^2 + (ml)^3/T}}{T} \right).
\]

Since the bootstrap sample is normally distributed and \( \sum_{h=1-l}^{l-1} h^2\omega_i(h)\hat{\gamma}_i(h) \) is bounded in probability in view of the fact that \( \sum_{h=-\infty}^{\infty} h^2\omega_i(h)E[\hat{\gamma}_i(h)] < \infty \), Theorem 2.2 is also applicable to the bootstrap sample, i.e.,

\[
\sup_{x \in [0,\infty)} |P(F_T(\infty) \leq x) - Q_\infty(x) - \hat{K}_T(x; \infty)| = o_p(1/T),
\]

where \( \hat{K}_T(x; \infty) = \frac{1}{\hat{\sigma}} \sum_{i=0}^{\infty} (\var^*(\xi^*_i) - \hat{\sigma}^2) E[(v_i^2 - 1)\mathbb{I}\{\mathcal{F}(v; \infty) \leq x\}] \). It is not hard to show that \( \hat{\sigma}^2 - \sigma^2 = O_p(\sqrt{1/T} + 1/T^2) \). Note that \( \var^*(\xi^*_i) - \hat{\sigma}^2 = \frac{1}{T} \sum_{h=1-l}^{l-1} g_{i,T}(h)\omega_i(h)\hat{\gamma}_i(h) \), which implies that \( \sup_{1 \leq i \leq m} \frac{|\var^*(\xi^*_i) - \hat{\sigma}^2|}{\sqrt{T(1+T^2)}} = O_p(1) \) (see e.g., (11)). Using the arguments in (12), we can show that

\[
\sup_{x \in [0,\infty)} \left| \frac{1}{2\hat{\sigma}^2} \sum_{i=m+1}^{\infty} (\var(\xi_i) - \var^*(\xi^*_i) + \hat{\sigma}^2 - \sigma^2) E[(v_i^2 - 1)\mathbb{I}\{\mathcal{F}(v; \infty) \leq x\}] \right| = O_p \left( \frac{1}{Tm^{a-\frac{3}{2}}} \right).
\]
Thus we get
\[
\sup_{x \in (0, \infty)} |N_T(x; \infty) - N_T^*(x; \infty)| \\
\leq \sup_{x \in (0, \infty)} \left| \frac{1}{2\sigma^2} \sum_{i=0}^{\infty} (\text{var}(\xi_i) - \text{var}^*(\xi_i^*) + \hat{\sigma}^2 - \sigma^2) E[(v_i^2 - 1)I\{F(v; \infty) \leq x\}] \right| \\
+ \sup_{x \in (0, \infty)} \left| \frac{1}{2\sigma^2} \sum_{i=1}^{\infty} (\text{var}^*(\xi_i^*) - \hat{\sigma}^2) E[(v_i^2 - 1)I\{F(v; \infty) \leq x\}] \right| \\
\leq \frac{1}{2\sigma^2} \sum_{1 \leq i \leq m} |\text{var}(\xi_i) - \sigma^2 - \text{var}^*(\xi_i^*) + \hat{\sigma}^2| \sup_{x \in (0, \infty)} \left| \sum_{i=1}^{m} E[(v_i^2 - 1)I\{F(v; \infty) \leq x\}] \right| \\
+ \sup_{x \in (0, \infty)} \frac{1}{2\sigma^2} \sum_{i=m+1}^{\infty} (\text{var}(\xi_i) - \text{var}^*(\xi_i^*) + \hat{\sigma}^2 - \sigma^2) E[(v_i^2 - 1)I\{F(v; \infty) \leq x\}] + O_p \left( \frac{\sqrt{1/T + 1/l^2}}{T} \right) \\
= O_p \left( \frac{\sqrt{m^3l^2 + (ml)^3/l^3}}{T} \right) + O_p \left( \frac{\sqrt{1/T + 1/l^2}}{T} \right) + O_p \left( \frac{1}{Tm^{a-2}} \right).
\]

It then follows that \( \sup_{x \in (0, \infty)} |P(F_T(\infty) \leq x) - P(F_T^*(\infty) \leq x)| \leq \sup_{x \in [0, \infty)} |N_T(x; \infty) - N_T^*(x; \infty)| + o_p(1/T) = o_p(1/T). \)

\[\Box\]

**Proof of Theorem 3.1.** The proof is similar to those of Lemma 1.13 and Theorem 3.2. The details are omitted. \[\Box\]

### 2. Simulation study.
We conduct a small simulation study to compare and contrast the finite sample performance of the small-\(b\) approximation, fixed-\(b\) approximation, MBB and Gaussian dependent bootstrap (GDB). Following the setup in Gonçalves and Vogelsang [2], we consider the AR(1) model,

\[
y_t = \rho y_{t-1} + \sqrt{1-\rho^2} \varepsilon_t, \quad t = 1, 2, \ldots, T,
\]

with \( \{\varepsilon_t\} \) being a sequence of iid \( N(0,1) \), \( t(3) \) or \( \exp(1) - 1 \) random variables. Consider the Wald statistic based on the HAC estimator with the Bartlett kernel and QS kernel for testing the null hypothesis \( E[y_t] = 0 \) versus the alternative that \( E[y_t] \neq 0 \) at 5% nominal level. Throughout the simulation we set \( T = 50 \) and the number of Monte Carlo replications to be 1000. The bootstrap tests are based on 1000 replications for each sample. We implement the MBB in a ‘naive’ fashion as described in Gonçalves and Vogelsang [2]. The simulation results for \( b = 0.04, 0.06, 0.08, 0.1, 0.2, \ldots, 1 \) and \( \rho = -0.7, 0.5 \) and 0.9 are summarized in Figures 1-3. We present the results for GDB with \( l = 5, 10 \) and MBB with block size equal to 5 and 10. It is seen from the figures that the GDB is more accurate than the small-\(b\) asymptotic approximation in most cases and improvement is often substantial especially for large \( b \). In the dependent cases (e.g., \( \rho = -0.7, 0.5 \) and 0.9), the GDB tends to provide a refinement over the fixed-\(b\) approximation for a proper bandwidth which is consistent with our theoretical findings. The improvement is apparent when the dependence is strong and \( b \) is small. In addition it is interesting
to note that the GDB not only provides an improvement when the innovations are Gaussian but also in the case of $t(3)$ distributed fat tailed innovations and $\exp(1) - 1$ distributed skewed innovation. The performance of GDB and MBB is in general comparable. MBB delivers slightly better size in most cases when the dependence is positive. When $\rho = 0.9$, the MBB with block size 10 apparently outperforms all the other methods for all three cases, suggesting that with a proper choice of block size, the MBB is capturing not only the asymptotic bias and variance of long run variance estimator but also the higher order moments. Since the GDB only captures the second order properties, it is not surprising that it can be inferior to MBB in some cases. Overall, the simulation results are consistent with those in Gonçalves and Vogelsang [2], and demonstrate the effectiveness of the proposed Gaussian dependent bootstrap in the Gaussian setting. The simulation results also suggest that our procedure may be useful in some non-Gaussian settings, though it can hardly be justified theoretically. The moving block bootstrap is expected to be second order accurate under the fixed-smoothing asymptotics, as seen from its empirical performance, but a rigorous theoretical justification seems very difficult.

References.


X. ZHANG
DEPARTMENT OF STATISTICS
UNIVERSITY OF MISSOURI-COLUMBIA
COLUMBIA, MO 65211.
E-MAIL: zhang104@illinois.edu

X. SHAO
DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
CHAMPAIGN, IL 61820.
E-MAIL: xshao@illinois.edu
Fig 1. Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with $N(0, 1)$ innovations.
Fig 2. Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with t(3) innovations.
Fig 3. Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with $\exp(1) - 1$ innovations.