# FIXED-SMOOTHING ASYMPTOTICS FOR TIME SERIES[1]

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 nonparametric 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 $x = (x_1, x_2, \ldots, x_{q_0}) \in \mathbb{R}^{q_0}$, we let $\|x\| = (\sum_{i=1}^{q_0} x_i^2)^{1/2}$ be the Euclidean norm. For a matrix $A = (a_{ij})_{i,j=1}^{q_0} \in \mathbb{R}^{q_0 \times q_0}$, denote by $\|A\|_2 = \sup_{\|x\|=1} \|Ax\|$ the spectral norm and $\|A\|_\infty = \max_{1 \le i,j \le q_0} |a_{ij}|$ the max norm. Denote by $\lfloor a \rfloor$ 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 $i \in \mathcal{M}_j = \{s \in \mathbb{Z} : (j-1)T/K < s \leq jT/K\}, j = 1, 2, \ldots, K$. Define the sample mean of the $k$th group as

$$\hat{\mu}_k = \frac{1}{|\mathcal{M}_k|} \sum_{i \in \mathcal{M}_k} X_i, \qquad k = 1, 2, \ldots, K,$$

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

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

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 $\sum_{h=-\infty}^{+\infty} \gamma_X(h) > 0$ and $\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty$. Further suppose that $|\mathcal{M}_1| = |\mathcal{M}_2| = \cdots = |\mathcal{M}_K|$ and $K$ is fixed. Then under $H_0$, we have*

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

*where $\Psi(x; K) = P(|t_{K-1}| \le x) - \frac{B}{2\sigma^2 T} \Upsilon(x; K)$ with*

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

$$- E\left[\chi_1^2 G_{K-1}\left(\frac{(K-1)\chi_1^2}{x^2}\right)\right] + 1$$

*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 $\Upsilon(x; K)$, 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 $\Upsilon(t_{K-1}(1-\alpha); K)/K$ 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 $\Upsilon(t_{K-1}(1-\alpha); K)/K$ increases rapidly for $K < 10$, and it becomes stable for relatively large $K$. For each $K \ge 2$, $\Upsilon(t_{K-1}(1-\alpha); K)/K$ 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 $K \le 4$, 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 $K \to +\infty$.
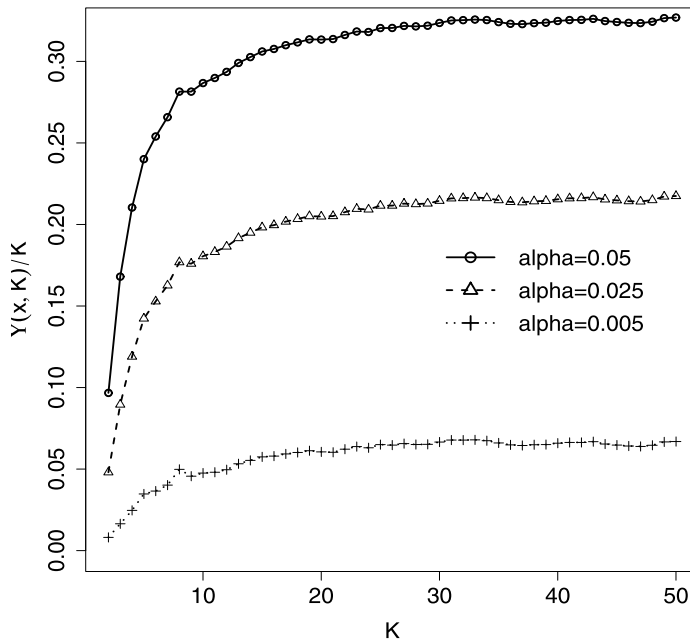
FIG. 1.    *Simulated values of* $\Upsilon(t_{K-1}(1-\alpha); K)/K$ *based on* 500,000 *replications.*

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* $x \in \mathbb{R}$.

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 $\mathcal{S}_{K-1} = \sqrt{\chi_{K-1}^2/(K-1)}$ with the $\chi_{K-1}^2$ distribution being independent with $Z$. Then the quantity $t_{K-1,\delta} = (Z + \delta)/\mathcal{S}_{K-1}$ follows a noncentral $t$ distribution with noncentral parameter $\delta$. Define $e_1(x) = E[\mathbf{I}\{|t_{K-1,\delta}| > x\}Z^2]$ and $e_2(x) = E[\mathbf{I}\{|t_{K-1,\delta}| > x\}\chi_{K-1}^2]$. Then under the local alternative, we have

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

where $\Upsilon_\delta(x; K) = K^2 P(|t_{K-1,\delta}| > x) - e_1(x) - (K+1)e_2(x)$. 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 $\Upsilon_\delta(t_{K-1}(1-\alpha); K)$ is roughly monotonic with respect to $K$ for $\delta \in (0, 4]$, 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 $-\frac{B}{2\sigma^2 T}\Upsilon(x, K)$. Specifically, let $\Sigma_T = (\sigma_{ij})_{i,j=1}^K$ with $\sigma_{ij} = q \operatorname{Cov}(\hat{\mu}_i, \hat{\mu}_j)$. 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*

$$(3) \quad P(|T_K| \leq x) = G_1(x^2) + \frac{1}{K-1}x^4 G_1''(x^2) - \frac{BK}{T\sigma^2}x^2 G_1'(x^2) + O(1/T).$$

REMARK 2.3. Since

$$P(|t_{K-1}| \leq x) = G_1(x^2) + \frac{1}{K-1}x^4 G_1''(x^2) + O(1/K^2)$$

(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., $\chi_1^2$) only captures the first term. In view of Proposition 2.1, it is not hard to see that

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

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 $\mathcal{G}(\cdot, \cdot)$ which satisfies the spectral decomposition

$$(4) \qquad \mathcal{G}(r, t) = \sum_{j=1}^{+\infty} \lambda_j \phi_j(r) \phi_j(t), \qquad 0 \le r, t \le 1,$$

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

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

where $\bar{X}_T = \sum_{i=1}^{T} X_i / T$ is the sample mean. For the convenience of presentation, we set $b = 1$. See Remark 2.4 for the case $b \in (0, 1)$. To illustrate the idea, we define the projection vectors $\xi_j = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \phi_j^0(i/T) X_i$ with $\phi_j^0(t) = \phi_j(t) - \frac{1}{T} \sum_{i=1}^{T} \phi_j(i/T)$ 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 $\bar{\mathcal{G}}(\cdot, \cdot)$, we can define the demeaned kernel,

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

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

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

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

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

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

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

with $\xi_0 = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} (X_i - \mu_0)$. The Wald test statistic with HAC estiamtor 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

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

(5)
$$K = 1, \ldots, \infty,$$

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 $\mathcal{G}(\cdot, \cdot)$ 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*

$$\sup_{1 \leq i \leq J} \sup_{t \in [0,1]} |\phi_i^{(j)}(t)| < CJ^j$$

*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 $\sup_{x \in [0,+\infty)} |\aleph_T(x; K)| = O(1/T)$ and*

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

*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 $\aleph_T(x; \infty)$ reflects the departure of $\{\xi_j\}_{j=0}^{+\infty}$ from the i.i.d. standard normal random variables $\{v_j\}_{j=0}^{+\infty}$. Specifically, the form of $\aleph_T(x; \infty)$ 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 $\int_0^1 \mathcal{G}(r, r) \, dr = \sum_{j=1}^{+\infty} \lambda_j = 1$. 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 $\aleph_T(x; \infty)$, which is a weighted sum of the relative difference of $\text{var}(\xi_i)$ and $\sigma^2$. Note that the difference $\text{var}(\xi_i) - \sigma^2$ relies on the second-order properties of the time series and the eigenfunctions of $\mathcal{G}(\cdot, \cdot)$, and the weight $E[(v_i^2 - 1)\mathbf{I}\{\mathcal{F}(\infty) \leq x\}]$ which depends on the eigenvalues of $\mathcal{G}(\cdot, \cdot)$ 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 $\mathcal{G}(\cdot, \cdot)$ is usually defined through a semi-positive definite univariate kernel $\mathcal{K}(\cdot)$, that is, $\mathcal{G}(r, t) = \mathcal{K}(r - t)$. In what follows, we make several remarks regarding this special case.

REMARK 2.4. For $0 < b \leq 1$, we define $\mathcal{G}_b(\cdot, \cdot) = \mathcal{G}(\cdot/b, \cdot/b)$. If $\mathcal{G}(\cdot, \cdot)$ is semi-positive definite on $[0, 1/b]^2$, then $\mathcal{G}_b(\cdot, \cdot)$ satisfies the spectral decomposition $\mathcal{G}_b(r, t) = \sum_{j=1}^{+\infty} \lambda_{j,b} \phi_{j,b}(r) \phi_{j,b}(t)$ with $0 \leq r, t \leq 1$. The eigencompoents of $\mathcal{G}_b(r, t)$ can be obtained by solving a homogenuous Fredholm integral equation of the second kind, where the solutions can be approximated numerically when analytical solutions are unavailable. When $\mathcal{G}(r, t) = \mathcal{K}(r - t)$, it was shown in Knessl and Keller [14] that under suitable assumptions on $\mathcal{K}(\cdot)$, $\lambda_{j,b} = b \int_{-\infty}^{+\infty} \mathcal{K}(r) \, dr - (\pi^2 j^2 b^3 / 2) \int_{-\infty}^{+\infty} r^2 \mathcal{K}(r) \, dr + o(b^3)$ 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 $\mathcal{G}(r, t) = \mathcal{K}(r - t)$ and the technical assumption $b < 1/(16 \int_{-\infty}^{+\infty} |\mathcal{K}(r)| \, dr)$ 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

TABLE 1
*Asymptotic comparison between the first and second-order approximations based on fixed-b and small-b asymptotics*

| Asymptotics | First order | Second order |
|---|---|---|
| Fixed-$b$ | $P(\mathcal{F}_b(\infty) \leq x)$ | $P(\mathcal{F}_b(\infty) \leq x) + \aleph_{T,b}(x; \infty)$ |
| Small-$b$ | $G_1(x)$ | $G_1(x) + (c_2 G_1''(x) x^2 - c_1 G_1'(x) x)b$ |
| | | $\quad - \frac{g_q \sum_{h=-\infty}^{+\infty} \lvert h \rvert^q \gamma_X(h)}{\sigma^2 (bT)^q} G_1'(x) x$ |

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 $\mathcal{K}(r - t) = \sum_{j=1}^{+\infty} \lambda_j \{\cos(2\pi jr) \cos(2\pi jt) + \sin(2\pi jr) \sin(2\pi jt)\} = \sum_{j=1}^{+\infty} \lambda_j \cos(2\pi j(r - t))$ with $\sum_{j=1}^{+\infty} \lambda_j = 1$ and $\lambda_j = O(1/j^{19+\epsilon})$ for some $\epsilon > 0$. Then the asymptotic expansion (6) holds for the Wald statistic based on the difference kernel $\mathcal{G}(r, t) = \mathcal{K}(r - t)$.

Define the Parzen characteristic exponent

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

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 $\mathcal{F}_b(\infty)$ and $\aleph_{T,b}(x; \infty)$ with $\phi_j$ and $\lambda_j$ being replaced with $\phi_{j,b}$ and $\lambda_{j,b}$ in the definition of $\mathcal{F}(\infty)$ and $\aleph_T(x; \infty)$. 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 $P(\mathcal{F}_b(\infty) \leq x) = G_1(x) + (c_2 G_1''(x) x^2 - c_1 G_1'(x) x)b + O(b^2)$ 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 expan-

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

PROPOSITION 2.3.    *Suppose* $\mathcal{K}(\cdot) : \mathbb{R} \to [0, 1]$ *is symmetric, semi-positive definite, piecewise smooth with* $\mathcal{K}(0) = 1$ *and* $\int_0^{+\infty} x\mathcal{K}(x)\, dx < \infty$. *The Parzen characteristic exponent of* $\mathcal{K}$ *is no less than one. Further assume that*

$$(7) \qquad \sup_{k \in \mathbb{N}} \left| \sum_{i=1}^k \tilde{\lambda}_{i,b}\big(\mathrm{var}(\tilde{\xi}_{i,b}) - \sigma^2\big) \right| = O\left( \sum_{i=1}^{+\infty} \tilde{\lambda}_{i,b}\big(\mathrm{var}(\tilde{\xi}_{i,b}) - \sigma^2\big) \right)$$

*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*

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

*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 $\tilde{\mathcal{G}}_{k,b}(r, t) = \sum_{i=1}^k \tilde{\lambda}_{j,b}\tilde{\phi}_{j,b}(r)\tilde{\phi}_{j,b}(t)$ is at the same or smaller order of the bias for the LRV estimator based on $\tilde{\mathcal{G}}_b(r, t)$. 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 $\aleph_T(x; \infty)$ [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 [33] for the use of a banded sample covariance matrix and McMurry and Politis [23] for a tapered version of the sample covariance matrix. In what follows, we shall consider the Bartlett kernel, that is, $\omega(x) = (1 - |x|)\mathbf{I}\{|x| < 1\}$, which guarantees to yield a semi-positive definite estimates, 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 $X_i^*$'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 [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*

$$(8) \qquad \sup_{x \in [0, +\infty)} \left| P(|T_K| \leq x) - P^*(|T_K^*| \leq x) \right| = o_p(1/T).$$

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*

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

*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* $\hat{D}_{T,1}^* = T^{-1} \sum_{i,j=1}^T \mathcal{G}(i/T, j/T)(X_i^* - \bar{X}_T^*)(X_j^* - \bar{X}_T^*)$ *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 Künsch [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-Guassian 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

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

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, 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 $\frac{\partial y}{\partial x}$ 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 $v = O(u)$. We first present the following lemmas whose proofs are given in the online supplement [34].

LEMMA 5.1.
$$\frac{\partial f}{\partial \operatorname{vec}(\Sigma)}(y, \Sigma) = \frac{f(y, \Sigma)}{2}\{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \operatorname{vec}(\Sigma^{-1})\}.$$

LEMMA 5.2.
$$\frac{\partial^2 f}{\partial \operatorname{vec}(\Sigma)\operatorname{vec}(\Sigma)}(y, \Sigma)$$
$$= \frac{1}{4}\{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \operatorname{vec}(\Sigma^{-1})\}\{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \operatorname{vec}(\Sigma^{-1})\}'$$
$$\times f(y, \Sigma)$$
$$- \frac{1}{2}\{(\Sigma^{-1}yy'\Sigma^{-1}) \otimes \Sigma^{-1} + \Sigma^{-1} \otimes (\Sigma^{-1}yy'\Sigma^{-1}) - \Sigma^{-1} \otimes \Sigma^{-1}\}$$
$$\times f(y, \Sigma).$$

LEMMA 5.3. *Let* $\{\Sigma_T\} \subset \mathbb{R}^{(K+1)\times(K+1)}$ *be a sequence of positive definite matrices with* $K + 1 \le T$. *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* $\tilde{\Sigma}_T(y)$ *be a* $(K + 1) \times (K + 1)$ *positive symmetric matrix which depends on* $y \in \mathbb{R}^{K+1}$. *Assume that* $\sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T(y) - \Sigma\|_2 \le \|\Sigma_T - \Sigma\|_2 = O(1/T)$ *for a positive definite matrix* $\Sigma$. *Let* $R_T = \Sigma_T - \Sigma$. *If* $K$ *is fixed with respect to* $T$, *we have*
$$\int_{y \in \mathbb{R}^{K+1}} \left|\operatorname{vec}(R_T)' \frac{\partial^2 f}{\partial \operatorname{vec}(\Sigma)\operatorname{vec}(\Sigma)}(y, \tilde{\Sigma}_T(y)) \operatorname{vec}(R_T)\right| dy = O(1/T^2).$$

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 $Y_i = \sqrt{q}(\hat{\mu}_i - \mu_0)$, and $\bar{Y} = \frac{1}{K}\sum_{i=1}^K Y_i$ and $S_Y^2 = \frac{1}{K-1}\sum_{i=1}^K(Y_i - \bar{Y})^2$ 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
$$\sigma_{ij} := \operatorname{Cov}(Y_i, Y_j) = \sum_{h=1-q}^{q-1}\left(\frac{q - |h|}{q}\right)\gamma_X(h - (j - i)q).$$

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

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

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

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

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

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

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

and

$$(10) \quad \begin{aligned} &\int_{y \in \mathbb{R}^K} \left|\operatorname{vec}(\Sigma_T - \sigma^2 I_K)' \frac{\partial^2 f}{\partial \operatorname{vec}(\Sigma) \operatorname{vec}(\Sigma)}(y, \tilde{\Sigma}_T(y)) \operatorname{vec}(\Sigma_T - \sigma^2 I_K)\right| dy \\ &= O\left(\frac{1}{T^2}\right), \end{aligned}$$

which imply that

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

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

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

where $g$ denotes the major term, $R(y)$ is the remainder term and $\delta_{ij} = \mathbf{I}\{i = j\}$ 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

$$J_1 = \sum_{i=1}^{K}(\sigma_{ii} - \sigma^2)E[\mathbf{I}\{|\tilde{T}_K(v)| > x\}v_i^2], \qquad J_2 = \sum_{i\neq j}\sigma_{ij}E[\mathbf{I}\{|\tilde{T}_K(v)| > x\}v_i v_j].$$

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 $U \sim \chi_1^2$, $D \sim \chi_{K-1}^2$ and $U$ and $D$ are independent. We define that

$$\begin{aligned}
&E[\mathbf{I}\{|\tilde{T}_K(v)| > x\}v_i^2] \\
&\quad = \frac{1}{K}E\left[\mathbf{I}\{|\tilde{T}_K(v)| > x\}\sum_{i=1}^{K}v_i^2\right] \\
&\quad = \frac{1}{K}E[\mathbf{I}\{|\tilde{T}_K(v)| > x\}U] + \frac{1}{K}E[\mathbf{I}\{|\tilde{T}_K(v)| > x\}D] \\
&\quad = \frac{1}{K}E\left[UG_{K-1}\left(\frac{(K-1)U}{x^2}\right)\right] + \frac{1}{K}E\left[D - DG_1\left(\frac{Dx^2}{K-1}\right)\right]
\end{aligned}$$

and

$$\begin{aligned}
&E[\mathbf{I}\{|\tilde{T}_K(v)| > x\}v_i v_j] \\
&\quad = \frac{1}{K(K-1)}E\left[\mathbf{I}\{|\tilde{T}_K(v)| > x\}\sum_{i\neq j}v_i v_j\right] \\
&\quad = \frac{1}{K-1}E[\mathbf{I}\{|\tilde{T}_K(v)| > x\}U] - \frac{1}{K(K-1)}E\left[\mathbf{I}\{|\tilde{T}_K(v)| > x\}\sum_{i=1}^{K}v_i^2\right] \\
&\quad = \frac{1}{K}E\left[UG_{K-1}\left(\frac{(K-1)U}{x^2}\right)\right] - \frac{1}{K(K-1)}E\left[D - DG_1\left(\frac{Dx^2}{K-1}\right)\right].
\end{aligned}$$

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] ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513

[2] ANDREWS, D. W. K. (2002). Higher-order improvements of a computationally attractive *k*-step bootstrap for extremum estimators. *Econometrica* **70** 119–162. MR1926257

[3] BESTER, C. A., CONLEY, T. G. and HANSEN, C. B. (2011). Inference with dependent data using cluster covariance estimators. *J. Econometrics* **165** 137–151.

[4] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

[5] CHEN, Y. and QU, Z. (2013). M tests with a new normalization matrix. *Econometric Rev*. To appear.

[6] GONÇALVES, S. and VOGELSANG, T. J. (2011). Block bootstrap HAC robust tests: The sophistication of the naive bootstrap. *Econometric Theory* **27** 745–791. MR2822364

[7] GÖTZE, F. and KÜNSCH, H. R. (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist*. **24** 1914–1933. MR1421154

[8] HALL, P. and HOROWITZ, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* **64** 891–916. MR1399222

[9] HÄRDLE, W., HOROWITZ, J. and KREISS, J. P. (2003). Bootstrap methods for time series. *International Statistical Review* **71** 435–459.

[10] IBRAGIMOV, R. and MÜLLER, U. K. (2010). *t*-statistic based correlation and heterogeneity robust inference. *J. Bus. Econom. Statist*. **28** 453–468. MR2732358

[11] INOUE, A. and SHINTANI, M. (2006). Bootstrapping GMM estimators for time series. *J. Econometrics* **133** 531–555. MR2252908

[12] JANSSON, M. (2004). The error in rejection probability of simple autocorrelation robust tests. *Econometrica* **72** 937–946. MR2051441

[13] KIEFER, N. M. and VOGELSANG, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* **21** 1130–1164. MR2200988

[14] KNESSL, C. and KELLER, J. B. (1991). Asymptotic properties of eigenvalues of integral equations. *SIAM J. Appl. Math.* **51** 214–232. MR1089139

[15] KUAN, C.-M. and LEE, W.-M. (2006). Robust *M* tests without consistent estimation of the asymptotic covariance matrix. *J. Amer. Statist. Assoc.* **101** 1264–1275. MR2328312

[16] LAHIRI, S. N. (1996). On Edgeworth expansion and moving block bootstrap for Studentized *M*-estimators in multiple linear regression models. *J. Multivariate Anal.* **56** 42–59. MR1380180

[17] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data.* Springer, New York. MR2001447

[18] LAHIRI, S. N. (2007). Asymptotic expansions for sums of block-variables under weak dependence. *Ann. Statist.* **35** 1324–1350. MR2341707

[19] LAHIRI, S. N. (2010). Edgeworth expansions for Studentized statistics under weak dependence. *Ann. Statist.* **38** 388–434. MR2589326

[20] LITTLE, G. and READE, J. B. (1984). Eigenvalues of analytic kernels. *SIAM J. Math. Anal.* **15** 133–136. MR0728688

[21] LOBATO, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.* **96** 1066–1076. MR1947254

[22] MCCULLOCH, C. E. (1982). Symmetric matrix derivatives with applications. *J. Amer. Statist. Assoc.* **77** 679–682. MR0675898

[23] MCMURRY, T. L. and POLITIS, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J. Time Series Anal.* **31** 471–482. MR2732601

[24] MÜLLER, U. K. (2007). A theory of robust long-run variance estimation. *J. Econometrics* **141** 1331–1352. MR2413504

[25] NEWEY, W. K. and WEST, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708. MR0890864

[26] PHILLIPS, P. C. B. (2005). HAC estimation by automated regression. *Econometric Theory* **21** 116–142. MR2153858

[27] SHAO, X. (2010). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 343–366. MR2758116

[28] SUN, Y. (2010). Let's fix it: Fixed-*b* asymptotics versus small-*b* asymptotics in heteroscedasticity and autocorrelation robust inference. Working paper, Dept. Economics, UCSD.

[29] SUN, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *J. Econometrics* **164** 345–366. MR2826775

[30] SUN, Y. (2013). A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. *Econom. J.* **16** 1–26.

[31] SUN, Y., PHILLIPS, P. C. B. and JIN, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* **76** 175–194. MR2374985

[32] VELASCO, C. and ROBINSON, P. M. (2001). Edgeworth expansions for spectral density estimates and Studentized sample mean. *Econometric Theory* **17** 497–539. MR1841819

[33] WU, W. B. and POURAHMADI, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768. MR2589209

[34] ZHANG, X. and SHAO, X. (2013). Supplement to "Fixed-smoothing asymptotics for time series." DOI:10.1214/13-AOS1113SUPP.

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

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
  AT URBANA-CHAMPAIGN
CHAMPAIGN, ILLINOIS 61820
USA
E-MAIL: xshao@illinois.edu