EDGEWORTH EXPANSIONS FOR SPECTRAL DENSITY ESTIMATES AND STUDENTIZED SAMPLE MEAN[1]

by

Carlos Velasco
Universidad Carlos III, Madrid

and

Peter M Robinson
London School of Economics and Political Science

Contents:

The Suntory Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020-7405 7686

Discussion Paper
No. EM/00/390
May 2000

[1] Research supported by ESRC Grant R000235892 and, in the case of the first author, by Fundacion Ramon Areces and Spanish Direccion General de Ensenanza Superior, Ref. No. PB98-0025 and, in the case of the second author, by the Lverhulme Personal Research Professorship. We wish to thank Masanubu Taniguchi for bringing to our attention the work of Bentkus and Rudzkis. We are grateful to the Editor and referees for very helpful comments.

Abstract

We establish valid Edgeworth expansions for the distribution of smoothed nonparametric spectral estimates, and of studentized versions of linear statistics such as the same mean, where the studentization employs such a nonparametric spectral estimate. Particular attention is paid to the spectral estimate at zero frequency and, correspondingly, the studentized sample mean, to reflect econometric interest in autocorrelation-consistent or long-run variance estimation. Our main focus is on stationary Gaussian series, though we discuss relaxation of the Gaussianity assumption. Only smoothness conditions on the spectral density that are local to the frequency of interest are imposed. We deduce empirical expansions from our Edgeworth expansions designed to improve on the normal approximation in practice, and also a feasible rule of bandwidth choice.

Keywords: Edgeworth expansions; nonparametric spectral estimates; stationary Gaussian series; studentized sample mean; bandwidth choice.

JEL No.: C22

1 Introduction

In this paper we analyze higher-order asymptotic properties of smoothed nonparametric estimates of the spectral density for a Gaussian stationary time series and of linear statistics studentized by such a nonparametric estimate. There is a large literature on the consistency and asymptotic normality of nonparametric spectral estimates and studentized linear statistics, but much less is known about higher-order properties, including the Edgeworth expansions we consider.

We focus principally on zero frequency and obtain Edgeworth expansions for the joint distribution of the spectral estimate and sample mean. These can be used to approximate the distribution and moments of smooth functions of these statistics, and we go on to analyze the higher-order asymptotic properties of the sample mean studentized by the spectral estimate. The studentization we employ is prompted by the fact that the variance of the sample mean is approximately proportional to the spectral density at zero frequency. Such studentization, with autocorrelated observations, goes back at least to Jowett (1954), whose work was developed by Hannan (1957), Brillinger (1979) and extended to more general circumstances, as recently reviewed by Robinson and Velasco (1997). In particular such ideas have been widely employed in econometric models, sometimes under the headings of "(heteroskedasticity and) autocorrelation-consistent variance estimation" and "long run variance estimation".

Spectral density estimation, and studentization of the sample mean, can be based on a parameterization of the spectral density, as when an autoregressive moving average model of given order is assumed. However, if the parameterization is incorrect (for example if one or both of the autoregressive or moving average orders is under-specified) or unidentified (as when both orders are overspecified) inconsistent spectral estimates result, and inferences based on the sample mean are invalidated. Nonparametric spectral estimation seeks to avoid these drawbacks. However, its implementation requires the user to specify a functional form (a kernel in our case), as well as a bandwidth, which determines the degree of smoothing. First-order asymptotic theory holds across a wide range of bandwidths, but the detail of Edgeworth expansions is more sensitive to bandwidth choice, reflecting finite sample practical experience. We use our Edgeworth expansions to approximate the moments of stochastic approximations whose distributions are very close to that of the original t-ratio and propose "optimal" choices of bandwidth, which can be proxied by data-dependent quantities. Also, we approximate our theoretical Edgeworth expansions, which involve population quantities, by empirical expansions for practical use. It is anticipated that our proposed corrections could outperform the normal approximation in highly autocorrelated processes, where nonparametric spectral estimates can be particularly biased, and thus severely influence the distribution of the studentized mean.

Spectral estimation and studentization at other frequencies is not essentially different from that at zero frequency and we discuss this extension explicitly. One important feature of our work is that smoothness, and indeed boundedness, of the spectral density is assumed only at the frequency of interest. This is natural because the variance of the sample mean is proportional to the Césaro sum of the Fourier series of the spectral density at zero frequency, which, by Fejér's theorem, converges if and only if this is a continuity point. These mild conditions are also practically desirable because they permit lack of smoothness, and even unboundedness, at remote frequencies, as can arise from long memory, cyclic or seasonal behaviour. Reliance on only local assumptions has recently been stressed in work by Robinson (1994, for example) on semiparametric analysis of long memory, and we employ similar truncation

techniques to achieve this. By contrast, the bulk of the literature on smoothed nonparametric spectral estimation imposes assumptions that imply at least boundedness of the spectral density at all frequencies. In particular, this is the case in the work of Bentkus (1976), Bentkus and Rudzkis (1982) and Rudzkis (1985) on higher-order asymptotic theory for nonparametric spectral estimates, whose approach we in other respects follow. It is also the case in the econometric work referred to above on consistency of autocorrelation-consistent or long run variance estimates and on the first-order limiting distribution of studentized statistics, which resorts to summability conditions on mixing numbers. On the other hand, the econometric literature typically avoids the Gaussianity assumption which we impose in the bulk of the paper, and the mixing conditions employed can cover a degree of heterogeneity across time, as well as dealing with far more general statistics, such as implicitly defined extremun estimates of vector-valued parameters. We suspect that in our higher-order treatment the stationarity assumption could to some extent be relaxed at cost of significantly more complicated conditions, while vector and other extensions should be possible, albeit notationally complex. Relaxation of our Gaussianity assumption which, as in much other work on higher-order expansions (see for example such time series references as Phillips (1977), Taniguchi (1991)), plays a considerable simplifying role, may lead to rather more complex expansions, which we investigate in Section 7. Though much recent higher-order asymptotic theory for non-Gaussian time series analysis has been based on the work of Götze and Hipp (1983) it is not known if their conditions allow a proof of the validity of the Edgeworth expansions for smoothed spectral estimates (see Remark 2.3 in Janas (1994)) though some ideas on nonparametric studentization are in Götze and Künsch (1996).

Mean-correction in spectral estimation does not affect first-order asymptotic distribution theory, but its effects may show up in terms of a smaller order of magnitude for the distribution of both spectral estimates and t-ratios. We study this correction in detail and our analysis could also be extended to residual-based nonparametric studentization of least squares estimates in a nonstochastically trending linear regression, possibly involving cosinusoidal regressors, whose variance may depend on the spectral density of the errors at various frequencies.

The paper is organized as follows. The following section provides the main assumptions used throughout. In Section 3 we establish a valid Edgeworth expansion for the distribution of the nonparametric estimate of the spectral density and analyze the joint distribution of the variance estimate and the sample mean. In Section 4 we establish a valid Edgeworth expansion for the studentized sample mean and consider the effects of mean-correction. Section 5 provides consistent estimates of higher-order correction terms and an empirical Edgeworth expansion. We extend our results to obtain a third-order approximation in Section 6. Finally in Sections 7 and 8 respectively, we analyze the effects on our approximations of higher-order cumulants for non-Gaussian series, and Edgeworth approximations for estimation at non-zero frequencies. Proofs, including some technical lemmas, appear in two appendices.

2 Nonparametric studentization of the sample mean

Let $\{X_t\}$ be a stationary Gaussian sequence with mean that is known (for the time being) to be zero, autocovariance function $\gamma(r)$ and spectral density $f(\lambda)$ defined by $\gamma(r) = \int_\Pi f(\lambda)e^{ir\lambda}d\lambda$, where $\Pi =$

$(-\pi, \pi]$, and satisfying $0 < f(0) < \infty$. Let $\overline{X} = N^{-1} \sum_{j=1}^{N} X_j$ and denote

$$V_N \stackrel{def}{=} \mathrm{Var}[\sqrt{N}\,\overline{X}] = \sum_{j=1-N}^{N-1} \left(1 - \frac{|j|}{N}\right) \gamma(j).$$

Then for all N such that $V_N > 0$,

$$u_1 \stackrel{def}{=} \frac{\sqrt{N}\,\overline{X}}{\sqrt{V_N}} \sim \mathcal{N}(0,1).$$

Since V_N is the Césaro sum of the Fourier coefficients of $f(\lambda)$ at $\lambda = 0$, if $f(\lambda)$ is continuous at $\lambda = 0$ then $\lim_{N\to\infty} V_N = 2\pi f(0)$ by Fejér's Theorem. If $\widehat{f}(0)$ is a consistent estimate, $\widehat{f}(0) \to_p f(0)$, then

$$Y_N \stackrel{def}{=} \frac{\sqrt{N}\,\overline{X}}{\sqrt{\widehat{V}}} \to_d \mathcal{N}(0,1),$$

where $\widehat{V} = 2\pi \widehat{f}(0)$. Defining

$$\widehat{\gamma}(\ell) = \frac{1}{N} \sum_{1 \le t, t+\ell \le N} X_t X_{t+\ell}, \qquad \ell = 0, \pm 1, \ldots, \pm(N-1),$$

consider the weighted-autocovariance nonparametric estimate of $f(0)$

$$\widehat{f}(0) = \frac{1}{2\pi} \sum_{\ell=1-N}^{N-1} \omega\left(\frac{\ell}{M}\right) \widehat{\gamma}(\ell) = \mathbf{X}' \frac{W_M}{2\pi N} \mathbf{X},$$

where $\mathbf{X} = (X_1, \ldots, X_N)'$ and W_M is the $N \times N$ matrix with (r,s)-th element

$$[W_M]_{r,s} = \omega\left(\frac{r-s}{M}\right) = \int_{\Pi} K_M(\lambda) e^{i(r-s)\lambda} d\lambda, \tag{1}$$

such that $K_M(\lambda)$ is a kernel function with smoothing or lag number M, which is a sequence of positive integers growing with N but more slowly. Then for an even, integrable function K which integrates to one, we set

$$K_M(\lambda) = M \sum_{j=-\infty}^{\infty} K(M[\lambda + 2\pi j]),$$

so $K_M(\lambda)$ is periodic of period 2π, even, integrable and $\int_{\Pi} K_M(\lambda) d\lambda = 1$. It follows that $\omega(r) = \int_{-\infty}^{\infty} e^{irx} K(x) dx$ and $\omega(0) = 1$ so we can write $\widehat{f}(0) = \int_{\Pi} K_M(\lambda) I(\lambda) d\lambda$, where $I(\lambda) = (2\pi N)^{-1} \times \left| \sum_{t=1}^{N} X_t \exp\{i\lambda t\} \right|^2$ is the periodogram of X_t, $t = 1, \ldots, N$. We restrict to this kernel class due to its simplicity for our analysis, though a variety of quadratic-form estimates could be considered (see e.g. Song and Schmeiser (1992)).

To analyze the joint distribution of the linear statistic $\overline{X}$ and the nonparametric estimate of its variance, it is convenient to work with standardized statistics with zero mean and unit variance. Suppose now that the estimate $\widehat{f}(0)$ is $\sqrt{N/M}$-consistent (cf. Hannan (1970, Chapter 5)). Write for $\mathbf{u} = (u_1, u_2)'$,

$$Y_N = Y_N(\mathbf{u}) = u_1 \left(1 + b_N + \sigma_N u_2 \sqrt{\frac{M}{N}}\right)^{-1/2}, \qquad u_2 = \sqrt{\frac{N}{M}} \left\{ \frac{\widehat{V} - E[\widehat{V}]}{V_N \sigma_N} \right\},$$

where $\sigma_N^2 = \mathrm{Var}[\sqrt{N/M}\widehat{V}/V_N]$ and $b_N = E[\widehat{V}]/V_N - 1$ are the "relative" variance and bias of $\widehat{V}$, and some of our notation suppresses the dependence on N. Then $u_2 = \mathbf{X}' Q_N \mathbf{X} - E[\mathbf{X}' Q_N \mathbf{X}]$ is a centered quadratic form in a Gaussian vector, where $Q_N = W_M (\sqrt{NM}\sigma_N V_N)^{-1}$ is a $N \times N$ matrix.

The joint characteristic function of $\mathbf{u}$ is

$$\psi_N(t_1,t_2) = |I - 2it_2\Sigma Q_N|^{-1/2} \exp\left\{-\frac{1}{2}t_1^2 \xi_N' \left(I - 2it_2\Sigma Q_N\right)^{-1} \Sigma\xi_N - it_2 E_N\right\},$$

where $E_N = E[\mathbf{X}'Q_N\mathbf{X}] = \text{Trace}[\Sigma Q_N]$, $\Sigma = E[\mathbf{X}\mathbf{X}']$, and $\xi_N = \mathbf{1}/\sqrt{NV_N}$, $\mathbf{1}$ being the $N \times 1$ vector $(1,1,\ldots,1)'$. Due to the normalizations $\mathbf{u}$ has identity covariance matrix and cumulant generating function

$$\varphi_N(t_1,t_2) \quad = \quad \log\psi_N(t_1,t_2) = \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} \kappa_N[r,s]\frac{(it_1)^r}{r!}\frac{(it_2)^s}{s!},$$

where the only non-zero bivariate cumulants are

$$\begin{aligned}
\kappa_N[0,s] &= 2^{s-1}(s-1)!\,\text{Trace}[(\Sigma Q_N)^s], \qquad s > 1,\\
\kappa_N[2,s] &= 2^s s!\,\xi_N'(\Sigma Q_N)^s\Sigma\xi_N, \qquad s > 0.
\end{aligned}$$

Phillips (1980) discusses these derivations and related literature for the analysis of the distribution of linear and quadratic forms under the normality assumption.

Here the Gaussianity assumption provides simple explicit expressions for the characteristic functions and cumulants of linear and quadratic forms, which otherwise would be very difficult to estimate for general dependent sequences. Furthermore these depend only on second-order properties of the time series, through Σ or f, which simplifies our set-up. We introduce the following assumptions about the Gaussian series X_t and $\widehat{f}(0)$.

Assumption 1 *$0 < f(0) < \infty$ and $f(\lambda)$ has d continuous derivatives ($d \geq 2$) in a neighbourhood of $\lambda = 0$, the dth derivative satisfying a Lipschitz condition of order ϱ, $0 < \varrho \leq 1$.*

Assumption 2 *The spectral density $f(\lambda) \in L_p$, for some $p > 1$, i.e. $\|f\|_p^p = \int_\Pi f^p(\lambda)d\lambda < \infty$.*

Assumption 3 *$K(x)$ is bounded, even and integrable on Π, and zero elsewhere, and integrates to one.*

Assumption 4 *$K(x)$ satisfies a uniform Lipschitz condition (of order 1) in $[-\pi,\pi]$.*

Assumption 5 *For $j = 0,1,\ldots,d$, $d \geq 2$ and $r = 1,2,\ldots$*

$$\mu_j(K^r) \stackrel{def}{=} \int_\Pi x^j \left[K(x)\right]^r dx = \begin{cases} = 0, & j < d;\\ \neq 0, & j = d.\end{cases}$$

Assumption 6 *$M^{-1} + MN^{-1} \to 0$, as $N \to \infty$.*

Assumption 7 *$M = C \cdot N^q$, with $0 < q < 1$ and $0 < C < \infty$.*

Assumption 1, which concerns bias, is implied by $\sum_{j=-\infty}^{\infty}|j|^{d+\varrho}|\gamma(j)| < \infty$, but this extends the smoothness assumption to all frequencies, whereas only local assumptions are natural for this problem. In particular, as in Robinson (1995a), for example, we allow, using truncation arguments, for lack of smoothness or even unboundedness (as arises from possibly cyclic long memory) at remote frequencies. The finite support requirement on K in Assumption 3 is helpful here, though undoubtedly it could be relaxed to a mild tail restriction. However, Assumption 2 imposes some restrictions on f beyond the origin, though in fact any $p > 1$ arbitrarily close to 1 will suffice for all our results.

From Assumption 3, the function $\omega(r)$ defined by (1) is even and bounded. Assumption 4 is needed to evaluate the cumulants of $\widehat{f}(0)$ and is satisfied for most kernels used in practice satisfying Assumption 3, but rules out kernels like the uniform. A modification of the proofs could permit kernels that have finitely many discontinuities. The second condition in Assumption 5 is designed for nonparametric bias reduction when $d > 2$ by means of higher-order kernels. Examples of kernels satisfying Assumptions 3, 4 and 5 are for $d = 2$, the Bartlett-Priestley or Epanechnikov window $K(\lambda) = \frac{3}{4\pi}(1 - \frac{\lambda^2}{\pi^2})$ and the triangular window $K(\lambda) = \frac{1}{\pi}\left(1 - \frac{|\lambda|}{\pi}\right)$; for $d > 2$, the following optimal kernels are taken from Gasser et al. (1985):

$$\text{for } \ d = 4, \qquad K_4(\lambda) = \frac{15}{32\pi}\left(7\frac{\lambda^4}{\pi^4} - 10\frac{\lambda^2}{\pi^2} + 3\right);$$

$$\text{for } \ d = 6, \qquad K_6(\lambda) = \frac{35}{256\pi}\left(-99\frac{\lambda^6}{\pi^6} + 189\frac{\lambda^4}{\pi^4} - 105\frac{\lambda^2}{\pi^2} + 15\right).$$

Assumption 6 on the bandwidth or lag number M is necessary for the consistency of $\widehat{f}(0)$, while we will sometimes wish to strengthen it by Assumption 7, possibly with restrictions on q.

3 Distribution of the nonparametric spectral estimate

In this section we analyze the asymptotic distribution of the nonparametric spectral estimate $\widehat{f}(0)$. Our results extend Bentkus and Rudzkis (1982) in that we do not assume boundedness of the spectral density at frequencies away from the origin. We give two lemmas about the bias of the estimate $\widehat{f}(0)$ for V_N. The first is standard in Fourier analysis (see Zygmund (1977), p. 91), and the logarithmic factor could be eliminated by assuming $\sum |\gamma(j)| < \infty$.

Lemma 1 *Under Assumption 1, with $d = 1$, $\varrho = 0$, $V_N - 2\pi f(0) = O\big(N^{-1}\log N\big)$, as $N \to \infty$.*

Lemma 2 *Under Assumptions 1, 3, 5 and 6, as $N \to \infty$,*

$$E[\widehat{f}(0)] - f(0) - \frac{f^{(d)}(0)}{d!}\mu_d(K)M^{-d} = O\big(N^{-1}\log N + M^{-d-\varrho}\big).$$

where $f^{(d)}(0)$ is the d-th derivative of $f(\lambda)$ evaluated at $\lambda = 0$.

From Lemmas 1 and 2 we estimate the relative bias b_N as $M \to \infty$

$$b_N = b_1 M^{-d} + O(M^{-d-\varrho} + N^{-1}\log N), \qquad b_1 = \frac{f^{(d)}(0)\,\mu_d(K)}{d!\,f(0)}.$$

We now study the cumulants of the normalized spectral estimate u_2.

Lemma 3 *Under Assumptions 1, 3, 4, $e_N(s) \stackrel{def}{=} M^{-d-\varrho} + N^{-1}M\log^{2s-1}N \to 0$ as $N \to \infty$, for $s > 2$,*

$$\bar{\kappa}_N[0,s] \ \stackrel{def}{=} \ \kappa_N[0,s]\left(\frac{N}{M}\right)^{(s-2)/2} = \sum_{j=0}^{d} \nabla_j[0,s]M^{-j} + O(e_N(s)),$$

where $\nabla_j[0,s]$ are bounded and depend on the moments of K and the derivatives of f at $\lambda = 0$, and do not depend on N or M.

For example $\nabla_0[0,s] = (4\pi)^{\frac{s-2}{2}}(s-1)!\|K\|_2^{-s}\|K\|_s^s$, $\nabla_1[0,s] = 0$ and the ∇ coefficients are scale-free as expected, but depend on the shape of f. If f is flat at $\lambda = 0$ then $\nabla_j[0,s] = 0$, $j \geq 1$. The proof of Lemma 3 employs a multivariate version of the Fejér kernel (see Appendix B) and uses the fact that, given the compact support of K, asymptotically we only smooth around zero frequency. Depending on the asymptotic relationship between M and N, some of the expansion can be included in the error term, since we have only assumed that $e_N(s)$ is $o(1)$ as $N \to \infty$, which in turn implies Assumption 6 for $s \geq 1$.

Due to the normalization $\kappa_N[0,2] = 1$ and if $e_N(2) \to 0$ as $N \to \infty$, we obtain for the asymptotic variance of $\sqrt{N/M}\widehat{f}(0)$, using the same techniques of the proof of Lemma 3 (see Appendix A), that

$$\frac{N}{M}\mathrm{Var}\left[\widehat{f}(0)\right] = 4\pi f^2(0)\|K\|_2^2 + O\big(e_N(2) + M^{-2}\big)\,,$$

and for some constants Θ_j,

$$\sigma_N = \sqrt{4\pi}\sum_{j=0}^{d}\Theta_j M^{-j} + O(e_N(2)) = \sqrt{4\pi}\|K\|_2 + O\big(M^{-2} + e_N(2)\big)\,,$$

as $M \to \infty$, with $\Theta_0 = \|K\|_2$, $\Theta_1 = 0$ and $\Theta_2 = \frac{1}{4}\|K\|_2^{-1}\mu_2(K^2)f^{-2}(0)\ddot{f}^2(0)$, $\ddot{f}^j(0) = \left(\frac{d}{d\lambda}\right)^2 f^j(\lambda)\Big|_{\lambda=0}$.

Then we can justify an optimal choice of M by minimizing the mean squared error (MSE) of $\widehat{f}(0)$, $E(\widehat{f}(0) - f(0))^2$ under Assumptions 1, 3, 4, 5 and $e_N(2) \to 0$ as $N \to \infty$ (cf. Lemmas 2 and 3), since if we are only interested in estimating f at the origin, it is natural to use local rules for bandwidth choice. Then the M which minimizes asymptotically the MSE is $M_{opt} = c_{opt} \cdot N^{1/(2d+1)}$, $0 < c_{opt} < \infty$, where

$$c_{opt} = c_{opt}(f,K) = \left[\frac{2d}{4\pi}\left(\frac{f^{(d)}(0)\,\mu_d(K)}{d!f(0)\,\|K\|_2}\right)^2\right]^{1/(2d+1)}, \tag{2}$$

which can be estimated by inserting consistent estimates of $f(0)$ and $f^{(d)}(0)$.

We now prove the validity of a second-order Edgeworth expansion to approximate the distribution of the vector $\mathbf{u}$, with error $o((N/M)^{-1/2})$, and including terms up to order $(N/M)^{-1/2}$ to correct the asymptotic normal distribution, which is the leading term of the expansion. Of course this will imply the validity of that expansion for the distribution of $\widehat{f}(0)$. We first study the cross cumulants of $\mathbf{u}$:

Lemma 4 *Under Assumptions 1, 3, 4, $e_N(s+2) \to 0$ as $N \to \infty$, for $s > 0$,*

$$\bar{\kappa}_N[2,s] \overset{def}{=} \kappa_N[2,s]\left(\frac{N}{M}\right)^{s/2} = \sum_{j=0}^{d}\nabla_j[2,s]M^{-j} + O(e_N(s+2))$$

where $\nabla_j[2,s]$ are bounded and depend on the moments of K and the derivatives of f at $\lambda = 0$, and do not depend on N or M.

For example we can obtain that $\nabla_0[2,s] = (4\pi)^{s/2}s!K^s(0)\,\|K\|_2^{-s}$ and $\nabla_1[2,s] = 0$.

For $B \in \mathcal{B}^2$, where $\mathcal{B}^2$ is any class of Borel sets in $\mathbb{R}^2$, set $Q_N^{(2)}\{B\} = \int_B \phi_{\mathbf{2}}(\mathbf{u})q_N^{(2)}(\mathbf{u})d\mathbf{u}$, where $\phi_{\mathbf{2}}(\mathbf{u}) = (2\pi)^{-1}\exp\left\{-\frac{1}{2}\|\mathbf{u}\|^2\right\}$ is the density of the bivariate standard normal distribution,

$$q_N^{(2)}(\mathbf{u}) = 1 + \frac{1}{3!}\left(\frac{M}{N}\right)^{-1/2}\{\nabla_0[0,3]H_3(u_2) + \nabla_0[2,1]H_2(u_1)H_1(u_2)\}\,,$$

and $H_j(\cdot)$ are the univariate Hermite polynomials of order j. Now we show that $Q_N^{(2)}$ is indeed a valid second-order Edgeworth expansion for the probability measure P_N of $\mathbf{u}$. For this we need Assumption 7,

but we do not assume yet the choice $q = 1/(1+2d)$ and/or $C = c_{opt}$ (see (2)) that would minimize the MSE of $\widehat{f}(0)$. This implies a rate of growth for M in terms of N, with Assumption 6 holding for this particular M. Define by $(\partial B)^\alpha$ a neighbourhood of radius α of the boundary of a set B.

Theorem 1 *Under Assumptions 1, 2 ($p > 1$), 3, 4, 7 ($0 < q < 1$), for $\alpha_N = (N/M)^{-\rho}$, $1/2 < \rho < 1$, and every class $\mathcal{B}^2$ of Borel sets in $\mathbb{R}^2$, as $N \to \infty$,*

$$\sup_{B\in\mathcal{B}^2} |P_N(B) - Q_N^{(2)}(B)| = o\left(\left(\frac{N}{M}\right)^{-1/2}\right) + \frac{4}{3}\sup_{B\in\mathcal{B}^2} Q_N^{(2)}\left\{(\partial B)^{2\alpha_N}\right\}.$$

The method of proof is based on first approximating the true characteristic function and then applying a smoothing lemma. Note that the second term on the right hand side is negligible if B is convex because α_N decreases as a power of N, and that the higher-order correction terms in $q_N^{(2)}$ depend only on K, but not on f. Naturally these terms only correct the marginal distribution of the spectral estimate but not that of the Gaussian sample mean. There is also a cross-term to deal with in the joint distribution, but none of these correct for the possible bias of the spectral estimate or for variance estimation since we have only dealt with exactly standardized statistics.

Using the results of Bhattacharya and Ghosh (1978) we can justify Edgeworth expansions for the distribution and moments of smooth functions of the spectral estimate and sample mean. We concentrate in the following section on the studentized mean Y_N.

4 Asymptotic expansion for the distribution of the studentized mean

The distribution of Y_N depends on such quantities as σ_N, b_N, $\kappa_N[r,s]$ etc., for which we have obtained expressions up to a certain degree of error in powers of N and M, the coefficients of the expansions depending on the unknown f and its derivatives at the origin and on the user-chosen kernel $K(\lambda)$. The accuracy of these approximations depends mainly on M and determines the error of the feasible Edgeworth expansion for the distribution of Y_N. In this section we impose Assumption 7 with $q = 1/(1+2d)$, but do not necessarily require that $C = c_{opt}$. Then $0 < M^{-d}/(N/M)^{-1/2} < \infty$ as $N \to \infty$ and the bias of $\widehat{f}(0)$ is of the same magnitude as the correction term obtained in $Q_N^{(2)}$, or as the standard deviation of $\widehat{f}(0)$. However this might not be the optimal choice for approximating the distribution or the MSE of the studentized statistic.

We first work out a linear stochastic approximation to $Y_N(\mathbf{u})$ and prove that its distribution is the same as Y_N up to order $o((N/M)^{-1/2})$. Then the asymptotic approximation for the distribution of the linear approximation is valid also for Y_N with that error. Expanding the bias b_N and the standard deviation σ_N we define

$$Y_N^L \stackrel{def}{=} u_1\left[1 - \frac{1}{2}\,b_1 M^{-d} - \frac{1}{2}\sqrt{4\pi}\|K\|_2\, u_2 (N/M)^{-1/2}\right].$$

Lemma 5 *Under Assumptions 1, 2 (for $p > 1$), 3, 4, 5 and 7, $q = 1/(1+2d)$, Y_N has the same Edgeworth expansion as Y_N^L uniformly for convex Borel sets up to the order $(N/M)^{-1/2}$.*

Note that under the conditions of the Lemma $\widehat{f}(0)$ is $\sqrt{N/M}$-consistent and the approximation we obtained in Section 2 for the distribution of Y_N is valid. The next step is to justify a valid Edgeworth expansion for the distribution of Y_N^L from that of $\mathbf{u}$.

Theorem 2 *Under Assumptions 1, 2 ($p > 1$), 3, 4, 5 and 7, $q = 1/(1+2d)$, for convex Borel sets C, as $N \to \infty$,*

$$\sup_C \left| Prob\{Y_N \in C\} - \int_C \phi(x)\left[1 + r_2(x)M^{-d}\right] dx \right| = o((N/M)^{-1/2}) \tag{3}$$

where $r_2(x) = -\frac{1}{2}b_1(x^2-1)$.

This expansion coincides with the formal Edgeworth expansion obtained estimating the first three cumulants of the linear approximation Y_N^L up to error $o((N/M)^{-1/2})$ as was shown by Bhattacharya and Ghosh (1978) for functions of sample moments of independent and identically distributed (i.i.d.) observations. The restriction to convex measurable sets in $\mathbb{R}$, i.e. intervals, could be avoided by proceeding as in that reference.

For the distribution function we set $C = (-\infty, y]$, and integrating and Taylor expanding the distribution function of the standard normal, $\Phi(y)$, we get, uniformly in y, under the conditions of Theorem 2:

$$\begin{aligned} \text{Prob}\{Y_N \le y\} &= \Phi(y) + \frac{1}{2}b_1 y\phi(y)\ M^{-d} + o((N/M)^{-1/2}) \\ &= \Phi\left(y\left[1 + \frac{1}{2}\,b_1 M^{-d}\right]\right) + o((N/M)^{-1/2}) \\ &= \Phi(y) + O((N/M)^{-1/2}), \end{aligned} \tag{4}$$

which shows that the normal approximation is correct up to order $O((N/M)^{-1/2})$ if $q = 1/(1+2d)$. On 'optimally' choosing $C = c_{opt}$ in Assumption 7 from (2), (4) becomes

$$\text{Prob}\{Y_N \le y\} = \Phi\left(y\left[1 + \ b_1'\ N^{-\frac{d}{1+2d}}\right]\right) + o\Big(N^{-\frac{d}{1+2d}}\Big)\,,$$

where

$$b_1' = \frac{b_1}{2}\left[\frac{2d}{4\pi}\left(\frac{f^{(d)}(0)\mu_d(K)}{d!f(0)\|K\|_2}\right)^2\right]^{\frac{-d}{2d+1}},$$

or equivalently, operating with the values of b_1 and c_{opt},

$$\begin{aligned} \text{Prob}\{Y_N \le y\} &= \Phi(y) + a_1\phi(y)(N/M_{opt})^{-1/2} + o\Big((N/M_{opt})^{-1/2}\Big) &(5)\\ &= \Phi\left(y\left[1 + a_1\ (N/M_{opt})^{-1/2}\right]\right) + o\Big((N/M_{opt})^{-1/2}\Big)\,, &(6) \end{aligned}$$

with $a_1 = \sqrt{\pi/(2d)}\,\|K\|_2\,\text{sign}\big[f^{(d)}(0)\,\mu_d(K)\big]$. When $d = 2$

$$b_1 = \frac{1}{2}f^{(2)}(0)\mu_2(K), \qquad a_1 = \frac{\sqrt{\pi}}{2}\|K\|_2\ \text{sign}\Big[f^{(2)}(0)\,\mu_2(K)\Big]\,,$$

and the approximations (4) and (6) have an immediate interpretation. Suppose that $\mu_2(K) = \int x^2K(x)dx > 0$ (e.g. if $K(x) \ge 0$, for all x). If $f(\lambda)$ has a peak at $\lambda = 0$ such that $f^{(2)}(0) < 0$ then, as is well known, the weighted autocovariance estimates $\widehat{f}(0)$ underestimate $f(0)$, and thus the variance of $\overline{X}$, consequently the confidence interval for $\sqrt{N/V_N}\,\overline{X}$ is too narrow for Y_N, and a corresponding test rejects too often since the ratio Y_N tends to increase. Our approximations tend to correct this problem, as in both cases they employ $\Phi(yk_N)$ where $k_N \le 0$, so for the same confidence level, the critical value y is larger (in absolute value) than the normal approximation. The same reasoning applies in the reverse direction, when there is a trough in $f(\lambda)$ at $\lambda = 0$. For $d > 2$ the interpretation is equivalent, but we have to take into account the sign of $K_d^{(1)}$, which can be negative, as for $K_4(x)$ and $d = 4$. The approximations (5)

and (6) are more attractive, since if we believe M is optimal, we need only estimate the sign of $f^{(d)}(0)$, not its value, to achieve second-order correctness.

Taniguchi and Puri (1996) obtained an Edgeworth expansion for the same t-statistic for possibly non-Gaussian AR(1) series when estimating $f(0)$ with the least squares estimate of the autoregressive coefficient θ. Their expansion is correct up to order $o(N^{-1/2})$ and depends on the kurtosis of the innovations but not on θ or f, by contrast to our nonparametric studentization.

We have assumed that EX_t is known in the spectral estimation. When EX_t is unknown, we can still take $EX_t = 0$, but replace $\widehat{\gamma}(\ell)$ by

$$\widetilde{\gamma}(\ell) = \frac{1}{N} \sum_{1 \le t, t+\ell \le N} (X_t - \overline{X})(X_{t+\ell} - \overline{X}), \qquad \ell = 0, \pm 1, \ldots, \pm(N-1),$$

and $\widehat{f}(0)$ by

$$\widetilde{f}(0) = \frac{1}{2\pi} \sum_{\ell=1-N}^{N-1} \omega\left(\frac{\ell}{M}\right) \widetilde{\gamma}(\ell) = (\mathbf{X} - \overline{X}\mathbf{1})' \frac{W_M}{2\pi N} (\mathbf{X} - \overline{X}\mathbf{1}).$$

The effect of mean correction is analyzed in the following lemma.

Lemma 6 *Under the assumptions of Theorem 2,* $NM^{-1}(\widetilde{f}(0) - \widehat{f}(0)) = \Delta_N$, *where* Δ_N *has bounded moments of all orders and* $E[\Delta_N] = -2\pi K(0) f(0) + O\left(MN^{-1} \log^2 N\right)$.

The distribution of $\sqrt{N/M}\widetilde{f}(0)$ is affected to a second order, $(M/N)^{1/2}$, by the mean correction so the studentized mean might be affected to order M/N. The bias is the same as found by Hannan (1958) in spectral estimation after trend removal. Of course, the asymptotic relationship of this bias with the smoothing bias studied in Lemma 2 depends on the degree of smoothing given by M. We substitute $\widetilde{f}(0)$ in all definitions involving $\widehat{f}(0)$ and denote the studentized mean using $\widetilde{f}(0)$ by

$$Y_N^\star = Y_N^\star(\mathbf{u}^\star) = u_1 \left(1 + b_N^\star + \sigma_N^\star u_2^\star \sqrt{\frac{M}{N}}\right)^{-1/2},$$

where $u_2^\star$, $b_N^\star$, $\sigma_N^\star$ and all quantities with a * superscript are as previously, but defined in terms of $\widetilde{V} = 2\pi\widetilde{f}(0)$.

Lemma 7 *Under the assumptions of Theorem 2,* $Y_N^\star$ *has the same Edgeworth expansion as* Y_N^L *for convex Borel sets, up to the order* $(N/M)^{-1/2}$.

It follows that the distribution of the sample mean studentized by the 'mean-corrected' spectral estimate $\widetilde{f}(0)$ can be approximated by the same Edgeworth approximation up to order $(N/M)^{-1/2}$ as when $\widehat{f}(0)$, based on a known mean, is used. However, the expansion for the distribution of $\mathbf{u}^\star$ can differ from that for the distribution of $\mathbf{u}$ in terms of order $(N/M)^{-1/2}$ as we investigate in Section 6.

5 Empirical approximation

The above approximations to the distribution of the studentized mean, and to optimal bandwidth choice, depend on the unknown $f(0)$ and derivative $f^{(d)}(0)$. These may be estimated in standard plug-in fashion (using an initial choice of bandwidth) to achieve an empirical Edgeworth approximation and approximately optimal bandwidth. This section proposes nonparametric estimates of the derivatives of

f and proves their consistency. Of course f has to be smoother than is necessary in estimation of $f(0)$, but again only around frequency zero.

We introduce the class of kernels $(\nu,\, r)$ $\nu = 0, 1, \ldots, r-1$ to estimate the ν-th derivative, following Gasser et al. (1985). Define the function V_ν of order (ν, r) such that

$$\int_\Pi V_\nu(x) x^j dx = \begin{cases} 0, & j = 0, \ldots, \nu-1, \nu+1, \ldots, r-1; \\ (-1)^\nu \nu!, & j = \nu; \\ \vartheta \neq 0, & j = r, \end{cases}$$

with support $[-\pi, \pi]$, and satisfying a Lipschitz condition of order 1. If $\nu = 0$ then we estimate the function itself and V_0 has equivalent properties to the kernel K we used to estimate f (compare this with Assumptions 3, 4 and 5). Examples of the class of kernels $(\nu,\, r)$ on $[-\pi, \pi]$ are,

$$\text{for } \nu{=}2,\, r{=}4, \qquad V_2(x) = \frac{105}{32\pi}\left(-5\frac{\lambda^4}{\pi^4} + 6\frac{\lambda^2}{\pi^2} - 1\right);$$
$$\text{for } \nu{=}2,\, r{=}6, \qquad V_2(x) = \frac{315}{64\pi}\left(77\frac{\lambda^6}{\pi^6} - 135\frac{\lambda^4}{\pi^4} + 63\frac{\lambda^2}{\pi^2} - 5\right).$$

We define $V_{m_\nu}(x) = m_\nu V_\nu(m_\nu x)$, $x \in [-\pi, \pi]$, for a sequence of integers $m_\nu = m_\nu(N)$, satisfying $m_\nu^{-1} + m_\nu N^{-1} \to 0$ as $N \to \infty$. We estimate $f^{(\nu)}(0)$ by

$$\widehat{f}^{(\nu)}(0) = (m_\nu)^\nu \int_\Pi V_{m_\nu}(\lambda) I(\lambda) d\lambda.$$

Lemma 8 *Under Assumption 1, $d = \nu{+}a$, $\varrho = 0$, and a kernel of order $(\nu, \nu{+}a)$, for some integer $a \geq 2$, and $(m_\nu)^{-1} + N^{-1}(m_\nu)^\nu \log N \to 0$ as $N \to \infty$, $E[\widehat{f}^{(\nu)}(0)] - f^{(\nu)}(0) = O\big((m_\nu)^\nu [N^{-1}\log N + {m_\nu}^{-\nu-a}]\big)$.*

Lemma 9 *Under the assumptions of Lemma 8, with $(m_\nu)^{-1} + N^{-1}(m_\nu)^{2\nu+1} + N^{-1} m_\nu \log^3 N \to 0$ as $N \to \infty$, $N {m_\nu}^{-2\nu-1} Var[\widehat{f}^{(\nu)}(0)] = 4\pi f^2(0) \|V_\nu\|_2^2 + o(1)$.*

Then with the conditions of these two Lemmas it is possible to obtain valid empirical Edgeworth expansions because the correction terms are of order $(M/N)^{1/2}$ and consistent estimates for f and $f^{(d)}$ introduce only an $o_p((M/N)^{1/2})$ error. Using the same techniques as for the cumulants of $\widehat{f}(0)$ (cf. Lemma 3) we can show that the fourth-order cumulant of $\widehat{f}^{(\nu)}(0)$, $\kappa_N(4)$, is of order of magnitude $N^{-2} m_\nu^{2(\nu+1)}$ and its fourth moment is therefore

$$\begin{aligned} E\left(\widehat{f}^{(\nu)}(0) - f^{(\nu)}(0)\right)^4 &= 3\mathrm{Var}\left[\widehat{f}^{(\nu)}(0)\right]^2 + 3E\left[\widehat{f}^{(\nu)}(0) - f^{(\nu)}(0)\right]^4 + \kappa_N(4) \\ &= O\left(m_\nu^{4\nu+2} N^{-2} + {m_\nu}^{4\nu} N^{-4} \log^4 N + {m_\nu}^{-4a}\right), \end{aligned} \tag{7}$$

from Lemmas 8 and 9. Then $\widehat{f}^{(\nu)}(0) \to f^{(\nu)}(0)$ almost surely from the Borel-Cantelli lemma and Markov's inequality if (7) is $O(N^{-1-\epsilon})$ for some $\epsilon > 0$. Given the MSE-optimal $m_\nu \sim C N^{1/(2\nu+2a+1)}$, this holds if $a > \nu + \frac{1}{2}$ and valid empirical Edgeworth expansions are thus available with $o((M/N)^{1/2})$ error, almost surely.

The same results hold if $\widehat{f}^{(\nu)}(0)$ is replaced by $\widetilde{f}^{(\nu)}(0)$ which employs mean-corrected quantities in the manner of $\widetilde{f}(0)$, while the distribution of derivative estimates can be studied in the same way as that of $\widehat{f}$. These estimates can also be used for plug-in rules of bandwidth choice, but estimates of M can affect higher-order properties of $\widehat{f}$ and t-ratios though first order asymptotics are likely to remain the same (cf. Robinson (1991)).

6 Third-order approximation

In this section we concentrate on obtaining a third-order approximation (that is, including terms of order M/N) to the distribution of the studentized sample mean. The previous results are insufficient to prove the validity when there is mean-correction in the nonparametric spectral estimate. As seen in Section 4, the mean-correction introduces a term of order $(M/N)^{1/2}$ in the expansion for $\sqrt{N/M}\widetilde{f}(0)$, so it will have an effect of order M/N in a third-order approximation for the studentized mean. As before, we denote by a star superscript, *, all quantities when $\widetilde{f}(0)$ is used instead of $\widehat{f}(0)$. First we study the bias, the following lemma simply extending Lemma 2 using Lemma 6:

Lemma 10 *Under Assumptions 1, 3, 4, 5, 6 and $M^{-1}+N^{-1}M\log N\to 0$ as $N\to\infty$,*

$$E[\widetilde{f}(0)]-f(0)=\frac{f^{(d)}(0)}{d!}\mu_d(K)M^{-d}-2\pi f(0)K(0)\frac{M}{N}+O\left(\frac{\log N}{N}+M^{-d-\varrho}+\left[\frac{M}{N}\right]^2\log^2 N\right).$$

The second term on the right hand side is due to the mean correction. To analyze the cumulants of $u_2^\star$ we can write it compactly as a quadratic form, $\mathbf{X}$, $u_2^\star=\mathbf{X}'Q_N^\star\mathbf{X}-E[\mathbf{X}'Q_N^\star\mathbf{X}]$, where $Q_N^\star=A_NQ_NA_N$, $A_N=I_N-\mathbf{1}\mathbf{1}'/N$, is the *mean-corrected* version of Q_N. We first analyze the cumulants of the joint distribution of $\mathbf{u}^\star$.

Lemma 11 *Under Assumptions 1, 3, 4, $e_N(s)\to 0$ as $N\to\infty$, for $s>2$,*

$$\bar\kappa_N^\star[0,s]\overset{def}{=}\kappa_N^\star[0,s]\left(\frac{N}{M}\right)^{(s-2)/2}=\sum_{j=0}^{d}\nabla_j[0,s]M^{-j}+O(e_N(s))\,,$$

$$\bar\kappa_N^\star[2,s-2]\overset{def}{=}\kappa_N^\star[2,s-2]\left(\frac{N}{M}\right)^{(s-2)/2}=O(e_N(s))\,,$$

where $\nabla_j[0,s]$ are defined as in Lemma 3.

The cumulants $\kappa_N^\star[0,s]$ of $u_2^\star$ thus have the same asymptotic approximations as the $\kappa_N[0,s]$, and all conclusions about the variance and optimal bandwidth with known mean assumed still go through. However the cross-cumulants $\kappa_N^\star[2,s]$ are asymptotically $o(1)$ after normalization. Therefore on the basis of cross-cumulants of any order, u_1 and $u_2^\star$ are asymptotically independent and variance estimation is asymptotically independent of mean estimation as if the sequence X_t were exactly independent.

We now fix the order needed for the expansions of the cumulants to obtain a third-order Edgeworth expansion for the distribution of $\mathbf{u}^\star$ when Assumption 7 holds. We need to consider terms in the expansion of $\kappa_N^\star[0,3]$ up to order $M^{-d^\star}$ such that if $d^\star<d$, then $M^{-d^\star-1}=o\left((M/N)^{1/2}\right)$, and if $d^\star=d$, then $M^{-d-\varrho}=o\left((M/N)^{1/2}\right)$, the errors being negligible if $q>1/(1+2d+2\varrho)$. The following theorem establishes validity of the third-order Edgeworth approximation $Q_N^{(3)\star}\{B\}=\int_B\phi_{\mathbf{2}}(\mathbf{u})q_N^{(3)\star}(\mathbf{u})d\mathbf{u}$, for the distribution $P_N^\star$ of $\mathbf{u}^\star$, where

$$q_N^{(3)\star}(\mathbf{u})=1+\frac{1}{6}\left(\frac{M}{N}\right)^{1/2}\sum_{j=0}^{d^\star}M^{-j}\nabla_j[0,3]H_3(u_2)+\frac{1}{72}\frac{M}{N}\left\{\nabla_0[0,3]^2H_6(u_2)+\frac{1}{4!}\nabla_0[0,4]H_4(u_2)\right\}.$$

Theorem 3 *Under Assumptions 1, 2 ($p>1$), 3, 4, 7 ($1/(1+2d+2\varrho)<q<1$), for $\alpha_N=(N/M)^{-\rho}$, $1<\rho<3/2$ and every class of Borel sets $\mathcal{B}^2$, as $N\to\infty$,*

$$\sup_{B\in\mathcal{B}^2}|P_N^\star(B)-Q_N^{(3)\star}(B)|=o\left(\left(\frac{N}{M}\right)^{-1}\right)+\frac{4}{3}\sup_{B\in\mathcal{B}^2}Q_N^{(3)}\left\{(\partial B)^{2\alpha_N}\right\}.$$

Next we consider the studentized sample mean $Y_N^\star$ using the nonparametric estimate $\widetilde{f}(0)$. To obtain a linear approximation for $Y_N^\star$, the main problem is the bias

$$b_N^\star = b_1 M^{-d} + b_2 \frac{M}{N} + O\left(N^{-1} \log N + M^{-d-\varrho} + \left[\frac{M}{N}\right]^2 \log^2 N \right),$$

with $b_2 = -2\pi K(0)$. To make $b_N^\star$ negligible up to order M/N we cannot employ the MSE$[\widetilde{f}(0)]$-optimal M, but instead require that

$$\lim_{N\to\infty} \frac{M}{N} M^d > 0, \tag{8}$$

which guarantees that the bias term of order M^{-d} is at most of order M/N, and that the term $O(M^{-d-\varrho})$ does not affect the third-order approximation under Assumption 7. This of course implies a significant undersmoothing, as M needs to increase much faster than $N^{1/(1+2d)}$, at least like $N^{1/(1+d)}$. Then incorporating the bias of order $O(MN^{-1})$, the third-order linear approximation to $Y_N^\star$ is

$$Y_N^{\star L} = u_1 \left[1 - \frac{1}{2} b_1 M^{-d} - \frac{1}{2} b_2 \frac{M}{N} - \frac{1}{2} \sigma_N u_2^\star \left(\frac{M}{N}\right)^{1/2} + \frac{3}{8} 4\pi \|K\|_2^2 (u_2^\star)^2 \frac{M}{N} \right], \tag{9}$$

and we justify the validity of a third-order Edgeworth expansion for the distribution of $Y_N^\star$ with

$$r_N(x) = \left[4\pi \|K\|_2^2 + 2\pi K(0) - b_1 N M^{-1-d} \right] \frac{x^2 - 1}{2} + 12\pi \|K\|_2^2 \frac{x^4 - 6x^2 + 3}{24}. \tag{10}$$

Theorem 4 *Under Assumptions 1, 2 ($p > 1$), 3, 4, 5, 7 and (8), for convex Borel sets C, as $N \to \infty$,*

$$\sup_C \left| Prob\{Y_N^\star \in C\} - \int_C \phi(x) \left[1 + r_N(x) \frac{M}{N} \right] dx \right| = o\left(\frac{M}{N}\right).$$

In particular, for the distribution function we obtain, uniformly in y,

$$\begin{aligned} \text{Prob}\{Y_N^\star \le y\} &= \Phi(y) - \frac{1}{2}\phi(y)\left(y^3 - 3y\right)\pi \|K\|_2^2 M N^{-1} \\ &+ \frac{1}{2} y \phi(y) \left[b_1 N M^{-d-1} - 4\pi \|K\|_2^2 - 2\pi K(0) \right] M N^{-1} + o(M/N). \end{aligned} \tag{11}$$

The coefficients of the polynomial $r_N(x)$ depend only on K, except for the term $b_1 N M^{-1-d}$, which involves $f(0)$ and $f^{(d)}(0)$. This is due to the moments of $\widetilde{f}(0)$ being proportional to $f(0)$, so the normalized distribution of $\mathbf{u}^\star$ has constant variance and higher-order cumulants (up to first order) with respect to $f(0)$. The term in b_1 disappears with sufficient undersmoothing, that is, if in (8) the left hand side is infinite. Of course, the larger M, the worse the approximation from the point of view of the M/N corrections. More informative expansions for the bias can be obtained, using higher-order derivatives of the spectral density at the origin and appropriate conditions on the kernel. Then (8) could be relaxed allowing the term in b_1 to be of larger order of magnitude than M/N and also permitting MSE-optimal M_{opt}.

To obtain the Edgeworth expansion of Theorem 4 we can simply calculate the formal expansion for the distribution of $Y_N^{\star L}$ based on the moments of $\mathbf{u}$ or we can proceed in an alternative way. Since we found in Lemma 11 that $\widetilde{f}(0)$ is asymptotically independent of $\overline{X}$, we can write

$$\Pr(Y_N^\star \le y) = \Pr\left(u_1 \le S^{1/2} y\right) \approx E_{u_1}\left[\Phi\left(S^{1/2} y \right) \right],$$

where $S = 1 + b_N^\star + u_2^\star \sigma_N^\star (M/N)^{1/2}$ and regarding u_1 and $u_2^\star$ as exactly independent. Then we can expand $\Phi\left(S^{1/2} y\right)$ around $\Phi(y)$,

$$\Phi\left(S^{1/2} y\right) = \Phi(y) + \phi(y)\, y \left(S^{1/2} - 1 \right) - \frac{1}{2} y^3 \phi(y) \left(S^{1/2} - 1 \right)^2 + \frac{1}{6}((y')^2 - 1) y^3 \phi(y') \left(S^{1/2} - 1 \right)^3, \tag{12}$$

where y' is in the line segment between y and $S^{1/2}y$. Now

$$S^{1/2} = 1 + \frac{1}{2}b_N^\star + \frac{1}{2}u_2^\star \sigma_N (M/N)^{1/2} - \frac{1}{8}u_2^2 \sigma_N^2 \frac{M}{N} + \xi_N,$$

where $E\left|\xi_N\right| = O((MN^{-1})^{3/2} + (b_N^\star)^2)$ and $b_N^\star \sim b_1 M^{-d} - 2\pi K(0) M N^{-1}$, obtaining

$$\begin{aligned} E\Big(S^{1/2} - 1\Big) &= \frac{1}{2}b_N^\star - \frac{1}{8}\sigma_N^2 \frac{M}{N} + o\Big(b_N^\star + \frac{M}{N}\Big) \\ E\Big(S^{1/2} - 1\Big)^2 &= \frac{1}{4}\sigma_N^2 \frac{M}{N} + o\Big((b_N^\star)^2 + \frac{M}{N}\Big). \end{aligned}$$

Therefore, taking expectations in (12) and grouping terms in powers of y, we obtain the same approximation for $\Pr\{Y_N^\star \le y\}$ as in (11),

$$E_{u_1}\Big[\Phi\left(S^{1/2}y\right)\Big] = \Phi(y) + y\phi(y)\left(\frac{1}{2}b_N^\star - \frac{1}{8}\sigma_N^2\frac{M}{N}\right) - \frac{y^3}{8}\phi(y)\,\sigma_N^2\frac{M}{N} + o\Big(b_N^\star + \frac{M}{N}\Big),$$

with a truncating error $O\Big(E\left|\left(S^{1/2}-1\right)\right|^3\Big)$, proceeding as in the Lemma of Robinson (1995b).

Following Hall (1992, Section 2.5) and using Theorem 4, we can also obtain a Cornish-Fisher approximation for the quantiles of the distribution of the studentized mean $Y_N^\star$ to construct, e.g., confidence intervals with improved asymptotic coverage by estimating the unknown terms in $r_N(x)$ as proposed in Section 5. Write $w_\alpha = w_\alpha(N,M)$ for the α-level quantile of $Y_N^\star$, determined by $w_\alpha = \inf\{x : \text{Prob}\{Y_N^\star \le x\} \ge \alpha\}$, and let z_α be the α-level standard normal quantile, given by $\Phi(z_\alpha) = \alpha$. Then immediately we have

Theorem 5 *Under Assumptions 1, 2 ($p > 1$), 3, 4, 5, 7 and (8), $w_\alpha = z_\alpha - r_N(z_\alpha)M/N + o(M/N)$, uniformly in $\epsilon < \alpha < 1 - \epsilon$ for each $\epsilon > 0$, where r_N is defined as before.*

7 Non-Gaussian time series

Though our development depends heavily on the Gaussianity assumption we here analyze informally the consequences up to third order of the Gaussianity relaxation. This may be achieved by considering distributions with Gram-Charlier representations incorporating corrections for skewness and kurtosis (see Phillips (1980) for related references).

The lack of Gaussianity affects in the first instance the joint characteristic function of $\mathbf{u}$, for which we would require some regularity conditions (cf. Lemmas 14 and 15 in Appendix B). This regularity involves the distribution of X_t and would also require summability conditions on higher-order cumulants or mixing type conditions as in Götze and Hipp (1983). Then the lack of Gaussianity shows up in the asymptotic approximations to the distributions in terms of the higher-order cumulants of the sequence X_t. It is well known (e.g. Hannan (1970, pp. 280)) that fourth-order cumulants do not affect (at first order) the asymptotic variance of smoothed estimates $\widehat{f}(0)$ and the same can be shown for higher-order cumulants of the normalized statistics u_1 and u_2.

Thus if we assume higher-order stationarity of X_t and that the higher-order spectral densities

$$f_k(\lambda_1, \ldots, \lambda_{k-1}) = (2\pi)^{1-k} \sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_{k-1}=-\infty}^{\infty} \text{cum}(X_o, X_{j_1}, \ldots, X_{j_{k-1}}),$$

$f_2(\lambda) = f(\lambda)$, are smooth enough at the origin in all their arguments, then simple results can be obtained. This condition on the higher-order spectral densities holds if for example X_t is a linear

process, $X_t = \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j}$, where the ϵ_t are i.i.d. with enough moments and the transfer function $\alpha(\lambda) = \sum_{j=0}^{\infty} \alpha_j \exp(i\lambda j)$ is sufficiently smooth at $\lambda = 0$; sufficiently strong summability conditions on α_j provide uniform smoothness. Then we can show that the normalized cumulants of $\mathbf{u}$, $\bar{\kappa}_N[a,b]$, are of the same magnitude as under Gaussianity, with identical leading terms, since higher-order cumulant spectra only appear in higher-order, $o(1)$, terms in their asymptotic expansions. Thus, up to errors of order $O(M^{-2} + e_N(a+b))$, we obtain

$$\begin{aligned} \bar{\kappa}_N[3,0] &= \sqrt{2\pi}\bar{f}_3(\mathbf{0})M^{-1/2}; \\ \bar{\kappa}_N[4,0] &= 2\pi\bar{f}_4(\mathbf{0})M^{-1}, \end{aligned}$$

see e.g. Götze and Hipp (1983), where $\bar{f}_k(\mathbf{0}) = f_k(\mathbf{0})f^{-k/2}(0)$. For the spectral density estimate we obtain that $\sigma_N^2 = 4\pi \|K\|_2^2 + 2\pi\bar{f}_4(\mathbf{0})M^{-1}$, using the techniques of Bentkus (1976), and with similar arguments the first cross cumulants of $\mathbf{u}$ are

$$\begin{aligned} \bar{\kappa}_N[1,1] &= \frac{1}{\sqrt{2}} \|K\|_2^{-1} \bar{f}_3(\mathbf{0})M^{-1/2}; \\ \bar{\kappa}_N[2,1] &= \sqrt{4\pi}K(0) \|K\|_2^{-1} + O(M^{-1}), \end{aligned}$$

and $\bar{\kappa}_N[1,2] = O(M^{-1/2})$, so higher-order spectra affect $\bar{\kappa}_N[a,b]$ at most to order $M^{-1/2}$. Then lack of Gaussianity affects neither the term in $(MN^{-1})^{1/2}$ of the Edgeworth approximation for the distribution of $\mathbf{u}$ (cf. $q_N^{(3)\star}$) nor the term in MN^{-1} for the distribution of Y_N (cf. r_N), as in this last case the approximation only depends on the leading terms of $\bar{\kappa}_N[2,1]$ and $\bar{\kappa}_N[1,1]$, (which remain the same) apart from the bias of $\widehat{f}(0)$, which does not depend on higher-order cumulants of X_t. In case of mean-corrected estimates some contributions cancel out, as the leading term of $\bar{\kappa}_N[2,1]$ (cf. Lemma 11).

We can also estimate the MSE of stochastic approximations to Y_N and analyze the higher-order effects of the bandwidth choice when Gaussianity is not assumed. From the third-order linear approximation to Y_N under condition (8),

$$Y_N^L = u_1 \left[1 - \frac{1}{2} b_1 M^{-d} - \frac{1}{2}\sigma_N u_2 \left(\frac{M}{N}\right)^{1/2} + \frac{3}{8}\, 4\pi \|K\|_2^2 u_2^2 \frac{M}{N} \right],$$

where $\sigma_N \sim \sqrt{4\pi}\|K\|_2$ can be expanded up to error $o((MN^{-1})^{1/2})$, and we obtain for non-Gaussian series that

$$E[Y_N^L] = -\frac{1}{2}\sigma_N \kappa_N[1,1] \left(\frac{M}{N}\right)^{1/2} + \frac{3}{2}\pi \|K\|_2^2 \kappa_N[1,2]\frac{M}{N} = -\sqrt{\frac{\pi}{2}}\bar{f}_3(\mathbf{0})N^{-1/2} + o\left(N^{-1/2}\right),$$

so $\mathrm{Bias}[Y_N^L]^2 = (\pi/2)\bar{f}_3^2(\mathbf{0})N^{-1} + o\left(N^{-1}\right)$, and

$$\begin{aligned} \mathrm{Var}[Y_N^L] &= 1 - b_1 M^{-d} - \sigma_N \kappa_N[2,1] \left(\frac{M}{N}\right)^{1/2} + 4\pi \|K\|_2^2 (1 + \kappa_N[2,2]) \frac{M}{N} + O(N^{-1}) \\ &= 1 - b_1 M^{-d} + 4\pi \left(\|K\|_2^2 - K(0) \right) \frac{M}{N} + o\left(M^{-d} + \frac{M}{N} \right). \end{aligned}$$

Similar conclusions can be obtained for mean-corrected spectral estimation, incorporating in the third-order stochastic approximation $Y_N^{\star L}$ (see (9)) the mean-correction bias of order MN^{-1}, $b_2 = -2\pi K(0)$, which remains the same up to that order. Thus bias $\mathrm{Bias}[Y_N^{\star L}]^2$ has the same expression as without mean-correction, because $\kappa_N^\star[1,1] = \kappa_N[1,1](1 + O(MN^{-1}))$, but the term in $K(0)$ in the variance is now different, since $\bar{\kappa}_N^\star[2,1] = o(1)$ from Lemma 11, and hence

$$\mathrm{Var}[Y_N^{\star L}] = 1 - b_1 M^{-d} + 4\pi \left(\|K\|_2^2 + \frac{1}{2}K(0) \right) \frac{M}{N} + o\left(M^{-d} + \frac{M}{N} \right).$$

Note that while the leading terms in the expansions for the variances depend on the properties of $f(\lambda)$ at $\lambda = 0$ and on K, the bias only depends on $\bar{f}_3(\mathbf{0})$, the relative skewness at zero frequency. From an MSE (of Y_N or $Y_N^\star$) point of view, the main focus is then on the variance contribution, and to make the two leading terms of its asymptotic expansion of same order of magnitude we can set $M \sim CN^{1/(1+d)}$ for some positive constant C (so M satisfies condition (8)). This implies a clear undersmoothing, to reduce the bias of $\widehat{f}(0)$, and that the normal approximation for the distributions of Y_N or $Y_N^\star$ is asymptotically correct up to error $O(MN^{-1})$, apart from the skewness correction by $\bar{\kappa}_N[3,0]$ which is of order $O(N^{-1/2})$ as for non-Gaussian standardized $\overline{X}$ and does not depend on spectral estimation (that is, on M or K).

8 Spectral estimation and studentization at non-zero frequencies

We consider in this section nonparametric spectral estimates at a frequency of interest $\lambda_o \in (0, \pi)$, since the case $\lambda_o = \pi$ is similar to estimation at the origin and we need not consider negative frequencies by symmetry. We suppose in this section that Assumption 1 holds in a neighbourhood of λ_o. Now all the arguments we have used for the analysis of $\widehat{f}(0)$ can be carried over to $\widehat{f}(\lambda_o)$,

$$\widehat{f}(\lambda_o) = \frac{1}{2\pi} \sum_{\ell=1-N}^{N-1} \omega\left(\frac{\ell}{M}\right) \widehat{\gamma}(\ell) \cos \ell\lambda_o = \mathbf{X}' \frac{W_M(\lambda_o)}{2\pi N} \mathbf{X},$$

$[W_M(\lambda_o)]_{r,s} = [W_M]_{r,s} \cos(r-s)\lambda_o$, if we keep the symmetry of the estimate $\widehat{f}(0)$ by writing $\widehat{f}(\lambda_o) = \int_\Pi K_M(\alpha-\lambda_o) I(\alpha) d\alpha = \int_\Pi H_M(\alpha) I(\alpha) d\alpha$, where $H_M(\alpha) = H_M(\alpha; \lambda_o) = \frac{1}{2}(K_M(\alpha - \lambda_o) + K_M(\alpha + \lambda_o))$. Now $H_M(\alpha)$ is even and periodic like $K_M(\alpha)$, and higher-order cumulants of $\widehat{f}(\lambda_o)$ are determined by the fact that for N large enough the kernels $K_M(\alpha - \lambda_o)$ and $K_M(\alpha + \lambda_o)$ do not overlap for $\lambda_o > 0$. However we cannot expect $f(\lambda)$ to be symmetric around non-zero λ_o as it automatically is around the origin, so existing odd derivatives of $f(\lambda)$ at λ_o are not zero in general and the expansion for moments of $\widehat{f}(\lambda_o)$ might contain additional terms. Furthermore, there is less reason in general to expect a spectral peak at an arbitrarily chosen non-zero frequency λ_o than at the origin, so interpretation of correction terms may be less immediate.

Define the discrete Fourier transform at λ as $w(\lambda) = N^{-1} \sum_{t=1}^N X_t \exp(i\lambda t)$, so $\overline{X} = w(0)$, and denote $w(\lambda) = w^R(\lambda) + i w^I(\lambda)$ for the real and complex components of $w(\lambda)$. Then for $\lambda_o > 0$

$$V_N^R(\lambda_o) \stackrel{def}{=} \mathrm{Var}[\sqrt{N}\, w^R(\lambda_o)] = \frac{1}{2} V_N(\lambda_o) + O(N^{-1} \log N),$$

where $V_N(\lambda_o) \stackrel{def}{=} \sum_{j=1-N}^{N-1} \left(1 - \frac{|j|}{N}\right) \gamma(j) \cos j\lambda_o = 2\pi f(\lambda_o) + O(N^{-1} \log N)$, using Assumption 1 as in Lemma 1. Then for any λ_o and N such that $V_N^R(\lambda_o) > 0$, we set

$$u_1^R(\lambda_o) \stackrel{def}{=} \frac{\sqrt{N}\, w^R(\lambda_o)}{\sqrt{V_N^R(\lambda_o)}} \sim \mathcal{N}(0,1),$$

and we can define $V_N^I(\lambda_o) \stackrel{def}{=} \mathrm{Var}[\sqrt{N}\, w^I(\lambda_o)] = \frac{1}{2} V_N(\lambda_o) + O(N^{-1} \log N)$ and $u_1^I(\lambda_o)$ similarly for $w^I(\lambda_o)$. The studentized statistic at frequency λ_o is

$$Y_N^R(\lambda_o) \stackrel{def}{=} \frac{\sqrt{N}\, w^R(\lambda_o)}{\sqrt{\widehat{V}(\lambda_o)}} = u_1^R(\lambda_o) \left(1 + b_N^R(\lambda_o) + \frac{\pi f(\lambda_o)}{V_N^R(\lambda_o)} \sigma_N(\lambda_o) u_2(\lambda_o) \sqrt{\frac{M}{N}}\right)^{-1/2}.$$

Here $\widehat{V}(\lambda_o) = \pi \widehat{f}(\lambda_o)$ and

$$u_2(\lambda_o) \stackrel{def}{=} \sqrt{\frac{N}{M}} \frac{\widehat{V}(\lambda_o) - E[\widehat{V}(\lambda_o)]}{\pi f(\lambda_o)\sigma_N(\lambda_o)}$$

are common for studentization of both $w^R(\lambda_o)$ and $w^I(\lambda_o)$, where $\sigma_N^2(\lambda_o)$ and $b_N^R(\lambda_o)$ are now the "relative" variance (with respect to $\pi f(\lambda_o)$) and bias of $\widehat{V}(\lambda_o)$, respectively. The bias estimation follows as for $\lambda_o = 0$ with $b_N^R(\lambda_o),\ b_N^I(\lambda_o) = b_1(\lambda_o)M^{-d} + O(M^{-d-\varrho} + N^{-1}\log N)$, and

$$b_1(\lambda_o) = \frac{f^{(d)}(\lambda_o)\mu_d(K)}{f(\lambda_o)d!}.$$

We can analyze the joint distribution of $\mathbf{u}(\lambda_o) = (u_1^R(\lambda_o), u_1^I(\lambda_o), u_2(\lambda_o))'$ under Gaussianity using the same definitions as for $\lambda_o = 0$, but in terms of the matrix $Q_N(\lambda_o) = (MN)^{-1/2}(\sigma_N(\lambda_o)\pi f(\lambda_o))^{-1}W_M(\lambda_o)$ and the vectors $\xi_N^R(\lambda_o) = (\cos\lambda_o, \ldots, \cos N\lambda_o)' / \sqrt{NV_N^R(\lambda_o)}$ and $\xi_N^I(\lambda_o) = (\sin\lambda_o, \ldots, \sin N\lambda_o)' / \sqrt{NV_N^I(\lambda_o)}$. The characteristic function of $\mathbf{u}(\lambda_o)$ is

$$\psi_N^{\lambda_o}(t_1^R, t_1^I, t_2) = |I - 2it_2\Sigma Q_N(\lambda_o)|^{-1/2} \exp\left\{-\frac{1}{2}\xi_N^{\lambda_o}(\mathbf{t})' \left(I - 2it_2\Sigma Q_N(\lambda_o)\right)^{-1} \Sigma \xi_N^{\lambda_o}(\mathbf{t}) - it_2 E_N(\lambda_o)\right\},$$

$\xi_N^{\lambda_o}(\mathbf{t}) = t_1^R\xi_N^R(\lambda_o) + t_1^I\xi_N^I(\lambda_o)$, and the only cumulants differing from zero are $\kappa_N^{\lambda_o}[a,b,s]$ for $a+b = 0, 2,\ s \geq 0$. Thus, for example, $\kappa_N^{\lambda_o}[1,1,s] = 2^s s!\,(\xi_N^R(\lambda_o))'(\Sigma Q_N(\lambda_o))^s \Sigma \xi_N^I(\lambda_o),\ s \geq 0$, and setting $\bar{\kappa}_N^{\lambda_o}[a,b,s] \stackrel{def}{=} \kappa_N^{\lambda_o}[a,b,s]\left(NM^{-1}\right)^{(s+a+b-2)/2}$ we obtain:

Lemma 12 *Under Assumptions 1, 3, 4, $e_N(s) \to 0$ as $N \to \infty$, for $s > 2$,*

$$\bar{\kappa}_N^{\lambda_o}[0,0,s] = \sum_{j=0}^{d} \nabla_j^{\lambda_o}[0,s]M^{-j} + O(e_N(s)),$$

$$\bar{\kappa}_N^{\lambda_o}[2,0,s-2],\ \ \bar{\kappa}_N^{\lambda_o}[0,2,s-2]\ = \sum_{j=0}^{d} \nabla_j^{\lambda_o}[2,s-2]M^{-j} + O(e_N(s)),$$

and $\bar{\kappa}_N^{\lambda_o}[1,1,s-2] = O(e_N(s))$, $s \geq 2$, where $\nabla_j^{\lambda_o}[0,s]$ and $\nabla_j^{\lambda_o}[2,s-2]$ are bounded and depend on K and the derivatives of f at λ_o, but not on N or M.

Now $\nabla_0^{\lambda_o}[0,s] = (2\pi)^{\frac{s-2}{2}}(s\text{–}1)!\|K\|_2^{-s}\|K\|_s^s$, $\sigma_N(\lambda_o) \sim \sqrt{2\pi}\|K\|_2$, and $\nabla_0^{\lambda_o}[2,s] = (2\pi)^{s/2}s!K^s(0)\,\|K\|_2^{-s}$ since $H_M(\lambda_o) = \frac{1}{2}K_M(0)$ for N large enough and $\lambda_o > 0$.

When EX_t is unknown we can use the sample mean-corrected statistic $\widetilde{f}(\lambda_o)$, and defining $\Delta_N(\lambda_o) = NM^{-1}\left(\widetilde{f}(\lambda_o) - \widehat{f}(\lambda_o)\right)$ we can follow the arguments of Lemma 6 to find that if $f(\lambda)$ is also smooth at $\lambda = 0$, $E\left[\Delta_N(\lambda_o)\right] = O\left(MN^{-1}\log^2 N\right)$, so $b_2(\lambda_o) = 0$, and $\mathrm{Var}[\Delta_N(\lambda_o)] = O(MN^{-1}\log^2 N)$, because $H_M(0) = K_M(\lambda_o) = 0$ for N large enough and $\lambda_o > 0$. Therefore, mean correction does not affect spectral estimation or studentization at $\lambda_o \neq 0$ at third-order MN^{-1}. However, a similar result to Lemma 6 holds if residuals from a least squares cosinusoidal regression at the same frequency λ_o are used. Also the expansions of Lemma 12 are still valid for the mean-corrected cumulants $\bar{\kappa}_N^{\star\lambda_o}[0,0,s]$, while the cross cumulants $\bar{\kappa}_N^{\star\lambda_o}[2,0,s]$ and $\bar{\kappa}_N^{\star\lambda_o}[0,2,s]$ are $o(1)$ as for $\lambda_o = 0$, leading again to asymptotic independence of $u_1^I(\lambda_o)$, $u_1^R(\lambda_o)$ and $u_2^\star(\lambda_o)$.

Using Lemma 12 we can construct a valid Edgeworth expansion for the distribution of $\mathbf{u}^\star(\lambda_o)$ under the assumptions of Theorem 3, and justify the validity of an Edgeworth approximation for the distribution of $Y_N^{R\star}(\lambda_o)$ in terms of that for $\mathbf{u}^\star(\lambda_o)$ under the conditions of Theorem 4 with

$$r_N^{\lambda_o}(x) = \left[2\pi\|K\|_2^2 - b_1(\lambda_o)NM^{-1-d}\right]\frac{x^2-1}{2} + 6\pi\|K\|_2^2\frac{x^4 - 6x^2 + 3}{24},$$

(cf. (10)) coinciding again with the formal Edgeworth expansion deduced from a linear approximation to $Y_N^{R\star}(\lambda_o)$. This approximation differs from estimation at $\lambda_o = 0$ with respect to the asymptotic variance and negligible bias effect of mean correction for spectral estimation at non-zero frequencies.

9 Appendix A: Proofs

We postpone the proofs of Lemmas 1 and 2 to Appendix B.

Proof of Lemma 3. We obtain for $s > 0$, $\kappa_N[0,s] = 2^{s-1}(s-1)!(\sigma_N V_N)^{-s}(MN)^{-s/2}\,\mathrm{Trace}\left[(\Sigma_N W_M)^s\right]$. Then, using Proposition 1 in Appendix B we have that

$$\bar{\kappa}_N[0,s] = \frac{2^{s-1}(s-1)!\,(2\pi)^{2s-1}}{(\sigma_N V_N)^s}\sum_{j=0}^{d} L_j(s)M^{-j} + O(e_N(s))\,. \tag{13}$$

Applying Proposition 1 to evaluate σ_N^2 under the same set of assumptions ($s = 2$),

$$\sigma_N^2\frac{V_N^2}{4\pi^2} = \frac{N}{M}\frac{2}{(2\pi N)^2}\mathrm{Trace}[(\Sigma_N W_M)^2] = 4\pi\sum_{j=0}^{d} L_j(2)M^{-j} + O(e_N(2))\,,$$

where for example $L_0(2) = f^2(0)\mu_0(K^2) = f^2(0)\|K\|_2^2$, $L_1(2) = 0$ and $L_2(2) = \frac{1}{2}\mu_2(K^2)\ddot{f}^2(0)$. Now as $0 < L_0(2) < \infty$ and all $L_j(2)$ are fixed constants independent of N or M, we can write for some constants $J_j(s)$

$$\left(\sigma_N\frac{V_N}{2\pi}\right)^{-s} = (4\pi)^{-s/2}\sum_{j=0}^{d} J_j(s)M^{-j} + O(e_N(2))\,, \tag{14}$$

where $J_0(s) = L_0(2)^{-s/2}$, etc. Denoting $C(0,s) = (4\pi)^{\frac{s-2}{2}}(s-1)!$ we can obtain from (13) and (14) the following expansion in powers of M^{-1} for the normalized cumulants, $\bar{\kappa}_N[0,s] = C(0,s)\sum_{j=0}^{d}\Gamma_j(s)M^{-j} + O(e_N(s))$, where $\Gamma_j(s) = \sum_{t=0}^{j} J_t(s)L_{j-t}(s)$ are constants not depending on N or M, and depending only on f and K, with $\Gamma_1(s) = 0$, $\Gamma_2(s) = J_0(s)L_2(s) + J_2(s)L_0(s)$, etc. Then the Lemma follows setting $\nabla_j[0,s] = C(0,s)\Gamma_j(s)$. ■

Proof of Lemma 4. We have $\kappa_N[2,s] = 2^s s!\,(MN)^{-s/2}\,N^{-1}V_N^{-s-1}\sigma_N^{-s}\mathbf{1}'(\Sigma_N W_M)^s\Sigma_N\,\mathbf{1}$. Then, using Proposition 2 the normalized cumulants are

$$\bar{\kappa}_N[2,s] = \left[\frac{2\pi}{V_N\sigma_N}\right]^s\frac{2\pi f(0)}{V_N}(4\pi)^s s!f(0)^s K(0)^s + O(e_N(s+2))\,,$$

as $K_M(0) = MK(0)$ given the compact support of K. Substituting the expansion for the value of $V_N\sigma_N$ and using Lemma 1, we obtain

$$\begin{aligned}\bar{\kappa}_N[2,s] &= \left[\frac{V_N\sigma_N}{2\pi}\right]^{-s}\left[1 + O(N^{-1}\log N)\right](4\pi)^s s!f(0)^s K(0)^s + O(e_N(s+2)) \\ &= (4\pi)^{-s/2}(4\pi)^s s!f(0)^s K(0)^s\sum_{j=0}^{d} J_j(s)M^{-j} + O(e_N(s+2))\,,\end{aligned}$$

where the $J_s(j)$ are as before. The Lemma follows with $\nabla_j[2,s] = (4\pi)^{-s/2}(4\pi)^s s!f(0)^s K(0)^s J_j(s)$. ■

Proof of Theorem 1. In order to prove the validity of an Edgeworth expansion for the distribution of $\mathbf{u}$ we check that the characteristic function of the expansion approximates well the true one. We first

construct the approximation for $\psi_N(\mathbf{t})$. We discuss the general case, since the same arguments will be used later for the proof of Theorem 4. As in Taniguchi (1987, pp. 11-14), using the fact that only the cumulants $\kappa_N[0,s]$ and $\kappa_N[2,s]$ are nonzero, the cumulant generating function is

$$\log\psi_N(\mathbf{t}) = \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1}\frac{(N/M)^{(2-s)/2}}{s!}\sum_{|\mathbf{r}|=s}\frac{s!}{r_1!r_2!}\bar{\kappa}_N[r_1,r_2](it_1)^{r_1}(it_2)^{r_2} + R_N(\tau), \tag{15}$$

where $\mathbf{r}=(r_1,r_2)'$, with $r_1\in\{0,2\}$ and $|\mathbf{r}|=r_1+r_2$, and

$$\begin{aligned}
R_N(\tau) &= \left(\frac{N}{M}\right)^{-\tau/2}\left[R_{0,\tau+2}(it_2)^{\tau+2} + R_{2,\tau}(it_1)^2(it_2)^{\tau}\right], \qquad \tau \text{ even},\\
R_N(\tau) &= \left(\frac{N}{M}\right)^{-\tau/2}\frac{1}{(\tau+2)!}\left[\bar{\kappa}_N[0,\tau+2](it_2)^{\tau+2} + \frac{(\tau+2)(\tau+1)}{2}\bar{\kappa}[2,\tau](it_1)^2(it_2)^{\tau}\right]\\
&\quad + \left(\frac{N}{M}\right)^{-(\tau+1)/2}\left[R_{0,\tau+3}(it_2)^{\tau+3} + R_{2,\tau+1}(it_1)^2(it_2)^{\tau+1}\right], \qquad \tau \text{ odd},
\end{aligned}$$

where the $R_{0,j}$ and $R_{2,j}$ are bounded. Thus, from Lemmas 3 and 4, $\log\psi_N(\mathbf{t})$ is

$$\begin{aligned}
&\frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1}\frac{(N/M)^{(2-s)/2}}{s!}\left[\bar{\kappa}_N[0,s](it_2)^s + \frac{s(s-1)}{2}\bar{\kappa}_N[0,s-2](it_1)^2(it_2)^{s-2}\right] + R_N(\tau)\\
= \;& \frac{1}{2}\|i\mathbf{t}\|^2 + \sum_{s=3}^{\tau+1}\left(\frac{N}{M}\right)^{(2-s)/2}\left[B_N(s,\mathbf{t}) + \left\{(it_1)^s + (it_1)^2(it_2)^{s-2}\right\}O(e_N(s))\right] + R_N(\tau),
\end{aligned}$$

where we have grouped terms in powers of M^{-1} in $B_N(s,\mathbf{t})$,

$$B_N(s,\mathbf{t}) = \frac{1}{s!}\sum_{j=0}^{d}M^{-j}\left\{\nabla_j[0,s](it_2)^s + \frac{s(s-1)}{2}\nabla_j[2,s-2](it_1)^2(it_2)^{s-2}\right\}.$$

The approximation of the characteristic function of $\mathbf{u}$ using its cumulant generating function, $A_N(\mathbf{t},\tau)$, has leading term $\exp\{\frac{1}{2}\|i\mathbf{t}\|^2\}$, multiplied by a polynomial in $\mathbf{t}$, depending on the cumulants of $\mathbf{u}$, and N and M,

$$A_N(\mathbf{t},\tau) = \exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}\left[1 + \sum_{j=3}^{\tau+1}\left(\frac{N}{M}\right)^{\frac{2-j}{2}}\sum_{\mathbf{r}}\prod_{n=3}^{\tau+1}\left[B_N(n,\mathbf{t})\right]^{r_n}\frac{1}{r_3!\cdots r_{\tau+1}!}\right],$$

where $\mathbf{r}=(r_3,\ldots,r_{\tau+1})'$, $r_n\in\{0,1,\ldots\}$ and the summation is over all $\mathbf{r}$ satisfying $\sum_{n=3}^{\tau+1}(n-2)r_n = j-2$. We need only keep terms up to a certain power of $(N/M)^{-1/2}$, so some terms in high powers of M^{-1} in $B_N(n,\mathbf{t})$ may be included in the general error term, without increasing its magnitude.

To obtain a second-order Edgeworth expansion we set $\tau=2$, including in $A_N(\mathbf{t},2)$ terms up to order $(N/M)^{-1/2}$,

$$A_N(\mathbf{t},2) = \exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}\left[\,1 \,+\, \bar{B}_N(3,\mathbf{t})\left(\frac{N}{M}\right)^{-1/2}\right], \tag{16}$$

where in $\bar{B}_N(3,\mathbf{t})$ only the leading term (in M^0) is kept in the expansion for the cumulants of order 3.

To measure the distance between the true distribution and its Edgeworth approximation, we apply the smoothing Lemma 13 due to Bhattacharya and Rao (1975, pp. 97-98, 113), with kernel Ψ. Lemma 14 studies the Edgeworth approximation for the characteristic function for $\|\mathbf{t}\|\le\delta_1\sqrt{N/M}$ (note that the characteristic function of the measure $Q_N^{(2)}\{\cdot\}$ is $A_N(\mathbf{t},2)$), whereas Lemma 15 analyzes its tail behaviour. First,

$$\|(P_N - Q_N^{(2)})\star\Psi_{\alpha_N}\| \;\le\; 2\sup_{B\subset B(0,r_N)^c}|(P_N - Q_N^{(2)})\star\Psi_{\alpha_N}| + 2\sup_{B\subset B(0,r_N)}|(P_N - Q_N^{(2)})\star\Psi_{\alpha_N}|,$$

where $r_N = (N/M)^{\beta}$, ($\beta>0$ to be chosen later), and here $\|\cdot\|$ denotes the variation norm of a measure, $\star$ means convolution and B^c the complementary set of B. For $B \subset B(0,r_N)^c$ we have uniformly

$$\begin{aligned} |(P_N - Q_N^{(2)}) \star \Psi_{\alpha_N}| &\leq |P_N \star \Psi_{\alpha_N}| + |Q_N^{(2)} \star \Psi_{\alpha_N}| \\ &\leq \text{Prob}\{\|\mathbf{u}\| \geq r_N/2\} \; + \; 2\Psi_{\alpha_N}\{B(0,r_N/2)^c\} + 2Q^{(2)}\{B(0,r_N/2)^c\}. \end{aligned}$$

Now $Q_N^{(2)}\{B(0,r_N/2)^c\} = o((N/M)^{-1/2})$ as this is the measure of a polynomial in Gaussian variables. Also $\text{Prob}\{\|\mathbf{u}\| \geq r_N/2\} = o((N/M)^{-1/2})$, because $\mathbf{u}$ has finite moments of all orders. Finally, from (44)

$$\Psi_{\alpha_N}\{B(0,r_N/2)^c\} = O([\alpha_N/r_N]^3) = O((N/M)^{-3(\rho+\beta)}) = o((N/M)^{-1/2}),$$

since $\rho + \beta > 1/6$. For $B \subset B(0,r_N)$ we have by Fourier Inversion

$$|(P_N - Q_N^{(2)}) \star \Psi_{\alpha_N}| \leq (2\pi)^{-2} 1\pi r_N^2 \int |(\widehat{P}_N - \widehat{Q}_N^{(2)})(\mathbf{t})\widehat{\Psi}_{\alpha_N}(\mathbf{t})| d\mathbf{t}, \tag{17}$$

where $\widehat{P}$ denotes the characteristic function of a probability measure P, so $\widehat{P}_N = \varphi_N(\mathbf{t})$ and $\widehat{Q}_N^{(2)} = A_N(\mathbf{t},2)$. Using Lemma 14, (17) is bounded by

$$O\left(\left(\frac{N}{M}\right)^{2\beta-1/2}\left[M^{-2} + e_N(3)\right]\right) \int_{\|\mathbf{t}\| \leq \delta_1\sqrt{N/M}} \left| e^{-d_1\|\mathbf{t}\|^2} F(\|\mathbf{t}\|)\right| |\widehat{\Psi}_{\alpha_N}(\mathbf{t})| d\mathbf{t} \tag{18}$$

$$+O((N/M)^{2\beta}) \int_{\delta_1\sqrt{N/M} < \|\mathbf{t}\| \leq a'(N/M)^{\rho}} |(\widehat{P}_N - \widehat{Q}_N^{(2)})(\mathbf{t})\widehat{\Psi}_{\alpha_N}(\mathbf{t})| d\mathbf{t}, \tag{19}$$

because from (45) $\widehat{\Psi}$ is zero for $\|\mathbf{t}\| > a'(N/M)^{\rho}$ and $a' = 8{\cdot}2^{4/3}\pi^{-1/3}$. Then for (18) to be $o((N/M)^{-1/2})$ it is necessary to choose $\beta \leq 1/4$ (due to the definition of $e_N(3)$ and $\beta < q/(1-q)$).

Finally, from Lemma 15, and for $\delta_1 m_N < \|\mathbf{t}\|$ and also for $\delta_1\sqrt{N/M} < \|\mathbf{t}\|$, since $m_N \leq \sqrt{N/M}$ for N large enough (from the first element in the minimum of the definition of m_N), we have that (19) is

$$O((N/M)^{2\beta}) \int_{\delta_1\sqrt{N/M} < \|\mathbf{t}\| \leq a'(N/M)^{\rho}} e^{-d_2 m_N^2} d\mathbf{t} + o((N/M)^{-1/2}),$$

and thus (19) is dominated by $O((N/M)^{2\beta+2\rho})e^{-d_2 m_N^2} + o((N/M)^{-1/2}) = o((N/M)^{-1/2})$, because with Assumption 7, $0 < q < 1$, we have that, for some $\varepsilon > 0$ depending on q and p, $m_N \geq \varepsilon N^{\varepsilon}$. Applying Lemma 13 the proof is complete. ■

Proof of Lemma 5. Set the neighbourhood of the origin $\Omega_N = \{\mathbf{u} : |u_i| < c_i N^{\mu},\ 0 < \mu < d/(3(1+2d)),\ i = 1,2\}$, where c_i are some fixed constants, and expand $Y_N(\mathbf{u})$ around $\mathbf{0}$ in Ω_N, with $|\theta| \leq 1$:

$$Y_N = \delta_N u_1 - \frac{1}{2}\delta_N^3 \sigma_N u_1 u_2 (N/M)^{-1/2} + Z_N(1)(N/M)^{-1}, \tag{20}$$

where $Z_N(1) = \frac{3}{8}\left(1 + b_N + \sigma_N\theta u_2 (N/M)^{-1/2}\right)^{-5/2}\sigma_N^2 u_1 u_2^2$ and $\delta_N = (1+b_N)^{-1/2}$. Substituting for σ_N and δ_N and their powers, we can write $Y_N = Y_N^L + Z_N(N/M)^{-1}$, where $Z_N = \sum_{j=1}^3 Z_N(j)$, $Z_N(2) = u_1 O(M\log N + NM^{-1-d-\varrho})$ and $Z_N(3) = u_1 u_2 O\big((N/M)^{1/2}[M^{-2} + e_N(2)]\big)$. Now we use Theorem 2 of Chibisov (1972) to prove that the error in the previous linear approximation can be neglected with error $o((M/N))^{1/2}$ if

$$\text{Prob}\left\{|Z_N| > \rho_N\sqrt{N/M}\right\} \leq \sum_{j=1}^{3} \text{Prob}\left\{|Z_N(j)| > \frac{1}{3}\rho_N\sqrt{N/M}\right\} = o((N/M)^{-1/2}) \tag{21}$$

for some positive sequence $\rho_N \to 0$ and $\rho_N\sqrt{N/M} \to \infty$. Choosing $\rho_N = 1/\log N$, writing

$$(N/M)^{-1/2} Z_N(2) = u_1 O((N/M)^{1/2}[N^{-1}\log N + M^{-d-\varrho}]) \quad (22)$$

$$(N/M)^{-1/2} Z_N(3) = u_1 u_2 O(M^{-2} + e_N(2)) \quad (23)$$

and applying Chebyshev's inequality, as u_1 and u_2 have finite moments of all orders we see that for (21) to hold it is sufficient that the error terms in the right hand sides of (22) and (23) be $O((N/M)^{-\mu})$, for some $\mu > 0$, which is true due to Assumption 7, $q = 1/(1+2d)$.

To check Chibisov condition (21) for $Z_N(1)$ we bound $\text{Prob}\{Z_N(1) > \rho_N(M/N)^{-1/2}\}$ by

$$\text{Prob}\left\{|R_N(1)|(M/N)^{1/4} > \rho_N^{1/2}\right\} + \text{Prob}\left\{|R_N(2)|(M/N)^{1/4} > \rho_N^{1/2}\right\} = P_1 + P_2,$$

say, where $R_N(2) = \frac{3}{8}\sigma_N^2 u_1 u_2^2$ has bounded moments of all orders. Now $P_2 = o((M/N)^{1/2})$ applying Chebyshev's inequality. Since $b_N = O(M^{-d} + N^{-1}\log N)$ and $(M/N)^{1/10}\rho_N^{-1/5} \to 0$ as $N \to \infty$, $P_1 = \text{Prob}\left\{\left|1 + b_N + \sigma_N\theta u_2 (N/M)^{-1/2}\right| < \rho_N^{-1/5}(M/N)^{1/10}\right\}$, and applying again Chebyshev's inequality this is less than $C\text{Prob}\left\{\left|u_2(M/N)^{1/2}\right| > c\right\} = o((M/N)^{1/2})$, for some positive constants C and c. ■

Proof of Theorem 2. We follow Taniguchi (1987). Consider the transformation $\mathbf{s} = (s_1, s_2)' = \left(Y_N^L(u_1,u_2), u_2\right)' = \Upsilon_N(\mathbf{u})$, say, and its inverse $\mathbf{u} = \Upsilon_N^{-1}(\mathbf{s}) = \left(u_1^\dagger(s_1,s_2), s_2\right)'$. Then we write, using $(1+x)^{-1} = 1 - x + x^2 - x^3 + \cdots$ for $|x| < 1$, uniformly in the set Ω_N, defined as in the proof of Lemma 5,

$$u_1^\dagger(\mathbf{s}) = s_1\left[1 + \frac{1}{2}b_1 M^{-d} + \frac{1}{2}\sqrt{4\pi}\|K\|_2 s_2 (N/M)^{-1/2}\right] + o((N/M)^{-1/2}),$$

where the truncation of the term in $s_1 s_2^2 O((N/M)^{-1})$ with error $o((N/M)^{-1/2})$ is allowed due to the definition of the set Ω_N. Writing for convex sets C, $\text{Prob}\{Y_N \in C\} = \text{Prob}\left\{\mathbf{u} \in \Upsilon_N^{-1}(C\times\mathbb{R})\right\}$, it follows from Lemma 1 that (as Υ_N is a C^∞ mapping on Ω_N),

$$\begin{aligned}&\sup_C \left|\text{Prob}\left\{\mathbf{u} \in \Upsilon_N^{-1}(C\times\mathbb{R})\right\} - Q_N^{(2)}\left\{\Upsilon_N^{-1}(C\times\mathbb{R})\right\}\right| \\ &= o((N/M)^{-1/2}) + \text{const.}\sup_C Q_N^{(2)}\left\{(\partial\Upsilon_N^{-1}(C\times\mathbb{R}))^{2\alpha_N}\right\}, \end{aligned} \quad (24)$$

where $\alpha_N = (N/M)^{-\rho}$, $1/2 < \rho < 1$. Also, from the continuity of Υ_N, we can obtain, for some $c > 0$,

$$Q_N^{(2)}\left\{(\partial\Upsilon_N^{-1}(C\times\mathbb{R}))^{2\alpha_N}\right\} \le Q_N^{(2)}\left\{(\Upsilon_N^{-1}((\partial C)^{c\alpha_N}\times\mathbb{R}))\right\} \quad (25)$$

and

$$\begin{aligned} Q_N^{(2)}\left\{(\Upsilon_N^{-1}(C\times\mathbb{R}))\right\} &= \int_{\Omega_N\cap\Upsilon_N^{-1}(C\times\mathbb{R})} \phi_{\mathbf{2}}(\mathbf{x}) q_N^{(2)}(\mathbf{x}) d\mathbf{x} + o((N/M)^{-1/2}) \\ &= \int_{\Omega_N^\star\cap\{C\times\mathbb{R}\}} \phi_{\mathbf{2}}(\Upsilon^{-1}(\mathbf{s})) q_N^{(2)}(\Upsilon_N^{-1}(\mathbf{s}))|J| d\mathbf{s} + o((N/M)^{-1/2}), \end{aligned}$$

where $\phi_{\mathbf{2}}(\cdot)$ is the bivariate standard normal density, $\Omega_N^\star = \Upsilon_N(\Omega_N)$ and $|J|$ is the Jacobian of the transformation. We can obtain, neglecting terms that contribute $o((N/M)^{-1/2})$ to the integrals, that

$$\begin{aligned} \phi_2(\Upsilon_N^{-1}(\mathbf{s})) &= \phi(s_1)\phi(s_2)\left(1 - \frac{1}{2}s_1^2\left[b_1 M^{-d} + \sqrt{4\pi}\,\|K\|_2\, s_2\left(\frac{M}{N}\right)^{1/2}\right]\right) \\ q_N^{(2)}(\Upsilon_N^{-1}(\mathbf{s})) &= 1 + \left(\frac{M}{N}\right)^{1/2}\frac{1}{3!}\left\{\nabla_0[0,3]H_3(s_2) + \nabla_0[2,1]H_2(s_1)H_1(s_2)\right\}, \end{aligned}$$

and $|J| = 1 + \frac{1}{2}b_1 M^{-d} + \frac{1}{2}\sqrt{4\pi}\,\|K\|_2\, s_2\,(M/N)^{\frac{1}{2}}$. Thus if $p_j(\mathbf{s})$ denote polynomials not depending on N or M,

$$\begin{aligned} Q_N^{(2)}\{\Upsilon_N^{-1}(C\times\mathbb{R})\} &= \int_{\Omega_N^\star\cap\{C\times\mathbb{R}\}} \phi_{\mathbf{2}}(\mathbf{s})\Big[1+p_1(\mathbf{s})(N/M)^{-1/2}+p_2(\mathbf{s})M^{-d}\Big]d\mathbf{s}+o((N/M)^{-1/2}) \\ &= \int_C \phi(s_1)\left\{\int_{\mathbb{R}}\Big[1+p_1(\mathbf{s})(N/M)^{-1/2}+p_2(\mathbf{s})M^{-d}\Big]\phi(s_2)ds_2\right\}ds_1+o((N/M)^{-1/2}) \\ &= \int_C \phi(s_1)\left[1+r_1(s_1)(N/M)^{-1/2}+r_2(s_1)M^{-d}\right]ds_1+o((N/M)^{-1/2}), \end{aligned}$$

where $r_j(s_1)$ are polynomials in s_1, with bounded coefficients in N. Integrating with respect to s_2 in $\mathbb{R}$ we obtain that $r_1(x)=0$ and $r_2(x) = -\frac{1}{2}b_1(x^2-1)$. The proof is completed by recalling (24), (25) and Lemma 5. As in Bhattacharya and Ghosh (1978) this expansion coincides with the formal Edgeworth expansion obtained calculating the first three cumulants of the linear approximation $Y_N^L = s_1$ to Y_N up to error $o((N/M)^{-1/2})$ because $E[s_1]$, $E[s_3] = o((N/M)^{-1/2})$ and $E[s_1^2] = 1 - b_1M^{-d} + o((N/M)^{-1/2})$. ■

Proof of Lemma 6. We obtain $\widetilde{f}(0) - \widehat{f}(0) = -2Z_N + R_N$, where $R_N = (2\pi N)^{-1}\overline{X}^2\mathbf{1}'W_M\mathbf{1}$ and $Z_N = (2\pi N)^{-1}X'W_M\mathbf{1}\overline{X} = (2\pi N^2)^{-1}X'W_M\mathbf{1}\,\mathbf{1}'X = X'\Lambda_N X$, with $\Lambda_N = (2\pi N^2)^{-1}W_M\mathbf{1}\,\mathbf{1}'$ a $N\times N$ matrix. The Lemma follows directly from Lemmas 17 and 18, because

$$\mathrm{Cum}_s[Z_N] = c_s\mathrm{Trace}[(\Sigma_N\Lambda_N)^s] = c_s\left(\frac{M}{N}\right)^s[2\pi f(0)K(0)]^s + O\left(\left(\frac{M}{N}\right)^{s+1}\log^2 N\right),$$

where $c_s = 2^{s-1}(s-1)!$ (so $(N/M)Z_N$ has bounded moments of all orders). Then, as $\overline{X}_N \sim \mathcal{N}(0, V_N/N)$ and from Lemma 1, under Assumption 1, $V_N = 2\pi f(0) + O(N^{-1}\log N)$, it follows that $(N/M)\,R_N$ has bounded moments of all orders too. ■

Proof of Lemma 7. We can write $u_2^\star = u_2 + (N/M)^{-1/2}\Delta_N'$, where the random variable Δ_N' has moments of all orders as Δ_N. Now $Y_N^\star = Y_N^L + [Z_N + \Delta_N'']\,(N/M)^{-1}$, where Δ_N'' depends on Δ_N, u_1 and u_2, and has moments of all orders, so it can be neglected when we approximate $Y_N^\star$ with Y_N^L. ■

The proofs of Lemmas 8 and 9 are postponed to Appendix B.

Proof of Lemma 11. Follows as for Lemmas 3 and 4 using Propositions 3 and 4. ■

Proof of Theorem 3. Follows as for Theorem 1. First, we approximate the joint characteristic function of $\mathbf{u}^\star = (u_1, u_2^\star)$. Define

$$A_N^\star(\mathbf{t},3) = \exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}\left[1 + B_N^\star(3,\mathbf{t})\left(\frac{M}{N}\right)^{1/2} + \left\{\overline{B}_N^\star(4,\mathbf{t}) + \frac{1}{2}\overline{B}_N^\star(3,\mathbf{t})^2\right\}\frac{M}{N}\right],$$

where we include in $B_N^\star$ the expansions for the corresponding cumulants up to the order $M^{-d^\star}$, but in $\overline{B}_N^\star$ only the leading terms are kept, so $B_N^\star(3,\mathbf{t}) = \frac{1}{3!}\sum_{j=0}^{d^\star} M^{-j}\nabla_j[0,3](it_2)^s$, $\overline{B}_N^\star(3,\mathbf{t}) = \frac{1}{3!}\nabla_0[0,3](it_2)^3$ and $\overline{B}_N^\star(4,\mathbf{t}) = \frac{1}{4!}\nabla_0[0,4](it_2)^4$. Now the Theorem follows as Theorem 1 using Lemmas 19, 20 and 21 instead of Lemmas 14, 15 and 16. ■

Proof of Theorem 4. We get, $\delta_N^\star = (1-b_N^\star)^{-1/2}$,

$$\delta_N^\star = 1 - \frac{1}{2}b_1M^{-d} - \frac{1}{2}b_2\frac{M}{N} + O\left(N^{-1}\log N + M^{-d-\varrho} + \left[\frac{M}{N}\right]^2\log^2 N\right),$$

and $\sigma_N^\star = \sigma_N + e_N = \sqrt{4\pi}\|K\|_2 + e_N$, where $e_N = O(M^{-2} + e_N(2))$. Therefore we can write $Y_N^\star = Y_N^{\star L} + Z_N(N/M)^{-3/2}$, where $Y_N^{\star L}$ is defined in (9) and $Z_N(N/M)^{-3/2}$ can be neglected in an approximation to the distribution of $Y_N^\star$ up to order M/N. Now we can use the same arguments as before to justify the Edgeworth approximation for $Y_N^\star$ in terms of that for $\mathbf{u}^\star$, since, under condition (8), $E[s_1]$, $E[s_1^3] = o(M/N)$, and neglecting terms $o(M/N)$,

$$E[s_1^2] = 1 - b_1 M^{-d} + \frac{M}{N}\left[-b_2 + 4\pi\|K\|_2^2\right], \quad E[s_1^4] = 3 - 6b_1 M^{-d} + \frac{M}{N}\left[-6b_2 + 36\pi\|K\|_2^2\right],$$

so the Theorem follows with the definition of $r_N(x)$. ■

10 Appendix B: Technical Lemmas

We first introduce the *Multiple Fejér Kernel* as in Bentkus (1972) or Dahlhaus (1983) for tapered series,

$$\Phi_N^{(n)}(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n-1}N}\frac{\sin Nx_1/2}{\sin x_1/2}\cdots\frac{\sin Nx_n/2}{\sin x_n/2} = \frac{1}{(2\pi)^{n-1}N}\sum_{t_1,\ldots,t_n=1}^{N}\exp\left\{i\textstyle\sum_{j=1}^n t_j x_j\right\},$$

with $x_n \equiv -\sum_{j=1}^{n-1} x_j$. For $n = 2$ this is Fejér's kernel. We have followed the same convention as in Keenan (1986, p. 137): although the functions $\Phi_N^{(n)}$ depend here on only $n-1$ arguments, we refer to n variables, with the restriction $\sum_1^n x_j \equiv 0(\mathrm{mod}2\pi)$. Then $\Phi_N^{(n)}(x_1, \ldots, x_n)$ is integrable in Π^{n-1}, integrates to one for all N and has the following properties:

- For $\delta > 0$, $N \geq 1$

$$\int_{D^c}\left|\Phi_N^{(n)}(x_1, \ldots, x_n)\right| dx_1 \cdots dx_{n-1} \;=\; O\left(\frac{\log^{n-1}N}{N\sin\delta/2}\right), \tag{26}$$

  where D^c is the complement in Π^{n-1} of the set $D = \{x \in \Pi^{n-1} : |x_j| \leq \delta,\ j = 1, \ldots, n-1\}$.

- For $j = 1, \ldots, n-1$,

$$\int_\Pi \cdots \int_\Pi |x_j|\left|\Phi_N^{(n)}(x_1, \ldots, x_n)\right| dx_1 \cdots dx_{n-1} \;=\; O\left(N^{-1}\log^{n-1}N\right). \tag{27}$$

- These properties follow due to

$$\left|\Phi_N^{(n)}(x_1, \ldots, x_n)\right| \leq \frac{1}{(2\pi)^{n-1}N}\left|\varphi_N(x_1)\right|\left|\varphi_N(x_2)\right|\cdots\left|\varphi_N(x_n)\right|, \tag{28}$$

  where $\varphi_N(x) = \sum_{t=1}^N \exp\{itx\}$ is the *Dirichlet Kernel*, which satisfies:

$$|\varphi_N(x)| \leq \min\left\{N, 2|x|^{-1}\right\}; \qquad \int_\Pi |\varphi_N(x)|\, dx = O(\log N). \tag{29}$$

Proof of Lemma 1. Applying the mean value theorem (MVT) for $f(\lambda)$ in an interval $[-\epsilon, \epsilon]$, $\epsilon > 0$, for some $|\theta| \leq 1$ depending on λ, since $V_N = 2\pi\int_\Pi f(\lambda)\Phi_N^{(2)}(\lambda)d\lambda$ and $\int_\Pi \Phi_N^{(2)}(\lambda)d\lambda = 1$,

$$\begin{aligned}
|V_N - 2\pi f(0)| &\leq 2\pi\left[\int_{|\lambda|\leq\epsilon} + \int_{|\lambda|>\epsilon}\right]\int_\Pi |f(\lambda) - f(0)|\left|\Phi_N^{(2)}(\lambda)\right| d\lambda \\
&= O\left(\int_{|\lambda|\leq\epsilon}|\lambda||f'(\lambda\theta)||\Phi_N^{(2)}(\lambda)|d\lambda + [\|f\|_1 + f(0)]N^{-1}\right),
\end{aligned}$$

which is $O\big(N^{-1}\log N\big)$ using the integrability of f (implied by stationarity), its differentiability around the origin and $|\Phi_N^{(2)}(\lambda)| = O(N^{-1}),$ if $|\lambda| \geq \epsilon > 0$, from (28) and (29). ■

Proof of Lemma 2. Writing the spectral estimate as $\widehat{f}(0) = \int_\Pi K_M(\lambda) I(\lambda) d\lambda$ where $I(\lambda)$ has expectation $E\left[I(\lambda)\right] = \int_\Pi \Phi_N^{(2)}(\lambda-\alpha) f(\alpha) d\alpha$ we obtain $E[\widehat{f}(0)] = \int_\Pi K_M(\lambda) \int_\Pi \Phi_N^{(2)}(\alpha) f(\lambda+\alpha) d\alpha d\lambda$. Then

$$\begin{aligned} E[\widehat{f}(0)] - f(0) - \frac{f^{(d)}(0)}{d!}\frac{\mu_d(K)}{M^d} &= \int_\Pi K_M(\lambda) \int_\Pi \Phi_N^{(2)}(\alpha) \left[f(\lambda+\alpha) - f(\lambda)\right] d\alpha d\lambda \\ &\quad + \int_\Pi K_M(\lambda) \left[f(\lambda) - f(0) - \frac{f^{(d)}(0)}{d!}\mu_d(K) M^{-d} \right] d\lambda \\ &= b_1 + b_2\,, \end{aligned}$$

say, where we have used the fact that K integrates to one. Introduce the sets $D = \{|\alpha|, |\lambda| \leq \epsilon/2\}$, and D^c, its complement in Π^2. Let b_{11} and b_{12} be the contributions to b_1 corresponding to D and D^c respectively. Then, for $|\theta| \leq 1$, depending on α, $b_{11} = \int_D K_M(\lambda) \Phi_N^{(2)}(\alpha) \left[f'(\lambda+\theta\alpha)\alpha\right] d\alpha d\lambda$ and

$$|b_{11}| \leq \sup_{|\lambda|\leq\epsilon} |f'(\lambda)| \int_{|\lambda|\leq\epsilon/2} |K_M(\lambda)| d\lambda \int_{|\alpha|\leq\epsilon/2} |\alpha| |\Phi_N^{(2)}(\alpha)| d\alpha = O\left(N^{-1}\log N\right).$$

To study b_{12} note first that $D^c \subset A_1 \cup A_2$ where $A_1 = \{|\alpha| > \epsilon/2\}$ and $A_2 = \{|\lambda| > \epsilon/2,\, |\alpha| \leq \epsilon/2\}$. Then the contribution to b_{12} from A_1 is

$$\begin{aligned} \left| \int_{|\alpha|>\epsilon/2} \int_\Pi K_M(\lambda) \left[f(\lambda+\alpha) - f(\lambda)\right] d\lambda \Phi_N^{(2)}(\alpha) d\alpha \right| &= O\left(N^{-1} \int_{\Pi^2} \left| K_M(\lambda) \left[f(\lambda+\alpha) - f(\lambda)\right] \right| d\lambda d\alpha \right) \\ &= O\left(N^{-1} \left[1 + \int_{|\lambda|\leq\epsilon} |K_M(\lambda) f(\lambda)|\, d\lambda \right] \right), \end{aligned} \tag{30}$$

which is $O\big(N^{-1}\int_\Pi |K_M(\lambda)|\, d\lambda\big) = O\big(N^{-1}\big)$, as the integral over $|\lambda| > \epsilon$ vanishes in (30) as $M \to \infty$. On the other hand, reasoning in a similar way, for M sufficiently large the contribution to b_{12} from A_2 is

$$\left| \int_{|\lambda|>\epsilon/2} \int_{|\alpha|\leq\epsilon/2} K_M(\lambda) \Phi_N^{(2)}(\alpha) \left[f(\lambda+\alpha) - f(\lambda)\right] d\alpha d\lambda \right| = 0, \tag{31}$$

because of the compact support of K. Thus $b_{12} = O\left(N^{-1}\right)$.

Now for b_2, splitting the integral in two parts for $|\lambda| \leq \epsilon$ and $|\lambda| > \epsilon$, denoted as b_{21} and b_{22} respectively, we have, constructing a Taylor expansion, (with $|\theta| \leq 1$, depending on λ),

$$\begin{aligned} b_{21} &= \int_{|\lambda|\leq\epsilon} K_M(\lambda) \left[\sum_{j=1}^{d-1} f^{(j)}(0) \frac{\lambda^j}{j!} + f^{(d)}(\theta\lambda) \frac{\lambda^d}{d!} - \frac{f^{(d)}(0)}{d!} \mu_d(K) M^{-d} \right] d\lambda \\ &= \sum_{j=1}^{d-1} f^{(j)}(0) \frac{1}{j!} \int_\Pi \lambda^j K_M(\lambda) d\lambda + \int_{|\lambda|\leq\pi/M} K_M(\lambda) \left[f^{(d)}(\theta\lambda) - f^{(d)}(0) \right] \lambda^d\, d\lambda \\ &= O\left(\int_\Pi |K_M(\lambda)| |\lambda|^{d+\varrho} d\lambda \right) = O(M^{-d-\varrho}), \end{aligned}$$

as all the integration is within $[-\epsilon, \epsilon]$ since $M \to \infty$ and using the Lipschitz property of $f^{(d)}$. As b_{22} is zero due to compact support of K, the Lemma is proved. ■

Proposition 1 *Under Assumptions 1, 3, 4, $e_N(2s) \to 0$, for $s \geq 2$,*

$$\mathit{Trace}\left[(\Sigma_N W_M)^s\right] = N(2\pi)^{2s-1} \sum_{j=0}^{d} L_j(s) M^{s-1-j} + O\big(N M^{s-1} e_N(2s)\big)\,,$$

where $e_N(s) = N^{-1}M\log^{2s-1}N$ *and* $L_j(s) = \frac{1}{j!}\mu_j(K^s)\ddot{f}^j(0)$ *with* $|L_j(s)| < \infty$ *and, as* $\mu_j(K^s)$*, the constants* $L_j(s)$ *only differ from zero for* j *even* ($j=0,\dots,d$).

Proof of Proposition 1. The proof is in two steps.

First step. We bound $A = \left|\text{Trace}\left[(\Sigma_N W_M)^s\right] - N(2\pi)^{2s-1}\int_\Pi f^s(\lambda)K_M^s(\lambda)d\lambda\right|$. First write, $r_{2s+1} \equiv r_1$,

$$\begin{aligned}\text{Trace}\left[(\Sigma_N W_M)^s\right] &= \sum_{1\le r_1,\dots,r_{2s}\le N}\prod_{j=1}^{s}\gamma(r_{2j-1}-r_{2j})\omega(\frac{r_{2j}-r_{2j+1}}{M}) \\ &= \sum_r\int_{\Pi^{2s}}\prod_{j=1}^{s}\{f(\lambda_{2j-1})K_M(\lambda_{2j})\}\exp\left\{i\textstyle\sum_{j=1}^{2s}\lambda_j(r_j-r_{j+1})\right\}d\lambda \\ &= N(2\pi)^{2s-1}\int_{\Pi^{2s}}G_M(\lambda,\mu)K_M(\lambda)\Phi_N^{(2s)}(\mu)d\lambda d\mu\end{aligned}$$

where $\Phi_N^{(2s)}(\mu) = \Phi_N^{(2s)}(\mu_1,\cdots,\mu_{2s})$, $G_M(\lambda,\mu) = f(\lambda-\mu_2-\cdots-\mu_{2s})K_M(\lambda-\mu_3-\cdots-\mu_{2s})\cdots f(\lambda-\mu_{2s})$, $d\mu = d\mu_2\cdots d\mu_{2s}$, $d\lambda = d\lambda_1\cdots\lambda_{2s}$ and we have made the change of variables

$$\begin{cases}\mu_1 = \lambda_1 - \lambda_{2s} \\ \mu_2 = \lambda_2 - \lambda_1 \\ \dots \\ \mu_{2s} = \lambda_{2s} - \lambda_{2s-1},\end{cases}\qquad\begin{cases}\lambda_{2s-1} = \lambda - \mu_{2s} \\ \lambda_{2s-2} = \lambda - \mu_{2s} - \mu_{2s-1} \\ \dots \\ \lambda_1 = \lambda - \mu_{2s} - \cdots - \mu_2 = \lambda - \mu_1,\end{cases}$$

($\sum_{j=1}^{2s}\mu_j = 0$), setting $\lambda = \lambda_{2s}$, and expressing all the λ_j in terms of λ and μ_j, $j = 2,\dots,2s$. Then

$$A \le N(2\pi)^{2s-1}\int_{\Pi^{2s}}\left|G_M(\lambda,\mu) - f^s(\lambda)K_M^{s-1}(\lambda)\right|\left|K_M(\lambda)\Phi_N^{(2s)}(\mu)\right|d\lambda d\mu. \tag{32}$$

We split the above integral into two sets, for small and for large μ_j. Define the set $D = \left\{\mu\in\Pi^{2s-1} : \sup_j|\mu_j| \le 1/(2sM)\right\}$. Taking into account that $|\lambda| \le \pi/M$ due to the compact support of K, in the set D all functions f are boundedly differentiable. Then we can use the inequality

$$|A_1\cdots A_r - B_1\cdots B_r| \le \sum_{q=0}^{r-1}|B_1\cdots B_q||B_{q+1} - A_{q+1}||A_{q+2}\cdots A_r| \tag{33}$$

and $\sup_\lambda|K_M(\lambda)| = O(M)$ to bound the integral of (32) over D by

$$O(NM^{s-1})\sum_{q=0}^{s-1}\int_\Pi\int_D|f(\lambda-\mu_{2+2q}\dots-\mu_{2s}) - f(\lambda)|\left|K_M(\lambda)\Phi_N^{(2s)}(\mu)\right|d\lambda d\mu \tag{34}$$

$$+O(NM^{s-1})\sum_{q=0}^{s-2}\int_\Pi\int_D|K_M(\lambda-\mu_{3+2q}\dots-\mu_{2s}) - K_M(\lambda)|\left|\Phi_N^{(2s)}(\mu)\right|d\lambda d\mu. \tag{35}$$

Then, applying the MVT and using (27) we obtain that (34) is

$$O(NM^{s-1})\int_\Pi|K_M(\lambda)|d\lambda\sum_{q=2}^{2s-1}\int_{\Pi^{2s-1}}|\mu_q|\left|\Phi_N^{(2s)}(\mu)\right|d\mu = O(M^{s-1}\log^{2s-1}N).$$

On the other hand, (35) is of order $O\big(M^s\log^{2s-1}N\big)$, using the Lipschitz property of K. Denote by D^c the complement of D in Π^{2s-1}. The contribution to A corresponding to the set D^c is bounded by

$$N(2\pi)^{2s-1}\int_\Pi\int_{D^c}|G_M(\lambda,\mu)K_M(\lambda)|\,|\Phi_N^{(2s)}(\mu)|d\lambda d\mu \tag{36}$$

$$+N(2\pi)^{2s-1}\int_\Pi|f^s(\lambda)K_M^s(\lambda)|d\lambda\int_{D^c}|\Phi_N^{(2s)}(\mu)|d\mu. \tag{37}$$

The expression in (37) is $O(M^s \log^{2s-1} N)$, by (26) and $\int |f^s(\lambda) K_M^s(\lambda)| d\lambda = O(M^{s-1})$, which follows from compact support of K. Now (36) is not larger than

$$\int_{D^\star} \prod_{j=1}^{s} |f(\lambda_{2j-1}) K_M(\lambda_{2j}) \varphi_N(\lambda_{2j} - \lambda_{2j-1}) \varphi_N(\lambda_{2j+1} - \lambda_{2j})| \, d\lambda_{2j} d\lambda_{2j-1}, \tag{38}$$

where $D^\star$ is the corresponding set to D^c with the former variables λ_j, $j = 1, \ldots, 2s$, defined by $D^\star = \{|\lambda_2 - \lambda_1| > \delta_N\} \cup \{|\lambda_3 - \lambda_2| > \delta_N\} \cup \ldots \cup \{|\lambda_{2s} - \lambda_{2s-1}| > \delta_N\}$, with $\delta_N = 1/(2sM)$, and a subindex $2s+1$ is to be interpreted as 1. Note that the last integral only differs from zero if $|\lambda_2|, |\lambda_4|, \ldots, |\lambda_{2s}| \leq \pi/M$. We consider only the case where just one of the events in $D^\star$ is satisfied, $|\lambda_{2j} - \lambda_{2j-1}| > \delta_N$ $(1 \leq j \leq s)$, say, the situation with an odd index or with more than one event being dealt with in a similar or simpler way.

First, if $|\lambda_{2j} - \lambda_{2j-1}| > \delta_N$, then $|\varphi_N(\lambda_{2j} - \lambda_{2j-1})| = O(M)$. Second, we can bound the integrals in λ_{2j} and λ_{2j-1}, with $\int_\Pi |\varphi_N(\lambda_{2j+1} - \lambda_{2j}) K_M(\lambda_{2j})| \, d\lambda_{2j} = O(M \log N)$, using (29), and

$$\int_\Pi |\varphi_N(\lambda_{2j-1} - \lambda_{2j-2}) f(\lambda_{2j-1})| \, d\lambda_{2j-1} = \int_{|\lambda_{2j-1}| \leq \epsilon} + \int_{|\lambda_{2j-1}| > \epsilon}. \tag{39}$$

If $|\lambda_{2j-1}| \leq \epsilon$ then $f(\lambda_{2j-1})$ is bounded and the corresponding integral is of order $O(\log N)$. If $|\lambda_{2j-1}| > \epsilon$, as $|\lambda_{2j-2}| < \pi/M$, we obtain that $|\lambda_{2j-1} - \lambda_{2j-2}| > \epsilon/2$, say, as $M \to \infty$, and then $|\varphi_N(\lambda_{2j-1} - \lambda_{2j-2})| = O(1)$. Thus the second integral is finite due to the integrability of f. Hence (39) is $O(\log N)$. There are $s - 1$ integrals of each type, which can be handled in the same way. Third, the remaining integral is of the general form

$$\int_\Pi \int_\Pi |K_M(\lambda_{2s}) f(\lambda_1) \varphi_N(\lambda_1 - \lambda_{2s})| \, d\lambda_1 d\lambda_{2s} = O(\log N),$$

since, as in (39), the integral in λ_1 is $O(\log N)$ for all λ_2, and $\int |K_M(\lambda_{2s})| d\lambda_{2s}$ is $O(1)$. Summarizing, the integral over $D^\star$ is $O(M^s \log^{2s-1} N)$, and compiling results we obtain that $A = O\big(M^{s-1} \log^{2s-1} N + M^s \log^{2s-1} N\big) = O\big(N M^{s-1} e_N(s)\big)$.

Second step. Defining $C_M(s) = \sum_{j=0}^{d} L_j(s) M^{s-1-j}$, we obtain, as $M \to \infty$,

$$\begin{aligned} \left| \int_\Pi K_M^s(\lambda) f^s(\lambda) d\lambda - C_M(s) \right| &\leq \int_\Pi |K_M(\lambda)|^{s-1} \left| f^s(\lambda) - \sum_{j=0}^{d} \frac{1}{j!} \left(\frac{d}{d\lambda} \right)^j f^s(0) \lambda^j \right| |K_M(\lambda)| d\lambda \\ &= O\left(\sup_\lambda |K_M(\lambda)|^{s-1} \int_\Pi |\lambda|^{d+\varrho} |K_M(\lambda)| d\lambda \right) = O(M^{s-1-d-\varrho}), \end{aligned}$$

using the Lipschitz property of $f^{(d)}(\lambda)$ in the same way as in the proof of Lemma 2. ■

Proposition 2 *Under Assumptions 1, 3, 4, $e_N(2s+2) \to 0$, for $s \geq 1$, $\mathbf{1}'(\Sigma_N W_M)^s \Sigma_N \mathbf{1} = N(2\pi)^{2s+1} \times [f(0)]^{s+1} [K_M(0)]^s + O\left(M^{s+1} \log^{2s+1} N\right)$.*

Proof of Proposition 2. We can write $\mathbf{1}'(\Sigma_N W_M)^s \Sigma_N \mathbf{1}$ as

$$\begin{aligned} &\sum_{0 \leq r_1, \ldots, r_{2s+2} \leq N} \gamma(r_{2s+1} - r_{2s+2}) \prod_{j=1}^{s} \left\{ \gamma(r_{2j-1} - r_{2j}) \omega\left(\frac{r_{2j} - r_{2j+1}}{M}\right) \right\} \\ = &\sum_r \int_{\Pi^{2s+1}} f(\lambda_{2s+1}) \prod_{j=1}^{s} \{f(\lambda_{2j-1}) K_M(\lambda_{2j})\} \exp\left\{ i \textstyle\sum_{j=1}^{2s+1} \lambda_j (r_j - r_{j+1}) \right\} d\lambda \\ = &(2\pi)^{2s+1} N \int_{\Pi^{2s+1}} S_M(\mu) \Phi_N^{(2s+2)}(\mu) d\mu, \end{aligned} \tag{40}$$

by change of variable, where $\Phi_N^{(2s+2)}(\mu) = \Phi_N^{(2s+2)}(\mu_1,\cdots,\mu_{2s+1},-\sum_{j=1}^{2s+1}\mu_j)$, $S_M(\mu) = f(\mu_1)K_M(\mu_1+\mu_2)\cdots K_M(\mu_1+\cdots+\mu_{2s})f(\mu_1+\cdots+\mu_{2s+1})$ and $d\mu = d\mu_1\cdots d\mu_{2s+1}$, $d\lambda = d\lambda_1\cdots d\lambda_{2s+1}$. To study the difference between the integral in (40) and $f^{s+1}(0)K_M^s(0)$ we divide the range of integration, Π^{2s+1}, into two sets, D and its complement D^c, where D is now defined by the condition $D = \{|\mu_j| \leq \pi/[M(2s+2)],\ j = 1,\ldots,2s+1\}$. In this case we only need the smoothness properties of K at the origin (inside D). For the difference in the set D, we can use inequality (33), the Lipschitz property of K and the differentiability of f:

$$\begin{aligned} &\left|\int_D S_M(\mu)\Phi_N^{(2s+2)}(\mu)d\mu - \int_D f^{s+1}(0)K_M^s(0)\Phi_N^{(2s+2)}(\mu)d\mu\right| \\ = \;& O(M^{s+1})\int_{\Pi^{2s+1}}\sum_{j=2}^{2s}|\mu_j|\left|\Phi_N^{(2s+2)}(\mu)\right|d\mu = O\big(M^{s+1}N^{-1}\log^{2s+1}N\big)\,, \end{aligned} \tag{41}$$

using (27). Focusing on the integral over the set D^c of (40) and using (26), this is bounded by

$$\int_{D^c}|S_M(\mu)|\left|\Phi_N^{(2s+2)}(\mu)\right|d\mu + O\big(M^{s+1}N^{-1}\log^{2s+1}N\big)\,. \tag{42}$$

As in the proof of the previous Proposition, the integral in (42) is less or equal than

$$\frac{N^{-1}}{(2\pi)^{2s+1}}\left|\int_{D^\star}\prod_{j=1}^{s}f(\lambda_{2j-1})K_M(\lambda_{2j})\varphi_N(\lambda_{2j}-\lambda_{2j-1})\varphi_N(\lambda_{2j+1}-\lambda_{2j})f(\lambda_{2s+1})\varphi_N(\lambda_1)\varphi_N(-\lambda_{2s+1})d\lambda\right| \tag{43}$$

where $D^\star = \{|\lambda_1| > \pi/[M(2s+2)]\}\cup\{|\lambda_2-\lambda_1| > \pi/[M(2s+2)]\}\cup\ldots\cup\{|\lambda_{2s-1}+\lambda_{2s}| > \pi/[M(2s+2)]\}$. Also, the integral in (43) is nonzero only if $|\lambda_2|,|\lambda_4|,\ldots,|\lambda_{2s}| \leq \pi/M$.

If $|\lambda_{j+1}-\lambda_j| > \pi/[M(2s+2)]$ for at least one index $j \in \{1,\ldots,2s\}$ we can repeat the procedure of Proposition 1 to obtain a bound of order $O(N^{-1}M^{s+1}\log^{2s+1}N)$ for this contribution in (43).

We now study the case in which $|\lambda_1| > \pi/[M(2s+2)]$. First, $|\varphi_N(\lambda_1)| = O(M)$. Truncating the integral at $|\lambda_1| = \epsilon$, $\int_\Pi f(\lambda_1)|\varphi_N(\lambda_2-\lambda_1)|d\lambda_1 = O(\log N)$, as $|\lambda_2-\lambda_1| > \epsilon/2$ if $|\lambda_1| > \epsilon$ and $|\lambda_2| \leq \epsilon/[M(2s+2)]$, since $M \to \infty$. Now $\int_\Pi |K_M(\lambda_2)\varphi_N(\lambda_3-\lambda_2)|\,d\lambda_2 = O(M\log N)$, and the integrals with respect to the remaining variables can be bounded in the same way, (43) being of order $O(N^{-1}M^{s+1}\log^{2s+1}N)$ again.

Therefore, from (41), (42) and the previous discussion for (43), the Proposition follows. ■

Lemma 13 (Bhattacharya and Rao, 1975, pp. 97-98, 113) *Let P and Γ be probability measures on $\mathbb{R}^2$ and $\mathcal{B}^2$ the class of all Borel subsets of $\mathbb{R}^2$. Let α be a positive number. Then there exists a kernel probability measure Ψ_α such that $\sup_{B\in\mathcal{B}^2}|P(B)-\Gamma(B)| \leq \frac{2}{3}\|(P-\Gamma)\star\Psi_\alpha\| + \frac{4}{3}\sup_{B\in\mathcal{B}^2}\Gamma\{(\partial B)^{2\alpha}\}$, where Ψ_α satisfies*

$$\Psi_\alpha(B(0,r)^c) = O\left(\left(\frac{\alpha}{r}\right)^3\right) \tag{44}$$

and its Fourier transform $\widehat{\Psi}_\alpha$ satisfies

$$\widehat{\Psi}_\alpha = 0 \;\; \textit{for } \|\mathbf{t}\| \geq 8\cdot 2^{4/3}/\pi^{1/3}\alpha. \tag{45}$$

$(\partial B)^{2\alpha}$ is a neighbourhood of radius 2α of the boundary of B, $\|\ \|$ is the variation norm of a measure in this case, and $\star$ means convolution. ■

Lemma 14 *Under Assumptions 1, 3, 4, $M^{-1} + N^{-1}M\log^5 N \to 0$, there exists $\delta_1 > 0$ such that, for $\|\mathbf{t}\| \le \delta_1\sqrt{N/M}$ and a number $d_1 > 0$:*

$$|\psi_N(\mathbf{t}) - A_N(\mathbf{t},2)| \le \exp\{-d_1\|\mathbf{t}\|^2\}F(\|\mathbf{t}\|)O\left(\left(\frac{N}{M}\right)^{-1/2}\left[M^{-2} + e_N(3)\right] + \frac{M}{N}\right),$$

where F is a polynomial in $\mathbf{t}$ with bounded coefficients and $A_N(\mathbf{t},2)$ is defined as in (16).

Proof of Lemma 14. Similarly to Feller (1971, p. 535) we have for complex α and β that $|e^\alpha - 1 - \beta| \le e^\gamma\left\{|\alpha-\beta| + \frac{|\beta|^2}{2}\right\}$, where $\gamma = \max\{|\alpha|,|\beta|\}$. We take (with $\tau = 2$ in (15)):

$$\alpha = \log\varphi(\mathbf{t}) - \frac{1}{2}\|i\mathbf{t}\|^2 = \left(\frac{M}{N}\right)^{1/2}\sum_{|\mathbf{r}|=3}\frac{1}{r_1!r_2!}\bar\kappa_N[r_1,r_2](it_1)^{r_1}(it_2)^{r_2} + R_N(2)$$

and $\beta = \left(MN^{-1}\right)^{1/2}\bar B_N(3,\mathbf{t})$. Then we have, using Lemmas 3 and 4 for $s = 3$,

$$\begin{aligned}|\alpha-\beta| &\le \left|\left(\frac{N}{M}\right)^{-1/2}O\big(M^{-2}+e_N(3)\big)\big[(it_2)^3+(it_1)^2(it_2)\big]+\frac{M}{N}\big[R_{04}(it_2)^4+R_{22}(it_1)^2(it_2)^2\big]\right| \\ &\le F_1(\|\mathbf{t}\|)O\left(\left(\frac{N}{M}\right)^{-1/2}\big[M^{-2}+e_N(3)\big]+\frac{M}{N}\right),\end{aligned}$$

where F_1 is a polynomial of degree 4. Now $\frac{1}{2}|\beta|^2 \le F_2(\|\mathbf{t}\|)O\big(\frac{M}{N}\big)$, where F_2 is a polynomial of degree 6. Then

$$|\alpha-\beta| + \frac{|\beta|^2}{2} \le F(\|\mathbf{t}\|)O\left(\left(\frac{N}{M}\right)^{-1/2}\big[M^{-2}+e_N(3)\big]+\frac{M}{N}\right) \tag{46}$$

for some polynomial F. Now to study γ, we first bound $|\beta|$ for $\|\mathbf{t}\| \le \delta_\beta\sqrt{N/M}$, $\delta_\beta > 0$:

$$\begin{aligned}|\beta| &\le \|\mathbf{t}\|^2\left\{\frac{1}{3!}\left(\frac{N}{M}\right)^{-1/2}\left[|\nabla_0[0,3]| + 3|\nabla_0[2,1]|\right]\|\mathbf{t}\|\right\} \\ &\le \|\mathbf{t}\|^2\left\{\frac{\delta_\beta}{3!}\left[|\nabla_0[0,3]| + 3|\nabla_0[2,1]|\right]\right\} \le \|\mathbf{t}\|^2 T_\beta,\end{aligned} \tag{47}$$

with $0 < T_\beta < 1/4$ on choosing δ_β sufficiently small. Now for α we can choose a $\delta_\alpha > 0$ so small that, for $\|\mathbf{t}\| \le \delta_\alpha\sqrt{N/M}$,

$$\begin{aligned}|\alpha| &\le \|\mathbf{t}\|^2\left\{\frac{1}{3!}\left(\frac{N}{M}\right)^{-1/2}\big[|\nabla_0[0,3]|+3|\nabla_1[2,1]| + O(M^{-2}+e_N(3)\big]\|\mathbf{t}\| + \frac{M}{N}\left[|R_{04}| + |R_{22}|\right]\|\mathbf{t}\|^2\right\} \\ &\le \|\mathbf{t}\|^2\left\{\frac{\delta_\alpha}{3!}\big[|\nabla_0[0,3]|+3|\nabla_0[2,1]|+O\big(M^{-2}+e_N(3)\big)\big]+\delta_\alpha^2[|R_{04}|+|R_{22}|]\right\} \\ &\le \|\mathbf{t}\|^2\left\{\frac{1}{4}+O\big(M^{-2}+e_N(3)\big)\right\}.\end{aligned} \tag{48}$$

From (47) and (48) we have that $e^\gamma \le \exp\left\{\|\mathbf{t}\|^2\left[\frac{1}{4}+O\big(M^{-2}+e_N(3)\big)\right]\right\}$ for $\|\mathbf{t}\| \le \delta_1\sqrt{N/M}$ where $\delta_1 = \min\{\delta_\alpha,\delta_\beta\}$. Then,

$$\exp\left\{-\frac{1}{2}\|\mathbf{t}\|^2+\gamma\right\} \le \exp\left\{\|\mathbf{t}\|^2\left[-\frac{1}{4}+O\big(M^{-2}+e_N(3)\big)\right]\right\} \le \exp\left\{-d_1\|\mathbf{t}\|^2\right\} \tag{49}$$

for one $d_1 > 0$, $\|\mathbf{t}\| \le \delta_1\sqrt{N/M}$. Since our approximation to $\varphi(\mathbf{t}) = \exp\{\frac{1}{2}\|i\mathbf{t}\|^2 + \alpha\}$ is $A_N(\mathbf{t},2) = \exp\left\{\frac{1}{2}\|i\mathbf{t}\|^2\right\}[1+\beta]$, using (46) and (49) the Lemma is proved. ■

Lemma 15 *Under Assumptions 1, 2 (some $p > 1$), 3, 4, $M^{-1} + N^{-1}M \log^3 N \to 0$ as $N \to \infty$ there exists $d_2 > 0$ such that for $\|\mathbf{t}\| > \delta_1 m_N$, $|\psi(t_1, t_2)| \leq \exp\left\{-d_2 m_N^2\right\}$, where $m_N \stackrel{def}{=} \min\left\{\left(MN^{-1}\right)^{1/2} \log N, N^{(p-1)/p}\right\} \to \infty$ as $N \to \infty$.*

Proof of Lemma 15. First, following Bentkus and Rudzkis (1982) we study the characteristic function of the spectral density estimate, which itself appears in the joint characteristic function. Define $\tau(t_2) = E\left[\exp\left\{it_2 u_2\right\}\right] = \tau'(t_2)\exp\left\{-it_2 E\right\}$, where

$$\tau'(t_2) = \left|I - \frac{2it_2}{\sqrt{NM}\sigma_N V_N}\Sigma_N W_M\right|^{-1/2} = \prod_{j=1}^{N}\left(1 - 2it_2\frac{\mu_j}{\sqrt{NM}\sigma_N V_N}\right)^{-1/2}$$

and μ_j are now the eigenvalues of the matrix $\Sigma_N W_M$. Obviously $|\tau(t_2)| = |\tau'(t_2)|$. Now as

$$1 = \mathrm{Var}[u_2] = \frac{1}{MN}\frac{1}{\sigma_N^2 V_N^2} 2\,\mathrm{Trace}[(\Sigma_N W_M)^2] = \frac{1}{MN}\frac{2}{\sigma_N^2 V_N^2}\sum_{j=1}^{N}\mu_j^2,$$

we obtain $\sum_{j=1}^{N}\mu_j^2 = \frac{1}{2}\sigma_N^2 V_N^2 MN = O(MN)$. Also we have that $\max_j |\mu_j| = \sup_{\|z\|=1}|(\Sigma_N W_M z, z)| = \|\Sigma_N W_M\|$. From Lemma 16, for a finite positive constant c_1 depending on f and K

$$\max_j |\mu_j| \leq c_1 \vartheta_N, \qquad \vartheta_N = \max\left\{M\log N, N^{\frac{2-p}{2p}} M^{1/2}\right\} \to \infty, \quad as \;\; N \to \infty.$$

Introduce now the notation $g_j = \mu_j\left[c_1\vartheta_N\right]^{-1}$ where $|g_j| \leq 1$. We have $\sum_{j=1}^{N} g_j^2 = \sigma_N^2 V_N^2 MN(2c_1^2\vartheta_N^2)^{-1}$, and (noting that $NM/\vartheta_N^2 \to \infty$, for all $p > 1$)

$$\begin{aligned}
|\tau(t_2)| &= \prod_{j=1}^{N}\left(1 + 4t_2^2\frac{c_1^2 g_j^2 \vartheta_N^2}{MN\sigma_N^2 V_N^2}\right)^{-1/4} \leq \prod_{j=1}^{N}\left(1 + t_2^2\frac{4c_1^2}{MN}\frac{\vartheta_N^2}{\sigma_N^2 V_N^2}\right)^{-\frac{1}{4}g_j^2} \\
&= \left(1 + t_2^2\frac{\vartheta_N^2}{NM}\frac{4c_1^2}{\sigma_N^2 V_N^2}\right)^{-\frac{1}{8}c_1^{-2}\sigma_N^2 V_N^2 NM\vartheta_N^{-2}} \\
&= \left(1 + t_2^2\frac{\vartheta_N^2}{NM}\left[c_2 + O(M^{-2} + e_N(2))\right]\right)^{-\frac{1}{2}\left[c_2^{-1} + O(M^{-2} + e_N(2))\right]NM\vartheta_N^{-2}},
\end{aligned}$$

where $c_2 = c_1^2/(\pi^2 4\pi f^2(0)\|K\|_2^2)$ is a constant from the expansion of $\sigma_N^2 V_N^2$ in powers of M^{-1}, and we have applied $(1 + at) \geq (1 + t)^a$, valid for $t \geq 0$, $0 \leq a \leq 1$. So for all $\eta > 0$, as $N, M \to \infty$ we have that

$$|\tau(t_2)| \leq \left(1 + \eta_1^2\right)^{-\eta_2\frac{NM}{\vartheta_N^2}} \tag{50}$$

for $|t_2| > \eta\sqrt{NM}/\vartheta_N$ and for $\eta_1 > 0$ and $\eta_2 > 0$ depending on η.

Then returning to the bivariate characteristic function, its modulus is equal to

$$|\varphi_N(t_1, t_2)| = |\tau(t_2)|\exp\left\{-\frac{1}{2}t_1^2\xi_N'\Re(I - 2it_2\Sigma_N Q_N)^{-1}\Sigma_N\xi_N\right\}, \tag{51}$$

where $\Re$ stands for real part. From Anderson (1958, p. 161) $\Re(\Sigma_N^{-1} - 2it_2 Q_N)^{-1} = \Re(I - 2it_2\Sigma_N Q_N)^{-1}\Sigma_N$ is positive definite as $t_2 Q_N$ is real (for every N). Then $\xi_N'\Re(I - 2it_2\Sigma_N Q_N)^{-1}\Sigma_N\xi_N > 0$ for all $t_2 \in \mathbb{R}$. Thus for $|t_2| \leq \delta\sqrt{NM}/\vartheta_N$, for all $\delta > 0$, $\xi_N'\Re(I - 2it_2\Sigma_N Q_N)^{-1}\Sigma_N\xi_N > \epsilon$ for some $\epsilon > 0$ fixed depending on δ, since we have that $\|\Sigma_N Q_N\| = O\big((MN)^{-1/2}\|\Sigma_N W_M\|\big) = O\big((MN)^{-1/2}\vartheta_N\big)$, and $\|\xi_N\| = 1/V_N$, with $V_N \to 2\pi f(0)$, $0 < f(0) < \infty$, as $N \to \infty$. Then,

$$\exp\left\{-\frac{1}{2}t_1^2\xi_N'\Re(I - 2it_2\Sigma_N Q_N)^{-1}\Sigma_N\xi_N\right\} \leq \exp\left\{-\frac{1}{2}t_1^2\epsilon_1\right\} \leq \exp\left\{-\frac{1}{4}\epsilon_1\delta_1^2\frac{NM}{\vartheta_N^2}\right\} \tag{52}$$

for $|t_1|\sqrt{2} > \delta_1\sqrt{NM}/\vartheta_N$ and $|t_2|\sqrt{2} \le \delta_1\sqrt{NM}/\vartheta_N$, and some $\epsilon_1 > 0$ depending on δ_1.

Thus from (50) and (52), there exists a $d_2 > 0$ such that $|\varphi(t_1,t_2)| \le \exp\left\{-d_2\frac{NM}{\vartheta_N^2}\right\}$ inside $\left\{\mathbf{t} : \|\mathbf{t}\| > \delta_1\sqrt{NM}/\vartheta_N\right\} \subset B_1 \cup B_2$ where $B_1 = \left\{\mathbf{t} \in \mathbb{R}^2 : |t_2| > \frac{\delta_1}{\sqrt{2}}\sqrt{NM}/\vartheta_N\right\}$ and $B_2 = \left\{\mathbf{t} \in \mathbb{R}^2 : |t_2| \le \frac{\delta_1}{\sqrt{2}}\sqrt{NM}/\vartheta_N \text{ and } |t_1| > \frac{\delta_1}{\sqrt{2}}\sqrt{NM}/\vartheta_N\right\}$ and the Lemma follows because

$$\frac{NM}{\vartheta_N^2} = MN\min\left\{\frac{1}{M^2\log^2 N}, N^{\frac{p-2}{p}}M^{-1}\right\} = m_N^2 \to \infty,$$

as $N \to \infty$. Note that $p > 2$ in 2 provides no further improvement in any bound, since the best rate in Lemma 16 below is already attained when f is in L_2. ■

Lemma 16 *Under the assumptions of Theorem 1, $\|\Sigma_N W_M\| \le c_1\vartheta_N$, where $0 < c_1 < \infty$ is a constant depending on f and K and $\vartheta_N \stackrel{def}{=} \max\left\{M\log N, N^{(2-p)/2p}M^{1/2}\right\} \to \infty$ as $N \to \infty$.*

Proof of Lemma 16. Write

$$\|\Sigma_N W_M\| = \sup_{\|z\|=1}\left|\sum_{j,h} z_j z_h \int_{\Pi^2} f(\lambda)K_M(\omega)\varphi_N(\lambda-\omega)e^{i(h\omega - j\lambda)}d\lambda d\omega\right| = \sup_{\|z\|=1}\left|\int_{\Pi^2} F_N(\lambda,\omega)d\lambda d\omega\right|, \quad (53)$$

say, where $F_N(\lambda,\omega) = f(\lambda)K_M(\omega)\varphi_N(\lambda-\omega)Z_N(-\lambda)Z_N(\omega)$ and $Z_N(\lambda) = \sum_{j=1}^N z_j e^{ij\lambda}$ for any vector z with $\|z\| = 1$. In the integral in (53) we need consider only the interval $w \in [-\pi/M, \pi/M]$, with $\pi/M \le \epsilon$ by $M \to \infty$. Denote the supremun of $f(\lambda)$ when $\lambda \in [-\epsilon,\epsilon]$ as $\|f_\epsilon\|_\infty$. Then the contribution from $|\lambda| \le \epsilon$ to (53) is bounded by

$$\begin{aligned}
& \sup_{\|z\|=1} M\|K\|_\infty\|f_\epsilon\|_\infty \int_\Pi\int_\Pi |\varphi_N(\lambda-\omega)Z_N(-\lambda)Z_N(\omega)|\, d\lambda d\omega \\
\le\ & \sup_{\|z\|=1} M\|K\|_\infty\|f_\epsilon\|_\infty \int_\Pi |\varphi_N(\alpha)|\left[\int_\Pi |Z_N(-\alpha-\omega)|^2\, d\omega \int_\Pi |Z_N(\omega)|^2\, d\omega\right]^{1/2} d\alpha \\
\le\ & 2\pi M\|K\|_\infty\|f_\epsilon\|_\infty \int_\Pi |\varphi_N(\alpha)|\, d\alpha \le c(f,K)M\log N, \qquad (54)
\end{aligned}$$

where $c(f,K)$ is a constant depending on f and K, and we have made the change of variable $\alpha = \lambda - \omega$ and used the fact that $\int_\Pi |Z_N(\omega)|^2 d\omega = 2\pi$. For other λ, we see that $|\lambda| > \epsilon$ and $|\omega| \le \pi/M$ imply $|\lambda - \omega| > \epsilon/2$, say, as $M \to \infty$, so $|\varphi_N(\lambda-\omega)| \le$ const. Then, for $1 < p \le 2$ and using $\sup_{z,\lambda}|Z_N(\lambda)| \le \sqrt{N}$ and Hölder inequality for $1 < p \le 2$, the contribution from $|\lambda| > \epsilon$ to (53) is bounded by

$$\begin{aligned}
& \text{const} \sup_{\|z\|=1} \int_\Pi\int_\Pi f(\lambda)\,|K_M(\omega)Z_N(\omega)Z_N(-\lambda)|\, d\lambda d\omega \\
\le\ & \text{const} \sup_{\|z\|=1} \left[\int_\Pi |K_M(\omega)|^2\, d\omega \int_\Pi |Z_N(\omega)|^2\, d\omega\right]^{1/2}\left[\int_\Pi f^p(\lambda)d\lambda\right]^{1/p}\left[\int_\Pi |Z_N(\lambda)|^{\frac{p}{p-1}}\, d\lambda\right]^{\frac{p-1}{p}} \\
\le\ & \text{const}\|K\|_\infty^{1/2}\|K\|_1\|f\|_p N^{\frac{2-p}{2p}}M^{1/2} = c'(f,K)N^{\frac{2-p}{2p}}M^{1/2}, \qquad (55)
\end{aligned}$$

using $\sup_{z,\lambda}|Z_N(\lambda)| \le \sqrt{N}$ and $\int_\Pi |Z_N(\lambda)|^2 d\lambda = 2\pi$. Then the Lemma follows from (54) and (55). ■

Lemma 17 *Under Assumptions 1, 3, 4, $M^{-1} + N^{-1}M\log^2 N \to 0$, $s = 1, 2, \ldots$*

$$\mathit{Trace}[(\Sigma_N W_M \mathbf{1}\,\mathbf{1}')^s] = (MN)^s\left[(2\pi)^2 f(0)K(0)\right]^s + O((NM)^{s-1}M^2\log^2 N).$$

Proof of Lemma 17. First we observe that $\text{Trace}[(\Sigma_N W_M \mathbf{1}\,\mathbf{1}')^s] = (\mathbf{1}'\Sigma_N W_M \mathbf{1})^s$ and

$$\mathbf{1}'\Sigma_N W_M \mathbf{1} = (2\pi)^2 N \int_{\Pi^2} f(\mu_1) K_M(\mu_1+\mu_2)\Phi_N^{(3)}(\mu_1,\mu_2)d\mu_1 d\mu_2. \tag{56}$$

Introduce the set $D = \{|\mu_j| \leq \pi/[2M],\ j=1,2\}$. Then, using Assumptions 1 and 4, for $d=1$,

$$\begin{aligned}
&\left|(2\pi)^2 N\int_D f(\mu_1)K_M(\mu_1+\mu_2)\Phi_N^{(3)}(\mu_1,\mu_2)d\mu_1 d\mu_2 - N(2\pi)^2 f(0)K_M(0)\int_D \Phi_N^{(3)}(\mu_1,\mu_2)d\mu_1 d\mu_2\right| \\
&= O(N)\int_D |f(\mu_1)K_M(\mu_1+\mu_2) - f(0)K_M(0)|\,|\Phi_N^{(3)}(\mu_1,\mu_2)|d\mu_1 d\mu_2 \\
&= O(NM)\sum_{j=1,2}\int_\Pi \left|\mu_j \Phi_N^{(3)}(\mu_1,\mu_2)\right| d\mu_1 d\mu_2 + O\big(NM^2\big)\sum_{j=1,2}\int_\Pi \left|\mu_j\Phi_N^{(3)}(\mu_1,\mu_2)\right| d\mu_1 d\mu_2,
\end{aligned}$$

which is $O(M^2\log^2 N)$. The contribution to (56) of the integral for the complement to the set D can be seen to be of order of magnitude $O(M^2\log^2 N)$, proceeding in the same way as in the proof of Proposition 1. ■

Lemma 18 *Under Assumptions 3, 4,* $M^{-1}+N^{-1}M\log N \to 0$, $(2\pi N)^{-1}\mathbf{1}'W_M\mathbf{1} = M\,K(0)+O\big(M^2N^{-1}\log N\big)$.

Proof of Lemma 18. It follows writing $(2\pi N)^{-1}\mathbf{1}'W_M\mathbf{1} = \int_\Pi K_M(\lambda)\Phi_N^{(2)}(\lambda)d\lambda$ and using the Lipschitz property of K and the properties of the Fejér's kernel. ■

Proof of Lemma 8. Following the proof of Lemma 2, we can write the bias $E[\widehat{f}^{(\nu)}(0)] - f^{(\nu)}(0)$ as

$$(m_\nu)^\nu \int_\Pi V_{m_\nu}(\lambda)\int_\Pi \Phi_N^{(2)}(\theta)\,[f(\lambda-\theta)-f(\lambda)]d\theta d\lambda + (m_\nu)^\nu\int_\Pi V_{m_\nu}(\lambda)\left[f(\lambda) - \frac{\lambda^\nu}{(-1)^\nu \nu!}f^{(\nu)}(0)\right]d\lambda.$$

Then employing the same methods of Lemma 2 with the properties of the kernel V_ν, this is $O\big((m_\nu)^\nu N^{-1}\log N + (m_\nu)^{-a}\big)$, and the Lemma follows. ■

Proof of Lemma 9. Likewise for the discussion of the cumulants of the spectral estimate contained in Proposition 1 we can write

$$\frac{N}{(m_\nu)^{2\nu+1}}\text{Var}[\widehat{f}^{(\nu)}(0)] \;=\; \frac{4\pi}{m_\nu}\int_{\Pi^4} f(\lambda - \textstyle\sum_{j=2}^4\mu_j)V_{m_\nu}(\lambda-\mu_3-\mu_4)f(\lambda-\mu_4)V_{m_\nu}(\lambda)\Phi_N^{(4)}(\mu)d\mu d\lambda. \tag{57}$$

As in Proposition 1 we have to take care of possible unboundedness of f away from the origin. We thus consider the set of integration $D = \{\mu\in[-\pi,\pi]^3 \,:\, |\mu_j| \leq 1/(4M),\ j=2,\ldots,4\}$. Then, the integral in (57) over the set D is

$$\frac{4\pi}{m_\nu}\int_\Pi f^2(\lambda)V_{m_\nu}^2(\lambda)d\lambda + O\big(N^{-1}m_\nu\log^3 N\big) = \frac{4\pi}{m_\nu}f^2(0)\int_\Pi V_{m_\nu}^2(\lambda)d\lambda + O\big(N^{-1}m_\nu\log^3 N + m_\nu^{-1}\big),$$

which is $4\pi f^2(0)\|V_\nu\|_2^2 + o(1)$, using evenness of f and its differentiability around $f(0)$. The integral in (57) over the complement to the set D can be seen to be $O(N^{-1}m_\nu\log^3 N)$, using the finite support of V_ν and the properties of $\Phi_N^{(4)}$, as in the proof of Proposition 1. ■

Proposition 3 *Under the assumptions of Proposition 1,* $\mathit{Trace}\,[(\Sigma_N W_M^\star)^s] = \mathit{Trace}\,[(\Sigma_N W_M)^s] + O(M^s)$, $W_M^\star = A_N W_M A_N$.

Proof of Proposition 3. Follows the proof of Proposition 1. The Fourier transform corresponding to the matrix A_N is

$$A_N(\lambda) = \frac{1}{2\pi}\left(1 - \frac{D_N(\lambda)}{N}\right), \qquad D_N(\lambda) = \sum_{j=1-N}^{N-1} e^{ij\lambda},$$

where $D_N(\lambda)$ is a version of Dirichlet kernel. Denote $a_N(j) = \delta(j=0) - N^{-1}$. We first rewrite Trace $[(\Sigma_N W_M^\star)^s]$ as, $r_{2s+1} \equiv r_1$,

$$\begin{aligned}
&\sum_{1\le r_1,r_1',\dots,r_{2s}\le N}\prod_{j=1}^{s}\gamma(r_{2j-1}-r_{2j-1}')a_N(r_{2j-1}'-r_{2j})\omega(\frac{r_{2j}-r_{2j}'}{M})a_N(r_{2j}'-r_{2j+1})\\
&= N\int_{\Pi^{4s}} f(\lambda-\mu_1'-\mu_2\cdots-\mu_{2s}')A_N(\lambda-\mu_2\cdots-\mu_{2s}')K_M(\lambda-\mu_2'\cdots-\mu_{2s}')\cdots f(\lambda-\mu_{2s-1}'\cdots-\mu_{2s}')\\
&\quad\times A_N(\lambda-\mu_{2s}-\mu_{2s}')K_M(\lambda-\mu_{2s}')A_N(\lambda)(2\pi)^{4s-1}\Phi_N^{(4s)}(\mu_1,\mu_1',\cdots,\mu_{2s})d\lambda d\mu,
\end{aligned}$$

using a change of variable as in the proof of Proposition 1, and $d\mu = d\mu_1' d\mu_2 d\mu_2'\cdots d\mu_{2s}$.

Now we deal with the cross products implicit in the functions $a_N(j)$ or $A_N(\lambda)$. The product containing no $D_N(\lambda)$ equals the integral in the case without mean-correction (cf. Proposition 1). Then all the remaining terms have 1, 2, ..., $2s$ functions $D_N(\lambda)$. We consider just one, and bound its contribution to the trace. From the proof it should be evident that similar bounds hold for the other terms. The typical term is

$$\begin{aligned}
&-\sum_{1\le r_1,r_1',\dots,r_{2s}\le N}\gamma(r_1-r_1')\frac{1}{N}\omega(\frac{r_2-r_3}{M})\cdots\gamma(r_{2s-1}-r_{2s})\omega(\frac{r_{2s}-r_1}{M})\\
&\quad= -\frac{1}{N}\sum_{1\le r_2,\dots,r_{2s},r_1,r_1'\le N}\omega(\frac{r_2-r_3}{M})\cdots\gamma(r_{2s-1}-r_{2s})\omega(\frac{r_{2s}-r_1}{M})\gamma(r_1-r_1')
\end{aligned}$$

which is $O(N^{-1}\,NM^s) = O(M^s)$, from Proposition 2, and there is no additional term of higher magnitude. ■

Proposition 4 *Under the assumptions of Proposition 2,* $\mathbf{1}'(\Sigma_N W_M^\star)^s\Sigma_N\,\mathbf{1} = O\big(M^{s+1}\log^{4s+1}N\big)$.

Proof of Proposition 4. We can write $\mathbf{1}'(\Sigma_N W_M^\star)^s\Sigma_N\,\mathbf{1}$ as

$$\begin{aligned}
&\sum_{0\le r_1,r_1',\dots,r_{2s+2}\le N}\gamma(r_1-r_1')a_N(r_1'-r_2)\cdots\omega(\frac{r_{2s}-r_{2s}'}{M})a_N(r_{2s}'-r_{2s+1})\gamma(r_{2s+1}-r_{2s+2})\\
&= (2\pi)^{4s+1}N\int_{\Pi^{4s+1}} f(\lambda_{2s+1})\prod_{j=1}^{s}\big\{f(\lambda_{2j-1})A_N(\lambda_{2j-1}')K_M(\lambda_{2j})A_N(\lambda_{2j}')\big\}\\
&\quad\times\Phi_N^{(4s+2)}(\lambda_1,\lambda_1'-\lambda_1,\lambda_2-\lambda_1',\cdots,\lambda_{2s+1}-\lambda_{2s}',-\lambda_{2s+1})d\lambda_1\cdots d\lambda_{2s+1}\\
&= (2\pi)^{4s+1}N\int_{\Pi^{4s+1}} H_N(\mu)A_N^{(2s)}(\mu)\Phi_N^{(4s+2)}(\mu)d\mu, \qquad (58)
\end{aligned}$$

say, where we have changed variables as in Proposition 2, $H_N(\mu) = f(\mu_1)K_M(\mu_1+\mu_1')\cdots K_M(\mu_1+\cdots+\mu_{2s}')f(\mu_1+\cdots+\mu_{2s+1})$, $A_N^{(2s)}(\mu) = A_N(\mu_1+\mu_1')\cdots A_N(\mu_1+\cdots+\mu_{2s}+\mu_{2s}')$ grouping all the functions A_N, and $d\mu = d\mu_1 d\mu_1'\cdots d\mu_{2s+1}$.

To study the difference between the integral in (58) and $f^{s+1}(0)K_M^s(0)\int A_N^{(2s)}(\mu)\Phi_N^{(4s+2)}(\mu)d\mu$ we divide the range of integration, Π^{4s+1}, into two sets, Ω and its complement Ω^c, where Ω is defined by the condition $\Omega = \{|\mu_j| \le \pi/[M(2s+2)],\ j = 1,\dots,2s+1\}$.

In this case we only need the smoothness properties of K at the origin (inside D). For the difference in the set Ω, we can use inequality (33), the Lipschitz property of K and the differentiability of f:

$$\begin{aligned}
&\left|\int_\Omega H_N(\mu)A_N^{(2s)}(\mu)\Phi_N^{(4s+2)}(\mu)d\mu - \int_\Omega f^{s+1}(0)K_M^s(0)A_N^{(2s)}(\mu)\Phi_N^{(4s+2)}(\mu)d\mu\right|\\
&= O(M^{s+1}\sup_\mu\left|A_N^{(2s)}(\mu)\right|)\int_{\Pi^{2s+1}}\sum_{j=2}^{2s}|\mu_j|\left|\Phi_N^{(4s+2)}(\mu)\right|d\mu = O\big(M^{s+1}N^{-1}\log^{4s+1}N\big), \qquad (59)
\end{aligned}$$

using the fact that $\sup_\mu |A_N^{(2s)}(\mu)| = O(1)$. Now, the integral over the set Ω^c can be bounded by

$$\int_{\Omega^c} |H_N(\mu)| \left| A_N^{(2s)}(\mu)\Phi_N^{(4s+2)}(\mu) \right| d\mu \; + \; O\big(M^{s+1}M^{-1}\log^{4s+1}N\big)\,. \tag{60}$$

As in the proof of Proposition 1, the integral over Ω^c in (60) is again of order $O(N^{-1}M^{s+1}\log^{4s+1}N)$, using boundedness of $A_N^{(2s)}$. Therefore, from (58) to (60) we have that

$$\mathbf{1}'(\Sigma_N W_M^\star)^s \Sigma_N \, \mathbf{1} = (2\pi)^{2s+1} N [f(0)]^{s+1} [K_M(0)]^s \int A_N^{(2s)}(\mu)\Phi_N^{(4s+2)}(\mu) d\mu + O\left(M^{s+1}\log^{4s+1}N\right),$$

which is just $O\left(M^{s+1}\log^{4s+1}N\right)$ because $\mathbf{1}'(A_N)^{2s}\,\mathbf{1} = \mathbf{1}'A_N\mathbf{1} = 0$ since $\int_{\Pi^{4s+1}} A_N^{(2s)}(\mu)\Phi_N^{(4s+2)}(\mu)d\mu$ $= \left((2\pi)^{4s+1}N\right)^{-1}\mathbf{1}'(A_N)^{2s}\mathbf{1} = 0$. ■

Lemma 19 *Under Assumptions 1, 3, 4, $M^{-1} + N^{-1}M\log^7 N \to 0$ as $N \to \infty$, there exists a positive number $\delta_1 > 0$ such that, for $\|\mathbf{t}\| \le \delta_1\sqrt{N/M}$ and a constant $d_1 > 0$:*

$$|\psi^\star(\mathbf{t}) - A_N^\star(\mathbf{t},3)| \le \exp\{-d_1\|\mathbf{t}\|^2\}F(\|\mathbf{t}\|)O\left(\left(\frac{N}{M}\right)^{-3/2} + \left(\frac{N}{M}\right)^{-1}\left[M^{-d^\star-1} + M^{-d-\varrho} + e_N(4)\right]\right),$$

where F is a polynomial in $\mathbf{t}$ with bounded coefficients.

Proof of Lemma 19. Follows as Lemma 14. ■

Lemma 20 *Under Assumptions 1, 2 ($p > 1$), 3, 4, $M^{-1} + N^{-1}M\log^3 N \to 0$ as $N \to \infty$, there exists a positive constant $d_2 > 0$ such that for $\|\mathbf{t}\| > \delta_1 m_N^\star$, $|\psi^\star(t_1,t_2)| \le \exp\left\{-d_2\left(m_N^\star\right)^2\right\}$ with $m_N^\star = m_N\log^{-2}N \to \infty$ as $N \to \infty$.*

Proof of Lemma 20. Follows as Lemma 15 using the fact that the asymptotic variance of the spectral estimate is unaffected by mean-correction, and using Lemma 21. ■

Lemma 21 *Under the assumptions of Theorem 3, $\|\Sigma_N W_M^\star\| \le c_1\vartheta_N^\star$, where $0 < c_1 < \infty$ is a constant depending on f and K, and $\vartheta_N^\star = \vartheta_N\log^2 N$.*

Proof of Lemma 21. Write as in the proof of Lemma 16, $\|\Sigma_N W_M^\star\| = \sup_{\|z\|=1}\left|\int_{\Pi^4} F_N(\lambda)d\lambda\right|$, where

$$F_N(\lambda) = Z_N(-\lambda_1)f(\lambda_1)A_N(\lambda_2)K_M(\lambda_3)A_N(\lambda_4)Z_N(\lambda_4)\varphi_N(\lambda_2-\lambda_1)\varphi_N(\lambda_3-\lambda_2)\varphi_N(\lambda_4-\lambda_3).$$

Then changing variables and using the periodicity of all functions,

$$\begin{aligned}
&\sup_{\|z\|=1}\int_{|\lambda_1|\le\epsilon}\int_{\Pi^3}|F_N(\lambda)|\,d\lambda \\
\le\;& \sup_{\|z\|=1} M\|K\|_\infty\|f_\epsilon\|_\infty\int_{\Pi^4}|Z_N(-\lambda_1)Z_N(\lambda_4)\varphi_N(\lambda_2-\lambda_1)\varphi_N(\lambda_3-\lambda_2)\varphi_N(\lambda_4-\lambda_3)|\,d\lambda \\
\le\;& \sup_{\|z\|=1} M\|K\|_\infty\|f_\epsilon\|_\infty\int_{\Pi^3}|\varphi_N(\mu_1)\varphi_N(\mu_2)\varphi_N(\mu_3)|\left[\int_\Pi|Z_N(\lambda)|^2\,d\lambda\int_\Pi\left|Z_N(\lambda-\textstyle\sum_{i=1}^3\mu_i)\right|^2 d\lambda\right]^{1/2}d\mu \\
\le\;& 2\pi M\|K\|_\infty\|f_\epsilon\|_\infty\left(\int_\Pi|\varphi_N(\alpha)|\,d\alpha\right)^3 \le c(f,K)M\log^3 N,
\end{aligned}$$

with $\int_\Pi |Z_N(\lambda)|^2 d\lambda = 2\pi$ and (29). For other values of λ_1, arguing as in the proof of Lemma 16 and since $|\lambda_3| \le \pi/M$, we obtain that

$$\begin{aligned}
&\sup_{\|z\|=1} \int_{|\lambda_1|>\epsilon} \left\{ \int_{|\lambda_2|>\epsilon/2} + \int_{|\lambda_2|\le\epsilon/2} \right\} \int_{\Pi^2} |F_N(\lambda)|\, d\lambda \\
&\le \quad \text{const} \sup_{\|z\|=1} \int_\Pi |Z_N(-\lambda_1)|\, f(\lambda_1) \left(\int_\Pi |\varphi_N(\lambda_2-\lambda_1)|\, d\lambda_2 \right) d\lambda_1 \int_{\Pi^2} |K_M(\lambda_3) Z_N(\lambda_4) \varphi_N(\lambda_4-\lambda_3)|\, d\lambda \\
&\qquad + \text{const} \sup_{\|z\|=1} \int_\Pi |Z_N(-\lambda_1) f(\lambda_1)|\, d\lambda_1 \int_{\Pi^2} \left(\int_\Pi |\varphi_N(\lambda_3-\lambda_2)|\, d\lambda_2 \right) |K_M(\lambda_3) Z_N(\lambda_4) \varphi_N(\lambda_4-\lambda_3)|\, d\lambda.
\end{aligned}$$

Now the Lemma follows using Hölder inequality, periodicity,

$$\sup_{\|z\|=1} \int_{\Pi^2} |K_M(\lambda_3) Z_N(\lambda_4) \varphi_N(\lambda_4-\lambda_3)|\, d\lambda \le \sup_{\|z\|=1} \|K_M\|_2 \|Z_N\|_2 \|\varphi_N\|_1 = O\left(M^{1/2} \log N \right),$$

$\sup_{\|z\|=1} \int_\Pi |Z_N(-\lambda_1) f(\lambda_1)|\, d\lambda_1 = O\left(N^{\frac{2-p}{2p}} \right)$, $\sup_z |Z_N| \le \sqrt{N}$ and $\|Z_N\|_2^2 = 2\pi$. ■

References

Anderson, T.W. (1958) *An Introduction to Multivariate Statistical Analysis.* New York: Wiley.

Bentkus, R.Y. (1972) On the error of the estimate of the spectral function of a stationary process. *Lithuanian Mathematical Journal* 12, 55-71.

Bentkus, R.Y. (1976) On cumulants of the spectrum estimates of a stationary time series. *Lithuanian Mathematical Journal* 16, 37-61.

Bentkus, R.Y. and R.A. Rudzkis (1982) On the distribution of some statistical estimates of spectral density. *Theory of Probability and its Applications* 27, 795-814.

Bhattacharya, R.N. and R.R. Rao (1975) *Normal Approximation and Asymptotic Expansions.* New York: Wiley.

Bhattacharya, R.N. and J.K. Ghosh (1978) On the validity of the formal Edgeworth expansion. *Annals of Statistics* 6, 434-451.

Brillinger, D.R. (1979) Confidence intervals for the crosscovariance function. *Selecta Statistica Canadiana* 5, 3-16.

Chibisov, D.M. (1972) An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. *Theory of Probability and its Applications* 17, 621-630.

Dahlhaus, R. (1983) Spectral analysis with tapered data. *Journal of Time Series Analysis* 4, 163-175.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications*, Vol. 2, 2nd Edition. New York: Wiley.

Gasser, T., H.-G. Muller and V. Mammitzsch (1985) Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society, Series B* 47, 238-252.

Götze, F. and C. Hipp (1983) Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlickheits Theorie und Verwandte Gebiete* 64, 211-239.

Götze, F. and H.R. Künsch (1996) Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics* 24, 1914-1933.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion.* New York: Springer-Verlag.

Hannan, E.J. (1957) The variance of the mean of a stationary process. *Journal of the Royal Statistical Society, Series B,* 19, 282-285.

Hannan, E.J. (1958) The estimation of the spectral density after trend removal. *Journal of the Royal Statistical Society, Series B* 20, 323-333.

Hannan, E.J. (1970) *Multiple Time Series.* New York: Wiley.

Janas, D. (1994) Edgeworth expansions for spectral mean estimates with applications to whittle estimates. *Annals of the Institute of Mathematical Statistics*, 46, 667-682.

Jowett, G.H. (1954) The comparison of means of sets of observations from sections of independent stochastic series. *Journal of the Royal Statistical Society, Series B* 17, 208-2 27.

Keenan, D.M. (1986) Limiting behavior of functionals of higher-order sample cumulant spectra. *Annals of Statistics* 14, 134-151.

Phillips, P.C.B. (1977) Approximations to some finite sample distributions associated with a first order stochastic difference equation. *Econometrica* 45, 463-485; erratum, 50, 274.

Phillips, P.C.B. (1980) Finite sample theory and the distributions of alternative estimators of the marginal propensity to consume. *Review of Economic Studies* 47, 183-224.

Robinson, P.M. (1991) Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica*, 59, 755-786.

Robinson, P.M. (1994) Semiparametric analysis of long-memory time series. *Annals of Statistics* 22, 515-539.

Robinson, P.M. (1995a) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23, 1048-1072.

Robinson, P.M. (1995b) The approximate distribution of nonparametric regression estimates. *Statistics and Probability letters* 23, 193-201.

Robinson, P.M. and C. Velasco (1997) Autocorrelation-robust inference. In G.S. Maddala and C.R. Rao (eds.), *Handbook of Statistics* 15, pp. 267-298. Amsterdam: North-Holland.

Rudzkis, R. (1985) On the distribution of the maximum deviation of the Gaussian stationary time series spectral density estimate. *Lithuanian Mathematical Journal* 25, 118-130.

Song, W.T. and B.W. Schmeiser (1992) Variance of the sample mean: properties and graphs of quadratic-form estimators. *Operations Research* 41, 501-517.

Taniguchi, M. (1987) Validity of Edgeworth expansions of minimum contrast estimators for Gaussian ARMA processes. *Journal of Multivariate Analysis* 21, 1-28.

Taniguchi, M. (1991) *Higher Order Asymptotic Theory for Time Series Analysis.* Lecture Notes in Statistics 68. Berlin: Springer-Verlag.

Taniguchi, M. and M.L. Puri (1996) Valid Edgeworth expansions of M-estimators in regression models with weakly dependent residuals. *Econometric Theory* 12, 331-346.

Zygmund, A. (1977) *Trigonometric Series.* Cambridge: Cambridge University Press.