*The Annals of Statistics*
2003, Vol. 31, No. 4, 1325–1375


# EDGEWORTH EXPANSIONS FOR SEMIPARAMETRIC WHITTLE ESTIMATION OF LONG MEMORY

BY L. GIRAITIS[1] AND P. M. ROBINSON[2]

*London School of Economics*

The semiparametric local Whittle or Gaussian estimate of the long memory parameter is known to have especially nice limiting distributional properties, being asymptotically normal with a limiting variance that is completely known. However in moderate samples the normal approximation may not be very good, so we consider a refined, Edgeworth, approximation, for both a tapered estimate and the original untapered one. For the tapered estimate, our higher-order correction involves two terms, one of order $m^{-1/2}$ (where $m$ is the bandwidth number in the estimation), the other a bias term, which increases in $m$; depending on the relative magnitude of the terms, one or the other may dominate, or they may balance. For the untapered estimate we obtain an expansion in which, for $m$ increasing fast enough, the correction consists only of a bias term. We discuss applications of our expansions to improved statistical inference and bandwidth choice. We assume Gaussianity, but in other respects our assumptions seem mild.

**1. Introduction.** First-order asymptotic statistical theory for certain semiparametric estimates of long memory is now well established and convenient for use in statistical inference. Let a stationary Gaussian process $X_t$, $t = 0, \pm 1, \ldots,$ have spectral density $f(\lambda)$, satisfying

$$\mathrm{Cov}(X_0, X_j) = \int_{-\pi}^{\pi} f(\lambda)\cos(j\lambda)\, d\lambda, \qquad j = 0, \pm 1, \ldots,$$

and for some $\alpha \in (-1, 1)$, $G \in (0, \infty)$,

$$f(\lambda) \sim G\lambda^{-\alpha} \qquad \text{as } \lambda \to 0^+, \tag{1.1}$$

where "$\sim$" means that the ratio of left and right sides tends to 1. Then (1.1) is referred to as a semiparametric model for $f(\lambda)$, specifying its form only near zero frequency, where $X_t$ can be said to have short memory when $\alpha = 0$, long memory when $\alpha \in (0, 1)$, and negative memory when $\alpha \in (-1, 0)$. The memory parameter $\alpha$ (like the scale parameter $G$) is typically unknown, and is of primary interest, being related to the fractional differencing parameter $d$ by $\alpha = 2d$ and to the self-similarity parameter $H$ by $\alpha = 2H - 1$. Equation (1.1) is satisfied by

Received September 2000; revised August 2002.
[1]Supported by ESRC Grant R000238212.
[2]Supported in part by a Leverhulme Trust Personal Research Professorship and ESRC Grant R000238212.
*AMS 2000 subject classifications.* Primary 62G20; secondary 62M10.
*Key words and phrases.* Edgeworth expansion, long memory, semiparametric estimation.

leading models for long and negative memory such as fractional autoregressive integrated moving averages (FARIMA) and fractional noise. The latter, however, are parametric, specifying $f(\lambda)$ up to finitely many unknown parameters over all frequencies $(-\pi, \pi]$. When $f(\lambda)$ is thus correctly parameterized, $\alpha$ (and other parameters) can then be precisely estimated, with rate $n^{1/2}$, where $n$ is the sample size. However, if the model is misspecified, inconsistent parameter estimates typically result. This is the case even for estimates of the long-run parameter $\alpha$ when (1.1) holds but the parameterization of higher frequencies is incorrect, in particular in a FARIMA model, if either or both the autoregressive or moving average orders are underspecified or both are overspecified.

Nevertheless, it is possible to find estimates of $\alpha$ and $G$ that can be shown to be consistent under (1.1), with $f(\lambda)$ unspecified away from zero frequency. Two classes of such, "semiparametric," estimates are based on the very well-established statistical principle of "whitening" the data and, as a consequence, have particularly neat asymptotic statistical properties which place them in the forefront for use in statistical inference on memory. This whitening occurs in the frequency domain. Let $w(\lambda)$ and $I(\lambda)$ be, respectively, the discrete Fourier transform and the periodogram of $X_t$ based on $n$ observations,

$$w(\lambda) = (2\pi n)^{-1/2} \sum_{t=1}^{n} X_t e^{it\lambda}, \qquad I(\lambda) = |w(\lambda)|^2. \tag{1.2}$$

Denote by $\lambda_j = 2\pi j/n$, for integer $j$, the Fourier frequencies. Then for certain sequences $l = l_n \geq 1$ and $m = m_n$, which increase slowly with $n$, under regularity conditions the ratios $r_j = I(\lambda_j)/f(\lambda_j)$, $l \leq j \leq m$ can be regarded as approximately independent and identically distributed (i.i.d.) in a sense that can be rigorously characterized. We call $l$ the *trimming* number and $m$ the *bandwidth* number.

A popular semiparametric estimate of $\alpha$ is the log-periodogram estimate of Geweke and Porter-Hudak (1983), defined here [in the manner of Robinson (1995a) that relates more directly to the form (1.1)] as the least squares estimate in the "linear regression model"

$$\log I(\lambda_j) = \log G - \alpha \log \lambda_j + u_j, \qquad j = l, \dots, m, \tag{1.3}$$

where the $u_j$ are "approximately" $\log r_j$, following (1.1). Denoting this estimate of $\alpha$ by $\widetilde{\alpha}$, Robinson (1995a) showed that under suitable conditions,

$$m^{1/2}(\widetilde{\alpha} - \alpha) \xrightarrow{d} N\left(0, \frac{\pi^2}{6}\right) \qquad \text{as } n \to \infty. \tag{1.4}$$

This is an extremely simple result to use in statistical inference, especially as the asymptotic variance $\pi^2/6$ is independent of $\alpha$. Hurvich and Brodsky (2001) showed that under slightly stronger conditions we can take $l = 1$ in the estimation, while Velasco (1999a) has shown that (1.4) can also hold, for a modified estimate, when $X_t$ is non-Gaussian but linear. In the asymptotic theory of Robinson (1995a) and Velasco (1999a), the conditions on $f(\lambda)$ away from zero frequency extend (1.1) only mildly, not requiring $f(\lambda)$ to be smooth or even bounded or bounded away from zero. However, under a global smoothness condition on $f(\lambda)/G\lambda^{-\alpha}$, similar results have been obtained by Moulines and Soulier (1999) for an alternative estimate originally proposed by Janacek (1982), in which increasingly many, $p$, trigonometric regressors are included in (1.3), and the regression is carried out over frequencies up to $j = n - 1$; the rate of convergence in (1.4) is then $p^{1/2}$, rather than $m^{1/2}$.

An efficiency improvement to $\tilde{\alpha}$ was proposed by Robinson (1995a), in which groups of finitely many, $J$, consecutive $I(\lambda_j)$ are pooled prior to logging. Asymptotic efficiency increases with $J$, but it turns out that the efficiency bound, as $J \to \infty$, can be achieved by an alternative estimate of $\alpha$, the Gaussian semiparametric or local Whittle estimate originally proposed by Künsch (1987). This is also based on periodogram ratios and as it is an implicitly defined extremum estimate, we henceforth distinguish between the true value, now denoted $\alpha_0$, and any admissible value, denoted $\alpha$. After eliminating $G$ from a narrow-band Whittle objective function, as in Robinson (1995b), we consider

$$\hat{\alpha} = \arg\min_{\alpha \in I} R(\alpha), \tag{1.5}$$

where

$$R(\alpha) = \log\left[\frac{1}{m}\sum_{j=1}^{m} j^{\alpha} I(\lambda_j)\right] - \frac{\alpha}{m}\sum_{j=1}^{m}\log j \tag{1.6}$$

and $I$ is a compact subset of $[-1, 1]$. Under regularity conditions, Robinson (1995b) showed that

$$m^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, 1) \qquad \text{as } n \to \infty. \tag{1.7}$$

These conditions are very similar to those employed by Robinson (1995a) for $\tilde{\alpha}$, except that $X_t$ need not be Gaussian, but only a linear process in martingale difference innovations, whose squares, centered at their expectation, are also martingale differences. Robinson and Henry (1999) showed that (1.7) can still hold when the innovations have autoregressive conditional heteroscedasticity. As in (1.4), the asymptotic variance in (1.7) is desirably constant over $\alpha_0$, while $\hat{\alpha}$ is clearly asymptotically more efficient than $\tilde{\alpha}$ for all $\alpha_0$ and the same $m$ sequence.

Semiparametric estimates have drawbacks, however. Due to the merely local specification (1.1), $m$ must increase more slowly than $n$, so that $\tilde{\alpha}$ and $\hat{\alpha}$ converge more slowly than ($n^{1/2}$ -consistent) estimates based on a fully parametric model. Indeed, too large a choice of $m$ entails an element of nonlocal averaging and is a source of bias. If $n$ is extremely large, as is possible in many financial time series, for example, then we may feel able to choose $m$ large enough to achieve acceptable precision without incurring significant bias. However, in series of moderate length, we have to think in terms of $m$ which may be small enough to prompt concern about the goodness of the normal approximation in (1.4) and (1.7).

Higher-order asymptotic theory is a means of improving on the accuracy of the normal approximation in many statistical models. This has been most extensively developed for parametric statistics, where in particular Edgeworth expansions of the distribution function and density function have been derived, such that the first term in the expansion corresponds to the normal approximation while later terms are of increasingly smaller order (in powers of $n^{-1/2}$) but improve on the approximation for moderate $n$. Taniguchi [(1991), e.g.] has extensively and rigorously analyzed Edgeworth expansions for Whittle estimates of parametric short memory Gaussian processes. Given this work, and Fox and Taqqu's (1986) extension to long memory of the central limit theorem (CLT) for Whittle estimates of Hannan (1973) under short memory, the existence and basic structure of Edgeworth expansions for Whittle estimates of parametric long memory models can be anticipated. Indeed, Liebermann, Rousseau and Zucker (2001) have developed valid Edgeworth expansions (of arbitrary order) for quadratic forms of Gaussian long memory series, with application to sample autocovariances and sample autocorrelations. Edgeworth expansions have also been developed for some statistics, which, like $\widetilde{\alpha}$ and $\widehat{\alpha}$, converge at slower, "nonparametric," rates. We note, for example, the work of Bentkus and Rudzkis (1982) on smoothed nonparametric spectral density estimates for short memory Gaussian time series, later developed by Velasco and Robinson (2001), while related results have also been obtained for smoothed nonparametric probability density estimates by Hall (1991) and for Nadaraya–Watson nonparametric regression estimates by Robinson (1995c). However, this literature seems small compared to the parametric one, and the development and study of Edgeworth expansions for semiparametric estimates of the memory parameter seems an especially distinctive problem, in view of the current interest in such estimates due to their flexibility, discussed above, the notational and expositional advantage of being able to focus on a single parameter $\alpha_0$, the simple parameter-estimate-free studentization afforded by (1.4) and (1.7), and the interesting role played by the bandwidth $m$ in a semiparametric set-up, in which terms due to the bias can compete with Edgeworth terms of a more standard character; indeed, our Edgeworth expansion provides a method of choosing $m$, proposed by Nishiyama and Robinson (2000) in another context, which seems more appropriate in the context of statistical inference than the usual minimum mean squared error rules.

We study here only $\widehat{\alpha}$, and trimmed and tapered versions of it, not so much because of its greater first-order efficiency than $\widetilde{\alpha}$, but because of its greater mathematical tractability. Though, unlike $\widetilde{\alpha}$, it is not defined in closed form, its higher-order properties can nevertheless be analyzed by making use of general results for the implicitly defined extremum estimates of Bhattacharya and Ghosh (1978), whereas the logged periodograms appearing in $\widetilde{\alpha}$ are technically difficult to handle. Our theory also requires development of Edgeworth expansions for quadratic forms of a type not covered by Lieberman, Rousseau and Zucker (2001) (due principally to the narrow-band nature of ours, in the frequency domain). Various other estimates of $\alpha_0$ that are also semiparametric in character have been studied, such as versions of the $R/S$ statistic, the averaged periodogram estimate, and the variance type estimate. However, not only do these also converge more slowly than $n^{1/2}$ under the semiparametric specification, but unlike $\widetilde{\alpha}$ and $\widehat{\alpha}$ they are not necessarily asymptotically normal, or they may be asymptotically normal only over a subset of $\alpha$ values, where they can have a complicated $\alpha$-dependent asymptotic variance; they have a nonstandard limit distribution elsewhere. Such estimates are thus much less convenient for use in statistical inference than $\widetilde{\alpha}$ and $\widehat{\alpha}$, and moreover do not lend themselves so readily to higher-order analysis. Though higher-order approximations to the distribution of $\widehat{\alpha}$ are of course more complicated than (1.7), they are, as we show, still usable, and indeed can be approximated by a normal distribution with a corrected mean and variance, so that normal-based inference is still possible.

We give greater stress to a (cosine bell) tapered version of $\widehat{\alpha}$, where the $m$ frequencies employed are not the adjacent Fourier ones, at $2\pi/n$ intervals, as used in (1.6), but are separated by $6\pi/n$ intervals, so that two $\lambda_j$ are "skipped." The skipping avoids the correlation across nearby frequencies that is induced by tapering, which otherwise improves the i.i.d. approximation of the periodogram ratios $r_j$, to enable a valid Edgeworth expansion with a correction term of order $m^{-1/2}$ (with desirably a completely known coefficient), along with a higher order "bias" term, which is increasing in $m$. The $m^{-1/2}$ correction term is what we would expect from the classical Edgeworth literature, obtaining in case of weighted periodogram spectral density estimates for short memory series. Without the tapering and skipping, the $m^{-1/2}$ term appears to be dominated by something which we estimate as of order $m^{-1/2}\log^4 m$, but if $m$ increases sufficiently fast this term is in any case dominated by the bias term. Tapering was originally used in nonparametric spectral analysis of short memory time series to reduce bias. More recently, to cope with possible nonstationarity, it has been used in the context of $\widetilde{\alpha}$ by Hurvich and Ray (1995) and in first-order asymptotic theory for both $\widetilde{\alpha}$ and $\widehat{\alpha}$ by Velasco (1999a, b); tapering has also been used in a stationary setting by Giraitis, Robinson and Samarov (2000) to improve the convergence rate of $\widetilde{\alpha}$ based on a data-dependent bandwidth. Trimming also plays a role in our Edgeworth expansion for the tapered estimate. This was used in first-order asymptotic theory for $\widetilde{\alpha}$ of Robinson (1995a), but not for $\widehat{\alpha}$ [Robinson (1995b)].

The following section describes our main results, with detailed definition of our estimates of $\alpha_0$, regularity conditions and Edgeworth expansions, including implications for improved inference and bandwidth choice. Section 3 develops our expansion to provide feasible improved inference, entailing data dependent estimation of the "higher-order bias." Section 4 presents the main steps of the proof, which depends on technical details developed in Sections 5–7, some of which may be of more general interest.

**2. Edgeworth expansions.** We define the statistics

$$w_h(\lambda) = \left(2\pi \sum_{t=1}^{n} h_t^2\right)^{-1/2} \sum_{t=1}^{n} h_t X_t e^{it\lambda}, \qquad I_h(\lambda) = |w_h(\lambda)|^2, \tag{2.1}$$

where $h_t = h(t/n)$, with

$$h(x) = \tfrac{1}{2}(1 - \cos 2\pi x), \qquad 0 \le x \le 1. \tag{2.2}$$

The function $h(x)$ is a *cosine bell taper*. We could establish results like those below with (2.2) replaced in (2.1) by alternative tapers $h(\cdot)$, which like (2.2), have the property of tending smoothly to zero as $x \to 0, x \to 1$. Tapers increase asymptotic variance unless a suitable degree, $\ell$, of skipping is implemented, such that only frequencies of form $\lambda_{\ell j}$ are included (so $\ell = 1$ in case of no skipping). We prefer not to incur this greater imprecision, but higher-order bias is seen to increase in $\ell$. For the cosine bell taper we have $\ell = 3$, while larger $\ell$ are needed for many members of the Kolmogorov class of tapers [see Velasco (1999a)], and on the other hand it seems $\ell = 2$ is possible in the complex-valued taper of Hurvich and Chen (2000). However, we in any case incorporate a method of bias correction, and since tapering is in our context just an (apparently) necessary nuisance, we fix on the familiar cosine bell (2.2). We call $w_h(\lambda)$ the tapered discrete Fourier transform and $I_h(\lambda)$ the tapered periodogram. Of course for $h(x) \equiv 1$, $0 \le x \le 1$, $w_h(\lambda)$ and $I_h(\lambda)$ reduce, respectively, to $w(\lambda)$ and $I(\lambda)$ in (1.2).

We consider alongside $\widehat{\alpha}$ (1.5) the tapered (and possibly trimmed) version,

$$\widehat{\alpha}_h = \arg\min_{\alpha \in I} R(\alpha), \tag{2.3}$$

where

$$R_h(\alpha) = \log\left[m^{-1} \sum_{j=l}^{m} j^{\alpha} I_h(\lambda_{3j})\right] - \frac{\alpha}{m - l + 1} \sum_{j=l}^{m} \log j, \tag{2.4}$$

the argument $\lambda_{3j}$ indicates that two $\lambda_j$ are successively skipped, and the lower limit of the summation indicates trimming for $l > 1$. Notice that $\widehat{\alpha}$ (1.5) is given by replacing $I_h(\lambda_{3j})$ by $I(\lambda_j)$, and $l$ by 1; we could allow for trimming also in (1.5) and (1.6) but it plays no useful role in our expansion for $\widehat{\alpha}$, unlike that for $\widehat{\alpha}_h$.

We now describe our regularity conditions. The first is standard.

ASSUMPTION $\alpha$. $\alpha_0$ is an interior point of $I = [a, b]$, where $a \ge -1, b \le 1$.

In the CLTs of Robinson (1995a, b), (1.1) was refined in order to describe the error in approximating the left side by the right. This error plays an even more prominent role in higher-order theory, and we introduce the following assumption.

ASSUMPTION $f$.

$$f(\lambda) = |\lambda|^{-\alpha_0} g(\lambda), \qquad \lambda \in [-\pi, \pi], \tag{2.5}$$

where for constants $c_0 \neq 0$, $c_1$ and $\beta \in (0, 2]$,

$$g(\lambda) = c_0 + c_1|\lambda|^{\beta} + o(|\lambda|^{\beta}) \qquad \text{as } \lambda \to 0. \tag{2.6}$$

In addition $f(\lambda)$ is differentiable in a neighborhood of the origin and

$$(\partial/\partial\lambda) \log f(\lambda) = O(|\lambda|^{-1}) \qquad \text{as } \lambda \to 0. \tag{2.7}$$

Under Assumption $f$, we have the following properties of the

$$v(\lambda_j) = \lambda_j^{\alpha_0/2} w(\lambda_j), \qquad v_h(\lambda_j) = \lambda_j^{\alpha_0/2} w_h(\lambda_j), \tag{2.8}$$

which are so important to the sequel that we present them here, without proof.

LEMMA 2.1 [Robinson (1995a)]. *Let Assumption $f$ be satisfied. Then uniformly in $1 \le k < j = o(n)$, as $n \to \infty$:*

(a) $Ev(\lambda_j)\overline{v(\lambda_j)} = g(\lambda_j) + O(j^{-1} \log j)$,
(b) $Ev(\lambda_j)v(\lambda_j) = O(j^{-1} \log j)$,
(c) $Ev(\lambda_j)\overline{v(\lambda_k)} = O(k^{-|\alpha_0|/2}|j|^{-1+|\alpha_0|/2} \log j)$,
(d) $Ev(\lambda_j)v(\lambda_k) = O(k^{-|\alpha_0|/2}|j|^{-1+|\alpha_0|/2} \log j)$.

This result was derived by Robinson (1995a), but in the actual statement of his Theorem 2, (c) and (d) were replaced by the weaker bound $k^{-|\alpha_0|/2}|j|^{-1+|\alpha_0|/2} \times \log j \le k^{-1} \log j$.

LEMMA 2.2 [Giraitis, Robinson and Samarov (2000)]. *Let Assumption $f$ be satisfied. Then uniformly in $1 \le k \le j - 3 = o(n)$, as $n \to \infty$:*

(a) $Ev_h(\lambda_j)\overline{v_h(\lambda_j)} = g(\lambda_j) + O(j^{-2})$,
(b) $Ev_h(\lambda_j)v_h(\lambda_j) = O(j^{-2})$,
(c) $Ev_h(\lambda_j)\overline{v_h(\lambda_k)} = O((j/n)^{\beta}|j-k|^{-2} + k^{-1}|j-k|^{-3/2})$,
(d) $Ev_h(\lambda_j)v_h(\lambda_k) = O((j/n)^{\beta}|j-k|^{-2} + k^{-1}|j-k|^{-3/2})$.

Note the requirement $k \le j - 3$ in Lemma 2.2, which corresponds to the skipping in $\widehat{\alpha}_h$.

In order to use our asymptotic expansions to improve statistical inference, it is generally necessary to specify $\beta$. Estimation of $\beta$ is discussed by Giraitis, Robinson and Samarov (2000). On the other hand, when $f(\lambda)$ is additive in a long memory spectrum and a short memory one, as can happen in case of measurement error or as a consequence of a stochastic volatility model, we typically have $\beta \le \alpha$. However, setting aside such structure, the leading parametric special cases of (1.1), such as FARIMA spectral densities, entail $\beta = 2$, and as this corresponds to the twice-differentiability condition stressed in much of the literature on smoothed nonparametric estimation of spectral and probability densities and regression functions, we explore this case in more detail, with a further refinement which also holds in the FARIMA case:

ASSUMPTION $f'$. Assumption $f$ holds with (2.6) replaced by

$$g(\lambda) = c_0 + c_1|\lambda|^2 + c_2|\lambda|^4 + o(|\lambda|^4) \qquad \text{as } \lambda \to 0. \tag{2.9}$$

The main assumption on the bandwidth $m$ also involves $\beta$.

ASSUMPTION $m$. For some $\eta > 0$,

$$n^{\eta} \le m = O(n^{2\beta/(2\beta+1)}). \tag{2.10}$$

Note that the CLT for $\widehat{\alpha}$, centered at $\alpha_0$, holds only for $m = o(n^{2\beta/(2\beta+1)})$ [Robinson (1995b)]. We allow the upper bound rate $n^{2\beta/(2\beta+1)}$ in (2.10) because we will also consider recentered estimation. The rate $n^{2\beta/(2\beta+1)}$ is the minimum mean squared error (MSE) one, and $K \in (0, \infty)$ in $m \sim Kn^{2\beta/(2\beta+1)}$ can be optimally chosen, in a data dependent fashion, on this basis [see Henry and Robinson (1996)].

For the trimming number $l$ we introduce the following.

ASSUMPTION $l$. If $|I| \le 1$, $\log^5 m \le l \le m^{1/3}$. If $|I| > 1$, $m^{\eta} \le l \le m^{1/3}$ for some $\eta > 0$.

Assumption $l$ implies that "less" trimming in $\widehat{\alpha}_h$ is needed when $|I| \le 1$, as is the case if we know $X_t$ has long memory and take $I \subset [0, 1]$. (In view of Assumption $\alpha$, this would not permit inference on short memory, $\alpha_0 = 0$, but $I = [-\varepsilon, 1 - \varepsilon]$ would.) Strictly speaking (see the proof of Lemma 5.7 below), this requirement $|I| \le 1$ can be relaxed to $I = [\alpha_0 - 1 + \varepsilon, 1]$ for any $\varepsilon > 0$, so that for $\alpha_0 < 0$, $I = [-1, 1]$ is possible, but of course $\alpha_0$ is unknown.

We establish Edgeworth expansions for the quantities

$$U_m = m^{1/2}(\widehat{\alpha} - \alpha_0), \qquad U_m^h = m^{1/2}(\widehat{\alpha}_h - \alpha_0).$$

These involve the parameter

$$\theta_\ell = \frac{c_1}{c_0}\frac{\beta}{(\beta+1)^2}\left(\frac{\ell}{2\pi}\right)^{\beta}, \tag{2.11}$$

for $\ell = 1$ and $\ell = 3$, respectively, and the sequence

$$q_m = m^{1/2}\left(\frac{m}{n}\right)^{\beta}, \tag{2.12}$$

where (2.11) and (2.12) represent, respectively, the coefficient and rate of a bias term. In connection with Assumption $m$ and the subsequent discussion, note that $q_m \to 0$ if $m = o(n^{2\beta/(2\beta+1)})$ whereas $q_m \sim K^{\beta+2}$ if $m \sim Kn^{2\beta/(2\beta+1)}$. We also introduce the standard normal distribution and density functions,

$$\Phi(y) = \int_{-\infty}^{y} \phi(y)\,dy, \qquad \phi(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

THEOREM 2.1. *Let Assumptions $\alpha$, $f$, $m$ and $l$ hold.*

(i) *If $m = o(n^{2\beta/(2\beta+1)})$ then as $n \to \infty$,*

$$\sup_{y \in R} \bigl| P\{U_m^h \le y\} - \Phi(y) - \phi(y)\bigl(\theta_3 q_m + m^{-1/2} p(y)\bigr)\bigr| = o(q_m + m^{-1/2}), \tag{2.13}$$

*where*

$$p(y) = \frac{2 + y^2}{3}. \tag{2.14}$$

*If $m \sim Kn^{2\beta/(2\beta+1)}$, $K \in (0, \infty)$, then as $n \to \infty$,*

$$\sup_{y \in R} \bigl| P\{U_m^h \le y\} - \Phi(y + \theta_3 K^{\beta+1/2}) \bigr| = o(1). \tag{2.15}$$

(ii) *If $\log^4 m/(m^{1/2} q_m) \to 0$,*

$$\sup_{y \in R} \bigl| P\{U_m \le y\} - \Phi(y) - \phi(y)\theta_1 q_m \bigr| = o(q_m). \tag{2.16}$$

There is no $m^{-1/2}$ term in the expansion (2.16) for the untapered estimate $\widehat{\alpha}$ because it is, in effect, dominated by a remainder term whose order of magnitude depends on the approximation errors in Lemma 2.1, so we are only able to obtain a useful asymptotic expansion by making $m$ increase faster than $n^{\beta/(\beta+1)}$ such that $q_m$ dominates. Our conditions are only sufficient, but we are unable to see a way of improving Lemma 2.1 to the extent of obtaining an expansion for $U_m$ involving both $m^{-1/2}$ and $q_m$, as in (2.13), explaining our resort to tapering. To conserve on space we focus the discussion which follows on the tapered results (2.13) and (2.15), though some consequences for the untapered case (2.16) can be inferred, dropping the $m^{-1/2}$ term and replacing $\theta_3$ by $\theta_1$.

There are three cases of interest in (2.13), which can be isolated and discussed similarly to Robinson (1995c) and Nishiyama and Robinson (2000), for different nonparametric and semiparametric statistics.

(i) When

$$m/n^{\beta/(\beta+1)} \to 0, \tag{2.17}$$

we deduce

$$P(U_m^h \le y) = \Phi(y) + p(y)\phi(y)m^{-1/2} + o(m^{-1/2}).$$

(ii) When

$$m \sim Kn^{\beta/(\beta+1)}, \qquad K \in (0,\infty), \tag{2.18}$$

we deduce

$$P(U_m^h \le y) = \Phi(y) + n^{-\beta/2(\beta+1)}\phi(y)\big(\theta_3 K^{\beta+1/2} + K^{-1/2}p(y)\big) + o\big(n^{-\beta/2(\beta+1)}\big). \tag{2.19}$$

(iii) When

$$m/n^{\beta/(\beta+1)} \to \infty, \tag{2.20}$$

we deduce

$$P(U_m^h \le y) = \Phi(y) + \theta_3\phi(y)q_m + o(q_m). \tag{2.21}$$

In case (i) $m$ is chosen so small that the bias does not enter. If we believe in (2.17) there is the benefit that $\theta_3$, which will be unknown in practice, is not involved in the refined approximation, only the known polynomial $p(y)$. In case (iii), on the other hand, $m$ is so large that the bias dominates; (2.20) permits only a slightly slower rate for $m$ than that required for (2.16) and in the region of $m = o(n^{2\beta/(2\beta+1)})$ that approaches the minimal MSE case

$$m \sim Kn^{2\beta/(2\beta+1)}. \tag{2.22}$$

Case (ii) is the one in which $m$ is chosen to minimize the error in the normal approximation. Note the difference between (2.22) and (2.18). Case (iii) has the advantage of entailing a smaller confidence interval. However, this is little comfort if the interval is not suitably centered and the normal interpretation is not appropriate. Robinson (1995c) and Nishiyama and Robinson (2000) suggested that it is $m$ that minimizes the deviation from the normal approximation, not $m$ that minimizes the MSE, that is most relevant in normal-based inference on $\alpha_0$, making (2.18) more relevant than (2.22).

We can go further and optimally estimate $K$ in (2.18). As in Nishiyama and Robinson (2000), consider, in view of (2.19),

$$K_{\text{opt}} = \arg\min_K \max_{y\in R} \big|\phi(y)\big(\theta_3 K^{\beta+1/2} + K^{-1/2}p(y)\big)\big|,$$

choosing $K_{\text{opt}}$ to minimize the maximal deviation from the usual normal approximation. We obtain the simple solution

$$K_{\text{opt}} = \big(3\theta_3(\beta+1/2)\big)^{-1/(\beta+1)}.$$

An alternative to carrying out inference using the Edgeworth approximation is to invert the Edgeworth expansion to get a new statistic whose distribution is closely approximated by the standard normal. From (2.13), uniformly in $y$,

$$P(U_m^h \le y) = \Phi\big(y + \theta_3 q_m + m^{-1/2} p(y)\big) + o(q_m + m^{-1/2}) \tag{2.23}$$

and hence

$$P\big(U_m^h + \theta_3 q_m + p(y) m^{-1/2} \le y\big) \sim \Phi(y).$$

It may be shown that (2.23) implies

$$P\big(U_m^h + \theta_3 q_m + p(y) m^{-1/2} \le y\big) = \Phi(y) + o(q_m + m^{-1/2})$$

uniformly in $y = o(m^{1/6})$. Indeed, by (2.13),

$$P\{U_m^h \le y\} - \Phi\big(y + \theta_3 q_m + m^{-1/2} p(y)\big) = o(q_m + m^{-1/2}).$$

Set $z = y + \theta_3 q_m + m^{-1/2} p(y) = y + m^{-1/2} y^2/3 + a$ where $a = \theta_3 q_m + 2m^{-1/2}/3$. Then

$$y = \frac{-1 + \sqrt{1 + 4m^{-1/2}(z - a)/3}}{2m^{-1/2}/3}.$$

Assuming that $z = o(m^{1/6})$, by Taylor expansion it follows that

$$y = z = a - m^{-1/2}(z - a)^2/3 + o(m^{-1/2}) = z - \theta_3 q_m - m^{-1/2} p(z) + o(m^{-1/2}).$$

The CLT (1.7) of Robinson (1995c) was established without Gaussianity and with only finite moments of order four assumed. The asymptotic variance in (1.7) is unaffected by cumulants of order three and more, and thus hypothesis tests and interval estimates based on (1.7) are broadly applicable. Looking only at our formal higher-order expansion, it is immediately apparent that the bias term (in $q_m$) will not be affected by non-Gaussianity, so nor will be the expansion when $m$ increases so fast that $q_m$ dominates [see (2.16) and (2.21)]. Moreover, preliminary investigations suggest that when $X_t$ is a linear process in i.i.d. innovations satisfying suitable moment conditions, the $m^{-1/2}$ term in the formal expansion is also generally unaffected. (Specifically, the leading terms in Corollary 7.1 are unchanged.) However, as proof of the validity of our expansions even in the linear case seems considerably harder and lengthier, we do not pursue the details here, adding that the estimates $\widehat{\alpha}, \widehat{\alpha}_h$ optimize narrow-band forms of Gaussian likelihoods and are thus in part motivated by Gaussianity, which in any case is frequently assumed in higher-order asymptotic theory.

**3. Empirical expansions and bias correction.** The present section develops our results to provide feasible improved statistical inference on $\alpha_0$. An approximate $100\gamma\%$ confidence interval for $\alpha_0$ based on the CLT is given by

$$(\widehat{\alpha}_h - z_{\gamma/2}m^{-1/2}, \widehat{\alpha}_h + z_{\gamma/2}m^{-1/2}), \tag{3.1}$$

where $1-\Phi(z_\gamma)=\gamma$. From (2.23) a more accurate confidence interval is

$$\begin{aligned}(&\widehat{\alpha}_h + \theta_3 q_m m^{-1/2} + p(z_{\gamma/2})m^{-1} - z_{\gamma/2}m^{-1/2},\\ &\widehat{\alpha}_h + \theta_3 q_m m^{-1/2} + p(z_{\gamma/2})m^{-1} + z_{\gamma/2}m^{-1/2}).\end{aligned} \tag{3.2}$$

Of course (3.1) and (3.2) correspond to level-$\gamma$ hypothesis tests on $\alpha_0$. We reject the null hypothesis $\alpha_0 = \alpha_0^0$, for given $\alpha_0^0$ (e.g., $\alpha_0^0 = 0$, corresponding to a test of short memory) if $\alpha_0^0$ falls outside (3.1) or, more accurately, (3.2).

An obvious flaw in the preceding discussion is that $\theta_3$ is unknown in practice. However, given an estimate $\widehat{\theta}_3$ such that

$$\widehat{\theta}_3 \to \theta_3 \qquad \text{a.s.}, \tag{3.3}$$

we deduce from (2.13) the empirical Edgeworth expansion

$$\begin{aligned}&\sup_{y\in R}\bigl|P\{U_m^h \le y\} - \Phi(y) - \phi(y)\bigl(\widehat{\theta}_3 q_m + m^{-1/2}p(y)\bigr)\bigr|\\ &\qquad = o(q_m + m^{-1/2}) \qquad \text{a.s.}\end{aligned} \tag{3.4}$$

We can likewise replace $\theta_3$ by $\widehat{\theta}_3$ in (2.15), (2.19), (2.21), (2.23) and (3.2).

We discuss two alternative estimates of $\theta_3$. Our first is

$$\widehat{\theta}_{3,1} = \left(\frac{n}{m'}\right)^{\beta} R_{m'}^{(1)}(\widehat{\alpha}_h), \tag{3.5}$$

where

$$R_{m'}^{(1)}(\alpha) = \frac{S_{1,m'}(\alpha)}{S_{0,m'}(\alpha)}, \tag{3.6}$$

in which

$$S_{k,m'}(\alpha) = \frac{1}{c_0 m'}\sum_{j=l}^{m'} \nu_{j,m'}^k \lambda_{3j}^{\alpha} I_h(\lambda_{3j}), \qquad k \ge 0, \tag{3.7}$$

and

$$\nu_{j,m'} = \log j - (m'-l+1)^{-1}\sum_{j=l}^{m'}\log j, \tag{3.8}$$

where $m'$ is another bandwidth, increasing faster than $m$. Note that $R_m^{(1)}(\alpha) = (d/d\alpha)R_m(\alpha)$ [see (2.4)], and so $R_m^{(1)}(\widehat{\alpha}_h) = 0$. Our second estimate of $\theta_3$ is

$$\widehat{\theta}_{3,2} = \frac{\widehat{c}_1}{\widehat{c}_0}\frac{\beta}{(\beta+1)^2}\left(\frac{3}{2\pi}\right)^{\beta}, \tag{3.9}$$

where, as in Henry and Robinson (1996), $\widehat{c}_0$ and $\widehat{c}_1$ are given by least squares regression based on (2.6), that is,

$$\begin{bmatrix}\widehat{c}_0\\ \widehat{c}_1\end{bmatrix} = \left[\sum_{j=l}^{m'}\begin{pmatrix}1 & \lambda_{3j}^{\beta}\\ \lambda_{3j}^{\beta} & \lambda_{3j}^{2\beta}\end{pmatrix}\right]^{-1}\sum_{j=l}^{m'}\begin{bmatrix}1\\ \lambda_{3j}^{\beta}\end{bmatrix}\lambda_{3j}^{\widehat{\alpha}_h} I_h(\lambda_{3j}). \tag{3.10}$$

Estimation of $\theta_1$ is relevant in connection with (2.16). We define $\widehat{\theta}_{1,1}$ by replacing $\lambda_{3j}$ by $\lambda_{1j}$, $I_h$ by $I$ and 3 by 1 in (3.7) and then $\widehat{\alpha}_h$ by $\widehat{\alpha}$ in (3.5). Likewise we can define $\widehat{\theta}_{1,2}$ by (3.9) with 3 replaced by 1 in (3.9) and 3, $\lambda_{3j}$, $I_h$ and $\widehat{\alpha}_h$ by 1, $\lambda_j$, $I$ and $\widehat{\alpha}$ in (3.10).

LEMMA 3.1. *Let Assumptions* $\alpha$, $f$, $l$ *and* $m$ *hold, and let*

$$nm^{-(1/2\beta)+\varepsilon} \le m' \le n^{1-\varepsilon} \tag{3.11}$$

*hold for some* $\varepsilon > 0$. *Then for* $\ell = 1, 3$,

$$\widehat{\theta}_{\ell,i} \to \theta_\ell \qquad a.s., \ i = 1, 2.$$

Note that (3.11) implies that $n^{2\beta/(2\beta+1)}/m' \to 0$ so that $m = o(m')$.

As discussed in the previous section, we now focus on the case $\beta = 2$. We first modify Theorem 2.1, noting that for $\beta = 2$,

$$q_m = m^{1/2}\left(\frac{m}{n}\right)^2, \qquad \theta_\ell = \frac{c_1}{c_0}\frac{2}{9}\left(\frac{\ell}{2\pi}\right)^2, \ \ell = 1, 3. \tag{3.12}$$

THEOREM 3.1. *Let* $\beta = 2$ *and Assumptions* $\alpha$, $f'$, $l$ *and* $m$ *hold.*

(i) *If* $m = o(n^{4/5})$ *then as* $n \to \infty$,

$$\sup_{y\in R}\left|P\{U_m^h \le y\} - \Phi\big(y + \theta_3 q_m + m^{-1/2}p(y)\big)\right| = o(m^{-1/2}). \tag{3.13}$$

*If* $m \sim Kn^{4/5}$, $K \in (0, \infty)$, *then as* $n \to \infty$,

$$\sup_{y\in R}\left|P\{U_m^h \le y\} - \Phi(y + \theta_3 K^{5/2})\right| = O(m^{-1/2}); \tag{3.14}$$

*indeed, more precisely,*

$$\begin{aligned}\sup_{y\in R}\Big|P\{U_m^h + \theta_3 K^{5/2} \le y\}\\ - \Phi\big(y + m^{-1/2}p(y) + m^{-1/2}K^5\gamma\big)\Big| = o(m^{-1/2}),\end{aligned} \tag{3.15}$$

*where*

$$\gamma = \left(\frac{c_2}{c_0}\frac{4}{25} - \left(\frac{c_1}{c_0}\right)^2\frac{22}{243}\right)\left(\frac{3}{2\pi}\right)^4.$$

(ii)

$$\sup_{y\in R}\big|P\{U_m+\theta_1 K^{5/2}\le y\}-\Phi(y)\big|=o(m^{-1/2}\log^4 m).$$

Prompted by (2.23) and (3.4), we now consider expansions for bias-corrected estimates,

$$\alpha_h^*=\widehat{\alpha}_h+(m/n)^2\widehat{\theta}_3^*,\qquad \alpha^*=\widehat{\alpha}+(m/n)^2\widehat{\theta}_1^*, \tag{3.16}$$

where

$$\widehat{\theta}_\ell^*=\frac{\widehat{\theta}_{\ell,1}}{1-(m/m')^2}. \tag{3.17}$$

(To conserve on space we consider only $\widehat{\theta}_{\ell,1}$ here and not $\widehat{\theta}_{\ell,2}$.) Define

$$U_m^{*h}=\sqrt{m}(\alpha_h^*-\alpha_0)=\sqrt{m}(\widehat{\alpha}_h-\alpha_0)+q_m\widehat{\theta}_3^*,$$
$$U_m^*=\sqrt{m}(\alpha^*-\alpha_0)=\sqrt{m}(\widehat{\alpha}-\alpha_0)+q_m\widehat{\theta}_1^*.$$

The following theorem shows that the distributions of $U_m^{*h},U_m^*$ converge to the normal limit faster than those of $U_m^h,U_m$ (albeit slower than the optimal rate pertaining to the infeasible statistics $U_m^h+q_m\theta_3,U_m+q_m\theta_1$). Set $k_m=\sqrt{m}(m/n)^2(m'/n)^2$, $v_m=(m/m')^2$, $r_m=k_m+v_m+m^{-1/2}$.

THEOREM 3.2. *Let $\beta=2$ and Assumptions $\alpha$, $f'$ and $l$ hold. Let*

$$l\le m\le n^{8/9-\varepsilon},\qquad m^{1+\varepsilon}\le m'=o\big(\min(n^{1-\varepsilon},n^2m^{-5/4})\big) \tag{3.18}$$

*for some $\varepsilon>0$. Then as $n\to\infty$,*

$$\sup_{y\in R}\big|P\{U_m^{*h}\le y\}-\Phi(y)\big|=O(r_m)\to 0 \tag{3.19}$$

*and, more precisely,*

$$\begin{aligned}&\sup_{y\in R}\big|P\{U_m^{*h}\le y\}-\Phi\big(y+(\theta_3a_3-b_3)k_m+m^{-1/2}p(y)-v_my\big)\big|\\&\quad=o(r_m)\to 0,\end{aligned} \tag{3.20}$$

*where*

$$a_\ell=\frac{c_1}{c_0}\frac{1}{3}\left(\frac{\ell}{2\pi}\right)^2,\qquad b_\ell=\frac{c_2}{c_0}\frac{4}{25}\left(\frac{\ell}{2\pi}\right)^2. \tag{3.21}$$

*Also,*

$$\sup_{y\in R}\big|P\{U_m^*\le y\}-\Phi(y)\big|=O(r_m+m^{-1/2}\log^4 m)\to 0. \tag{3.22}$$

On choosing $m \sim Kn^{8/11}, m' \sim K'n^{10/11}$, for $K, K' \in (0, \infty)$, it follows that $k_m, v_m$ both increase like $m^{-1/2}$ and so the term $r_m$ in the error bounds of (3.19) and (3.20) is minimized by the rate $r_m = n^{-4/11}$. Moreover, it may be shown that we can then invert (3.20) to get

$$\sup_{\{y:|y|=o(m^{1/6})\}} \big|P\{U_m^* + m^{-1/2}\big(K^{5/2}K'^2(\theta a - v) + p(y) - (K/K')^2 y\big) \le y\} - \Phi(y)\big| = o(m^{-1/2}).$$

On the other hand, if $m/n^{8/11} + m/m'^{4/5} + m'm^{3/2}/n^2 \to 0$ (as is true if $m'$ increases either like $nm^{-1/8}$ or $n^{10/11}$) then again we have $r_m = m^{-1/2}$ in (3.19) and (3.20), but this converges more slowly than $n^{-4/11}$; note that here $k_m, v_m = o(m^{-1/2})$ so the correction terms of orders $k_m, v_m$ on the left-hand side of (3.20) can be omitted. On the other hand, if $m' = o(n^{8/11})$ then for any choice of $m$ satisfying our conditions, we have $k_m = o(r_m)$, so the correction term in $k_m$ can be omitted. Finally, if $n^{8/11}/m \to 0$ and $m = O(n^{8/9-\varepsilon})$ then $r_m$ in (3.19) and (3.20) converges more slowly than $m^{-1/2} = o(\max(k_m, v_m))$; indeed on equating the rates of $k_m, v_m$ (so $m'$ increases like $nm^{-1/8}$), we obtain $r_m = m^{9/4}/n^2$, which decreases more slowly than $n^{-4/11}$ but no more slowly than $n^{-9\varepsilon/4}$. Of course in this case the correction term of order $m^{-1/2}$ on the right-hand side of (3.20) can be omitted. For example, in the case $m \sim Kn^{4/5}$ discussed in Theorem 3.1 (where there is not even a central limit theorem for $\widehat{\alpha}_h$ centered at $\alpha_0$), we must have $m' \sim K'n^{9/10}$, and hence $r_m = n^{-1/5}$ in (3.19) and (3.20), while we can invert (3.20) to get

$$\sup_y \big|P\{U_m^*\big(1 - (K/K')^2 n^{-1/5}\big) + K^{5/2}K'^2 n^{-1/5}(\theta a - v) \le y\} - \Phi(y)\big| = o(n^{-1/5}).$$

With regard to (3.22) for the untapered estimate, the error $r_m + m^{-1}\log^4 m$ is minimized, for large $n$, by $m = K(n\log^2 n)^{8/11}, m' = K'n^{10/11}\log^{2/11} n$, whence it decays like $n^{-4/11}\log^{36/11} n$. However, it must be stressed that the $m^{-1/2}\log^4 m$ component of (3.22) is just an upper bound.

We stress that the choices of $m, m'$ discussed above are designed to minimize the error in the normal approximation, but the upper bound choice $m = n^{8/9-\varepsilon}$ in (3.18) entails an asymptotically smaller confidence interval. Moreover, from the standpoint of minimum mean-squared error estimation, the methods of Andrews and Guggenberger (2003), Andrews and Sun (2001) and Robinson and Henry (2003) provide optimal choices of $m$ of order $n^{1/2-\eta}$ for arbitrarily small $\eta > 0$, while those of Moulines and Soulier (1999) and Hurvich and Brodsky (2001) provide an optimal choice of order $(n/\log n)^{1/2}$.

**4. Proofs for Sections 2 and 3.** To avoid repetition we attempt to cover both the tapered estimate, $\widehat{\alpha}_h$, and the untapered one, $\widehat{\alpha}$, simultaneously in the proofs, for brevity denoting both $\widehat{\alpha}$. Likewise, except in Section 7, we use $R(\alpha), U_m, I(\lambda), \ell, \theta, l$, to denote, respectively, $R_h(\alpha), U_m^h, I_h(\lambda), 3, \theta_3, l$ in the tapered case, and $R(\alpha), U_m, I(\lambda), 1, \theta_1, 1$ in the untapered case. We also introduce

$$\kappa_{m,l} = \log^4 m \mathbb{1}_{\{\ell=1\}} + l^{-1/2} \log^2 m \mathbb{1}_{\{\ell=3\}}, \tag{4.1}$$

meaning that we have $\kappa_{m,l} = l^{-1/2} \log^2 m$ with tapering and $\kappa_{m,l} = \log^4 m$ without tapering, and the remainder terms

$$\Delta_m = \max(m^{-1/2}, (m/n)^{\beta}), \qquad \widetilde{\Delta}_m = (m/n)^{\beta} + m^{-1/2} + m^{-1/2}\kappa_{m,l}, \tag{4.2}$$

$\widetilde{\Delta}_m$ being the remainder in our final expansions and $\Delta_m$ $(= O(\widetilde{\Delta}_m))$ that in our auxiliary expansion. Note that $\Delta_m = m^{-1/2}$ when $m = O(n^{2\beta/(2\beta+1)})$ (as in Theorem 2.1) and $\Delta_m = (m/n)^{\beta}$ otherwise. Throughout, $C$ denotes a generic, arbitrarily large constant.

By the mean value theorem,

$$\begin{aligned} R^{(1)}(\widehat{\alpha}) &= R^{(1)}(\alpha_0) + (\widehat{\alpha} - \alpha_0) R^{(2)}(\alpha_0) \\ &\quad + \frac{(\widehat{\alpha} - \alpha_0)^2}{2} R^{(3)}(\alpha_0) + \frac{(\widehat{\alpha} - \alpha_0)^3}{3!} R^{(4)}(\bar{\alpha}), \end{aligned} \tag{4.3}$$

where

$$R^{(j)}(\alpha) = \frac{d^j}{d\alpha^j} R(\alpha).$$

Writing $S_k(\alpha) = \frac{1}{c_0 m} \sum_{j=l}^m \nu_j^k \lambda_{\ell j}^{\alpha} I(\lambda_{\ell j})$ [cf. (3.7)], with $\nu_j = \nu_{j,m}$ [see (3.8)],

$$R^{(1)}(\alpha) = \frac{(d/d\alpha) S_0(\alpha)}{S_0(\alpha)} - \widetilde{m} = \frac{\sum_{j=l}^m \nu_j j^{\alpha} I(\lambda_{\ell j})}{\sum_{j=l}^m j^{\alpha} I(\lambda_{\ell j})},$$

where $\widetilde{m} = (m - l + 1)^{-1} \sum_{j=l}^m \log j$. Then with $S_k = S_k(\alpha_0)$, $R^{(k)} = R^{(k)}(\alpha_0)$, we have

$$R^{(1)} = \frac{S_1}{S_0}, \qquad R^{(2)} = \frac{S_2 S_0 - S_1^2}{S_0^2}, \qquad R^{(3)} = \frac{S_3 S_0^2 - 3 S_2 S_1 S_0 + 2 S_1^3}{S_0^3}.$$

Note that under the above assumptions $P(S_0 > 0) = 1$.

REMARK 4.1. $R^{(1)}$, $R^{(2)}$ and $R^{(3)}$ are invariant with respect to the scaling constant in $S_k$. Therefore, without loss of generality, we can replace (2.6) in the proofs below by

$$g(\lambda) = 1 + \frac{c_1}{c_0} |\lambda|^{\beta} + o(|\lambda|^{\beta}) \tag{4.4}$$

and (2.9) by

$$g(\lambda) = 1 + \frac{c_1}{c_0} |\lambda|^2 + \frac{c_2}{c_0} |\lambda|^4 + o(|\lambda|^4). \tag{4.5}$$

PROOF OF THEOREM 2.1. Define

$$Z_j = m^{1/2}(S_j - ES_j), \qquad j = 0, 1, 2, \dots. \tag{4.6}$$

By Lemma 5.6 we have

$$U_m = -B_m + V_m + \widetilde{\Delta}_m^{1+\delta}\xi_m,$$

where

$$B_m = m^{1/2}ES_1(2 - ES_2) - m^{1/2}(ES_1)^2 \tag{4.7}$$

and

$$V_m = -Z_1(2 - ES_2) + \frac{Z_1Z_2 + Z_1^2}{m^{1/2}} + (2Z_1 + Z_2)ES_1, \tag{4.8}$$

where $\delta > 0$ and $\xi_m$ denotes a remainder term. Set $V_m' = V_m + \widetilde{\Delta}_m^{1+\delta}$. Thus

$$P(U_m \le y) = P(V_m' \le y + B_m). \tag{4.9}$$

By Lemma 6.3,

$$\sup_{y \in R}\left|P(V_m' \le y) - \Phi(y) - m^{-1/2}\phi(y)p(y)\right| = o(\widetilde{\Delta}_m). \tag{4.10}$$

It remains to derive an expansion for $B_m$. By Lemma 7.1, bearing in mind that $(m/n)^\beta = O(m^{-1/2})$, we have $ES_1 = \theta(m/n)^\beta + o(m^{-1}) + O(m^{-1}\kappa_{m,l})$, $ES_2 = 1 + o(m^{-1/2})$, so that

$$m^{1/2}ES_1 = \theta q_m + o(q_m + m^{-1/2}) + O(m^{-1/2}\kappa_{m,l}),$$

$$m^{1/2}ES_1(2 - ES_2) = \theta q_m + o(q_m).$$

If $m = o(n^{2\beta/(2\beta+1)})$, then $q_m \to 0$, and

$$B_m = \theta q_m + o(q_m + m^{-1/2}) + O(m^{-1/2}\kappa_{m,l}). \tag{4.11}$$

If $m \sim Kn^{2\beta/(2\beta+1)}$, then $q_m \sim K^{\beta+1/2}$, and we obtain

$$B_m = K^{\beta+1/2}\theta + o(1). \tag{4.12}$$

Relations (4.9)–(4.12) imply (2.13) and (2.15) of Theorem 2.1. □

PROOF OF THEOREM 3.1. This follows the lines of that of Theorem 2.1. Relations (4.9) and (4.10) remain valid. Recall that $q_m = (m/n)^2 m^{1/2}$, and under Assumption $m$, $q_m = O(1)$. To expand $B_m$, we use (7.15) and (7.16) with $k = 2$ to deduce

$$B_m = \theta q_m + \gamma q_m (m/n)^2 + o(m^{-1/2}) + O(m^{-1/2}\kappa_{m,l}), \tag{4.13}$$

with

$$\theta = e(1,\ell,2) = \left(\frac{c_1}{c_0}\right)^2 \frac{2}{9}\left(\frac{\ell}{2\pi}\right)^2,$$

$$\gamma = d(1,\ell,4) - e(1,\ell,2)e(2,\ell,2) - e^2(1,\ell,2)$$

$$= \left(\frac{c_2}{c_0}\frac{4}{25} - \left(\frac{c_1}{c_0}\right)^2 \frac{22}{243}\right)\left(\frac{\ell}{2\pi}\right)^4,$$

where $e(k,\ell,\beta)$ and $d(k,\ell,\beta)$ are defined in (7.5) and (7.17). If $m = o(n^{4/5})$, then $q_m = o(1)$ so that $B_m = \theta q_m + o(m^{-1/2}) + O(m^{-1/2}\kappa_{m,l})$, and (3.13) follows from (4.9) and (4.10). If $m \sim K n^{4/5}$, then $(m/n)^2 \sim K^{5/2} m^{-1/2}$ and thus $B_m = \theta K^{5/2} + m^{-1/2} K^5 \gamma + o(m^{-1/2}) + O(m^{-1/2}\kappa_{m,l})$. Therefore from (4.9) and (4.10), it follows that

$$\sup_{y\in R}\Big|P(U_m \le y) - \Phi(y + \theta K^{5/2} + m^{-1/2}K^5\gamma)$$
$$- m^{-1/2}\phi(y+\theta K^{5/2})p(y+\theta K^{5/2})\Big|$$
$$= o(m^{-1/2}) + O(m^{-1/2}\kappa_{m,l}),$$

which implies (3.15). □

Denote by $\mathcal{X}$ the set of all sequences $\xi_m$ satisfying

$$P(|\xi_m| \ge m^{\varepsilon}) = o(m^{-p}) \qquad \text{all } \varepsilon > 0, \text{ all } p \ge 1. \tag{4.14}$$

Note that $\xi_m \in \mathcal{X}$, $\eta_m \in \mathcal{X}$ implies $\xi_m\eta_m \in \mathcal{X}$. For ease of exposition we denote by $\xi_m$ a generic member of $\mathcal{X}$.

PROOF OF LEMMA 3.1. Set $p_m = \widehat{\theta}_{\ell,1} - \theta = (n/m')^{\beta} R_{1,m'}(\widehat{\alpha}) - \theta$. [Here and below we index some quantities by $m$ even if they depend on $m'$ also, noting from (3.11) that $m'$ depends on $m$.]

By the Borel–Cantelli lemma it suffices to show that, for all $\delta > 0$,

$$\sum_{m=1}^{\infty} P\{|p_m| \ge \delta\} < \infty. \tag{4.15}$$

We show that

$$p_m = o(1) + m^{-\varepsilon}\xi_m \qquad \text{a.s.} \tag{4.16}$$

for some $\varepsilon > 0$ where $\xi_m \in \mathcal{X}$. Then

$$P\{|p_m| \ge \delta\} \le P\{o(1) \ge \delta/2\} + P\{m^{-\varepsilon}\xi_m \ge \delta/2\} = o(m^{-2})$$

by (4.14), and thus (4.15) holds. By the mean value theorem, we have

$$R^{(1)}_{m'}(\widehat{\alpha}) = R^{(1)}_{m'}(\alpha_0) + (\widehat{\alpha} - \alpha_0)\frac{d}{d\alpha}R^{(1)}_{m'}(\bar{\alpha}),$$

where $|\bar{\alpha}-\alpha_0| \le |\widehat{\alpha}-\alpha_0|$. We show that

$$p_{1,m} := \left|(n/m')^{\beta} R^{(1)}_{m'}(\alpha_0) - \theta\right| = o(1) + m^{-\varepsilon}\xi_m, \tag{4.17}$$

$$p_{2,m} := (n/m')^{\beta}\left|(\widehat{\alpha}-\alpha_0)\frac{d}{d\alpha}R^{(1)}_{m'}(\bar{\alpha})\right| = m^{-\varepsilon}\xi_m, \tag{4.18}$$

which yields (4.16). To prove (4.17), note that, writing $Z_{i,m'}(\alpha) = m'^{1/2}(S_{i,m'}(\alpha) - ES_{i,m'}(\alpha))$,

$$S_{1,m'}(\alpha_0) = ES_{1,m'}(\alpha_0) + m'^{-1/2} Z_{1,m'}(\alpha_0) = \theta(m'/n)^{\beta} + o\big((m'/n)^{\beta}\big) + m'^{-1/2}\xi_m$$

by Lemmas 5.3 and 7.1. Observe that (3.11) implies

$$(n/m')^{\beta} m^{-1/2} \le m^{-\varepsilon} \tag{4.19}$$

for some $\varepsilon > 0$. Thus

$$(n/m')^{\beta} S_{1,m'}(\alpha_0) = \theta + o(1) + m^{-\varepsilon}\xi_m.$$

Applying Lemma 5.4 to $S_{0,m'}(\alpha_0)^{-1}$, we get

$$S_{0,m'}(\alpha_0)^{-1} = 1 + O\big((m'/n)^{\beta}\big) + m'^{-1/2}\xi_{m'} = 1 + O\big((m'/n)^{\beta}\big) + m'^{-1/2}\xi_m.$$

Thus

$$\begin{aligned}(n/m')^{\beta} R^{(1)}_{m'}(\alpha_0) &= (n/m')^{\beta} S_{1,m'}(\alpha_0) S_{0,m'}(\alpha_0)^{-1} \\ &= \big(\theta + o(1) + m^{-\varepsilon}\xi_m\big)\big(1 + O(m'/n)^{\beta} + m'^{-\varepsilon}\xi_m\big) \\ &= \theta + o(1) + m^{-\varepsilon}\xi_m.\end{aligned}$$

Hence $p_{1,m} = o(1) + m^{-\varepsilon}\xi_m$ and (4.17) holds.

To prove (4.18), note that $|\frac{d}{d\alpha}R^{(1)}_{m'}(\bar{\alpha})| \le C\log^2 m'$ [see the proof of (5.12)]. Then, by (4.19),

$$p_{2,m} \le C(n/m')^{\beta} m^{-1/2}|U_m|\log^2 m' \le Cm^{-\varepsilon}\xi_m,$$

since by Lemma 5.7, $U_m \log^2 m' = m^{1/2}\Delta_m \xi_m \log^2 m' = \xi_m \log^2 m' \in \mathcal{X}$, bearing in mind that under Assumption $m$, $\Delta_m = O(m^{-1/2})$. Thus (4.18) holds and the proof for $\widehat{\theta}_{\ell,1}$ is complete.

The proof for $\widehat{\theta}_{\ell,2}$ follows on showing that $d_{i,m}(\widehat{\alpha}) = \widehat{c}_i - c_i \to 0$ a.s., $i = 1, 2$. As before, it suffices to show that, for all $\delta > 0$,

$$\sum_{m=1}^{\infty} P\{|d_{i,m}(\widehat{\alpha})| \ge \delta\} < \infty, \qquad i = 0, 1. \tag{4.20}$$

By Lemma 5.8 with $k = p = 2$, $P\{|\widehat{\alpha}-\alpha_0| \ge (\log n)^{-2}\} = o(m^{-2})$, it suffices to prove (4.20) in case $|\widehat{\alpha}-\alpha_0| \le (\log n)^{-2}$. Similarly to the proof of Lemma 3.1 it remains to show that

$$|d_{i,m}(\widehat{\alpha})| = o(1) + m^{-\varepsilon}\xi_m, \qquad i = 1, 2, \tag{4.21}$$

for some $\varepsilon > 0$ where $\xi_m \in \mathcal{X}$. By the mean value theorem,

$$d_{i,m}(\widehat{\alpha}) = d_{i,m}(\alpha_0) + (\widehat{\alpha} - \alpha_0)\frac{d}{d\alpha}d_{i,m}(\bar{\alpha}),$$

where $|\bar{\alpha} - \alpha_0| \le |\widehat{\alpha} - \alpha_0] \le (\log m)^{-2}$. We show that, as $n \to \infty$ for $i = 1, 2$,

$$Ed_{i,m}(\alpha_0) = o(1), \tag{4.22}$$

$$p'_{i,m} := d_{i,m}(\alpha_0) - Ed_{i,m}(\alpha_0) = m^{-\varepsilon}\xi_m, \tag{4.23}$$

$$p''_{i,m} := |\widehat{\alpha} - \alpha_0|\left|\frac{d}{d\alpha}d_{i,m}(\bar{\alpha})\right| = m^{-\varepsilon}\xi_m, \tag{4.24}$$

which yield (4.21).

First, (4.22) follows by approximating sums by integrals and Lemma 7.1. We have

$$|p'_{1,m}| \le C\bigl(m'^{-1/2}|Z_{0,m'}(\alpha_0)| + (m'/n)^{-\beta}m'^{-1/2}|Z_{0,m'}(\alpha_0+\beta)|\bigr),$$

$$p'_{2,m} \le C\bigl((m'/n)^{-\beta}m'^{-1/2}|Z_{0,m'}(\alpha_0)| + (m'/n)^{-2\beta}m'^{-1/2}|Z_{0,m'}(\alpha_0+\beta)|\bigr).$$

By (3.11), $(m'/n)^{-\beta}m'^{-1/2} \le m^{-\varepsilon}$, and by Lemma 5.3, $Z_{0,m'}(\alpha_0) \in \mathcal{X}$. Using Lemma 7.3, it is easy to show that $E|(m'/n)^{-\beta}Z_{0,m'}(\alpha_0+\beta)|^k < \infty$ as $n \to \infty$ for any $k \ge 1$, so $(m'/n)^{-\beta}Z_{0,m'}(\alpha_0+\beta) \in \mathcal{X}$, to imply (4.23). It remains to show (4.24). Since $|\bar{\alpha} - \alpha_0| \le \log^{-2} m$, it is easy to see that

$$\left|\frac{d}{d\alpha}S_{0,m'}(\bar{\alpha})\right| \le C(\log m)S_{0,m'}(\alpha_0),$$

$$\left|(m'/n)^{-\beta}\frac{d}{d\alpha}S_{0,m'}(\bar{\alpha}+\beta)\right| \le C(\log m)S_{0,m'}(\alpha_0).$$

Thus

$$\begin{aligned} p''_{i,m} &\le C|\widehat{\alpha}_m - \alpha_0|(m'/n)^{-\beta}(\log m)S_{0,m'}(\alpha_0) \\ &\le C(m'/n)^{-\beta}m^{-1/2}|U_m|(\log m)S_{0,m'}(\alpha_0). \end{aligned}$$

Since under Assumption $m$, $\Delta_m = O(m^{-1/2})$, from (4.19) and Lemma 5.7 it follows that

$$(m'/n)^{-\beta}m^{-1/2}|U_m|\log m = m^{-\varepsilon}\xi_m \log m \in \mathcal{X}.$$

This and $S_{0,m'}(\alpha_0) \in \mathcal{X}$ imply that $p''_{i,m} = m^{-\varepsilon}\xi_m$. □

PROOF OF THEOREM 3.2. By Lemma 5.6,

$$U_m = -B_m + V_m + \widetilde{\Delta}_m^{1+\delta}\xi_m. \tag{4.25}$$

By (4.13),

$$B_m = \theta q_m + o(k_m + m^{-1/2}) + O(m^{-1/2}\kappa_{m,l}), \tag{4.26}$$

since $q_m(m/n)^2 = k_m v_m = o(k_m)$. This and Lemma 4.1 give

$$\begin{aligned} m^{1/2}(\alpha^* - \alpha_0) &\equiv U_m + q_m\widehat{\theta}^* \\ &= -(\theta a - b)k_m + V_m - v_m Z_1 \\ &\quad + o(r_m) + O(m^{-1/2}\kappa_{m,l}) + (\widetilde{\Delta}_m + v_m)^{1+\delta}\xi_m, \end{aligned}$$

where $\delta > 0$, writing $\widehat{\theta}^* = \widehat{\theta}_3^*$. This and Lemma 6.4 imply (3.20). □

LEMMA 4.1. *Suppose that the assumptions of Theorem* 3.2 *hold. Then*

$$\widehat{\theta}^* q_m = \theta q_m - (\theta a - b)k_m - Z_1 v_m + o(r_m) + v_m^{1+\delta}\xi_m \tag{4.27}$$

*for some* $\delta > 0$, $\xi_m \in \mathcal{X}$.

PROOF. Set $j_m = (1 - v_m)q_m\widehat{\theta}^*$ and $t_m = (m/n)^2$. Then $q_m = m^{1/2}t_m$. We show that

$$\begin{aligned} j_m &= \theta q_m(1 - v_m) - (\theta a - b)k_m - Z_1 v_m \\ &\quad + o(k_m + v_m + m^{-1/2}) + v_m^{1+\delta}\xi_m, \end{aligned} \tag{4.28}$$

where $\theta_\ell, a_\ell, b_\ell$ are given in (3.12) and (3.21). Dividing both sides of (4.28) by $1 - v_m$ and taking into account that $v_m \to 0$ gives (4.27). By Taylor's theorem,

$$\begin{aligned} j_m &= \sqrt{m} v_m R_{m'}^{(1)}(\widehat{\alpha}) \\ &= \sqrt{m} v_m \biggl( R_{m'}^{(1)}(\alpha_0) + (\widehat{\alpha} - \alpha_0)R_{m'}^{(2)}(\alpha_0) + \frac{(\widehat{\alpha} - \alpha_0)^2}{2!} R_{m'}^{(3)}(\bar{\alpha}) \biggr). \end{aligned}$$

Thus

$$j_m = d_{1,m} + d_{2,m} + d_{3,m},$$

where

$$\begin{aligned} d_{1,m} &= m^{1/2} v_m R_{m'}^{(1)}(\alpha_0), \qquad d_{2,m} = m^{1/2} v_m (\widehat{\alpha} - \alpha_0) R_{m'}^{(2)}(\alpha_0), \\ d_{3,m} &= m^{1/2} v_m \frac{(\widehat{\alpha} - \alpha_0)^2}{2!} R_{m'}^{(3)}(\bar{\alpha}). \end{aligned}$$

We shall show that

$$d_{1,m} = \theta\sqrt{m}t_m + (v - \theta a)k_m + o(r_m) + v_m^{1+\delta}\xi_m, \tag{4.29}$$

$$d_{2,m} = -\theta\sqrt{m}t_m v_m - Z_1 v_m + o(r_m) + v_m^{1+\delta}\xi_m, \tag{4.30}$$

$$d_{3,m} = v_m^{1+\delta}\xi_m \tag{4.31}$$

for some $\delta > 0$ where $\theta = e(1, \ell, 2),\ v = d(1, \ell, 4),\ a = e(0, \ell, 2)$ are given by (7.5) and (7.17); then (4.28) follows.

Note that $d_{1,m} = m^{1/2}(m/m')^2 S_{1,m'}(\alpha_0) S_{0,m'}(\alpha_0)^{-1}$. We have

$$S_{1,m'}(\alpha_0) = ES_{1,m'}(\alpha_0) + m'^{-1/2} Z_{1,m'}(\alpha_0) = \theta t_{m'} + v t_{m'}^2 + o(t_{m'}^2) + m'^{-1/2}\xi_m$$

since $Z_{1,m'}(\alpha_0) \in \mathcal{X}$ by Lemma 5.3, and Lemma 7.2 implies that

$$ES_{1,m'}(\alpha_0) = \theta t_{m'} + v t_{m'}^2 + o(t_{m'}^2 + m'^{-1/2}). \tag{4.32}$$

Lemma 5.4 applied to $S_{0,m'}(\alpha_0)^{-1}$ implies that

$$S_{0,m'}(\alpha_0)^{-1} = 2 - S_{0,m'}(\alpha_0) + o(t_{m'}) + m'^{-1/2}\xi_m,$$

whereas by Lemmas 7.2 and 5.3,

$$S_{0,m'}(\alpha_0) = ES_{0,m'}(\alpha_0) + m'^{-1/2} Z_{0,m'}(\alpha_0) = 1 + a t_{m'} + o(t_{m'}) + m'^{-1/2}\xi_m.$$

Since $(m/m')^2 t_{m'} = t_m$, we get

$$\begin{aligned} d_{1,m} &= m^{1/2}(m/m')^2\big(\theta t_{m'} + v t_{m'}^2 + o(t_{m'}^2) + m'^{-1/2}\xi_m\big) \\ &\quad \times \big(1 - a t_{m'} + o(t_{m'}) + m'^{-1/2}\xi_m\big) \\ &= \theta\sqrt{m} t_m + (v - \theta a)\sqrt{m} t_m t_{m'} + o(k_m) + v_m^{1+\delta}\xi_m \end{aligned}$$

for some $\delta > 0$. Thus (4.29) holds.

We now bound $d_{2,m}$. By Lemmas 5.4 and 7.2,

$$\begin{aligned} R_{m'}^{(2)}(\alpha_0) &\equiv \big(S_{2,m'}(\alpha_0) S_{0,m'}(\alpha_0) - S_{1,m'}(\alpha_0)^2\big) S_{0,m'}(\alpha_0)^{-2} \\ &= 1 + O\big((m'/n)^\beta + m'^{-1/2}\big)\xi_m, \end{aligned}$$

and by (4.25) and (4.26),

$$U_m = \sqrt{m}(\widehat{\alpha} - \alpha_0) = -\theta\sqrt{m} t_m + V_m + o(r_m) + O(m^{-1/2}\kappa_{m,l}) + \Delta_m^{1+\delta}\xi_m.$$

From (4.8), Lemma 5.3 and (4.32) it follows that $V_m = -Z_1 + \widetilde{\Delta}_m \xi_m$. Thus $U_m = -\theta\sqrt{m} t_m - Z_1 + o(r_m) + O(m^{-1/2}\kappa_{m,l}) + \widetilde{\Delta}_m\xi_m$, and

$$\begin{aligned} d_{2,m} &= v_m\big(-\theta\sqrt{m} t_m - Z_1 + o(r_m) + O(m^{-1/2}\kappa_{m,l}) + \widetilde{\Delta}_m\xi_m\big)(1 + \widetilde{\Delta}_m\xi_m) \\ &= -\theta\sqrt{m} t_m v_m - v_m Z_1 + o(k_m + v_m) + v_m^{1+\delta}\xi_m \end{aligned}$$

since $\widetilde{\Delta}_m \le v_m^\eta$ for some $\delta > 0$. This proves (4.30).

Finally, note that

$$|d_{3,m}| \le v_m m^{-1/2} U_m^2 \big|R_{m'}^{(3)}(\bar{\alpha})\big| = v_m m^{1/2} \Delta_m^2 \xi_m^2 \big|R_{m'}^{(3)}(\bar{\alpha})\big|$$

by Lemma 5.7. Similarly to (5.12) we can show that $|R_{m'}^{(3)}(\bar{\alpha})| \le C(\log m')^4$. Since under (3.18) $m^{1/2}\Delta_m^2 \le v_m^\delta$ for some $\delta > 0$, we get $|d_{3,m}| \le v_m^{1+\delta}\xi_m$. Thus (4.31) holds. $\square$

**5. Approximation lemmas.** To characterize negligible terms of our expansions, we shall use the following version of Chibisov's (1972) Theorem 2, which we present without proof.

LEMMA 5.1. *Let* $Y_m = V_m + \delta_m^2 \xi_m$, *where* $\delta_m \to 0$ *as* $m \to 0$,

$$P(|\xi_m| \geq \delta_m^{-\varepsilon}) = o(\delta_m) \qquad \text{for some } 0 < \varepsilon < 1, \tag{5.1}$$

*and* $V_m$ *have the asymptotic expansion*

$$P(V_m \leq y) = \Phi(y) - \phi(y)\big(\delta_{m,1} p_1(y) + \delta_{m,2} p_2(y)\big) + o(\delta_m) \tag{5.2}$$

*uniformly in* $y \in R$ *where* $p_1(y), p_2(y)$ *are polynomials and* $\delta_{m,i} = O(\delta_m)$, $i = 1, 2$. *Then*

$$P(Y_m \leq y) = \Phi(y) - \phi(y)\big(\delta_{m,1} p_1(y) + \delta_{m,2} p_2(y)\big) + o(\delta_m) \tag{5.3}$$

*uniformly in* $y \in R$.

We shall use the following corollary of Lemma 5.1 for the remainder terms $\xi_m \in \mathcal{X}$ defined in Section 4.

LEMMA 5.2. *Suppose that* $\delta_m^{-1} \geq m^{\varepsilon}$ *and* $Y_m = V_m + \delta_m^{1+\varepsilon'} \xi_m$ *for some* $\varepsilon > 0$, $\varepsilon' > 0$ *as* $m \to 0$, *and* $\xi_m \in \mathcal{X}$. *Then* (5.2) *implies* (5.3).

We first discuss properties of the sums $S_i$, $Z_i$ for a wider range of $m$ than in Assumption $m$.

ASSUMPTION $m^*$. $m \geq l$ is such that $n^{\varepsilon} \leq m \leq n^{1-\varepsilon}$ for some $\varepsilon > 0$.

LEMMA 5.3. *Suppose that Assumptions* $f, l$ *and* $m^*$ *hold. Then for* $Z_j$ *given by* (4.6) *and any* $j = 0, 1, 2, \ldots$,

$$Z_j \in \mathcal{X}. \tag{5.4}$$

PROOF. By Lemma 7.4, for any $k \geq 1$, $E|Z_j|^k < \infty$ uniformly in $m \to \infty$. Thus (5.4) follows by the Chebyshev inequality. □

We consider following lemma, which is related to the lemma of Robinson (1995c) and concerns general functions $Y_m = f_m(S_0, S_1, \ldots, S_3)$ of the variables $S_k$ (3.7).

LEMMA 5.4. *Let* (4.4) *and Assumptions* $f, l$ *and* $m^*$ *hold. Let* $Y_m = f_m(S_0, S_1, \ldots, S_3)$ *for a function* $f_m$ *such that for some* $\delta > 0$ *the partial* *derivatives* $(\partial^2/\partial x_i x_j) f_m(x_0, \dots, x_3)$, $i, j = 0, \dots, 3$, *are bounded in* $|x_j - e_j| \le \delta$, $j = 0, \dots, 3$. *Then as* $n \to \infty$,

$$Y_m = f_m(e_0, e_1, \dots, e_3) + \sum_{j=0}^{3} \frac{\partial}{\partial x_j} f_m(e_0, e_1, \dots, e_3)(S_j - e_j) + O\big((m/n)^{2\beta}\big) + m^{-1}\xi_m, \tag{5.5}$$

$$Y_m = f_m(e_0, e_1, \dots, e_3) + O\big((m/n)^{\beta}\big) + m^{-1/2}\xi_m \tag{5.6}$$

*and*

$$Y_m \in \mathcal{X}, \tag{5.7}$$

*where* $\xi_m \in \mathcal{X}$.

PROOF. Observe that for any $j = 0, \dots, 3$, we can write $Y_m \mathbb{1}(m^{-1/2}|Z_j| > \delta/2) \equiv m^{-1}\xi_m$ where $\xi_m \in \mathcal{X}$. Indeed, for all $p \ge 1$,

$$P\{|\xi_m| \ge m^{\varepsilon}\} \le P\{m^{-1/2}|Z_j| > \delta/2\} = o(m^{-p}),$$

since $Z_i \in \mathcal{X}$. Therefore,

$$Y_m = f_m(S_0, S_1, \dots, S_3)\mathbb{1}(m^{-1/2}|Z_j| \le \delta/2, j = 0, \dots, 3) + m^{-1}\xi_m.$$

We have by Lemma 7.1 below that

$$E(S_i) = e_i + O\big((m/n)^{\beta} + m^{-1/2}\big), \qquad i = 0, 1, 2, 3, \tag{5.8}$$

where by (7.5), $e_0 = 1, e_1 = 0, e_2 = 1, e_3 = -2$. By (5.8),

$$S_j = ES_j + m^{-1/2}Z_j = e_j + O\big((m/n)^{\beta} + m^{-1/2}\big) + m^{-1/2}Z_j.$$

Thus if $|m^{-1/2}Z_j| \le \delta/2$ for any $j = 0, \dots, 3$ then $|S_j - e_j| \le \delta$ for large $m$, and by Taylor's theorem we get

$$Y_m = f_m(e_0, e_1, \dots, e_3) + \sum_{j=0}^{3} \frac{\partial}{\partial x_j} f_m(e_0, e_1, \dots, e_3)(S_j - e_j) + O\left(\sum_{j=0}^{3}(S_j - e_j)^2\right) + O\big((m/n)^{2\beta}\big) + m^{-1}\xi_m.$$

In view of (5.8) and (5.4),

$$\begin{aligned}(S_j - e_j)^2 &\le 2\big((ES_j - e_j)^2 + m^{-1}Z_j^2\big) = O\big((m/n)^{2\beta}\big) + 2m^{-1}Z_1^2 \\ &= O\big((m/n)^{2\beta}\big) + 2m^{-1}\xi_m\end{aligned} \tag{5.9}$$

since $Z_1 \in \mathcal{X}$. This proves (5.5). Equation (5.9) implies that $|S_j - e_j| = O((m/n)^{\beta}) + m^{-1/2}\xi_m$, and therefore from (5.5) it follows that (5.6) holds. Equation (5.6) implies (5.7). □

LEMMA 5.5. *Let* (4.4) *and Assumptions* $\alpha$, $f$, $l$ *and* $m^*$ *hold*, *and let*

$$m \le n^{4\beta/(4\beta+1)-\varepsilon} \qquad \text{for some } \varepsilon > 0. \tag{5.10}$$

*Then*

$$m^{1/2}R^{(1)}(\alpha_0) + U_m R^{(2)}(\alpha_0) + \frac{U_m^2}{2m^{1/2}}R^{(3)}(\alpha_0) = \Delta_m^{1+\delta}\xi_m \tag{5.11}$$

*for some* $\delta > 0$ *and* $\xi_m \in \mathcal{X}$.

PROOF. Multiplying both sides of (4.3) by $m^{1/2}$, the left-hand side of (5.11) can be written as $\Delta_m^{1+\delta}\xi_m$ with

$$\xi_m = O\big(m^{1/2}\Delta_m^{-1-\delta}|R^{(1)}(\widehat{\alpha})| + m^{-1}\Delta_m^{-1-\delta}|U_m|^3|R^{(4)}(\bar{\alpha})|\big),$$

where $|\bar{\alpha} - \alpha_0| \le |\widehat{\alpha} - \alpha_0|$. It remains to show that $\xi_m \in \mathcal{X}$. We show first that

$$|R^{(4)}(\bar{\alpha})| \le C \log^4 m. \tag{5.12}$$

Now $R^{(4)}(\bar{\alpha})$ is a linear combination of terms

$$\frac{F_0^{l_0}(\bar{\alpha})F_1^{l_1}(\bar{\alpha})\cdots F_4^{l_4}(\bar{\alpha})}{F_0^4(\bar{\alpha})}, \qquad l_0 + \cdots + l_4 = 4, 0 \le l_0, \dots, l_4 \le 4, \tag{5.13}$$

where

$$F_k(\alpha) = \frac{d^k}{d\alpha^k}F_0(\alpha), \qquad F_0(\alpha) = \sum_{j=l}^{m} j^\alpha I(\lambda_{\ell j}), \qquad k \ge 0.$$

Now

$$|F_k(\alpha)| \le \log^k m \sum_{j=1}^{m} j^\alpha I(\lambda_{\ell j}) = (\log m)^k F_0(\alpha).$$

Therefore the terms in (5.13) are bounded by $\log^4 m$, and (5.12) holds. By Lemma 5.7, $U_m = m^{1/2}\Delta_m\widetilde{\xi}_m$ where $\widetilde{\xi}_m \in \mathcal{X}$. Assumption (5.10) implies that

$$m^{1/2}\Delta_m^2 \le \Delta_m^\delta \qquad \text{for some } \delta > 0. \tag{5.14}$$

Therefore

$$\begin{aligned}\xi_m &\le C\big(m^{1/2}\Delta_m^{-1-\delta}|R^{(1)}(\widehat{\alpha})| + m^{1/2}\Delta_m^{2-\delta}\widetilde{\xi}_m^3(\log m)^4\big)\\ &\le C\big(m^{1/2}\Delta_m^{-1-\delta}|R^{(1)}(\widehat{\alpha})| + \widetilde{\xi}_m^3(\log m)^4\big).\end{aligned}$$

Thus, for large $m$,

$$\begin{aligned}P\{|\xi_m| \ge m^\varepsilon\} &\le P\{|R^{(1)}(\widehat{\alpha})| > 0\} + P\{\widetilde{\xi}_m \ge m^{\varepsilon/8}\}\\ &\le P\{\widehat{\alpha} = \pm 1\} + P\{\widetilde{\xi}_m \ge m^{\varepsilon/8}\} = o(m^{-p})\end{aligned}$$

for all $p \ge 1$, since $P\{\widehat{\alpha} = \pm 1\} = o(m^{-p})$ by Lemma 5.8 below, using the fact that $R^{(1)}(\widehat{\alpha}) = 0$ if $\widehat{\alpha} \in (-1, 1)$ and $P\{\widetilde{\xi}_m \ge m^{\varepsilon/8}\} = o(m^{-p})$ by $\xi_m \in \mathcal{X}$. □

LEMMA 5.6. *Let* (4.4), (5.10) *and Assumptions* $\alpha$, $f$, $l$ *and* $m^*$ *hold. Then for some* $\delta > 0$,

$$U_m = -B_m + V_m + \Delta_m^{1+\delta}\xi_m, \tag{5.15}$$

*where* $B_m$ *and* $V_m$ *are given by* (4.7) *and* (4.8), *and* $\xi_m \in \mathcal{X}$.

PROOF. We deduce from (5.11) that

$$U_m = -m^{1/2}R^{(1)} - U_m(R^{(2)} - 1) - m^{-1/2}U_m^2 R^{(3)}/2 + \Delta_m^{1+\delta}\xi_m.$$

By definition of $R^{(1)}$, $R^{(2)}$ and $R^{(3)}$, we can write

$$U_m = -m^{1/2}S_1 h(S_0) - U_m f(S_0, S_1, S_2) - m^{-1/2}U_m^2 g(S_0, S_1, S_2, S_3) + \Delta_m^{1+\delta}\xi_m,$$

where

$$h(x_0) = x_0^{-1},$$

$$f(x_0, x_1, x_2) = \frac{x_2 x_0 - x_1^2}{x_0^2} - 1,$$

$$g(x_0, x_1, x_2, x_3) = \frac{x_3 x_0^2 - 3x_0 x_1 x_2 + 2x_1^3}{2x_0^3}.$$

We apply now Lemma 5.4. Since $h(e_0) = h(1) = 1$, $(\partial/\partial x_0)h(e_0) = -1$, we get by (5.5),

$$h(S_0) = 1 - (S_0 - 1) + m^{-1}\xi_m = 2 - S_0 + \Delta_m^2\xi_m.$$

Similarly, since

$$f(e_0, e_1, e_2) = f(1, 0, 1) = 0,$$

$$\frac{\partial}{\partial x_0} f(1, 0, 1) = -1, \qquad \frac{\partial}{\partial x_1} f(1, 0, 1) = 0, \qquad \frac{\partial}{\partial x_2} f(1, 0, 1) = 1,$$

by (5.5) of Lemma 5.4,

$$f(S_0, S_1, S_2) = (S_0 - 1) - (S_2 - 1) + m^{-1}\xi_m = S_2 - S_0 + \Delta_m^2\xi_m.$$

Finally, since $g(e_0, e_1, e_2, e_3) = g(1, 0, 1, -2) = -1$ and $g(e_0, e_1, e_2, e_4) = 1$, by (5.6) of Lemma 5.4, we have $g(S_0, S_1, S_2, S_3) = -1 + \Delta_m\xi_m$. Thus

$$U_m = -\sqrt{m}S_1(2 - S_0) - U_m(S_2 - S_0) + m^{-1/2}U_m^2 + y_m + \Delta_m^{1+\delta}\xi_m, \tag{5.16}$$

where $|y_m| \le \Delta_m^2\xi_m(\sqrt{m}|S_1| + |U_m| + m^{-1/2}|U_m^2|)$. Relations (5.8) and (5.4) imply that

$$|S_1| \le C\Delta_m + m^{-1/2}|Z_1| = \Delta_m\xi_m, \tag{5.17}$$

where $\xi_m \in \mathcal{X}$. By Lemma 5.7, $U_m = m^{1/2}\Delta_m^2\xi_m$. Thus $|y_m| = m^{1/2}\Delta_m^3\xi_m$ and, using (5.14), $|y_m| = \Delta_m^{1+\delta}\xi_m$. Next, by (5.8) and Lemma 5.3,

$$S_2 - S_0 = (ES_2 - ES_0) + m^{-1/2}(Z_2 - Z_0) = O(\Delta_m) + m^{-1/2}(Z_2 - Z_0) = \Delta_m\xi_m,$$

and $S_1(2 - S_0) = \Delta_m\xi_m$, $S_1(1 - S_0) = \Delta_m^2\xi_m$ where $\xi_m \in \mathcal{X}$. Thus, repeatedly applying the recurrence relation (5.16) and taking into account (5.14), we get

$$\begin{aligned} U_m &= -\sqrt{m}S_1(2 - S_0) + \sqrt{m}S_1(S_2 - S_0) + m^{-1/2}(\sqrt{m}S_1)^2 + \Delta_m^{1+\delta}\xi_m \\ &= -\sqrt{m}S_1(2 - S_2) + m^{-1/2}(\sqrt{m}S_1)^2 + \Delta_m^{1+\delta}\xi_m \\ &= -\sqrt{m}ES_1(2 - ES_2) - Z_1(2 - ES_2) + Z_2ES_1 + m^{-1/2}Z_1Z_2 \\ &\quad + \big(m^{1/2}(ES_1)^2 + 2Z_1ES_1 + m^{-1/2}Z_1^2\big) + \Delta_m^{1+\delta}\xi_m \\ &= -B_m + V_m + \Delta_m^{1+\delta}\xi_m. \end{aligned}$$

This completes the proof of (5.15). □

LEMMA 5.7. *Let* (4.4) *and Assumptions* $\alpha$, $f$, $l$ *and* $m^*$ *hold. Then*

$$U_m = m^{1/2}\Delta_m\xi_m, \tag{5.18}$$

*where* $\xi_m \in \mathcal{X}$.

PROOF. By Lemma 5.8, $U_m 1(|\widehat{\alpha} - \alpha_0| > \log^{-4} n) = \xi_m \in \mathcal{X}$. It remains to show that $U_m(|\widehat{\alpha} - \alpha_0| \le \log^{-4} n) = \xi_m \in \mathcal{X}$. Let $|\widehat{\alpha} - \alpha_0| \le \log^{-4} n$. Then as $n \to \infty$, $\widehat{\alpha} \in (-1, 1)$ and, consequently, $R^{(1)}(\widehat{\alpha}) = 0$, so that

$$0 = m^{1/2}R^{(1)} + U_mR^{(2)} + \frac{U_m(\widehat{\alpha} - \alpha_0)}{2}R^{(3)}(\bar{\alpha})$$

for $|\bar{\alpha} - \alpha_0| \le |\widehat{\alpha} - \alpha_0|$. Similarly to the proof of (5.12) it can be shown that $|R^{(3)}(\bar{\alpha})| \le C(\log n)^3$, so that $|\widehat{\alpha} - \alpha_0||R^{(3)}(\bar{\alpha})| \le \log^{-1} n$ and

$$0 = m^{1/2}R^{(1)} + U_m\big\{R^{(2)} + O(\log^{-1} n)\big\}.$$

Thus

$$U_m = -m^{1/2}R^{(1)}\big\{R^{(2)} + O(\log^{-1} n)\big\}^{-1} = -m^{1/2}S_1f_n(S_0, S_1, S_2),$$

where, by definition of $R^{(1)}$, $R^{(2)}$,

$$f_n(x_0, x_1, x_1) = x_0^{-1}\left(\frac{x_2x_0 - x_1^2}{x_0^2} + O(\log^{-1} n)\right)^{-1}.$$

Since

$$f_n(e_0, e_1, e_2) = f_n(1, 0, 1) = \big(1 + O(\log^{-1} n)\big)^{-1} < \infty,$$

Lemma 5.4 implies that $f_n(S_0, S_1, S_2) = \xi_m \in \mathcal{X}$, whereas by (5.17), $S_1 = \Delta_m\xi_m$, to give (5.18). □

LEMMA 5.8. *Let* (4.4) *and Assumptions* $\alpha$, $f$, $l$ *and* $m^*$ *hold. Then for any* $s \geq 1$,

$$P\{|\widehat{\alpha} - \alpha_0| \geq (\log n)^{-s}\} = o(m^{-p}),$$

*for all* $p \geq 1$.

PROOF. Let $\varepsilon > 0$ be arbitrarily small. Set $F_1 = \{\alpha \in [-1, 1] : (\log n)^{-s} \leq |\alpha - \alpha_0|, \alpha_0 - \alpha \leq 1 - \varepsilon\}$, $F_2 = \{1 - \varepsilon \leq \alpha_0 - \alpha \leq 1 + \varepsilon\}$, $F_3 = \{\alpha \in [-1, 1] : \alpha_0 - \alpha \geq 1 + \varepsilon\}$. If $|I| \leq 1$ then $I \subset F_1$ when $\varepsilon > 0$ is small enough. In that case $F_2 = F_3 = \varnothing$. Hence we shall consider $F_2$, $F_3$ only for $|I| > 1$ when $l \geq n^\eta$ holds for some $\eta > 0$ by Assumption $l$.

It suffices to show that

$$d_i := P\{\widehat{\alpha} \in F_i\} = o(m^{-p}), \qquad i = 1, 2, 3. \tag{5.19}$$

We have

$$\begin{aligned} d_i &\leq P\{R(\widehat{\alpha}) \leq R(\alpha_0), \widehat{\alpha} \in F_i\} \\ &= P\left(\log\left(\frac{F_0(\widehat{\alpha})}{F_0(\alpha_0)}\right) \leq (\widehat{\alpha} - \alpha_0)\widetilde{m}, \widehat{\alpha} \in F_i\right) \\ &= P\left(\frac{F_0(\widehat{\alpha})}{F_0(\alpha_0)} \leq e^{(\widehat{\alpha} - \alpha_0)\widetilde{m}}, \widehat{\alpha} \in F_i\right). \end{aligned}$$

Set

$$f_m(\alpha) = m^{-1}\lambda_\ell^{\alpha_0} F_0(\alpha) e^{-(\alpha - \alpha_0)\widetilde{m}}$$

and define $s_m(\alpha) = f_m(\alpha) - Ef_m(\alpha)$, $e_m(\alpha) = Ef_m(\alpha) - Ef_m(\alpha_0)$. Then

$$d_i \leq P\{f_m(\widehat{\alpha}) \leq f_m(\alpha_0) : \widehat{\alpha} \in F_i\} \leq P\{e_m(\widehat{\alpha}) \leq |s_m(\widehat{\alpha})| + |s_m(\alpha_0)| : \widehat{\alpha} \in F_i\}.$$

We show below that

$$e_m(\alpha) \geq p_{i,\alpha} \tag{5.20}$$

uniformly in $\alpha \in F_i$, $i = 1, 2, 3$, where $p_{1,\alpha} = c \log^{-2s} n$, $p_{2,\alpha} = c$, $p_{3,\alpha} = c(m/l)^{\alpha_0 - \alpha - 1}$ for some $c > 0$. Using (5.20), we get

$$\begin{aligned} d_i &\leq P\left(1 < \sup_{\alpha \in F_i} p_{i,\alpha}^{-1}\big(|s_m(\alpha)| + |s_m(\alpha_0)|\big)\right) \\ &\leq CE\left(\sup_{\alpha \in F_i} p_{i,\alpha}^{-2k}\big(|s_m(\alpha)|^{2k} + |s_m(\alpha_0)|^{2k}\big)\right), \end{aligned} \tag{5.21}$$

$k \geq 1$. If we show that for large enough $k$,

$$E \sup_{\alpha \in F_i} |p_{i,\alpha}^{-1} s_m(\alpha)|^{2k} = o(m^{-p}), \qquad i = 1, 2, 3, \tag{5.22}$$

and

$$E|s_m(\alpha_0)|^{2k} = o(m^{-p}), \tag{5.23}$$

then (5.19) follows from (5.21). Since $s_m(\alpha_0) = m^{-1/2} Z_0$ and $E|Z_0|^k < \infty$ for any $k \geq 1$ by Lemma 7.4, (5.23) follows by the Chebyshev inequality.

Before proving (5.22) we show (5.20). With

$$d_j(\alpha) = j^{\alpha-\alpha_0} e^{-(\alpha-\alpha_0)\widetilde{m}} \tag{5.24}$$

we can write

$$e_m(\alpha) = m^{-1} \sum_{j=l}^{m} \big(d_j(\alpha) - 1\big) E\big[\lambda_{\ell j}^{\alpha_0} I(\lambda_{\ell j})\big].$$

By (4.4) and (a) of Lemma 2.1 or Lemma 2.2,

$$E\big[\lambda_{\ell j}^{\alpha_0} I(\lambda_{\ell j})\big] = 1 + O(j/n)^{\beta} + O(j^{-1}\log j)$$

so that

$$e_m(\alpha) = m^{-1} \sum_{j=l}^{m} (d_j(\alpha) - 1)\big\{1 + O(j/n)^{\beta} + O(j^{-1}\log j)\big\}. \tag{5.25}$$

Since

$$\widetilde{m} = \log m - 1 + o(m^{-1/2}), \tag{5.26}$$

by $\sum_{j=1}^{m} \log j = m\log m - m + O(\log m)$, from (5.24) it follows that

$$d_j(\alpha) = (ej/m)^{\alpha-\alpha_0}\big(1 + o(m^{-1/2})\big). \tag{5.27}$$

*Case* (a). Let $\alpha \in F_1$. Then (5.25) and (5.27) imply

$$\begin{aligned} e_m(\alpha) &= m^{-1} \sum_{j=l}^{m} d_j(\alpha) - 1 + m^{-1} \sum_{j=l}^{m} (m/j)^{1-\varepsilon} O\big((m/n)^{\beta} + j^{-1}\log j\big) \\ &= m^{-1} \sum_{j=l}^{m} d_j(\alpha) - 1 + O\big(m^{-\eta'}\big) \end{aligned}$$

for some $\eta' > 0$ when $\varepsilon > 0$ is chosen small enough. Hence for large $m$,

$$\begin{aligned} e_m(\alpha) &= m^{-1} \sum_{j=l}^{m} \left(\frac{m}{ej}\right)^{\alpha_0-\alpha} - 1 + O\big(m^{-\eta'}\big) \\ &= \frac{e^{\alpha-\alpha_0}}{\alpha - \alpha_0 + 1} - 1 + O\big(m^{-\eta'}\big) \geq c|\alpha - \alpha_0|^2 + O\big(m^{-\eta'}\big) \geq c(\log n)^{-2s}/2 \end{aligned}$$

using the inequality $e^y - (1+y) \geq cy^2$ for $y \geq -1 + \varepsilon$, $c > 0$, and $|\alpha_0 - \alpha| \geq (\log n)^{-s}$. This proves (5.20) in case $i = 1$.

*Case* (b). Let $\alpha \in F_2$. Then $1-\varepsilon \le \alpha_0 - \alpha \le 1+\varepsilon$. Then

$$m^{-1}\sum_{j=l}^{m} d_j(\alpha) \ge C^{-1}m^{-1}\sum_{j=l}^{m}(j/m)^{\alpha-\alpha_0} = C^{-1}m^{-1}\sum_{j=l}^{m}(m/j)^{\alpha_0-\alpha}$$

$$\ge C^{-1}m^{-1}\sum_{j=l}^{m}(m/j)^{1-\varepsilon} \ge (C\varepsilon)^{-1}.$$

Choosing $\varepsilon > 0$ small, (5.25) and the assumption $l \ge n^{\eta}$ imply (5.20) for $i = 2$.

*Case* (c). Let $\alpha \in F_3$. Then $\alpha_0 - \alpha \ge 1+\varepsilon$, and thus

$$m^{-1}\sum_{j=l}^{m} d_j(\alpha) \ge C^{-1}m^{-1}\sum_{j=l}^{m}(j/m)^{\alpha-\alpha_0} \ge C^{-1}(m/l)^{\alpha_0-\alpha-1},$$

which together with (5.25) and the assumption $l \ge n^{\eta}$ implies (5.20) for $i = 3$.

To prove (5.22), set

$$\zeta_i(\alpha) = p_{i,\alpha}^{-1} s_m(\alpha) = m^{-1}\sum_{j=l}^{m} p_{i,\alpha}^{-1} d_j(\alpha)\lambda_{\ell j}^{\alpha_0}\big(I(\lambda_{\ell j}) - EI(\lambda_{\ell j})\big), \qquad i = 1, 2, 3.$$

By Lemma 7.3,

$$E|\zeta_i(\alpha) - \zeta_i(\alpha')|^{2k} \le CD_m(\alpha,\alpha')^k, \qquad E|\zeta_i(\alpha)|^{2k} \le CD_m(\alpha)^k, \tag{5.28}$$

where

$$D_m(\alpha,\alpha') = m^{-2}\sum_{j=l}^{m}\big|p_{i,\alpha}^{-1}d_j(\alpha) - p_{i,\alpha'}^{-1}d_j(\alpha')\big|^2,$$

$$D_m(\alpha) = m^{-2}\sum_{j=l}^{m}|p_{i,\alpha}^{-1}d_j(\alpha)|^2.$$

We show that

$$D_m(\alpha,\alpha') \le |\alpha-\alpha'|^2 h, \tag{5.29}$$

$$D_m(\alpha) \le h, \tag{5.30}$$

uniformly in $\alpha, \alpha' \in F_i$, $i = 1, 2, 3$, with some $h = cn^{-\gamma}$ where $\gamma > 0$, $c > 0$ do not depend on $m$. Then from (5.31) of Lemma 5.9, by (5.28)–(5.30) it follows that

$$E\Big(\sup_{t\in F_i}|\zeta_i(t)|^{2k}\Big) \le B_0 h^k = O(n^{-k\gamma}) = O(n^{-p}),$$

choosing $k$ such that $k\gamma > p$ to prove (5.22).

We prove first (5.29). Let $\alpha, \alpha' \in F_i$, $i = 1, 2$. Setting $h_j = je^{-\widetilde{m}}$ we can write $d_j(\alpha) = h_j^{\alpha-\alpha_0}$. By the mean value theorem,

$$|d_j(\alpha) - d_j(\alpha')| = \big|h_j^{\alpha-\alpha_0} - h_j^{\alpha'-\alpha_0}\big| \le C|\log(h_j)|h_j^{\bar{\alpha}-\alpha_0}|\alpha-\alpha'|,$$

where $\bar{\alpha} \in [\alpha, \alpha'] \subset F_i$. By (5.26), $h_j = C(j/m)(1 + O(m^{-1/2}))$ and $|\log h_j| \le C \log n$, uniformly in $l \le j \le m$. If $\alpha, \alpha' \in F_1$ then

$$|h_j|^{\bar{\alpha}-\alpha_0} \le C(m/j)^{\alpha_0-\bar{\alpha}} \le C(m/j)^{1-\varepsilon},$$

and since $p_{1,\alpha} = c(\log n)^{-2k}$,

$$D_m(\alpha, \alpha') \le C(\log n)^{4k+2} m^{-2} \sum_{j=l}^{m} (m/j)^{2(1-\varepsilon)} |\alpha - \alpha'|^2 \le C m^{-\varepsilon} |\alpha - \alpha'|^2.$$

If $\alpha, \alpha' \in F_2$ then $|h_j|^{\bar{\alpha}-\alpha_0} \le C(m/j)^{1+\varepsilon}$ and, since $l \ge n^{\eta}$, $p_{s,\alpha} = c$,

$$D_m(\alpha, \alpha') \le C(\log n)^2 m^{-2} \sum_{j=l}^{m} (m/j)^{2(1+\varepsilon)} |\alpha - \alpha'|^2 \le C m^{-\eta/2} |\alpha - \alpha'|^2$$

when $\varepsilon$ is small enough.

If $\alpha, \alpha' \in F_3$ then

$$p_{3,\alpha}^{-1} d_j(\alpha) = C(m/l) \left( \frac{e^{\tilde{m}} l}{jm} \right)^{\alpha_0-\alpha} = (m/l) \left( \frac{l}{ej} (1 + o(1)) \right)^{\alpha_0-\alpha},$$

so that

$$\begin{aligned} |p_{3,\alpha}^{-1} d_j(\alpha) - p_{3,\alpha'}^{-1} d_j(\alpha')| &\le C(m/l) \left( \log\left( \frac{e^{\tilde{m}} l}{jm} \right) \right) \left( \frac{e^{\tilde{m}} l}{jm} \right)^{\alpha_0-\bar{\alpha}} |\alpha - \alpha_0| \\ &\le C(m/l)(\log m)(l/j)^{\alpha_0-\bar{\alpha}} |\alpha - \alpha_0|. \end{aligned}$$

Since $\bar{\alpha} \in F_3$ implies $\alpha_0 - \bar{\alpha} \ge 1 + \varepsilon$, we obtain

$$\begin{aligned} D_m(\alpha, \alpha') &\le C|\alpha - \alpha'|^2 l^{-2} \log^2 m \sum_{j=l}^{m} (l/j)^{2(1+\varepsilon)} \\ &\le C|\alpha - \alpha'|^2 l^{-1} \log^2 m \le C n^{-\eta/2} |\alpha - \alpha'|^2 \end{aligned}$$

since $l \ge n^{\eta}$. This proves (5.29). The proof of (5.30) in cases $i = 1, 2, 3$ is similar. □

The following lemma is a modified version of Theorem 19 of Ibragimov and Has'minskii [(1981), page 372] and follows by the same argument.

LEMMA 5.9. *Let the random process $\zeta(t)$ be defined and continuous with probability 1 on the closed set $F$. Assume that there exist integers $m \ge r \ge 2$ and a number $H$ such that for all $t, s \in F$,*

$$E|\zeta(t) - \zeta(s)|^m \le h|t - s|^r, \qquad E|\zeta(t)|^m \le h.$$

*Then*

$$E\left( \sup_{t \in F} |\zeta(t)|^m \right) \le B_0 h, \tag{5.31}$$

*where $B_0$ depends on $m, r$ and does not depend on $\zeta$.*

**6. Second-order expansions.** Since by Lemma 5.6, $U_m = -B_m + V_m + \Delta_m^{1+\delta}\xi_m$, the expansion for $U_m$ requires one for $V_m$ (4.8). This in turn requires one for $Z = (Z_1, Z_2)'$, where $Z_i$ are given in (4.6) and defined with $\ell = 3$ in case of tapering and $\ell = 1$ in case of no tapering. We assume in this section that (4.4) and Assumption $m^*$ are satisfied. We shall derive the expansion of $V_m$ in terms of $\widetilde{\Delta}_m (\geq \Delta_m)$.

We shall approximate the distribution function $P(Z \leq x)$, $x = (x_1, x_2) \in \mathbb{R}^2$, by

$$F(x) = \int_{y \leq x} \phi(y : \Omega) K(y)\, dy, \tag{6.1}$$

where

$$\phi(y : \Omega) = (2\pi)^{-1} |\Omega|^{-1/2} \exp\left(-\tfrac{1}{2} y' \Omega^{-1} y\right), \qquad y \in \mathbb{R}^2,$$

is the density of a zero-mean bivariate Gaussian vector with covariance matrix

$$\Omega = \begin{pmatrix} e_{1+1} & e_{1+2} \\ e_{2+1} & e_{2+2} \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ -2 & 9 \end{pmatrix},$$

where the elements of $\Omega$ are defined by (7.5) and related to $Z_1, Z_2$ by

$$E[Z_p Z_v] = e_{p+v} + 2(m/n)^{\beta} e(p + v, \ell, \beta) + o(\Delta_m) \tag{6.2}$$

(see Lemma 7.5). The polynomial $K(y)$ is given by

$$K(y) = 1 + \left(\frac{m}{n}\right)^{\beta} \frac{1}{2!} P^{(2)}(y) + m^{-1/2} \frac{1}{3!} P^{(3)}(y), \tag{6.3}$$

where $P^{(2)}(y)$, $P^{(3)}(y)$ are polynomials defined by

$$P^{(2)}(y) = 2 \sum_{i,j=1}^{2} e(i + j, \ell, \beta) H_{ij}(y), \qquad P^{(3)}(y) = 2 \sum_{i,j,k=1}^{2} e_{i+j+k} H_{ijk}(y),$$

$$H_{ij}(y) = \phi(y : \Omega)^{-1} \frac{\partial^2}{\partial y_i\, \partial y_j} \phi(y : \Omega),$$

$$H_{ijk}(y) = -\phi(y : \Omega)^{-1} \frac{\partial^3}{\partial y_i\, \partial y_j\, \partial y_k} \phi(y : \Omega), \qquad i, j, k = 1, 2.$$

THEOREM 6.1. *Suppose that* (4.4) *and Assumptions* $\alpha$, $f$, $l$ *and* $m^*$ *hold. Then*

$$\sup_B |P(Z \in B) - F(B)| = \tfrac{4}{3} \sup_B F\left((\partial B)^{2\varepsilon}\right) + o(\widetilde{\Delta}_m) \tag{6.4}$$

*for any* $\varepsilon = m^{-1-\rho}$ $(0 \leq \rho < 1/2)$, *where* $\sup_B$ *is taken over all Borel sets* $B$ *in* $\mathbb{R}^2$, $F(B) = \int_B \phi(y : \Omega) K(y)\, dy$ *and* $(\partial B)^{\varepsilon}$ *is an* $\varepsilon$ *neighborhood of* $B$. *In particular*,

$$\sup_{x \in R^2} |P(Z \leq x) - F(x)| = o(\widetilde{\Delta}_m). \tag{6.5}$$

PROOF. Set

$$F^*(B) = \int_B \phi(y:\Omega) K_m(y)\,dy, \tag{6.6}$$

where

$$K_m(y) = 1 + \frac{1}{2!} P_m^{(2)}(y) + m^{-1/2} \frac{1}{3!} P^{(3)}(y),$$

$$P_m^{(2)}(y) = \sum_{i,j=1}^{2} (E[Z_i Z_j] - e_{i+j}) H_{ij}(y).$$

We show below that

$$\sup_B |P(Z \in B) - F^*(B)| = \tfrac{4}{3} \sup_B F\big((\partial B)^{2\varepsilon}\big) + o(\widetilde{\Delta}_m). \tag{6.7}$$

By (6.2) it follows that

$$P_m^{(2)}(y) = (m/n)^{\beta} P^{(2)}(y) + o(\Delta_m)\|y\|^2,$$

where $\Delta_m = \max((m/n)^{\beta}, m^{-1/2}) \le \widetilde{\Delta}_m$. Therefore

$$\sup_B |F^*(B) - F(B)| = o(\widetilde{\Delta}_m), \tag{6.8}$$

and (6.4) follows from (6.8) and (6.7).

When $\sup_B$ is taken over the sets $B = \{z : z \le x\}$, $x \in \mathbb{R}^2$, (6.5) follows from (6.4), noting that $\sup_B F((\partial B)^{2\varepsilon}) = o(\Delta_m)$.

To prove (6.7) we obtain first an asymptotic expansion for the characteristic function

$$\tau(t) = \tau(t_1, t_2) = \exp(itZ), \qquad t = (t_1, t_2), t_1, t_2 \ge 0.$$

Set $Q = tZ = t_1 Z_1 + t_2 Z_2$. We shall show that

$$\log \tau(t) = \frac{i^2}{2!} \mathrm{Cum}_2(Q) + \frac{i^3}{3!} \mathrm{Cum}_3(Q) + O\big(\mathrm{Cum}_4(Q)\big), \tag{6.9}$$

where $\mathrm{Cum}_j(Q)$ denotes the $j$th cumulant of $Q$. Since

$$tZ = t_1 Z_1 + t_2 Z_2 = m^{-1/2} \sum_{j=l}^{m} (t_1 \nu_j + t_2 \nu_j^2) \lambda_{\ell j}^{\alpha_0} \big(I(\lambda_{\ell j}) - EI(\lambda_{\ell j})\big),$$

we can write $tZ = X' B_n X - E[X' B_n X]$ where $X = (X_1, \dots, X_n)$ and $B_n = (b_{i,j})_{i,j=1,\dots,n}$ is a symmetric matrix defined by

$$m^{-1/2} \sum_{j=l}^{m} (t_1 \nu_j + t_2 \nu_j^2) \lambda_{\ell j}^{\alpha_0} I(\lambda_{\ell j}) = X' B_n X.$$

Then [see (3.2.36) of Taniguchi (1991)] $\tau(t) = |I - 2iS|^{-1/2} \exp(-i \operatorname{Tr}(S))$, where $S = R_n^{1/2} B_n R_n^{1/2}$, $R_n = (r(i-j))_{i,j=1,\dots,n}$, being the covariance matrix of $X$, with $r(t) = \operatorname{Cov}(X_t, X_0)$.

Since $S$ is symmetric, it has real eigenvalues, denoted $\rho_j$, $j = 1, \dots, n$. Therefore as in (3.2.36) of Taniguchi (1991), we can write

$$\log \tau(t) = -(1/2) \sum_{j=1}^n \log(1 - 2i\rho_j) - i \sum_{j=1}^n \rho_j.$$

Using Lemma 8.1 of Bhattacharya and Rao [(1976), page 57], we get

$$\log(1 - ih) = -ih + \frac{h^2}{2} + \frac{ih^3}{3} + (ih)^4 \int_0^1 \frac{(1-v)^3}{(1-ivh)^4} \, dv,$$

where

$$\left| \int_0^1 \frac{(1-v)^3}{(1-ivh)^4} \, dv \right| \le \int_0^1 \frac{1}{(|1 + |vh|^2)^2} \, dv \le \int_0^1 1 \, dv = 1.$$

Thus

$$\begin{aligned} \log \tau(t) &= \sum_{j=1}^n i^2 \rho_j^2 + \tfrac{4}{3} \sum_{j=1}^n i^3 \rho_j^3 + O\left( \sum_{j=1}^n \rho_j^4 \right) \\ &= i^2 \operatorname{Tr}(S^2) + \tfrac{4}{3} i^3 \operatorname{Tr}(S^3) + O(\operatorname{Tr}(S^4)). \end{aligned} \tag{6.10}$$

Since

$$\begin{aligned} \operatorname{Cum}_2(Q) &= 2 \operatorname{Tr}([B_n R_n]^2), \\ \operatorname{Cum}_3(Q) &= 8 \operatorname{Tr}([B_n R_n]^3), \\ \operatorname{Cum}_4(Q) &= 48 \operatorname{Tr}([B_n R_n]^4), \end{aligned} \tag{6.11}$$

(6.10) implies (6.9). Note now that

$$\operatorname{Cum}_2(Q) = \sum_{p,v=1}^2 t_p t_v E[Z_p Z_v] = t' \Omega t + p_m^{(2)}(t)$$

by (6.2), where

$$p_m^{(2)}(t) = \sum_{p,v=1}^2 t_p t_v (E[Z_p Z_v] - e_{p+v}).$$

Since $\Omega$ is positive definite, such that $t' \Omega t \ge \|t\|^2/4$, in view of (6.2),

$$|p_m^{(2)}(t)| \le C \Delta_m \|t\|^2,$$

so it follows that for large enough $m$,

$$\operatorname{Var}(Q) = \operatorname{Cum}_2(Q) \ge \|t\|^2/8. \tag{6.12}$$

By Lemma 7.5,

$$\mathrm{Cum}_3(Q) = \sum_{i,j,k=1}^{2} E[Z_i Z_j Z_k] t_i t_j t_k = m^{-1/2} p^{(3)}(t) + O(\widetilde{\Delta}_m^2)\|t\|^3,$$

where

$$p^{(3)}(t) = 2 \sum_{i,j,k=1}^{2} e_{i+j+k} t_i t_j t_k.$$

Finally, since $EQ = 0$,

$$\mathrm{Cum}_4(Q) = EQ^4 - 3(EQ^2)^2 = O(\widetilde{\Delta}_m^2 \|t\|^4), \tag{6.13}$$

using Lemma 7.5. Hence by (6.9),

$$\log \tau(t) = -\frac{1}{2} t' \Omega t + \frac{p_m^{(2)}(it)}{2!} + \frac{m^{-1/2} p^{(3)}(it)}{3!} + O(\widetilde{\Delta}_m^2 \|t\|_+^4), \tag{6.14}$$

where $\|t\|_+ = \max(\|t\|, 1)$. Set

$$\tau^*(t) = \exp\left(-\frac{1}{2} t' \Omega t\right)\left(1 + \frac{p_m^{(2)}(it)}{2!} + m^{-1/2} \frac{p^{(3)}(it)}{3!}\right),$$

which corresponds to the Fourier transform of the measure $F^*$ in $\mathbb{R}^2$ given by (6.6) [see, e.g., Taniguchi (1991), page 14]. Equation (6.7) now follows from Lemma 6.1 below, using the same argument as in the proof of Lemma 3.2.8 in Taniguchi (1991). □

Lemma 6.1 corresponds to Lemmas 3.2.5 and 3.2.6 of Taniguchi (1991).

LEMMA 6.1. *There exists* $\delta > 0$ *such that, as* $n \to \infty$, *for all* $t$ *satisfying* $\|t\| \le \delta \widetilde{\Delta}_m^{-1}$,

$$|\tau(t) - \tau^*(t)| \le m^{-1} \exp(-a\|t\|^2) P(t), \tag{6.15}$$

*where* $a > 0$ *and* $P(t)$ *is a polynomial, and for all* $\|t\| > \delta \widetilde{\Delta}_m^{-1}$,

$$|\tau(t)| \le \exp(-a_1 m^{\varepsilon}), \tag{6.16}$$

*where* $a_1 > 0$, $\varepsilon > 0$.

PROOF. By (6.14), $\log \tau(t) = -\frac{1}{2} t' \Omega t + k(t)$, where

$$|k(t)| \le C(\widetilde{\Delta}_m \|t\|_+^3 + \widetilde{\Delta}_m^2 \|t\|_+^4) \le \|t\|_+^2 / 16 \tag{6.17}$$

for $\|t\| \le \delta \widetilde{\Delta}_m^{-1}$ where $\delta > 0$ is chosen sufficiently small.

Using (6.14) and the inequality $|e^z - 1 - z| \le \frac{1}{2}|z|^2 e^{|z|}$, we see that

$$\begin{aligned}|\tau(t) - \tau^*(t)| &= \exp\bigl(-\tfrac{1}{2}t'\Omega t\bigr)\bigl|\exp(k(t)) - \bigl\{1 + k(t) + O(\widetilde{\Delta}_m^2 \|t\|_+^4)\bigr\}\bigr| \\ &\le C \exp\bigl(-\tfrac{1}{8}\|t\|^2\bigr) \exp\bigl(|k(t)|\bigr)\bigl(|k(t)|^2 + O(\widetilde{\Delta}_m^2 \|t\|_+^4)\bigr) \\ &\le C \exp\bigl(-\tfrac{1}{8}\|t\|^2\bigr) \exp(\|t\|^2/16)\|t\|_+^4.\end{aligned}$$

This proves (6.15). To show (6.16) note that

$$|\tau(t)| \le \prod_{j=1}^n (1 + 4\rho_j^2)^{-1/4}.$$

By the inequality $\log(1+x) \ge x/(1+x)$, $x > 0$, we get

$$\begin{aligned}\log|\tau(t)| &\le -\frac{1}{4}\sum_{j=1}^n \log(1 + 4\rho_j^2) \le -\frac{1}{4}\sum_{j=1}^n \frac{4\rho_j^2}{1 + 4\rho_j^2} \\ &\le -\frac{1}{4(1+\rho_*^2)}\sum_{j=1}^n \rho_j^2 = -\frac{\operatorname{Tr}(S^2)}{4(1+\rho_*^2)},\end{aligned}$$

where $\rho_*^2 = \max_j \rho_j^2$. Note that

$$\rho_*^2 \le [\operatorname{Tr}(S^4)]^{1/2} = \bigl(\operatorname{Cum}_4(Q)/48\bigr)^{1/2} \le C\widetilde{\Delta}_m \|t\|_+^2$$

by (6.11) and (6.13). The assumption $\|t\| \ge \delta \widetilde{\Delta}_m^{-1} > 1$ implies that

$$(1 + \rho_*^2)^{-1} \ge (2\rho_*^2)^{-1} \ge \frac{1}{C\widetilde{\Delta}_m \|t\|_+^2}.$$

For large $m$, by (6.12), $\operatorname{Tr}(S^2) = \operatorname{Var}(Q)/2 \ge \|t\|^2/16$. Thus, since $\widetilde{\Delta}_m^{-1} \ge m^\varepsilon$ for some $\varepsilon > 0$,

$$\log|\tau(t)| \le -C^{-1}(\widetilde{\Delta}_m \|t\|^2)^{-1}\|t\|^2/16 = -C^{-1}m^\varepsilon/16,$$

to prove (6.16). □

LEMMA 6.2. *Let* (4.4) *and Assumptions* $\alpha$, $f$, $l$ *and* $m^*$ *hold. Then, with* $V_m$ *given by* (4.8),

$$\sup_{y \in R}\bigl|P(V_m \le y) - \Phi(y) - m^{-1/2}\phi(y)p(y)\bigr| = o(\widetilde{\Delta}_m),$$

*where* $p(y)$ *is given by* (2.14).

PROOF. We shall derive the second-order expansion

$$P(V_m \le y) = \Phi(y) - m^{-1/2}\phi(y)\widetilde{p}(y) + o(\widetilde{\Delta}_m) \tag{6.18}$$

uniformly in $y \in \mathbb{R}$ where

$$\widetilde{p}(y) = a_1 + a_2 \frac{y}{2!} + a_3 \frac{y^2 - 1}{3!} \tag{6.19}$$

and the coefficients $a_1, a_2, a_3$ are defined [cf. (2.1.16), page 15 of Taniguchi (1991)] by

$$\mathrm{Cum}_j(V_m) = \mathbb{1}_{\{j=2\}} + m^{-1/2} a_j + o(\widetilde{\Delta}_m), \qquad j = 1, 2, 3. \tag{6.20}$$

In fact we shall show that (6.20) holds with $\Delta_m (\leq \widetilde{\Delta}_m)$ instead of $\widetilde{\Delta}_m$. We first show that

$$a_1 = -1, \qquad a_2 = 0, \qquad a_3 = -2. \tag{6.21}$$

Write $V_m = -P + m^{-1/2} Q + R$, where $P = Z_1(2 - ES_2)$, $Q = Z_1 Z_2 + Z_1^2$, $R = (2Z_1 + Z_2) ES_1$.

Since $EP = ER = 0$ and by (7.39), (7.38), $EQ = EZ_1Z_2 + EZ_1^2 = -1 + o(1)$, we obtain

$$\mathrm{Cum}_1(V_m) \equiv EV_m = m^{-1/2} EQ = -m^{-1/2} + o(m^{-1/2}) \tag{6.22}$$

and therefore $a_1 = -1$. Now, by (6.22),

$$\begin{aligned} &\mathrm{Cum}_2(V_m) \\ &\quad = E(V_m - EV_m)^2 = EV_m^2 + o(m^{-1/2}) \\ &\quad = EP^2 - 2EP(m^{-1/2}Q + R) + E(m^{-1/2}Q + R)^2 + o(m^{-1/2}). \end{aligned} \tag{6.23}$$

We show that

$$EP^2 = 1 + o(\Delta_m), \qquad EPQ = O(\Delta_m), \qquad EPR = o(\Delta_m) \tag{6.24}$$

and

$$E|m^{-1/2}P + Q|^i = O(\Delta_m^i), \qquad i = 2, 3, 4, \tag{6.25}$$

which with (6.23) implies

$$\mathrm{Cum}_2(V_m) = 1 + o(\Delta_m), \tag{6.26}$$

and thus $a_2 = 0$.

By Lemma 7.1,

$$ES_1 = O(\Delta_m), \qquad ES_2 = 1 + \nu(m/n)^\beta + o(\Delta_m), \tag{6.27}$$

where $\nu = e(2, \ell, \beta)$, and (7.38) implies

$$\begin{aligned} EP^2 &= EZ_1^2(2 - ES_2)^2 = \big(1 + 2\nu(m/n)^\beta\big)\big(1 - \nu(m/n)^\beta\big)^2 + o(\Delta_m) \\ &= \big(1 + 2\nu(m/n)^\beta\big)\big(1 - 2\nu(m/n)^\beta\big) + o(\Delta_m) = 1 + o(\Delta_m), \end{aligned}$$

while from (6.27), (7.38) and (7.39), it follows that

$$EPQ = (2 - ES_2)(E[Z_1^2 Z_2] + EZ_1^3) = \bigl(1 + o(1)\bigr)O(\Delta_m) = O(\Delta_m),$$

and

$$EPR = (2 - ES_2)ES_1(2EZ_1^2 + EZ_1 Z_2) = \bigl(1 + o(1)\bigr)O(\Delta_m)o(1) = o(\Delta_m).$$

Thus (6.24) holds and (6.25) follows using (6.27) and $E|Z_j|^k < \infty$, shown in Lemma 7.4.

Next, from $EV_m^3 = \mathrm{Cum}_3(V_m) + 3\,\mathrm{Cum}_2(V_m)\,\mathrm{Cum}_1(V_m) + \mathrm{Cum}_1(V_m)^3$, by (6.22) and (6.26) it follows that

$$\mathrm{Cum}_3(V_m) = EV_m^3 + 3m^{-1/2} + o(\Delta_m),$$

where

$$\begin{aligned} EV_m^3 &= E(-P + [m^{-1/2}Q + R])^3 \\ &= -EP^3 + 3EP^2(m^{-1/2}Q + R) \\ &\quad - 3EP(m^{-1/2}Q + R)^2 + E(m^{-1/2}Q + R)^3 \\ &= -EP^3 + 3EP^2(m^{-1/2}Q + R) + o(\Delta_m), \end{aligned}$$

in view of (6.25) and (6.24). From (6.27), (7.38) and (7.39) it follows that

$$\begin{aligned} EP^3 &= EZ_1^3(2 - ES_2)^3 \\ &= \bigl(-4m^{-1/2} + o(\Delta_m)\bigr)\bigl(1 + o(1)\bigr) = -4m^{-1/2} + o(\Delta_m), \\ EP^2Q &= (2 - ES_2)^2\{EZ_1^3 Z_2 + EZ_1^4\} \\ &= \bigl(1 + o(1)\bigr)\bigl(-6 + 3 + o(1)\bigr) = -3 + o(1), \\ EP^2R &= (2 - ES_2)^2 ES_1\{EZ_1^2 Z_2 + 2EZ_1^3\} \\ &= \bigl(1 + o(1)\bigr)O(\Delta_m)\{O(\Delta_m)\} = o(\Delta_m) \end{aligned}$$

which yields $EV_m^3 = -5m^{-1/2} + o(\Delta_m)$. Thus $\mathrm{Cum}_3(V_m) = -2m^{-1/2} + o(\Delta_m)$ and $a_3 = -2$.

It remains to establish the validity of the expansion (6.18). The proof is based on the expansion for $(Z_1, Z_2)$ of Lemma 6.2 and follows by a similar argument to the proof of Lemma 3.2.9 of Taniguchi (1991) or the proof of Bhattacharya and Ghosh (1978). Denote by

$$f(z_1, z_2) = \frac{\partial^2}{\partial z_1\,\partial z_2} F(z_1, z_2) = \phi(z_1, z_2 : \Omega)K(z_1, z_2)$$

the density of $(Z_1, Z_2)'$, where $F$ is defined in (6.1) and $K$ is defined in (6.3). Set $B_y = \{v(z_1, z_2) \le y\}$. Then by (6.4) of Theorem 6.1,

$$\sup_y |P\{V_m \le y\} - F(B_y)| = \tfrac{4}{3}\sup_y F\bigl((\partial B_y)^{\varepsilon}\bigr) + o(\widetilde{\Delta}_m),$$

where $\varepsilon = m^{-1-\rho}$ for some $0 < \rho < 1/2$. We will show that

$$F(B_y) = \int_{x \le y} \phi(x)\big(1 + p(x)\big)\,dx + o(\widetilde{\Delta}_m), \tag{6.28}$$

$$F\big((\partial B_y)^{\varepsilon}\big) = o(\widetilde{\Delta}_m) \tag{6.29}$$

uniformly in $y$, to prove (6.18). Setting

$$v(x_1, x_2) = -x_1[(2 - ES_2) - ES_1] + m^{-1/2}x_1x_2 + x_2ES_1 \tag{6.30}$$

and $f^*(x_1, x_2) = f(x_1, x_2 - x_1)$, we can write $V_m$ in (4.8) as $V_m = v(Z_1, Z_1 + Z_2)$, and

$$F(B_y) = \int_{v(x_1,x_2)\le y} f^*(x_1, x_2)\,dx_1\,dx_2.$$

Denote $v = v(x_1, x_2)$. Then $x_1 = (-v + x_2ES_1)D^{-1}$ where $D = (2 - ES_1 - ES_2) - m^{-1/2}x_2$. Since $|f^*(x_1, x_2)| \le C\exp(-c(x_1^2 + x_2^2))$ with some $c > 0$, then for any $\delta > 0$,

$$\begin{aligned} &P(V_m \le y)\\ &\quad= \int_{v\le y\,:\,|x_1|,|x_2|\le m^{\delta}} f^*\big((-v + x_2ES_1)D^{-1}, x_2\big)(-D^{-1})\,dv\,dx_2 + o(\widetilde{\Delta}_m). \end{aligned}$$

When $|x_1|, |x_2| \le m^{\delta}$ and $\delta > 0$ is small, Lemma 7.1 implies

$$\begin{aligned} D^{-1} &= 1 + h_m(x_2) + o(\widetilde{\Delta}_m),\\ h_m(x_2) &= \big(e(1,\ell,\beta) + e(2,\ell,\beta)\big)(m/n)^{\beta} + m^{-1/2}x_2, \end{aligned}$$

$$\begin{aligned} x_1 &= -v - vh_m(x_2) + e(1,\ell,\beta)x_2(m/n)^{\beta} + o(\widetilde{\Delta}_m)(|v| + |x_2|)\\ &= -v + o(1). \end{aligned} \tag{6.31}$$

This and Taylor's theorem imply

$$\begin{aligned} &P(V_m \le y)\\ &\quad= \int_{v\le y\,:\,|v|,|x_2|\le m^{\delta}} \bigg(f^*(-v, x_2) + \big(vh_m(x_2) - e(1,\ell,\beta)x_2(m/n)^{\beta}\big)\\ &\qquad\qquad \times \frac{\partial}{\partial v} f^*(-v, x_2)\bigg)\,dv\,dx_2\\ &\qquad + o(\widetilde{\Delta}_m), \end{aligned}$$

and integrating out $x_2$, we arrive at the second-order expansion

$$F(V_m \le v) = \int_{v\le y} \phi(v)P_m(v)\,dv + o(\widetilde{\Delta}_m),$$

where $P_m(y)$ is quadratic in $y$. Comparing this expansion with (6.18) we conclude that $P_m(x) \equiv 1 - m^{-1/2}\widetilde{p}(x)$, where $\widetilde{p}$ (6.19), as already shown, has coefficients (6.21), so that $P_m(x) = 1 + m^{-1/2}(2 + x^2)/3$, to prove (6.28).

To show (6.29) note that

$$F\big((\partial B_y)^\varepsilon\big) = \int_{(\partial B_y)^\varepsilon : |x_i| \le m^\delta} f^*(x_1, x_2)\, dx_1\, dx_2 + o(m^{-1/2})$$

for any $\delta > 0$. By (6.30), from $y = v(x_1, x_2)$ we can solve for $x_2 = h(y, x_1)$. Thus

$$\begin{aligned} F\big((\partial B_y)^\varepsilon\big) &\le C \int_{(\partial B_y)^\varepsilon : |x_i| \le m^\delta} dx_1\, dx_2 + o(m^{-1/2}) \\ &\le C \int_{|x_1| \le m^\delta} \int_{x_2 \in [h(y,x_1) - \varepsilon, h(y,x_1) + \varepsilon]} dx_2\, dx_1 + o(m^{-1/2}) \\ &\le C 2\varepsilon m^\delta + o(m^{-1/2}) = o(m^{-1/2}) \end{aligned}$$

when $\varepsilon = m^{-1-\rho}$ and $\rho > \delta$. □

LEMMA 6.3. *Let*

$$V_m' = V_m + \widetilde{\Delta}_m^{1+\delta} \xi_m,$$

*where $V_m$ is given by* (4.8), $\xi_m \in \mathcal{X}$ *and* $0 < \delta < 1$. *Then under the assumptions of Lemma* 6.2,

$$\sup_{y \in R} \big| P(V_m' \le y) - \Phi(y) - m^{-1/2}\phi(y) p(y) \big| = o(\widetilde{\Delta}_m). \tag{6.32}$$

PROOF. By Chibisov's Lemma 5.2 and Lemma 6.2 of this paper,

$$P(V_m' \le y) = P(V_m \le y) + o(\widetilde{\Delta}_m),$$

implying (6.32). □

LEMMA 6.4. *Let the assumptions of Lemma* 6.2 *hold and $V_m^* = V_m - v_m Z_1$ where $V_m$ is as in Lemma* 6.2 *and $v_m$ is a sequence of real numbers such that $v_m = o(m^{-\varepsilon})$ for some $\varepsilon > 0$. Then as $n \to \infty$*,

$$\sup_{y \in R} \big| P(V_m^* \le y) - \Phi(y) - \phi(y)\{m^{-1/2} p(y) - v_m y\} \big| = o(\widetilde{\Delta}_m + v_m). \tag{6.33}$$

PROOF. Similarly to the proof of relations (6.20) in Lemma 6.2, it can be shown that

$$\mathrm{Cum}_1(V_m^*) = -m^{-1/2} + o(\Delta_m), \qquad \mathrm{Cum}_2(V_m^*) = 1 + 2v_m + o(\Delta_m + v_m),$$

$$\mathrm{Cum}_3(V_m^*) = -2m^{-1/2} + o(\Delta_m + v_m),$$

whence (6.33) follows using the same argument as in the proof of (6.18) of Lemma 6.2. □

**7. Technical lemmas.** The present section provides approximations for

$$S_{k,1} = \frac{1}{c_0 m} \sum_{j=1}^{m} \nu_j^k \lambda_j^{\alpha_0} I(\lambda_{\ell j}), \tag{7.1}$$

$$S_{k,3} = \frac{1}{c_0 m} \sum_{j=l}^{m} \nu_j^k \lambda_{3j}^{\alpha_0} I_h(\lambda_{3j}), \tag{7.2}$$

and related quantities. Note that $S_{k,3}$ is $S_{k,m}(\alpha_0)$ (3.7) and is relevant in case of tapering, but to discuss the untapered estimate of $\alpha_0$ we need to study (7.1) also. Recall the definition (4.1), $\kappa_{m,l}$, which likewise differs between tapering (synonymous with $\ell = 3$ in our setup) and no tapering (synonymous with $\ell = 1$).

Put

$$t(k,\beta) = \int_0^1 (\log x + 1)^k x^\beta \, dx, \qquad k \geq 1, \beta \geq 0.$$

LEMMA 7.1. *Let Assumptions* $\alpha$, $f$, $l$ *and* $m^*$ *and* (4.4) *hold. Then as* $m \to \infty$, *for* $\ell = 1, 3$,

$$ES_{1,\ell} = (m/n)^\beta e(1,\ell,\beta) + o\big((m/n)^\beta\big) + O(m^{-1}\kappa_{m,l}) \tag{7.3}$$

*and*

$$\begin{aligned} ES_{k,\ell} &= e_k + (m/n)^\beta e(k,\ell,\beta) \\ &\quad + o\big((m/n)^\beta + m^{-1/2}\big), \qquad k = 0, 2, 3, 4, \dots, \end{aligned} \tag{7.4}$$

*where*

$$e_k = t(k,0), \qquad e(k,\ell,\beta) = \frac{c_1}{c_0}\left(\frac{\ell}{2\pi}\right)^\beta t(k,\beta), \qquad k = 0, 1, \dots. \tag{7.5}$$

REMARK 7.1. In Lemma 7.1, $e_0 = 1$, $e_1 = 0$, $e_2 = 1$, $e_3 = -2$, $e_4 = 9$,

$$t(0,\beta) = \frac{1}{\beta+1}, \qquad t(1,\beta) = \frac{\beta}{(\beta+1)^2}, \qquad t(2,\beta) = \frac{\beta^2+1}{(\beta+1)^3},$$

$$t(3,\beta) = \frac{\beta^3 + 3\beta - 2}{(\beta+1)^4}, \qquad t(4,\beta) = \frac{\beta^4 + 6\beta^2 - 8\beta + 9}{(\beta+1)^5}.$$

PROOF OF LEMMA 7.1. We show that

$$\begin{aligned} ES_{k,\ell} &= e_k + \lambda_{\ell m}^\beta c_1 t(k,\beta) + o\big((m/n)^\beta\big) + O\big(m^{-1}\kappa_{m,l}(\log m)^{k-1}\big) \\ &\quad + O(m^{-1} l \log^k m)\mathbb{1}_{\{k \geq 2\}}, \qquad k = 0, 1, \dots, \end{aligned} \tag{7.6}$$

which implies (7.3). Then (7.4) follows from (7.6) and Assumption $l$. To prove (7.6) note that (4.4) and assumption (a) of Lemma 2.2 or Lemma 2.1 imply

$$E[\lambda_j^{\alpha_0} I(\lambda_j)] = 1 + (c_1/c_0)\lambda_j^{\beta} + r_j(1), \tag{7.7}$$

$$E[\lambda_{3j}^{\alpha_0} I(\lambda_{3j})] = 1 + (c_1/c_0)\lambda_{3j}^{\beta} + r_j(3), \tag{7.8}$$

where $r_j(1) = o((j/n)^{\beta}) + O(j^{-2})$ (without tapering), and $r_j(3) = o((j/n)^{\beta}) + O(j^{-1}\log j)$ (with tapering). Setting

$$t_m(k,\beta) = m^{-1}\sum_{j=l}^{m} \nu_j^k (j/m)^{\beta}, \qquad R_m(k,\ell,\beta) = m^{-1}\sum_{j=l}^{m} \nu_j^k r_j(\ell), \tag{7.9}$$

we can write

$$ES_{k,\ell} = t_m(k,0) + (c_1/c_0)\lambda_{\ell m}^{\beta} t_m(k,\beta) + R_m(k,\ell,\beta). \tag{7.10}$$

Note that

$$t_m(k,\beta) = t(k,\beta) + O(l\log^k m/m), \qquad k \ge 0, \beta \ge 0, \tag{7.11}$$

and

$$R_m(k,\ell,\beta) = o\big((m/n)^{\beta}\big) + O\big(m^{-1}\kappa_{m,l}(\log m)^{k-1}\big), \qquad k \ge 0. \tag{7.12}$$

Equations (7.10)–(7.12) and $t_m(1,0) = 0$ imply (7.6).

To prove (7.11) note that $\sum_{j=1}^{m} \log j = m\log m - m + O(\log m)$ implies

$$\nu_j = \log(j/m) + 1 + O(l\log m/m) \tag{7.13}$$

and therefore

$$\begin{aligned} t_m(k,\beta) &= m^{-1}\sum_{j=l}^{m} \big(\log(j/m)+1\big)^k (j/m)^{\beta} + O(l\log^k m/m) \\ &= t(k,\beta) + o(m^{-1/2}), \qquad k \ge 0, \beta \ge 0, \end{aligned}$$

under Assumption $l$. To show (7.12) note that (7.13) and (7.11) imply

$$|\nu_j| \le C\log m, \qquad \sum_{j=l}^{m} |\nu_j^k| \le m^{1/2}\left(\sum_{j=l}^{m} \nu_j^{2k}\right)^{1/2} \le Cm, \qquad k \ge 1. \tag{7.14}$$

Therefore

$$\begin{aligned} |R_m(k,\ell,\beta)| &\le o\big((m/n)^{\beta}\big) m^{-1}\sum_{j=l}^{m} |\nu_j^k| + Cm^{-1}\log^k m \sum_{j=l}^{m} |r_j(\ell)| \\ &= o\big((m/n)^{\beta}\big) + O\big(m^{-1}\kappa_{m,l}(\log m)^{k-1}\big), \qquad k \ge 1. \qquad \square \end{aligned}$$

LEMMA 7.2. *Let* (4.5) *and Assumptions* $\alpha$, $f'$ *and* $m^*$ *hold. Then*

$$ES_{1,\ell} = e(1,\ell,2)(m/n)^2 + d(1,\ell,4)(m/n)^4 + o\big((m/n)^4 + m^{-1}\big) + O(m^{-1}\kappa_{m,l}), \tag{7.15}$$

$$ES_{k,\ell} = 1 + e(k,\ell,2)(m/n)^2 + o\big((m/n)^2 + m^{-1/2}\big), \qquad k = 0, 2, 3, \ldots, \tag{7.16}$$

*where*

$$d(k,\ell,\beta) = (c_2/c_0)(\ell/2\pi)^\beta t(k,\beta), \qquad k \geq 0. \tag{7.17}$$

PROOF. Similarly to (7.10), under (4.5), we get

$$ES_{k,\ell} = t_m(k,0) + (c_1/c_0)\lambda_{\ell m}^2 t_m(k,2) + (c_2/c_0)\lambda_{\ell m}^4 t_m(k,4) + R_m(k,\ell,4), \qquad k \geq 0. \tag{7.18}$$

Since $t_m(1,0) = 0$, this and (7.11) and (7.12) imply that

$$ES_{1,\ell} = (c_1/c_0)\lambda_{\ell m}^2 t(k,2) + (c_2/c_0)\lambda_{\ell m}^4 t(k,4) + o\big((m/n)^4\big) + O\big((m/n)^2 l \log m/m + m^{-1}\kappa_{m,l}\big),$$

where

$$(m/n)^2 l \log m/m \leq (m/n)^4/\log^2 m + (l \log^2 m/m)^2 = o\big((m/n)^4 + m^{-1}\big)$$

by Assumption $l$, and thus (7.15) holds. Equation (7.16) repeats (7.4). □

LEMMA 7.3. *Let Assumptions* $\alpha$, $f$, $l$ *and* $m^*$ *hold. For* $p = 1, \ldots, k$, *let*

$$S_m^{(p)} = \sum_{j=l}^{m} a_j^{(p)} |v^{(\ell)}(\lambda_{\ell j})|^2, \qquad p = 1, \ldots, m,$$

*where* $(a_j^{(p)})_{l=l,\ldots,m}$, $p = 1, \ldots, k$, *are real numbers, and* $v^{(\ell)}(\lambda_{\ell j}) = v_h(\lambda_{3j})$ *with tapering*, $v^{(\ell)}(\lambda_{\ell j}) = v(\lambda_j)$ *without tapering, where* $v(\lambda_j)$, $v_h(\lambda_{3j})$ *are given in* (2.8). *Then for any* $k \geq 1$ ,

$$\left| E \prod_{p=1}^{k} [S_m^{(p)} - ES_m^{(p)}] \right| \leq C \prod_{p=1}^{k} \|a^{(p)}\|_2, \tag{7.19}$$

*where* $\|a^{(p)}\|_2 = \{\sum_{j=l}^m (a_j^{(p)})^2\}^{1/2}$ *and* $C < \infty$ *does not depend on* $m$ *or* $a^{(p)}$, *but depends on* $k$.

PROOF. We have

$$\Sigma_m := E\prod_{p=1}^{k}[S_m^{(p)} - ES_m^{(p)}]$$
$$= \sum_{j_1,\dots,j_p=l}^{m}\prod_{p=1}^{k} a_{j_p}^{(p)} E\prod_{p=1}^{k}[|v^{(\ell)}(\lambda_{\ell j_p})|^2 - E|v^{(\ell)}(\lambda_{\ell j_p})|^2].$$

We introduce the table

$$T = \begin{pmatrix} (1,1) & (1,2) \\ \cdots & \cdots \\ (k,1) & (k,2) \end{pmatrix}, \tag{7.20}$$

and define $\eta_{p,1}(j) = v^{(\ell)}(\lambda_{\ell j})$, $\eta_{p,2}(j) = \overline{v^{(\ell)}(\lambda_{\ell j})}$, for $p = 1,\dots,k$. We denote by $\gamma = (V_1,\dots,V_k)$ partitions of $T$ into nonintersecting sets $V_s$ of the form $V_s = \{(p,v),(p',v')\}$ $(p \neq p')$, and write $V_s \in \mathcal{V}_0$ if $v \neq v'$ and $V_s \in \mathcal{V}_1$ if $v = v'$. Denote by $\Gamma = \{\gamma\}$ the set of all partitions $\gamma$ and by $\Gamma_0$ the set of $\gamma = (V_1,\dots,V_k)$ such that $V_s \in \mathcal{V}_0$, $s = 1,\dots,k$. By Gaussianity, we can write, using diagram formalism [see, e.g., Brillinger (1975), page 21],

$$E\Sigma_m = \sum_{\gamma\in\Gamma} Q_\gamma, \tag{7.21}$$

where

$$Q_\gamma = \sum_{j_1,\dots,j_k=l}^{m}\left(\prod_{p=1}^{k} a_{j_p}^{(p)}\right) q_{V_1}\cdots q_{V_k}, \tag{7.22}$$

where, for $V_s = ((p,v),(p',v'))$, $q_{V_s} \equiv q_{V_s}(j_p, j_{p'}) = E[\eta_{p,v}(j_p)\eta_{p',v'}(j_{p'})]$. Set

$$q_{V_s}^* \equiv |a_{j_p}^{(p)} a_{j_{p'}}^{(p')}|^{1/2} |q_{V_s}(j_p, j_{p'})|.$$

Clearly

$$|Q_\gamma| \le \sum_{j_1,\dots,j_k=l}^{m} q_{V_1}^* \cdots q_{V_k}^*. \tag{7.23}$$

Each argument $j_1,\dots,j_k$ in (7.23) belongs to exactly two functions, $q_{V_s}^*$. Therefore, by the Cauchy inequality, we get

$$|Q_\gamma| \le \sum_{j_1,\dots,j_k=l}^{m} q_{V_1}^* \cdots q_{V_k}^* \le \|q_{V_1}^*\|_2 \cdots \|q_{V_k}^*\|_2, \tag{7.24}$$

where

$$\|q_{V_s}^*\|_2 = \left(\sum_{j,k=l}^{m}\{q_{V_s}^*(j,k)\}^2\right)^{1/2}, \qquad s = 1,\dots,k.$$

We now show that

$$\|q^*_{V_s}\|_2 \le C\big(\|a^{(p)}\|_2\|a^{(p')}\|_2\big)^{1/2}, \tag{7.25}$$

which together with (7.24) and (7.21) implies (7.19).

We have

$$\begin{aligned}\|q^*_{V_s}\|_2^2 &\le \sum_{j=l}^{m}|a_j^{(p)}a_j^{(p')}||q_{V_s}(j,j)|^2 + \sum_{l\le k,j\le m\,:\,k\ne j}|a_k^{(p)}a_j^{(p')}||q_{V_s}(j,k)|^2 \\ &=: \|q^*_{V_s,1}\|_2^2 + \|q^*_{V_s,2}\|_2^2.\end{aligned} \tag{7.26}$$

From (a) and (b) of Lemma 2.2 or Lemma 2.1, it follows that $|q_{V_s}(j,j)| \le C$. Thus

$$\|q^*_{V_s,1}\|_2^2 \le C\sum_{j=l}^{m}|a_j^{(p)}a_j^{(p')}| \le C\|a^{(p)}\|_2\|a^{(p')}\|_2. \tag{7.27}$$

With tapering, from (c) and (d) of Lemma 2.2 it follows that

$$\big|q_{V_s}(k,j)\big| \le C\big((m/n)^{\beta}|j-k|^{-2}+\big(\min(k,j)\big)^{-1}|j-k|^{-3/2}\big), \qquad l\le k\ne j\le m.$$

Therefore

$$\begin{aligned}\|q^*_{V_s,2}\|_2^2 &\le C\Bigg(\sum_{l\le k,j\le m\,:\,k\ne j}|a_j^{(p)}a_k^{(p')}|\big((m/n)^{2\beta}|j-k|^{-4} \\ &\qquad\qquad + \big(\min(k,j)\big)^{-2}|j-k|^{-3}\big)\Bigg) \\ &\le C\bigg((m/n)^{2\beta}\|a^{(p)}\|_2\|a^{(p')}\|_2 + \max_{l\le j,k\le m}|a_j^{(p)}a_k^{(p')}|l^{-1}\bigg) \\ &\le C\|a^{(p)}\|_2\|a^{(p')}\|_2.\end{aligned} \tag{7.28}$$

Without tapering, by (c) and (d) of Lemma 2.1,

$$\big|q_{V_s}(k,j)\big| \le Ck^{-|\alpha|/2}|j|^{-1+|\alpha|/2}\log j, \qquad 1\le k<j\le m,$$

and

$$\begin{aligned}\|q^*_{V_s,2}\|_2^2 &\le C\sum_{l\le k,j\le m\,:\,k\ne j}|a_j^{(p)}a_k^{(p')}||q_{V_s}(k,j)|^2 \\ &\le C\|a^{(p)}\|_2\|a^{(p')}\|_2\Bigg(\sum_{l\le k<j\le m}\big|q_{V_s}(k,j)\big|^4\Bigg)^{1/2} \\ &\le C\|a^{(p)}\|_2\|a^{(p')}\|_2.\end{aligned} \tag{7.29}$$

The proof of (7.29) implies also the relation

$$\|q^*_{V_s,2}\|_2^2 \le \max_{l\le j,k\le m} |a_j^{(p)} a_k^{(p')}| \sum_{l\le k<j\le m} k^{-|\alpha|} |j|^{-2+|\alpha|} \log^2 j \tag{7.30}$$
$$\le C \max_{l\le j,k\le m} |a_j^{(p)} a_k^{(p')}| \log^3 m,$$

which we shall use in the proof of Lemma 7.5 below. Equations (7.26)–(7.30) imply (7.25). □

LEMMA 7.4. *Let Assumptions $\alpha$, $f$, $l$ and $m^*$ hold and $Z_q$ be given by* (4.6) *with $\ell = 3$ under tapering and $\ell = 1$ without tapering. Then for any fixed $q \ge 0$ and $k \ge 1$,*

$$E|Z_q|^{2k} < \infty$$

*uniformly in $m$.*

PROOF. Applying Lemma 7.3 to (4.6) with $a_j^{(q)} = m^{-1/2}\nu_j^q$, $p = 1, \dots, 2k$, we get

$$E|Z_q|^{2k} \le C\left(m^{-1}\sum_{j=l}^{m} \nu_j^{2q}\right)^k \le C$$

in view of (7.11). □

LEMMA 7.5. *Let* (4.4) *and Assumptions $\alpha$, $f$, $l$ and $m^*$ hold. Then for any $1 \le q_1, q_2, q_3, q_4 \le 2$,*

$$E[Z_{q_1} Z_{q_2}] = e_{q_1+q_2} + e(q_1+q_2, \ell, \beta)(m/n)^\beta + o(\Delta_m), \tag{7.31}$$

$$E[Z_{q_1} Z_{q_2} Z_{q_3}] = 2m^{-1/2} e_{q_1+q_2+q_3} + O(\widetilde{\Delta}_m^2), \tag{7.32}$$

$$\operatorname{Cum}(Z_{q_1}, Z_{q_2}, Z_{q_3}, Z_{q_4}) = O(\widetilde{\Delta}_m^2), \tag{7.33}$$

$$E[Z_{q_1} Z_{q_2} Z_{q_3} Z_{q_4}] = e_{q_1+q_2} e_{q_3+q_4} + e_{q_1+q_3} e_{q_2+q_4} \tag{7.34}$$
$$+ e_{q_1+q_4} e_{q_2+q_3} + o(\Delta_m),$$

*where $\Delta_m$, $\widetilde{\Delta}_m$ are given by* (4.2).

PROOF. Note first that $\operatorname{Cum}(Z_{q_1}, Z_{q_2}) = E[Z_{q_1} Z_{q_2}]$ and $\operatorname{Cum}(Z_{q_1}, Z_{q_2}, Z_{q_3}) = E[Z_{q_1} Z_{q_2} Z_{q_3}]$. Let $a_j^{(p)} = m^{-1/2}\nu_j^{q_p}$, $p = 1, \dots, k$. Then $Z_{q_p} = S_m^{(p)} - ES_m^{(p)}$. From (7.14) and (7.11),

$$\max_{l\le j\le m} |a_j^{(p)}| \le Cm^{-1/2}\log^2 m, \qquad \|a^{(p)}\|_2 \le C, \qquad p = 1, \dots, k. \tag{7.35}$$

By diagram formalism [see, e.g., Brillinger (1975), page 21], we can write

$$c_k := \operatorname{Cum}(Z_{q_1}, \dots, Z_{q_k}) = \sum_{\gamma \in \Gamma^c} Q_\gamma,$$

where $\Gamma^c \subset \Gamma$ denotes a subset of connected partitions $\gamma = (V_1, \dots, V_k)$ of the table (7.20), $T$, and $Q_\gamma$ is given in (7.22). We show that

$$c_k = \sum_{\gamma \in \Gamma_0^c} Q'_\gamma + O(\widetilde{\Delta}_m^2), \tag{7.36}$$

where $\Gamma_0^c \subset \Gamma_0$ denotes the subset of connected partitions and

$$Q'_\gamma = \sum_{j_1, \dots, j_k = l}^{m} \left( \prod_{p=1}^{k} a_{j_p}^{(p)} \right) q'_{V_1} \cdots q'_{V_k}$$

with $q'_{V_s}(j_{p_1}, j_{p_2}) = \mathbb{1}_{\{j_{p_1} = j_{p_2}\}} q_{V_s}(j_{p_1}, j_{p_2})$ for $V_s = ((p_1, v_2), (p_2, v_2))$.

The derivations (7.26)–(7.30) imply that for any connected partition $\gamma \in \Gamma^c$, $Q_\gamma = Q'_\gamma + r_\gamma$, where

$$|r_\gamma| \le \sum_{p, v=1 : p \ne v}^{k} \|q^*_{V_p,2}\|_2 \|q^*_{V_v,2}\|_2 \prod_{j=1 : j \ne p, v}^{k} \|q^*_{V_p}\|_2.$$

By (7.25), $\|q^*_{V_p}\|_2 \le C$. With tapering, (7.28) and (7.35) imply

$$\|q^*_{V_p,2}\|_2^2 \le C\big((m/n)^{2\beta} + m^{-1} \log^4 m / l^{-1}\big) \le C \widetilde{\Delta}_m^2.$$

Without tapering, from (7.30) it follows that

$$\|q^*_{V_p,2}\|_2^2 \le C\big((m/n)^{2\beta} + m^{-1} \log^7 m\big) \le C \widetilde{\Delta}_m^2.$$

Thus $r_\gamma = O(\widetilde{\Delta}_m^2)$. Then (7.36) follows if we show that $Q'_\gamma = O(\widetilde{\Delta}_m^2)$ for $\gamma \in \Gamma^c \backslash \Gamma_0^c$.

In that case $\gamma$ has at least two different $V_p, V_s \in \mathcal{V}_1$. By the Cauchy inequality,

$$|Q'_\gamma| \le \|q^*_{V_p,1}\|_2 \|q^*_{V_s,1}\|_2 \prod_{j=1 : j \ne p, l}^{k} \|q^*_{V_p}\|_2 = O(\widetilde{\Delta}_m^2)$$

since $\|q^*_{V_p}\|_2 \le C$, $j = 1, \dots, k$, and, for $V_p \in \mathcal{V}_1$, $\|q^*_{V_p,1}\|_2^2 = O(\widetilde{\Delta}_m^2)$. Indeed, if $V_s \in \mathcal{V}_1$ then

$$\|q^*_{V_s,1}\|_2^2 = \max_{1 \le i, j \le m} |a_i^{(p)} a_j^{(p')}| \sum_{j=l}^{m} |q_{V_s}(j, j)| \le C m^{-1} (\log m)^4 \sum_{j=l}^{m} |q_{V_s}(j, j)|.$$

With tapering, from (b) of Lemma 2.2 it follows that $|q_{V_s}(j, j)| \le C j^{-2}$, so $\|q^*_{V_s,1}\|_2^2 \le C m^{-1} \log^4 m / l^{-1} \le C \widetilde{\Delta}_m^2$.

Without tapering, by (b) of Lemma 2.1, $|q_{V_s}(j,j)| \le Cj^{-1}\log j$ and $\|q^*_{V_s,1}\|_2^2 \le Cm^{-1}\log^6 m \le C\widetilde{\Delta}_m^2$. This proves (7.36).

We derive now (7.31)–(7.33) using (7.36).

Let $k=2$. Then $\Gamma_0^c$ consists of one $\gamma=(V_1,V_2)$ such that $V_1=((1,1),(2,2))$, $V_2=((1,2),(2,1))$. By (a) of Lemma 2.2 or Lemma 2.1, under (4.4),

$$q_{V_1,1}(j,j) = 1 + (c_1/c_0)\lambda_{\ell j}^{\beta} + r_j(\ell), \tag{7.37}$$

where $r_j(\ell)$ are as in the proof of Lemma 7.1. Hence

$$\begin{aligned} Q'_\gamma &= m^{-1}\sum_{j=l}^{m} \nu_j^{q_1+q_2}\bigl(q_{V_1,1}(j,j)\bigr)^2 \\ &= m^{-1}\sum_{j=l}^{m} \nu_j^{q_1+q_2}\bigl[1+2(c_1/c_0)\lambda_{\ell j}^{\beta}\bigr] + o(\Delta_m) \\ &= e_{q_1+q_2} + e(q_1+q_2,\ell,\beta)(m/n)^{\beta} + o(\Delta_m) \end{aligned}$$

by (7.11). This and (7.36) imply (7.31) since $\widetilde{\Delta}_m^2 = o(\Delta_m)$.

Let $k=3$. Then $\Gamma_0^c$ consists of two partitions,

$$\begin{aligned} \gamma &= (V_1,V_2,V_3), & V_1 &= \bigl((1,1),(2,2)\bigr), \\ V_2 &= \bigl((2,1),(3,2)\bigr), & V_3 &= \bigl((3,1),(1,2)\bigr) \end{aligned}$$

and

$$\begin{aligned} \gamma &= (V_1,V_2,V_3), & V_1 &= \bigl((1,1),(3,2)\bigr), \\ V_2 &= \bigl((2,1),(1,2)\bigr), & V_3 &= \bigl((3,1),(2,2)\bigr). \end{aligned}$$

For each of these $\gamma$, by (7.37) and (7.11),

$$\begin{aligned} Q'_\gamma &= m^{-3/2}\sum_{j=1}^{m} \nu_j^{q_1+q_2+q_3}\bigl(q_{V_1,1}(j,j)\bigr)^3 \\ &= m^{-3/2}\sum_{j=1}^{m} \nu_j^{q_1+q_2+q_3}\bigl(1+(c_1/c_0)\lambda_{\ell j}^{\beta}+r_j(\ell)\bigr)^3 \\ &= m^{-3/2}\sum_{j=1}^{m} \nu_j^{q_1+q_2+q_3} + O\bigl((m/n)^{2\beta}+m^{-1}\bigr) = e_{q_1+q_2+q_3} + O(\widetilde{\Delta}_m^2). \end{aligned}$$

This and (7.36) prove (7.32).

Let $k=4$. Then

$$Q'_\gamma \le Cm^{-4/2}\sum_{j=1}^{m} \nu_j^{q_1+q_2+q_3+q_4}\bigl(q_{V_1,1}(j,j)\bigr)^4 \le Cm^{-1} = O(\widetilde{\Delta}_m^2)$$

by (7.14) since $|q_{V_1,1}(j,j)| \le C$.

Finally, by Isserlis' formula,

$$E[Z_{q_1}\cdots Z_{q_4}] = E[Z_{q_1}Z_{q_2}]E[Z_{q_3}Z_{q_4}] + E[Z_{q_1}Z_{q_3}]E[Z_{q_2}Z_{q_4}] + E[Z_{q_1}Z_{q_4}]E[Z_{q_2}Z_{q_3}] + \operatorname{Cum}(Z_{q_1}, Z_{q_2}, Z_{q_3}, Z_{q_4}),$$

and (7.31) and (7.33) imply (7.34). □

From Lemma 7.5 and Remark 7.1 we have the corollary.

COROLLARY 7.1. *Let* (4.4) *and Assumptions* $\alpha$, $f$, $l$ *and* $m^*$ *hold. Then as* $n \to \infty$,

$$\begin{aligned} EZ_1^2 &= 1 + 2e(2,\ell,\beta)\left(\frac{m}{n}\right)^{\beta} + o(\Delta_m),\\ EZ_1^3 &= -4m^{-1/2} + o(\Delta_m),\\ EZ_1^2 Z_2 &= 9m^{-1/2} + o(\Delta_m), \end{aligned} \tag{7.38}$$

$$EZ_1Z_2 \to -2, \qquad EZ_1^3 Z_2 \to -6, \qquad EZ_1^4 \to 3. \tag{7.39}$$

**Acknowledgment.** We are grateful to a referee for very perceptive and constructive comments, which have led to a significant improvement of the paper.

## REFERENCES

ANDREWS, D. W. K. and GUGGENBERGER, P. (2003). A bias-reduced log-periodogram regression estimator for the long memory parameter. *Econometrica* **71** 675–712.

ANDREWS, D. W. K. and SUN, Y. (2001). Local polynomial Whittle estimation of long range dependence. Preprint. Cowles Foundation for Research in Economics, Yale Univ.

BENTKUS, R. Y. and RUDZKIS, R. A. (1982). On the distribution of some statistical estimates of a spectral density. *Theory Probab. Appl.* **27** 795–814.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

BRILLINGER, D. R. (1975). *Time Series*: *Data Analysis and Theory*. Holden Day, San Francisco.

CHIBISOV, D. M. (1972). An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. *Theory Probab. Appl.* **17** 620–630.

FOX, R. and TAQQU, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532.

GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.* **4** 221–238.

GIRAITIS, L., ROBINSON, P. M. and SAMAROV, A. (1997). Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. Time Ser. Anal.* **18** 49–60.

GIRAITIS, L., ROBINSON, P. M. and SAMAROV, A. (2000). Adaptive semiparametric estimation of the memory parameter. *J. Multivariate Anal.* **72** 183–207.

HALL, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications. *Statistics* **22** 215–232.

HANNAN, E. J. (1973). The asymptotic theory of linear time series models. *J. Appl. Probab.* **10** 130–145.

HENRY, M. and ROBINSON, P. M. (1996). Bandwidth choice in Gaussian semiparametric estimation of long range dependence. *Athens Conference on Applied Probability and Time Series Analysis. Time Series Analysis. Lecture Notes in Statist.* **115** 220–232. Springer, New York.

HURVICH, C. M. and BRODSKY, J. (2001). Broadband semiparametric estimation of the memory parameter of a long memory time series using fractional exponential models. *J. Time Ser. Anal.* **22** 221–249.

HURVICH, C. M. and CHEN, W. W. (2000). An efficient taper for potentially overdifferenced long memory time series. *J. Time Ser. Anal.* **21** 155–180.

HURVICH, C. M. and RAY, B. (1995). Estimation of the memory parameter for nonstationary or noninvertable fractionally integrated processes. *J. Time Ser. Anal.* **16** 17–42.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation. Asymptotic Theory.* Springer, New York.

JANACEK, G. J. (1982). Determining the degree of differencing for time series via the log spectrum. *J. Time Ser. Anal.* **3** 177–183.

KÜNSCH, H. (1987). Statistical aspects of self-similar processes. In *Proc. First World Congress of the Bernoulli Society* (Yu. A. Prokhorov and V. V. Sazonov, eds.) **1** 67–74. VNU Science Press, Utrecht.

LIEBERMAN, O., ROUSSEAU, J. and ZUCKER, D. M. (2001). Valid Edgeworth expansion for the sample autocorrelation function under long range dependence. *Econometric Theory* **17** 257–275.

MOULINES, E. and SOULIER, P. (1999). Broadband log-periodogram regression of time series with long range dependence. *Ann. Statist.* **27** 1415–1439.

NISHIYAMA, Y. and ROBINSON, P. M. (2000). Edgeworth expansions for semiparametric averaged derivatives. *Econometrica* **68** 931–979.

ROBINSON, P. M. (1995a). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23** 1048–1072.

ROBINSON, P. M. (1995b). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661.

ROBINSON, P. M. (1995c). The approximate distribution of nonparametric regression estimates. *Statist. Probab. Lett.* **23** 193–201.

ROBINSON, P. M. and HENRY, M. (1999). Long and short memory conditional heteroscedasticity in estimating the memory parameter of levels. *Econometric Theory* **15** 299–336.

ROBINSON, P. M. and HENRY, M. (2003). Higher-order kernel semiparametric $M$-estimation of long memory. *J. Econometrics* **114** 1–27.

TANIGUCHI, M. (1991). *Higher-Order Asymptotic Theory for Time Series Analysis. Lecture Notes in Statist.* **68**. Springer, New York.

VELASCO, C. (1999a). Nonstationary log-periodogram regression. *J. Econometrics* **91** 325–371.

VELASCO, C. (1999b). Gaussian semiparametric estimation of nonstationary time series. *J. Time Ser. Anal.* **20** 87–127.

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

DEPARTMENT OF ECONOMICS
LONDON SCHOOL OF ECONOMICS
HOUGHTON STREET
LONDON WC2A 2AE
UNITED KINGDOM
E-MAIL: L.Giraitis@lse.ac.uk
P.M.Robinson@lse.ac.uk