Econometric Theory, **28**, 2012, 457–470.
doi:10.1017/S0266466611000399

ASYMPTOTIC THEORY FOR MAXIMUM LIKELIHOOD ESTIMATION OF THE MEMORY PARAMETER IN STATIONARY GAUSSIAN PROCESSES

OFFER LIEBERMAN
University of Haifa

ROY ROSEMARIN
London School of Economics

JUDITH ROUSSEAU
CEREMADE, University Paris Dauphine

Consistency, asymptotic normality, and efficiency of the maximum likelihood estimator for stationary Gaussian time series were shown to hold in the short memory case by Hannan (1973, *Journal of Applied Probability* 10, 130–145) and in the long memory case by Dahlhaus (1989, *Annals of Statistics* 34, 1045–1047). In this paper we extend these results to the entire stationarity region, including the case of antipersistence and noninvertibility.

1. INTRODUCTION

Let X_t, $t \in \mathbb{Z}$, be a stationary Gaussian time series with mean μ and spectral density $f_\theta(\omega)$, $\omega \in \Pi \equiv [-\pi, \pi]$, and denote the true values of the parameters by μ_0 and θ_0. We are concerned with spectral densities $f_\theta(\omega)$ that belong to the parametric family $\{f_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}$, such that for all $\theta \in \Theta$

$$f_\theta(\omega) \sim |\omega|^{-\alpha(\theta)} L_\theta(\omega) \quad \text{as } \omega \to 0, \tag{1}$$

where $\alpha(\theta) < 1$ and $L_\theta(\omega)$ is a positive function that varies slowly at $\omega = 0$. X_t is said to have long memory (or long-range dependence) if $0 < \alpha(\theta) < 1$, short memory (or short-range dependence) if $\alpha(\theta) = 0$, and antipersistence if $\alpha(\theta) < 0$. The range $\alpha(\theta) \le -1$ corresponds to noninvertibility, and our results cover this case as well. Two examples of parametric models that are consistent with (1) are the fractional Gaussian noise (Mandelbrot and Van Ness, 1968) and the autoregressive fractionally integrated moving average (ARFIMA) models (Granger and Joyeux, 1980; Hosking, 1981).

Address correspondence to Offer Lieberman, Department of Economics, University of Haifa, Haifa 31905, Israel; e-mail: offerl@econ.haifa.ac.il.

The asymptotic properties of the Gaussian maximum likelihood estimator (MLE) for short memory dependent observations were derived by Hannan (1973). For the Gaussian ARFIMA$(0,d,0)$ model, the memory parameter is d, which corresponds to $\alpha(\theta)/2$ in (1). Yajima (1985) proved consistency and asymptotic normality of the MLE when $0 < d < \frac{1}{2}$ and asymptotic normality of the least squares estimator when $0 < d < \frac{1}{4}$. Dahlhaus (1989, 2006) established consistency, asymptotic normality, and efficiency for general Gaussian stationary processes with long memory satisfying (1) and $0 < \alpha < 1$. Similar results for the parametric Gaussian MLE under antipersistence and noninvertibility do not appear to be documented in the literature.

In the semiparametric framework, Robinson (1995b) established consistency and asymptotic normality of the log-periodogram estimator when $-1 < \alpha < 1$. Velasco (1999b) extended these results by showing that consistency still holds for the range $-1 < \alpha < 2$ and asymptotic normality for $-1 < \alpha < 3/2$. Moreover, with a suitable choice of data taper, a modified version of this estimator was shown to be consistent and asymptotically normal for any real α.

For the Whittle MLE, Fox and Taqqu (1986) proved consistency and asymptotic normality under the condition $0 < \alpha < 1$. Velasco and Robinson (2000) extended these results to the range $-1 < \alpha < 2$ and with adequate data tapers, to any degree of nonstationary. Lately, Shao (2010) considered a nonstationarity-extended Whittle estimation that is shown to be consistent and asymptotically normal for any $\alpha > -1$ (except $\alpha = 1, 3, 5, \ldots$) and to enjoy higher efficiency than Velasco and Robinson's (2000) tapered Whittle estimator in the nonstationary case.

The local Whittle estimator was shown by Robinson (1995a) to be asymptotically normal for $-1 < \alpha < 1$, and Velasco (1999a) extended these results by proving consistency for $-1 < \alpha < 2$ and asymptotic normality for $-1 < \alpha < 3/2$. As with the "ordinary" Whittle MLE, with suitable tapering, the results are extended to any $\alpha \geq 1$. Abadir, Distaso, and Giraitis (2007) developed an untapered nonstationarity-extended local Whittle estimation and proved consistency and asymptotic normality when the generating process is linear, for any $\alpha > -3$ (except $\alpha = -1, 1, 3, \ldots$) with higher efficiency than Velasco's (1999a) tapered local Whittle estimator.

For the exact local Whittle estimator, Shimotsu and Phillips (2004) proved asymptotic normality for any real α, if the true mean of the series is known, and Shimotsu (2010) showed that similar results hold in the case where the process has an unknown mean and a linear time trend, for $\alpha \in (-1, 4)$.

The purpose of this paper is to continue this line of literature and fill the gap concerning the asymptotic properties of the Gaussian-MLE by extending it to the case $\alpha < 1$.

Noninvertible processes may arise in practice by over-differencing to eliminate stochastic and deterministic trends; see Beran, Feng, Franke, Hess, and Ocker (2003). Antipersistence behavior was also noticed as a feature of financial time series including, for example, Peters (1994) and Shiryaev (1999), who modeled implied and realized volatility of the Standard & Poor's 500 index, and Karuppiah and Los (2005), who investigated nine foreign exchange rates and concluded that most rates are antipersistent. For other examples of antipersistent processes, we refer the reader to Tsai (2009) and the references therein.

Although there are simulations studies that analyze the performance of these estimators in long- and short memory and antipersistence (see Sowell, 1992; Cheung and Diebold, 1994; Hauser, 1999; Nielsen and Frederiksen, 2005), we emphasize that to date, consistency, asymptotic normality, and efficiency of the Gaussian MLE antipersistence and noninvertibility case have not been established. We prove these properties without making a priori assumptions on the memory type of the series. By this it is meant that the researcher is free to find the MLE over the entire range $\alpha < 1$.

In practice, to date, if the MLE for the memory of a given data set was found to be negative and the process was assumed to have positive memory, the value of the MLE was censored to zero. In various simulation experiments, this resulted in a pileup of MLE values at zero, and essentially this amounts to restricted maximum likelihood estimation, rather than the unrestricted analogue. See, for instance, Lieberman and Phillips (2004). By establishing a theory for the range $\alpha < 1$, the pileup at zero is avoided.

Our set of assumptions is not stronger than those of Dahlhaus (1989, 2006) and is satisfied in the stationary ARFIMA (p, d, q) model, allowing for the possibility that $d \leq -1/2$.

The outline of the paper is as follows. Section 2 states the model and main results of the paper. Section 3 concludes. The Appendix gives the main proof. Auxilliary results are contained in Appendix B of the full version of this paper, available at http://www.ceremade.dauphine.fr/˜rousseau/LRRlongmemo.html.

2. ASSUMPTIONS AND MAIN RESULTS

As in Dahlhaus's (1989) notation, let

$$\nabla g_\theta = \left(\frac{\partial}{\partial\theta_j} g_\theta\right)_{j=1,\dots,p} \quad \text{and} \quad \nabla^2 g_\theta = \left(\frac{\partial^2}{\partial\theta_j \partial\theta_k} g_\theta\right)_{j,k=1,\dots,p}.$$

We denote by $\|A\|$ the spectral norm of an $N \times N$ matrix A and by $|A|$ the Euclidean norm of A, that is,

$$\|A\| = \sup_{x\in\mathbb{C}^n} \left(\frac{x^* A^* A x}{x^* x}\right)^{1/2}, \qquad |A| = \left[\operatorname{tr}\left(A^* A\right)\right]^{1/2},$$

where A^* is the conjugate transpose of A. We require the following assumptions.

Assumption 0. (a) X_t, $t \in \mathbb{Z}$, is a stationary Gaussian sequence with mean $\mu \in \mathbb{R}$ and spectral density $f_\theta(\omega)$, $\omega \in \Pi \equiv [-\pi, \pi]$. The true values of the parameters of the process are μ_0 and $\theta_0 \in \Theta \subseteq \mathbb{R}^p$. If θ and θ' are distinct elements of Θ, we assume that the set $\{\omega | f_\theta(\omega) \neq f_{\theta'}(\omega)\}$ has a positive Lebesgue measure.

(b) The parameter θ_0, lies in the interior of Θ, and Θ is compact.

There exists $\alpha : \Theta \to (-\infty, 1)$ such that for each $\delta > 0$, we have the following.

Assumption 1. $f_\theta(\omega)$, $f_\theta^{-1}(\omega)$, $\partial/\partial\omega f_\theta(\omega)$ are continuous at all (ω, θ), $\omega \neq 0$, and

$$f_\theta(\omega) = O\left(|\omega|^{-\alpha(\theta)-\delta}\right); \qquad f_\theta^{-1}(\omega) = O\left(|\omega|^{\alpha(\theta)-\delta}\right);$$
$$\frac{\partial}{\partial\omega} f_\theta(\omega) = O\left(|\omega|^{-\alpha(\theta)-1-\delta}\right).$$

Assumption 2. $\partial f_\theta(\omega)/\partial\theta_j$ and $\partial^2 f_\theta(\omega)/\partial\theta_j\partial\theta_k$ are continuous at all (ω, θ), $\omega \neq 0$, and

$$\frac{\partial}{\partial\theta_j} f_\theta(\omega) = O\left(|\omega|^{-\alpha(\theta)-\delta}\right), \qquad 1 \le j \le p,$$
$$\frac{\partial^2}{\partial\theta_j\partial\theta_k} f_\theta(\omega) = O\left(|\omega|^{-\alpha(\theta)-\delta}\right), \qquad 1 \le j, k \le p,$$
$$\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} f_\theta(\omega) = O\left(|\omega|^{-\alpha(\theta)-\delta}\right), \qquad 1 \le j, k, l \le p.$$

Assumption 3. $\partial^2 f_\theta(\omega)/\partial\omega\partial\theta_k$ are continuous at all (ω, θ), $\omega \neq 0$, and

$$\frac{\partial^2}{\partial\omega\partial\theta_k} f_\theta(\omega) = O\left(|\omega|^{-\alpha(\theta)-1-\delta}\right), \qquad 1 \le j \le p.$$

Assumption 4. The function $\alpha(\theta)$ is continuous, and the constants appearing in the $O(\cdot)$ above can be chosen independently of θ (not of δ).

We also assume that $\hat{\mu}_N$, the estimator of μ_0, fulfills the following condition.

Assumption 5. For each $\delta > 0$,

$$\hat{\mu}_N = \mu_0 + o_p\left(N^{\{\alpha(\theta_0)-1\}/2+\delta}\right).$$

Assumptions 0–4 are modifications of Dahlhaus's (1989) assumptions (A0), (A2), (A3), and (A7)–(A9). The most important aspect of the assumptions is that $\alpha(\theta)$ may have values in the interval $(-\infty, 1)$, extending Dahlhaus's (1989) assumptions, which limited $\alpha(\theta)$ to the interval $(0, 1)$. Assumption 5 corresponds to the assumption on $\hat{\mu}_N$ in Theorem 3.2 of Dahlhaus (1989). This condition is fulfilled, for example, by the arithmetic mean and linear M-estimates (see Beran, 1991), for $\alpha(\theta_0) \in (-1, 1)$, but Samarov and Taqqu (1988) showed that it does not hold for the arithmetic mean when $\alpha(\theta_0) < -1$. Adenstedt (1974) proved that

Assumption 5 is in fact satisfied for the generalized least squares (GLS) estimator for all $\alpha(\theta_0) < 1$, which is not a feasible estimator, but we can easily extend his result for any estimator $\hat{\mu}_N$ of the form

$$\hat{\mu}_N = \left(\mathbf{1}'\Sigma_N\left(f^*\right)\mathbf{1}\right)^{-1}\mathbf{1}'\Sigma_N\left(f^*\right)\mathbf{X},$$

where $\mathbf{1}$ is an $N \times 1$ vector of 1's, $\mathbf{X} = (X_1, ..., X_N)'$, $\Sigma_N(f_\theta)$ is the covariance matrix of $\mathbf{X}$, given by

$$\Sigma_N(f) = \left[\int_{-\pi}^{\pi} e^{i(r-s)\omega} f(\omega)\, d\omega\right]_{r,s=1,\ldots,N}, \tag{2}$$

$f^* = f_{\theta^*}$, with θ^* any value in Θ satisfying $\alpha(\theta^*) = \inf_{\theta\in\Theta}\alpha(\theta)$ (by compactness of Θ there exists at least one such value), or even $f^*(\lambda) = (1-\cos\lambda)^{-\alpha^*/2}$, where $\alpha^* \le \inf_{\theta\in\Theta}\alpha(\theta)$. Indeed, we can then bound

$$\begin{aligned}
\mathrm{E}\left(\hat{\mu}_N - \mu_0\right)^2 &= \mathrm{E}\left(\left(\mathbf{1}'\Sigma_N\left(f^*\right)^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'\Sigma_N\left(f^*\right)^{-1}(\mathbf{X}-\mu_0\mathbf{1})\right)^2 \\
&\le \left(\mathbf{1}'\Sigma_N\left(f^*\right)^{-1}\mathbf{1}\right)^{-2}\left|\mathbf{1}'\Sigma_N\left(f^*\right)^{-1}\Sigma_N\left(f_{\theta_0}\right)\Sigma_N\left(f^*\right)^{-1}\mathbf{1}\right| \\
&\le \left(\mathbf{1}'\Sigma_N\left(f^*\right)^{-1}\mathbf{1}\right)^{-1}\left\|\Sigma_N\left(f^*\right)^{-1/2}\Sigma_N\left(f_{\theta_0}\right)\Sigma_N\left(f^*\right)^{-1/2}\right\| \\
&\le K N^{\alpha^*-1+(\alpha(\theta_0)-\alpha^*)+\delta} \le K N^{-1+\alpha(\theta_0)+\delta}, \forall\delta > 0,
\end{aligned}$$

where the last line is deduced from Theorem 5.2 of Adenstedt (1974) for the term $\mathbf{1}'\Sigma_N(f^*)^{-1}\mathbf{1}$ and from Lemma 2 of Lieberman, Rosemarin, and Rousseau (2010). Note that this result could be guessed from Theorem 7.2 of Adenstedt, which proved that underestimating α does not change the rate at which the BLUE estimator of μ converges.

Assumptions 0(a) and 1–4 hold if $X_t - \mu_0$ is a fractional Gaussian noise with self-similarity parameter $0 < H < 1$, or a Gaussian ARFIMA process with a differencing parameter $d < \frac{1}{2}$. Finally, note that as in Dahlhaus (1989), our Assumption 1 allows neither a pole nor a zero outside the origin, which excludes processes such as seasonally (possibly fractionally) differenced series.

Denote by $\hat{\theta}_N$ the estimator obtained by minimizing the $-1/N$-normalized Gaussian plug-in log-likelihood function

$$\mathcal{L}_N(\theta) = \frac{1}{2N}\log\det\Sigma_N(f_\theta) + \frac{1}{2N}\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right)'\Sigma_N(f_\theta)^{-1}\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right)$$

with respect to θ. The main results of the paper are stated in the following theorem. It establishes consistency, asymptotic normality, and efficiency of the Gaussian MLE, $\hat{\theta}_N$.

THEOREM 1. *Under Assumptions 0–2, 4, and 5:*

(i) $\hat{\theta}_N \to_p \theta_0$.

(ii) $\sqrt{N}\left(\hat{\theta}_N - \theta_0\right) \underset{d}{\to} N\left(0, \Gamma(\theta_0)^{-1}\right),$ (3)

where $\Gamma(\theta)$ *is the Fisher information matrix, given by*

$$\Gamma(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} (\nabla \log f_\theta(\omega))(\nabla \log f_\theta(\omega))' d\omega.$$

The main effort in the proof is in the establishment of consistency. Because of the nonuniform behavior of the quadratic form $\left(\mathbf{X} - \hat{\mu}_N \mathbf{1}\right)' \Sigma_N(f_\theta)^{-1}\left(\mathbf{X} - \hat{\mu}_N \mathbf{1}\right)$ around $\alpha(\theta_0) - \alpha(\theta) = 1$, implied by Theorem 5 of Lieberman et al. (2010), in our proof we consider separately the regions of θ with $\alpha(\theta_0) - \alpha(\theta) < 1$ and with $\alpha(\theta_0) - \alpha(\theta) \geq 1$. A similar distinction between the two cases was made by Fox and Taqqu (1987), Terrin and Taqqu (1990), Robinson (1995a), and Velasco and Robinson (2000). We derive a uniform limit for the plug-in log-likelihood function that is valid on any compact parameter subspace of Θ in which $\max_\theta(\alpha(\theta_0) - \alpha(\theta)) < 1$. To handle the region of $\theta's$ on which $\alpha(\theta_0) - \alpha(\theta) \geq 1$, we adopt a similar idea to that of Velasco and Robinson (Thm. 1), who proved that in this region, the discrete $-1/N$-normalized Whittle log-likelihood converges to $+\infty$ a.s. as $N \to \infty$.

3. CONCLUSIONS

There is a very large body of literature on long memory processes and, in particular, on the asymptotic properties of various estimators in this context under different conditions. The main contribution that this paper makes is in the establishment of consistency, asymptotic normality, and efficiency of the Gaussian MLE when the memory parameter satisfies $\alpha(\theta_0) < 1$. This range includes all types of memory under stationarity and allows for the possibility of noninvertibility. This work therefore extends Dahlhaus's (1989, 2006) seminal contribution, which was done under long memory only, i.e., under the condition $0 < \alpha(\theta) < 1$. Similar progress has already been made in the semiparametric literature recently (e.g., Velasco, 1999a; 1999b; Velasco and Robinson, 2000; Shimotsu, 2010), but up to this point in time, the results for the parametric Gaussian case were confined to the long memory range only.

REFERENCES

Abadir, K.M., W. Distaso, & L. Giraitis (2007) Nonstationarity-extended local Whittle estimation. *Journal of Econometrics* 141, 1353–1384.

Adenstedt, R.K. (1974) On large-sample estimation for the mean of a stationary random sequence. *Annals of Statistics* 2, 1095–1107.

Beran, J. (1991) M-estimator of location for data with slowly decaying serial correlations. *Journal of the American Statistical Association* 86, 704–708.

Beran, J., Y. Feng, G. Franke, D. Hess, & D. Ocker (2003) Semiparametric modeling of stochastic and deterministic trends and fractional stationarity. In G. Rangarajan &

M. Ding, (eds.) *Processes with Long Range Correlations: Theory and Applications*, pp. 225–250. Springer.

Cheung, Y.W. & F.X. Diebold (1994) On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. *Journal of Econometrics* 62, 301–316.

Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Annals of Statistics* 17, 1749–1766.

Dahlhaus, R. (2006) Correction note: Efficient parameter estimation for self-similar processes. *Annals of Statistics* 34, 1045–1047.

Fox, R. & M.S. Taqqu (1986) Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* 14, 157–532.

Fox, R. & M.S. Taqqu (1987) Central limit theorems for quadratic forms in random variables having long-range dependence. *Probability Theory and Related Fields* 74, 213–240.

Granger, C.W.J. & R. Joyeux (1980) An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–31.

Hannan, E.J. (1973) The asymptotic theory of linear time series models. *Journal of Applied Probability* 10, 130–145.

Hauser, M.A. (1999) Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study. *Journal of Statistical Planning and Inference* 80, 229–255.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165–176.

Karuppiah, J. & C.A. Los (2005) Wavelet multiresolution analysis of high-frequency Asian FX rates, summer 1997. *International Review of Financial Analysis* 14, 211–246.

Lieberman, O. & P.C.B. Phillips (2004) Expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter. *Econometric Theory* 20, 464–484.

Lieberman, O., R. Rosemarin, & J. Rousseau (2010) Asymptotic Theory for Maximum Likelihood Estimation of the Memory Parameter in Stationary Gaussian Processess (full version). Available at http://www.ceremade.dauphine.fr/~rousseau/LRRlongmemo.html.

Mandelbrot, B.B. & J.W. Van Ness (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10, 422–37.

Nielsen, M. & P.H. Frederiksen (2005) Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews* 24, 405–443.

Peters, E.E. (1994) *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*. Wiley.

Polard, D. (1984) *Convergence of Stochastic Processes*. Wiley.

Robinson, P.M. (1995a) Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics* 23, 1630–1661.

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

Samarov, A. & M.S. Taqqu (1988) On the efficiency of the sample mean in long-memory noise. *Journal of Time Series Analysis* 9, 191–200.

Shao, X. (2010) Nonstationarity-extended Whittle estimation. *Econometric Theory* 26, 1060–1087.

Shimotsu, K. (2010) Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory* 26, 501–540.

Shimotsu, K. & P.C.B. Phillips (2004) Exact local Whittle estimation of fractional integration. *Annals of Statistics* 33, 1890–1933.

Shiryaev, A.N. (1999) *Essentials of Stochastic Finance: Facts, Models, Theory*. World Scientific.

Sowell, F.B. (1992) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* 53, 165–188.

Terrin, N. & M.S. Taqqu (1990) A noncentral limit theorem for quadratic forms of Gaussian stationary sequences. *Journal of Theoretical Probability* 3, 449–475.
Tsai, H. (2009) On continuous-time autoregressive fractionally integrated moving average processes. *Bernoulli* 15, 178–194.
Velasco, C. (1999a) Gaussian semiparametric estimation of nonstationary time series. *Journal of Time Series Analysis* 20, 87–127.
Velasco, C. (1999b) Non-stationary log-periodogram regression. *Journal of Econometrics* 91, 325–371.
Velasco, C. & P.M. Robinson (2000) Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association* 95, 1229–1243.
Yajima, Y. (1985) On estimation of long-memory time series models. *Australian Journal of Statistics* 27, 303–20.

APPENDIX

Throughout the Appendix, whenever no confusion occurs, we shall use α to denote $\alpha(\theta)$ with the relevant $\theta \in \Theta$ and α_0 to denote $\alpha(\theta_0)$. Also, K denotes a generic positive bounding constant that may vary from step to step.

Proof of Theorem 1. (i) Set $\mathbf{Y} = \mathbf{X} - \mu_0 \mathbf{1}$ and $\mathbf{Z} = \Sigma_N(\theta_0)^{-1/2}\mathbf{Y}$, so that $\mathbf{Z} \sim \mathcal{N}(0, I_N)$. Using the positive and negative parts of ∇f_θ together with Chebyshev's inequality, Theorem 5.2 of Adenstedt (1974), and Lemma 2 of Lieberman et al. (2010), we have for any $\theta \in \Theta$

$$\left|\mathbf{1}'\Sigma_N^{-1}(\theta)\Sigma_N(|\nabla f_\theta|)\Sigma_N^{-1}(\theta)\mathbf{Y}\right| \le K(\mathbf{Z}'\mathbf{Z})^{1/2}N^{(\alpha(\theta_0)-\alpha(\theta))_+/2+1/2-\alpha(\theta)/2+\delta},$$
$$\forall \delta > 0. \tag{A.1}$$

Using the mean value theorem with mean value θ^* and applying (A.1), we obtain

$$\begin{aligned} A(\theta,\theta') &\equiv |\mathcal{L}_N(\theta) - \mathcal{L}_N(\theta')| \qquad \textbf{(A.2)} \\ &= \frac{1}{2N}\left|\mathbf{Y}'[\Sigma_N(f_\theta)^{-1} - \Sigma_N(f_{\theta'})^{-1}]\mathbf{Y} + \log\det\left[\Sigma_N(f_\theta)\Sigma_N(f_{\theta'})^{-1}\right]\right| \\ &\quad + \frac{(\mu_0 - \hat{\mu}_N)^2}{2N}\left|1'[\Sigma_N(f_\theta)^{-1} - \Sigma_N(f_{\theta'})^{-1}]1\right| \\ &\quad + \frac{|\mu_0 - \hat{\mu}_N|}{N}\left|1'[\Sigma_N(f_\theta)^{-1} - \Sigma_N(f_{\theta'})^{-1}]\mathbf{Y}\right| \le \frac{|\theta-\theta'|}{2N} \\ &\quad \times \Big[|\mathbf{Y}'[\Sigma_N(f_{\theta^*})^{-1}\Sigma_N(\nabla f_{\theta^*})\Sigma_N(f_{\theta^*})^{-1}]\mathbf{Y}| \\ &\qquad + |\mathrm{tr}\left[\Sigma_N(f_{\theta^*})^{-1}\Sigma_N(\nabla f_{\theta^*})\right]|\Big] \\ &\quad + K\frac{|\theta-\theta'|}{2N}\Big(N^{-1+\alpha_0+\delta}\mathbf{1}'\Sigma_N(f_{\theta^*})^{-1}\Sigma_n(|\nabla f_{\theta^*}|)\Sigma_N(f_{\theta^*})^{-1}\mathbf{1} \\ &\quad + KN^{(\alpha_0-\alpha(\theta^*))_+/2+\alpha_0/2-\alpha(\theta^*)/2+\delta}(\mathbf{Z}'\mathbf{Z})^{1/2}\Big), \qquad \forall \delta > 0. \end{aligned}$$

Thus, using Lemma 2 of Lieberman et al. (2010) for the first two terms, there exists a $\gamma \in (1,\infty)$ such that

$$A(\theta,\theta') \le \frac{|\theta-\theta'|K}{2N}\left[\mathbf{Z}'\mathbf{Z}N^{\gamma} + N^{\gamma+1} + N^{\gamma}(\mathbf{Z}'\mathbf{Z})^{1/2}\right], \qquad \forall\delta > 0.$$

Hence, setting $c_N \equiv N^{-\gamma-\epsilon/2}$, for any $\epsilon > 0$

$$P_{\theta_0}\left[\sup_{|\theta-\theta'|<c_N} |A(\theta,\theta')| > \epsilon\right] \le P_{\theta_0}\left(\frac{\mathbf{Z}'\mathbf{Z}}{N} > \epsilon N^{\epsilon/2}\right) + o(1) = o(1). \tag{A.3}$$

Let $L_N(\theta) = \mathcal{L}_N(\theta,\mu_0)$. We see that

$$\begin{aligned}|\mathcal{L}_N(\theta) - L_N(\theta)| &= \frac{1}{2N}\left|\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right)'\Sigma_N(f_\theta)^{-1}\left(\mathbf{X}-\hat{\mu}_N\mathbf{1}\right) - \mathbf{Y}'\Sigma_N(f_\theta)^{-1}\mathbf{Y}\right| \qquad \text{(A.4)}\\ &\le \frac{1}{N}\left|\mu_0-\hat{\mu}_N\right|\left|\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{Y}\right| + \frac{1}{2N}\left|\mu_0-\hat{\mu}_N\right|^2\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}.\end{aligned}$$

Let $\Theta_+(\delta) = \{\theta \in \Theta; \alpha(\theta) \ge \alpha_0; |\theta-\theta_0| \ge \delta\}$, $\Theta_-(\delta) = \{\theta \in \Theta; \alpha(\theta) \le \alpha_0; |\theta-\theta_0| \ge \delta\}$, $\Theta_+ = \Theta_+(0)$, $\Theta_- = \Theta_-(0)$. Using Theorem 5.2 of Adenstedt (1974), we obtain $\sup_{\theta\in\Theta_+}\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1} \le KN^{1-\alpha+\delta} \le KN^{1-\alpha_0+\delta}$, $\forall\delta > 0$ so that together with Assumption 5, this implies

$$\frac{1}{2N}\left|\mu_0-\hat{\mu}_N\right|^2\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1} = o_P(N^{-1+\delta}), \qquad \forall\delta > 0, \tag{A.5}$$

uniformly in $\theta \in \Theta_+$. Similarly, with probability going to one, uniformly in $\theta \in \Theta_+$ and $\forall\delta > 0$,

$$\begin{aligned}\frac{1}{N}\left|\mu_0-\hat{\mu}_N\right|\left|\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{Y}\right| &\le KN^{-3/2+\alpha_0/2+\delta}\left(\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}\right)^{1/2}\left(\mathbf{Z}'\mathbf{Z}\right)^{1/2}|\\ &\quad\times\left\|\Sigma_N(\theta_0)^{1/2}\Sigma_N(\theta)^{-1/2}\right\|\\ &= o_P(1). \qquad \text{(A.6)}\end{aligned}$$

Equations (A.5) and (A.6) imply that (A.4) is $o_P(1)$ uniformly on Θ_+. Together with (A.3), we have for all $\epsilon > 0$

$$\begin{aligned}&P_{\theta_0}\left[\sup_{|\theta'-\theta|<c_N,\theta',\theta\in\Theta_+}|L_N(\theta') - L_N(\theta)| > \epsilon\right]\\ &\quad\le P_{\theta_0}\left[\sup_{|\theta-\theta'|<c_N,\theta',\theta\in\Theta_+}|A(\theta,\theta')| > \epsilon/2\right] + o(1) = o(1). \qquad \text{(A.7)}\end{aligned}$$

We now prove that, for all $\epsilon > 0$, $P_{\theta_0}\left[\inf_{\theta\in\Theta_+(\delta)} L_N(\theta) - L_N(\theta_0) < \epsilon\right] \to_{N\to\infty} 0$. Consider a covering of Θ_+, with balls of radii c_N and centers θ_j, $j = 1,\ldots,J_N$, where $J_N \le KN^{pK_1}$. Such a covering is possible because of the compactness of Θ_+. Applying the chaining lemma (Polard, 1984) and using (A.3), for all $\epsilon > 0$,

$$P_{\theta_0}\left[\inf_{\theta\in\Theta_+(\delta)} L_N(\theta) - L_N(\theta_0) < \epsilon\right] \le P_{\theta_0}\left[\sup_{|\theta'-\theta|<c_N,\theta',\theta\in\Theta_+} |L_N(\theta') - L_N(\theta)| > \epsilon/2\right]$$
$$+ \sum_{j=1}^{J_N} P_{\theta_0}\left[L_N(\theta_j) - L_N(\theta_0) < \epsilon/2\right]$$
$$\le \sum_{j=1}^{J_n} P_{\theta_0}\left[L_N(\theta_j) - L_N(\theta_0) < \epsilon/2\right] + o(1). \quad \textbf{(A.8)}$$

Continuing, each term in (A.8) can be written as

$$P_{\theta_0}\left[L_N(\theta_1) - L_N(\theta_0) < \epsilon/2\right] = P_{\theta_0}\left[\mathbf{Y}'[\Sigma_N(\theta_0)^{-1} - \Sigma_N(\theta_1)^{-1}]\mathbf{Y} - \Delta_N > 0\right],$$

where Δ_N is given by $\Delta_N = -\epsilon N + \log\det[\Sigma_N(\theta_1)\Sigma_N(\theta_0)^{-1}]$. Since $\alpha_0 \le \alpha(\theta_1)$ on Θ_+, an application of Chernoff's inequality, with $0 < s < 1$, yields

$$P_{\theta_0}\left[Y'[\Sigma_N(\theta_0)^{-1} - \Sigma_N(\theta_1)^{-1}]Y - \Delta_N > 0\right] \quad \textbf{(A.9)}$$
$$\le \exp\left\{s\epsilon N/2 - sNK(f_{\theta_0}, f_{\theta_1}) + \frac{s^2}{2}\mathrm{tr}\left[(I_N - \Sigma_N(\theta_1)^{-1}\Sigma_N(\theta_0))^2\right]\right\},$$

where

$$K(f_1, f_2)N = \left(\mathrm{tr}[\Sigma_N(f_1)\Sigma_N(f_2)^{-1} - I_N] - \log\det[\Sigma_N(f_2)^{-1}\Sigma_N(f_1)]\right)/2.$$

As in Dahlhaus (1989, p. 1755), uniformly in Θ_+,

$$K(f_{\theta_0}, f_{\theta_1}) \ge \frac{K}{N}\mathrm{tr}\left[(I_N - \Sigma_N(\theta_0)\Sigma_N(\theta_1)^{-1})^2\right], \quad \textbf{(A.10)}$$

implying that

$$P_{\theta_0}\left[Y'[\Sigma_N(\theta_0)^{-1} - \Sigma_N(\theta_1)^{-1}]Y - \Delta_N > 0\right]$$
$$\times \le \exp\left\{\epsilon N/2 - K\,\mathrm{tr}\left[(I_N - \Sigma_N(\theta_1)^{-1}\Sigma_N(\theta_0))^2\right]\right\}. \quad \textbf{(A.11)}$$

By Theorem 5 of Lieberman et al. (2010), for any $0 < u < 1$, uniformly in $\Theta_+(\delta) \cap \{\theta; \alpha(\theta) \ge -1+u\}$, there exists a $b_1(\delta) > 0$ such that, for a large enough N,

$$\mathrm{tr}\left[(I_N - \Sigma_N(\theta_1)^{-1}\Sigma_N(\theta_0))^2\right] = \frac{N}{4\pi}\int\left(\frac{f_{\theta_0}(\omega)}{f_{\theta_1}(\omega)} - 1\right)^2 d\omega \ge Nb_1(\delta). \quad \textbf{(A.12)}$$

Further, uniformly in $\Theta_+(\delta) \cap \{\theta; \alpha(\theta) < -1+u\}$, because $\alpha(\theta) < 0$, we have $\Sigma_N(\theta_1)^{-1} \ge KI_N$, and

$$\mathrm{tr}\left[(I_N - \Sigma_N(\theta_1)^{-1}\Sigma_N(\theta_0))^2\right] \ge NC^2\int\left(f_{\theta_1}(\omega) - f_{\theta_0}(\omega)\right)^2 d\omega \ge Nb_2(\delta), \quad \textbf{(A.13)}$$

for some $b_2(\delta) > 0$. It follows from (A.9)–(A.13) that we can choose $\epsilon > 0$ small enough such that

$$P_{\theta_0}\left[L_N(\theta_1) - L_N(\theta_0) < \epsilon/2\right] \leq e^{-NKb(\delta)/2}. \tag{A.14}$$

Combining (A.8) with (A.14), we see that, for some constant $K' > 0$,

$$P_{\theta_0}\left[\inf_{\theta\in\Theta_+(\delta)} L_N(\theta) - L_N(\theta_0) < 0\right] \leq K'^{-NKb(\delta)}N^{pK_1} + o(1) = o(1). \tag{A.15}$$

To proceed, we decompose $\Theta_-(\delta)$ as $\Theta_-(\delta) = \Theta_{1-} \cup \Theta_{2-}$, with $\Theta_{1-} = \{\theta \in \Theta_-(\delta); \alpha(\theta) \geq -1 + \alpha_0 + \epsilon'\}$ and $\Theta_{2-} = \Theta_-(\delta) \setminus \Theta_{1-}$, for some small $\epsilon' > 0$. With very similar calculations to those leading to (A.15), we obtain

$$P_{\theta_0}\left[\inf_{\theta\in\Theta_{1-}} \mathcal{L}_N(\theta) - L_N(\theta_0) < \epsilon\right] = o(1). \tag{A.16}$$

We now study the behavior of $\mathcal{L}_n(\theta)$ over Θ_{2-}. Let $c > 0$ and $b \in \mathbb{R}$ be such that $g(x) = cx^{-b} \leq \inf_{\Theta_{2-}} f_\theta(x)$, and $f_2(x) = C|x|^{-\alpha_0+1-\epsilon'}$ such that $f_2(x) \geq \sup_{\Theta_{2-}} f_\theta(x)$. Such functions exists by the compactness of Θ. Note that for all $\theta \in \Theta_{2-}$,

$$\mathcal{L}_N(\theta) \geq \frac{1}{2N}[\mathbf{Y}'\Sigma_N(f_2)^{-1}\mathbf{Y} - 2(\hat{\mu}_N - \mu_0)\mathbf{1}\Sigma_N(f_2)^{-1}\mathbf{Y} + \log|\Sigma_N(g)|].$$

Because $(\hat{\mu}_N - \mu_0)\mathbf{1}\Sigma_N(f_2)^{-1}\mathbf{Y} = o_p(1)$ and the fact that

$$\frac{1}{N}\log|\Sigma_N(g)\Sigma_N(f_0)^{-1}| \to_{N\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi}(\log g(\omega) - \log f_0(\omega))d\omega,$$

uniformly in Θ_{2-}, we have

$$\mathcal{L}_N(\theta) - L_N(\theta_0) \geq \frac{1}{2N}\left[\mathbf{Y}'(\Sigma_N(f_2)^{-1} - \Sigma_N(\theta_0)^{-1})\mathbf{Y}\right] - K,$$

with probability going to 1. If $\alpha_0 \geq 0$, then $\alpha_0 - 1 + \epsilon' > -1$, and by Theorem 5 of Lieberman et al. (2010),

$$\frac{1}{N}\mathrm{tr}\left[\Sigma_N(\theta_0)\Sigma_N(f_2)^{-1} - I_N\right] \geq \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[K\omega^{-1+\epsilon'} - 1\right]d\omega \geq \frac{K}{\epsilon'}.$$

If $\alpha_0 \leq 0$, then Lemma 7 of Lieberman et al. (2010) implies that for all $A > 0$ if ϵ' is sufficiently small

$$\frac{1}{N}\mathrm{tr}\left[\Sigma_N(\theta_0)\Sigma_N(f_2)^{-1} - I_N\right] \geq A.$$

Hence for any $\epsilon > 0$, by setting $A > 2(K+\epsilon)$ and ϵ' small enough, we get

$$\begin{aligned}
&P_{\theta_0}\left[\inf_{\theta\in\Theta_{2-}} \mathcal{L}_N(\theta) - L_N(\theta_0) < \epsilon\right]\\
&\quad \le P_{\theta_0}\left[\frac{1}{2N}\left[\mathbf{Y}'(\Sigma_N(f_2)^{-1} - \Sigma_N(\theta_0)^{-1})\mathbf{Y}\right] \le \epsilon + K\right]\\
&\quad \le P_0\left[\mathbf{Y}'(\Sigma_N(\theta_0)^{-1} - \Sigma_N(f_2)^{-1})\mathbf{Y} + \mathrm{tr}\left[\Sigma_N(\theta_0)\Sigma_N(f_2)^{-1} - I_N\right]\right.\\
&\qquad\qquad \left.\times \ge \frac{1}{2}\mathrm{tr}\left[\Sigma_N(\theta_0)\Sigma_N(f_2)^{-1} - I_N\right]\right]\\
&\quad \le \frac{4\mathrm{tr}\left[\left(\Sigma_N(\theta_0)^{1/2}\Sigma_N(f_2)^{-1}\Sigma_N(\theta_0)^{1/2} - I_N\right)^2\right]}{\left(\mathrm{tr}\left[\Sigma_N(\theta_0)^{1/2}\Sigma_N(f_2)^{-1}\Sigma_N(\theta_0)^{1/2} - I_N\right]\right)^2}\\
&\quad \le \frac{8[|\Sigma_N(\theta_0)^{1/2}\Sigma_N(f_2)^{-1/2}|^2\|\Sigma_N(f_2)^{-1/2}\Sigma_N(\theta_0)^{1/2}\|^2 + N]}{|\Sigma_N(\theta_0)^{1/2}\Sigma_N(f_2)^{-1/2}|^4}\\
&\quad = o(1).
\end{aligned} \tag{A.17}$$

Equations (A.15), (A.16), and (A.17) complete the proof of consistency.

(ii) By the mean value theorem,

$$\nabla\mathcal{L}_N\left(\hat{\theta}_N\right) - \nabla\mathcal{L}_N(\theta_0) = \nabla^2\mathcal{L}_N\left(\overline{\theta}_N, \hat{\mu}_N\right)\left(\hat{\theta}_N - \theta_0\right), \tag{A.18}$$

with $\left|\overline{\theta}_N - \theta_0\right| \le \left|\hat{\theta}_N - \theta_0\right|$. Since θ_0 lies in the interior of Θ, for all $\varepsilon > 0$, $\left(\sqrt{N}\nabla\mathcal{L}_N\left(\hat{\theta}_N\right) > \varepsilon\right) \to_p 0$. Also,

$$\begin{aligned}
\sqrt{N}\nabla\mathcal{L}_N(\theta_0) = &\frac{1}{2\sqrt{N}}\mathrm{tr}\left\{\Sigma_{\theta_0}^{-1}\Sigma_{\nabla,\theta_0}\right\}\\
&-\frac{1}{2\sqrt{N}}\left(\mathbf{X} - \hat{\mu}_N\mathbf{1}\right)'\Sigma_N(f_\theta)^{-1}\Sigma_N(\nabla f_\theta)\Sigma_N(f_\theta)^{-1}\left(\mathbf{X} - \hat{\mu}_N\mathbf{1}\right).
\end{aligned}$$

Using similar decompositions to (A.4), $\sup_{\theta\in\Theta}\sqrt{N}\left|\nabla\mathcal{L}_N(\theta_0) - \nabla L_N(\theta_0)\right| \to_p 0$ and

$$\sqrt{N}\nabla L_N(\theta_0) = \sqrt{N}\nabla^2\mathcal{L}_N\left(\overline{\theta}_N, \hat{\mu}_N\right)\left(\hat{\theta}_N - \theta_0\right) + o_p(1).$$

We now prove that

$$\mathcal{L}_N\left(\overline{\theta}_N, \hat{\mu}_N\right) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\nabla f_{\theta_0}\nabla f_{\theta_0}}{f_{\theta_0}^2}(\omega)d\omega + o_p(1). \tag{A.19}$$

Set $J_N = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\nabla f_{\theta_0} \nabla f_{\theta_0}}{f_{\theta_0}^2}(\omega)d\omega$, and write $\Sigma_\theta = \Sigma_N(f_\theta)$, $\Sigma_{\nabla,\theta} = \Sigma_N(\nabla f_\theta)$, $\Sigma_{\nabla^2,\theta} = \Sigma_N\left(\nabla^2 f_\theta\right)$,

$$
\begin{aligned}
\nabla^2 \mathcal{L}_N(\theta) = & -\frac{1}{2N}\operatorname{tr}\left(\Sigma_\theta^{-1}\Sigma_{\nabla,\theta}\Sigma_\theta^{-1}\Sigma_{\nabla,\theta}\right) + \frac{1}{2N}\operatorname{tr}\left(\Sigma_\theta^{-1}\Sigma_{\nabla^2,\theta}\right) \\
& + \frac{1}{N}\left(\mathbf{X} - \hat{\mu}_N \mathbf{1}\right)' \Sigma_\theta^{-1}\Sigma_{\nabla,\theta}\Sigma_\theta^{-1}\Sigma_{\nabla,\theta}\Sigma_\theta^{-1}\left(\mathbf{X} - \hat{\mu}_N \mathbf{1}\right) \\
& - \frac{1}{2N}\left(\mathbf{X} - \hat{\mu}_N \mathbf{1}\right)' \Sigma_\theta^{-1}\Sigma_{\nabla^2,\theta}\Sigma_\theta^{-1}\left(\mathbf{X} - \hat{\mu}_N \mathbf{1}\right),
\end{aligned}
$$

so that

$$
\nabla^2 \mathcal{L}_N(\theta) - \nabla^2 L_N(\theta) = \frac{1}{N}\left[2\mathbf{Y}' A_\theta \mathbf{1}(\mu_0 - \hat{\mu}_N) + (\mu_0 - \hat{\mu}_N)^2 \mathbf{1}' A_\theta \mathbf{1}\right], \tag{A.20}
$$

where A_θ is a linear combination of matrices of the form $\Sigma_\theta^{-1}\Sigma_{\nabla,\theta}\Sigma_\theta^{-1}\Sigma_{\nabla,\theta}\Sigma_\theta^{-1}$ and $\Sigma_\theta^{-1}\Sigma_{\nabla^2,\theta}\Sigma_\theta^{-1}$. On an application of Lemma 2 of Lieberman et al. (2010), the absolute value of (A.20) is less than or equal to

$$
K\left[N^{-1+\alpha_0/2+(\alpha_0-\alpha(\theta))_+ +\delta}\left(\mathbf{Z}'\mathbf{Z}\right)^{1/2} + N^{-1+\alpha_0+\delta}\right] = o(1),
$$

uniformly on $U_\epsilon(\theta_0) = \{\theta; |\theta - \theta_0| \le \epsilon\}$, with $\epsilon > 0$ small. By similar calculations to those involving (A.2), letting $c_N = N^{-\gamma}$ for some $\gamma > 0$, it can be seen that, for all $\epsilon' > 0$,

$$
P_{\theta_0}\left[\sup_{|\theta-\theta'|\le c_N}\left|\nabla^2 L_N(\theta) - \nabla^2 L_N\left(\theta'\right)\right| > \epsilon'\right] = o(1) \tag{A.21}
$$

and

$$
P_{\theta_0}\left[\left|\nabla^2 \mathcal{L}_N(\theta, \mu_0) - \nabla^2 \mathcal{L}_N(\theta_0, \mu_0)\right| > u\right] \le e^{-cN^{1-2\delta}u^2}, \tag{A.22}
$$

for some $c > 0$ and $\delta < 1/2$, which can be chosen as small as need be. Inequalities (A.21) and (A.22) imply that

$$
P_{\theta_0}\left[\sup_{|\theta-\theta_0|<\epsilon}\left|\nabla^2 \mathcal{L}_N(\theta, \mu_0) - \nabla^2 \mathcal{L}_N(\theta_0, \mu_0)\right| > \epsilon'\right] = o(1).
$$

Lemma 8 or Theorem 5 of Lieberman et al. (2010) implies (A.19). Note that $J_N \ge cI_N$ for some positive constant $c > 0$. Therefore, we set $Z_N = \sqrt{N} J_N^{-1/2} \nabla \mathcal{L}_N(\theta_0, \mu_0)$. Since $\| \Sigma_{\theta_0}^{-1/2} \Sigma_{|\nabla_j|,\theta_0}^{1/2} \|^2 \le CN^\delta$ and since $J_N \ge cI_N$ for N large enough, the following Laplace transform satisfies

$$
\begin{aligned}
\mathrm{E}(t) &= \mathrm{E}_0^N\left[e^{t'Z_N}\right]\\
&= \exp\left\{\frac{t'J_N^{-1/2}\operatorname{tr}\left[\Sigma_{\theta_0}^{-1}\Sigma_{\nabla,\theta_0}\right]}{2\sqrt{N}}\right\}\left|I_N+\frac{t'J_N^{-1/2}\Sigma_{\theta_0}^{-1/2}\Sigma_{\nabla,\theta_0}\Sigma_{\theta_0}^{-1/2}}{\sqrt{N}}\right|^{-1/2}\\
&= \exp\left\{\frac{1}{4N}\operatorname{tr}\left[\left\{\Sigma_{\theta_0}^{-1}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\right\}^2\right]\right.\\
&\qquad +\frac{1}{6N^{3/2}}\operatorname{tr}\left[\left\{\left(I_N+2u\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right)^{-1}\right.\right.\\
&\qquad\qquad \left.\left.\left.\times\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right\}^3\right]\right\},
\end{aligned}
$$

for $u\in(0,1)$.

It is quite easy to verify that

$$
\frac{\operatorname{tr}\left[\left\{\Sigma_{\theta_0}^{-1}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\right\}^2\right]}{4N}=\frac{\sum_{j=1}^k t_j^2}{2}+o(1).
$$

We thus need only prove that the second term is $o(1)$. We have already proved that

$$
\left(I_N+2u\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right)>I_N/2.
$$

Thus the second term is bounded by

$$
\begin{aligned}
&\frac{1}{6N^{3/2}}\operatorname{tr}\left[\left\{\left(I_N+2u\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right)^{-1}\right.\right.\\
&\qquad\left.\left.\times\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right\}^3\right]\\
&\quad\le\frac{4}{3N^{3/2}}\operatorname{tr}\left[\left\{\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right\}^3\right]\\
&\quad\le CN^{-1/2}\|\Sigma_{\theta_0}^{-1/2}\Sigma_{|\nabla_j|,\theta_0}^{1/2}\|^6\\
&\quad\le CN^{\delta-1/2},\qquad\qquad \forall\delta>0=o(1).
\end{aligned}
$$

This leads to $\mathrm{E}(t)=e^{|t|^2/2}(1+o(1))$ for all t, so that $Z_n\to\mathcal{N}(0,I_p)$, and (ii) of Theorem 1 is proved. ■