Asymptotic Theory for Maximum Likelihood Estimation of the Memory Parameter in Stationary Gaussian Processess

By

Offer Lieberman[1]

University of Haifa

Roy Rosemarin

London School of Economics

and

Judith Rousseau

CEREMADE, University Paris Dauphine

---

[1]Address for correspondence: Department of Economics, University of Haifa, Haifa 31905, Israel. E-mail: offerl@econ.haifa.ac.il

Abstract

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) and in the long memory case by Dahlhaus (1989). In this paper, we extend these results to the entire stationarity region, including the case of antipersistence and noninvertibility. In the process of proving the main results, we provide a useful theorem on the limiting behavior of a product of Toeplitz matrices under strictly weaker conditions than those employed by Dahlhaus (1989).

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 $\alpha(\theta) \leq -1$ 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 ARFIMA models (Granger and Joyeux 1980, Hosking 1981).

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 nor-

mality 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 (1995a) established consistency and asymptotic normality of the log-periodogram estimator when $-1 < \alpha < 1$. Velasco (1999a) 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 (2009) considered a nonstationarity-extended Whittle estimation which is shown to be consistent and asymptotically normal for any $\alpha > -1$ (except $\alpha = 1, 3, 5, ...$) 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 (1995b) to be asymptotically normal for $-1 < \alpha < 1$, while Velasco (1999b) 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

2

$\alpha \geq 1$. 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 $\alpha = -1, 1, 3, ...$) with higher efficiency than Velasco's (1999b) 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 (2006) 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 the 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 $\alpha < 1$.

Non-invertible processes may arise in practice by over-differencing to eliminate stochastic and deterministic trends, see Beran et al. (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 S&P500 index, and Karuppiah and Los (2005) who investigated nine FX 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.

While there are simulations studies that analyze the performance of these estimators in long-, short memory and antipersitence, (see Sowell 1992, Cheung and Diebold 1994, Hauser 1999, and Nielsen and Frederiksen 2005), we emphasize that

todate, consistency, asymptotic normality and efficiency of the Gaussian MLE antipersistence and noninvertibility case have not been established. We prove these properties without making *apriori* 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, todate, 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 pile-up 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 (2004a). By establishing a theory for the range $\alpha < 1$, the pile-up at zero is avoided.

Our set of assumptions are not stronger than those of Dahlhaus (1989, 2006) and are 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. Appendix A gives the main proof, while all auxiliary results and their proofs are collected in Appendix B.

2 Assumptions and Main Results

As in Dahlhaus's (1989) notation, let

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

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

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

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

(A.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 $\delta > 0$:

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

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

(A.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

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

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

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

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

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

(A.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.

(A.5) For each $\delta > 0$

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

Assumptions (A.0)-(A.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 (A.5) corresponds to the assumption on $\hat{\mu}_N$ in Theorem 3.2 of Dahlhaus (1989). This condition is

6

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's (1988) showed that it does not hold for the arithmetic mean when $\alpha(\theta_0) < -1$. Adenstedt (1974) proved that (A.5) is in fact satisfied for the 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

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

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

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

$f^* = f_{\theta^*}$, 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 $\alpha^* \leq \inf_{\theta \in \Theta} \alpha(\theta)$. Indeed, we can then bound

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

where the last line is deduced from Theorem 5.2 of Adenstedt (1974) for the term

7

$\mathbf{1}'\Sigma_N\left(f^*\right)^{-1}\mathbf{1}$ and from Lemma 2 in Appendix B. It is to be noted that this result could be guessed from Theorem 7.2 of Adenstedt (1974) where he proved that underestimating α does not change the rate at which the BLUE estimator of μ converges.

Assumptions (A.0)(a) and (A.1)-(A.4) hold if $X_t - \mu_0$ is a fractional Gaussian noise with self-similarity parameter $0 < H < 1$, or Gaussian ARFIMA process with a differencing parameter $d < \frac{1}{2}$. Finally, note that as in Dahlhaus (1989), our Assumption (A.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

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

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 (A.0)-(A.2) and (A.4)-(A.5):*

(i)

$$\hat{\theta}_N \to_p \theta_{0.}$$

(ii)

$$\sqrt{N}\left(\hat{\theta}_N - \theta_0\right) \xrightarrow[d]{} N\left(0, \Gamma\left(\theta_0\right)^{-1}\right),$$ (3)

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

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

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

The reason for the aforementioned break is that the limiting behavior of the trace of a product of Toeplitz matrices is very different when $\alpha\left(\theta_0\right) - \alpha\left(\theta\right) < 1$ and when $\alpha\left(\theta_0\right) - \alpha\left(\theta\right) \geq 1$. These terms appear in the cumulants of the log-likelihood and its derivatives. In the process of proving the main results, we generalize Theorem

9

5.1 of Dahlhaus (1989) and prove it under strictly weaker conditions. This result is of use and interest by its own right. It continues a very long tradition on the limiting behavior of Toeplitz matrices. See, among others, Kac (1954), Grenander and Szegö (1958), Taniguchi (1983), Fox and Taqqu (1987), Avram (1988), Bercu *et. al.* (1997) and Lieberman and Phillips (2004b).

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 2006), but up to this point in time, the results for the parametric Gaussian case were confined to the long memory range only.

References

[1] Abadir, K.M., Distaso, W., Giraitis, L., 2007. Nonstationarity-extended local Whittle estimation. Journal of Econometrics 141, 1353–1384.

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

[3] Avram, F., 1988. On bilinear forms in Gaussian random variables and Toeplitz matrices. Probability Theory and Related Fields 79, 37-45.

[4] Beran, J., 1991. M-estimator of location for data with slowly decaying serial correlations. Journal of the American Statistical Association 86, 704-708.

[5] Beran, J., Feng, Y., Franke, G., Hess, D., Ocker, D., 2003. Semiparametric modeling of stochastic and deterministic trends and fractional stationarity. In Processes with Long Range Correlations: Theory and Applications (G. Rangarajan and M. Ding, eds.) 225–250, Berlin: Springer.

[6] Bercu, B., Gamboa F., Rouault, A., 1997. Large deviations for quadratic forms of Gaussian stationary processes. Stochastic Processes and Their Applications 71, 75-90.

[7] Cheung, Y.W., Diebold, F.X., 1994. On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. Journal of Econometrics 62, 301-316.

[8] Dahlhaus, R., 1989. Efficient parameter estimation for self-similar processes. Annals of Statistics 17, 1749-1766.

[9] Dahlhaus, R., 2006. Correction note: Efficient parameter estimation for self-similar processes. Annals of Statistics 34, 1045-1047.

[10] Fox, R., Taqqu, M.S., 1986. Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Annals of Statistics 14, 157-532.

[11] Fox, R., Taqqu, M.S., 1987. Central limit theorems for quadratic forms in random variables having long-range dependence. Probability Theory and Related Fields 74, 213-240.

[12] Granger, C.W.J., Joyeux, R., 1980. An introduction to long-range time series models and fractional differencing. Journal of Time Series Analysis 1, 15-31.

[13] Grenander, U., Szegö, G., 1958. Toeplitz forms and their applications. University of California Press, Berkeley and Los Angeles.

[14] Hannan, E.J., 1973. The asymptotic theory of linear time series models. Journal of Applied Probability 10, 130-145.

[15] 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.

[16] Hosking, J.R.M., 1981. Fractional differencing. Biometrika 68, 165-176.

[17] Hurvich, C.M., Ray, B.K., 1995. Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processe. Journal of Time Series Analysis 16, 17-41.

[18] Kac, M., 1954. Toeplitz matrices, translation kernels and related problem in probability theory. Duke Mathematics Journal 21, 501-509.

[19] Karuppiah, J., Los, C.A., 2005. Wavelet multiresolution analysis of high-frequency Asian FX rates, Summer 1997. International Review of Financial Analysis 14, 211-246.

[20] Lieberman, O., Phillips, P.C.B., 2004a. Expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter. Econometric Theory 20, 464–484.

[21] Lieberman, O., Phillips, P.C.B., 2004b. Error bounds and asymptotic expansions for Toeplitz product functionals of unbounded spectra. Journal of Time Series Analysis 25, 733-753.

[22] Mandelbrot, B.B. Van Ness, J.W., 1968. Fractional Brownian motions, fractional noises and applications, SIAM Review 10, 422-37.

[23] Nielsen, M., Frederiksen, P.H., 2005. Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. Econometric Reviews 24, 405-443.

[24] Peters, E.E., 1994. Fractal Market Analysis: Applying Chaos Theory to Investment and Economics. New York: Wiley.

[25] Polard, D., 1984. Convergence of Stochastic Processes. Wiley, New-York.

[26] Robinson, P.M., 1995a. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048-1072.

[27] Robinson, P.M., 1995b. Gaussian semiparametric estimation of long-range dependence. Annals of Statistics 23, 1630–1661.

[28] Samarov, A., Taqqu, M.S., 1988. On the efficiency of the sample mean in tong-memory noise. Journal of Time Series Analysis 9, 191-200.

[29] Shao, X., 2009. Nonstationarity-Extended Whittle Estimation. forthcoming in Econometric Theory.

[30] Shimotsu, K., 2006. Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend. Working Papers no. 1061, Queen's University. Econometric Theory, forthcoming in Econometric Theory.

[31] Shimotsu, K., Phillips, P.C.B., 2004. Exact Local Whittle Estimation of Fractional Integration. Annals of Statistics 33, 1890-1933.

[32] Shimotsu K., Phillips, P. C. B., 2006. Local Whittle Estimation of Fractional Integration and Some of its Variants. Journal of Econometrics 130, 209-233.

[33] Shiryaev, A.N., 1999. Essentials of Stochastic Finance: Facts, Models, Theory. River Edge, NJ. World Scientific.

[34] Smith, A., Sowell, F., Zin, S., 1997. Fractional Integration with Drift: Estimation in Small Samples. Empirical Economics 22, 103-116.

[35] Sowell, F.B., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics 53, 165-188.

[36] Taniguchi, M., 1983. On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. Annals of Statistics 11, 157-169.

[37] Terrin, N., Taqqu, M.S., 1990. A noncentral limit theorem for quadratic forms of Gaussian stationary sequences. Journal of Theoretical Probability 3, 449-475.

[38] Tsai, H., 2009. On continuous-time autoregressive fractionally integrated moving average processes. Bernoulli 15, 178-194.

[39] Velasco, C., 1999a. Non-stationary log-periodogram regression. Journal of Econometrics 91, 325-371.

[40] Velasco, C., 1999b. Gaussian semiparametric estimation of nonstationary time series. Journal of Time Series Analysis 20, 87-127.

[41] Velasco, C., Robinson, P.M., 2000. Whittle Pseudo-Maximum Likelihood Estimation for Nonstationary Time Series. Journal of the American Statistical Association 95, 1229-1243.

[42] Yajima, Y., 1985. On estimation of long-memory time series models. Australian Journal of Statistics 27, 303-20.

Appendix A: Main Results

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 which 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 and Lemma 2, we have for any $\theta \in \Theta$

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

$$(4)$$

Using the mean value theorem with mean value θ^* and applying (4), we obtain

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

Thus, using Lemma 2 for the first two terms, there exists a $\gamma \in (1, \infty)$ such that

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

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

$$P_{\theta_0}\left[\sup_{|\theta-\theta'|<c_N} |A(\theta, \theta')| > \epsilon \right] \leq P_{\theta_0}\left(\frac{\mathbf{Z}'\mathbf{Z}}{N} > \epsilon N^{\epsilon/2} \right) + o(1) = o(1). \quad (6)$$

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

$$
\begin{aligned}
|\mathcal{L}_N(\theta) - L_N(\theta)| &= \frac{1}{2N}\left|(\mathbf{X} - \hat{\mu}_N\mathbf{1})' \Sigma_N(f_\theta)^{-1}(\mathbf{X} - \hat{\mu}_N\mathbf{1}) - \mathbf{Y}'\Sigma_N(f_\theta)^{-1}\mathbf{Y}\right| \quad (7) \\
&\leq \frac{1}{N}|\mu_0 - \hat{\mu}_N|\left|\mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{Y}\right| + \frac{1}{2N}|\mu_0 - \hat{\mu}_N|^2 \mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1}.
\end{aligned}
$$

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

$$\frac{1}{2N}|\mu_0 - \hat{\mu}_N|^2 \mathbf{1}'\Sigma_N(f_\theta)^{-1}\mathbf{1} = o_P(N^{-1+\delta}), \quad \forall \delta > 0, \quad (8)$$

17

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

$$
\begin{aligned}
\frac{1}{N} \left| \mu_0 - \hat{\mu}_N \right| \left| \mathbf{1}' \Sigma_N \left(f_\theta \right)^{-1} \mathbf{Y} \right| &\leq KN^{-3/2 + \alpha_0/2 + \delta} \left(\mathbf{1}' \Sigma_N \left(f_\theta \right)^{-1} \mathbf{1} \right)^{1/2} \left(\mathbf{Z}' \mathbf{Z} \right)^{1/2} \left| \right. \\
&\quad \times \left\| \Sigma_N (\theta_0)^{1/2} \Sigma_N (\theta)^{-1/2} \right\| \\
&= o_P(1). \qquad (9)
\end{aligned}
$$

Equations (8) and (9) imply that (7) is $o_P(1)$ uniformly on Θ_+. Together with (6), we have for all $\epsilon > 0$

$$
\begin{aligned}
P_{\theta_0} \left[\sup_{|\theta' - \theta| < c_N, \theta', \theta \in \Theta_+} |L_N(\theta') - L_N(\theta)| > \epsilon \right] &\leq P_{\theta_0} \left[\sup_{|\theta - \theta'| < c_N, \theta', \theta \in \Theta_+} |A(\theta, \theta')| > \epsilon/2 \right] \\
&\quad + o(1) = o(1). \qquad (10)
\end{aligned}
$$

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, $j = 1, ..., J_N$, where $J_N \leq KN^{pK_1}$. Such a covering is possible, because of the compactness of Θ_+.

Applying the chaining lemma (Polard (1984)) and using (6), for all $\epsilon > 0$,

$$P_{\theta_0} \left[\inf_{\theta \in \Theta_+(\delta)} L_N(\theta) - L_N(\theta_0) < \epsilon \right] \leq P_{\theta_0} \left[\sup_{|\theta' - \theta| < c_N, \theta', \theta \in \Theta_+} |L_N(\theta') - L_N(\theta)| > \epsilon/2 \right]$$

$$+ \sum_{j=1}^{J_N} P_{\theta_0} \left[L_N(\theta_j) - L_N(\theta_0) < \epsilon/2 \right]$$

$$\leq \sum_{j=1}^{J_n} P_{\theta_0} \left[L_N(\theta_j) - L_N(\theta_0) < \epsilon/2 \right] + o(1). \quad (11)$$

Continuing, each term in (11) 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 $\alpha_0 \leq \alpha(\theta_1)$ on Θ_+, an application of Chernoff's inequality, with $0 < s < 1$, yields

$$P_{\theta_0} \left[Y'[\Sigma_N(\theta_0)^{-1} - \Sigma_N(\theta_1)^{-1}]Y - \Delta_N > 0 \right] \quad (12)$$

$$\leq \exp \left\{ s\epsilon N/2 - sNK(f_{\theta_0}, f_{\theta_1}) + \frac{s^2}{2} tr \left[(I_N - \Sigma_N(\theta_1)^{-1}\Sigma_N(\theta_0))^2 \right] \right\},$$

where

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

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

$$K(f_{\theta_0}, f_{\theta_1}) \geq \frac{K}{N} tr \left[(I_N - \Sigma_N(\theta_0) \Sigma_N(\theta_1)^{-1})^2 \right], \qquad (13)$$

implying that

$$P_{\theta_0} \left[Y'[\Sigma_N(\theta_0)^{-1} - \Sigma_N(\theta_1)^{-1}]Y - \Delta_N > 0 \right] \leq \exp \left\{ \epsilon N/2 - K tr \left[(I_N - \Sigma_N(\theta_1)^{-1}\Sigma_N(\theta_0))^2 \right] \right\}.$$

$$(14)$$

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

$$tr \left[(I_N - \Sigma_N(\theta_1)^{-1}\Sigma_N(\theta_0))^2 \right] = \frac{N}{4\pi} \int \left(\frac{f_{\theta_0}(\omega)}{f_{\theta_1}(\omega)} - 1 \right)^2 d\omega \geq N b_1(\delta). \qquad (15)$$

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

$$tr \left[(I_N - \Sigma_N(\theta_1)^{-1}\Sigma_N(\theta_0))^2 \right] \geq N C^2 \int \left(f_{\theta_1}(\omega) - f_{\theta_0}(\omega) \right)^2 d\omega \geq N b_2(\delta), \qquad (16)$$

for some $b_2(\delta) > 0$. It follows from (12)-(16) that we can choose $\epsilon > 0$ small enough such that

$$P_{\theta_0} \left[L_N(\theta_1) - L_N(\theta_0) < \epsilon/2 \right] \leq e^{-NKb(\delta)/2}. \qquad (17)$$

20

Combining (11) with (17), we see that for some constant $K' > 0$

$$P_{\theta_0}\left[\inf_{\theta \in \Theta_+(\delta)} L_N(\theta) - L_N(\theta_0) < 0\right] \le K'^{-NKb(\delta)} N^{pK_1} + o(1) = o(1). \qquad (18)$$

To proceed, we decompose $\Theta_-(\delta)$ as $\Theta_-(\delta) = \Theta_{1-} \cup \Theta_{2-}$, with $\Theta_{1-} = \{\theta \in \Theta_-(\delta); \alpha(\theta) \ge -1 + \alpha_0 + \epsilon'\}$ and $\Theta_{2-} = \Theta_-(\delta) \setminus \Theta_{1-}$, for some small $\epsilon' > 0$. With very similar calculations to those leading to (18), we obtain

$$P_{\theta_0}\left[\inf_{\theta \in \Theta_{1-}} \mathcal{L}_N(\theta) - L_N(\theta_0) < \epsilon\right] = o(1). \qquad (19)$$

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

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

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

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

21

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

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

If $\alpha_0 \leq 0$, then Lemma 7 implies that for all $A > 0$ if ϵ' is sufficiently small

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

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

$$
\begin{aligned}
P_{\theta_0} & \left[\inf_{\theta \in \Theta_{2-}} \mathcal{L}_N(\theta) - L_N(\theta_0) < \epsilon \right] \qquad\qquad (20) \\
\leq\ & P_{\theta_0} \left[\frac{1}{2N} \left[\mathbf{Y}'(\Sigma_N(f_2)^{-1} - \Sigma_N(\theta_0)^{-1}) \mathbf{Y} \right] \leq \epsilon + K \right] \\
\leq\ & P_0 \left[\mathbf{Y}'(\Sigma_N(\theta_0)^{-1} - \Sigma_N(f_2)^{-1}) \mathbf{Y} + tr \left[\Sigma_N(\theta_0) \Sigma_N(f_2)^{-1} - I_N \right] \geq \frac{1}{2} tr \left[\Sigma_N(\theta_0) \Sigma_N(f_2)^{-1} - I_N \right] \right] \\
\leq\ & \frac{4 tr \left[\left(\Sigma_N(\theta_0)^{1/2} \Sigma_N(f_2)^{-1} \Sigma_N(\theta_0)^{1/2} - I_N \right)^2 \right]}{(tr \left[\Sigma_N(\theta_0)^{1/2} \Sigma_N(f_2)^{-1} \Sigma_N(\theta_0)^{1/2} - I_N \right])^2} \\
\leq\ & \frac{8 [|\Sigma_N(\theta_0)^{1/2} \Sigma_N(f_2)^{-1/2}|^2 \|\Sigma_N(f_2)^{-1/2} \Sigma_N(\theta_0)^{1/2}\|^2 + N]}{|\Sigma_N(\theta_0)^{1/2} \Sigma_N(f_2)^{-1/2}|^4} \\
=\ & o(1).
\end{aligned}
$$

Equations (18), (19) and (20) complete the proof of consistency.

22

(ii) By the mean value theorem,

$$\nabla \mathcal{L}_N\left(\hat{\theta}_N\right) - \nabla \mathcal{L}_N\left(\theta_0\right) = \nabla^2 \mathcal{L}_N\left(\overline{\theta}_N, \hat{\mu}_N\right)\left(\hat{\theta}_N - \theta_0\right), \tag{21}$$

with $\left|\overline{\theta}_N - \theta_0\right| \leq \left|\widehat{\theta}_N - \theta_0\right|$. 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,

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

Using similar decompositions to (7), $\sup_{\theta\in\Theta}\sqrt{N}\left|\nabla \mathcal{L}_N\left(\theta_0\right) - \nabla L_N\left(\theta_0\right)\right| \to_p 0$ and

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

We now prove that

$$\mathcal{L}_N\left(\overline{\theta}_N, \hat{\mu}_N\right) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\nabla f_{\theta_0}\nabla f_{\theta_0}}{f_{\theta_0}^2}(\omega)d\omega + o_p(1). \tag{22}$$

23

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

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

so that

$$
\nabla^2 \mathcal{L}_N(\theta) - \nabla^2 L_N(\theta) = \frac{1}{N}\left[2\mathbf{Y}' A_\theta \mathbf{1}(\mu_0 - \hat{\mu}_N) + (\mu_0 - \hat{\mu}_N)^2 \mathbf{1}' A_\theta \mathbf{1}\right], \tag{23}
$$

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, the absolute value of (23) is less than or equal to $K\left[N^{-1+\alpha_0/2+(\alpha_0-\alpha(\theta))_+ + \delta}(\mathbf{Z}'\mathbf{Z})^{1/2} + N^{-1+\alpha_0+\delta}\right] = o(1)$, uniformly on $U_\epsilon(\theta_0) = \{\theta; |\theta - \theta_0| \leq \epsilon\}$, with $\epsilon > 0$ small. By similar calculations to those involving (5), letting $c_N = N^{-\gamma}$ for some $\gamma > 0$, it can be seen that for all $\epsilon' > 0$,

$$
P_{\theta_0}\left[\sup_{|\theta - \theta'| \leq c_N} \left|\nabla^2 L_N(\theta) - \nabla^2 L_N(\theta')\right| > \epsilon'\right] = o(1) \tag{24}
$$

and

$$
P_{\theta_0}\left[\left|\nabla^2 \mathcal{L}_N(\theta, \mu_0) - \nabla^2 \mathcal{L}_N(\theta_0, \mu_0)\right| > u\right] \leq e^{-cN^{1-2\delta}u^2}, \tag{25}
$$

24

for some $c > 0$ and $\delta < 1/2$, which can be chosen as small as need be. Inequalities (24) and (25) imply that

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

Lemma 8 or Theorem 5 imply (22). Note that $J_N \geq cI_N$ 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 $\|\Sigma_{\theta_0}^{-1/2}\Sigma_{|\nabla_j|,\theta_0}^{1/2}\|^2 \leq CN^\delta$ and since $J_N \geq cI_N$ for N large enough, the following Laplace transform satisfies

$$
\begin{aligned}
E(t) &= E_0^N\left[e^{t'Z_N}\right] \\
&= e^{\frac{t'J_N^{-1/2}tr\left[\Sigma_{\theta_0}^{-1}\Sigma_{\nabla,\theta_0}\right]}{2\sqrt{N}}}\left|id + 2\frac{t'J_N^{-1/2}\Sigma_{\theta_0}^{-1/2}\Sigma_{\nabla,\theta_0}\Sigma_{\theta_0}^{-1/2}}{2\sqrt{N}}\right|^{-1/2} \\
&= e^{\frac{tr\left[\left\{\Sigma_{\theta_0}^{-1}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\right\}^2\right]}{4N} + \frac{1}{6N^{3/2}}tr\left[\left\{\left(I_N+2u\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right)^{-1}\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right\}^3\right]}.
\end{aligned}
$$

It is quite easy to verify that

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

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

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

25

Thus the second term is bounded by

$$
\frac{1}{6N^{3/2}} tr\left[\left\{\left(I_N + 2u\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right)^{-1}\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right\}^3\right]
$$

$$
\leq \frac{4}{3N^{3/2}} tr\left[\left\{\Sigma_{\theta_0}^{-1/2}\left(t'J_N^{-1/2}\Sigma_{\nabla,\theta_0}\right)\Sigma_{\theta_0}^{-1/2}\right\}^3\right]
$$

$$
\leq CN^{-1/2}\|\Sigma_{\theta_0}^{-1/2}\Sigma_{|\nabla_j|,\theta_0}^{1/2}\|^6
$$

$$
\leq CN^{\delta-1/2}, \quad \forall\delta > 0 = o(1)
$$

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

Appendix B: Auxiliary Results

The following Lemma generalizes Lemma 5.3 of Dahlhaus (1989) to the negative exponents case. The proof is similar to that of Dahlhaus (1989) and is omitted for brevity.

Lemma 2 *Let $f(\omega)$ and $g(\omega)$ be positive, nonnegative and symmetric functions such that $f^{-1}(\omega) = O(|\omega|^\alpha)$, $g(\omega) = O\left(|\omega|^{-\beta}\right)$, and $\alpha, \beta < 1$. Then*

$$
\left\|\Sigma_N(f)^{-1/2}\Sigma_N(g)^{1/2}\right\|^2 = \left\|\Sigma_N(g)^{1/2}\Sigma_N(f)^{-1/2}\right\|^2 = O\left(N^{(\beta-\alpha)_+}\right).
$$

We shall make extensive use of the following Lemma, which is Theorem 2 of Lieberman and Phillips (2004b).

Lemma 3 *Let $f_j(\omega)$ and $g_j(\omega)$ be positive symmetric functions such that $g_j(\lambda) =$*

$O\left(|\lambda|^{-\beta}\right)$, $f_j\left(\lambda\right) = O\left(|\lambda|^{-\alpha}\right)$, $\alpha < 1$, $\beta < 1$, $j = 1, ..., p$, and $p\left(\alpha + \beta\right) < 1$. As-sume that $\forall t > 0$, $\exists M_{t_1}\ M_{t_2}$, such that $\sup_{|\lambda|>t}|f'\left(\lambda\right)| \leq M_{t_1}$ and $\sup_{|\lambda|>t}|g'\left(\lambda\right)| \leq M_{t_2}$. Then $\forall \varepsilon > 0$

$$\left|\frac{1}{N}tr\left(\Pi_{j=1}^{p}\Sigma_N\left(g_j\right)\Sigma_N\left(f_j\right)\right) - \left(2\pi\right)^{2p-1}\int_{-\pi}^{\pi}\Pi_{j=1}^{p}\left(g_j\left(\lambda\right)f_j\left(\lambda\right)\right)d\lambda\right| = O\left(N^{-1+p(\alpha+\beta)_++\varepsilon}\right).$$

We remark that Lemma 3 is a stronger version of Theorem 1(a) of Fox and Taqqu (1987), who gave an $o\left(1\right)$ upper bound instead.

The following result is the main building block in the proof of Theorem 5, presented below.

Lemma 4 *Let $f_j\left(\omega\right)$ and $g_j\left(\omega\right)$ be positive symmetric functions such that $g_j\left(\lambda\right) = O\left(|\lambda|^{-\beta-\delta}\right)$, $f_j\left(\lambda\right) = O\left(|\lambda|^{-\alpha-\delta}\right)$, $f_j^{-1}\left(\lambda\right) = O\left(|\lambda|^{\alpha-\delta}\right)$, $\forall \delta > 0$, $-1 < \alpha < 1$, $\beta < 1$, $j = 1, ..., p$. Assume that $\forall t > 0$, $\exists M_{t_1}\ M_{t_2}$, such that $\sup_{|\lambda|>t}\left|f'_j\left(\lambda\right)\right| \leq M_{t_1}$ and $\sup_{|\lambda|>t}\left|g'_j\left(\lambda\right)\right| \leq M_{t_2}$, $j = 1, ..., p$. Let $h_j\left(\omega\right) = \left(2\pi\right)^2 f_j\left(\omega\right)$. Then $\forall \delta > 0$*

$$\left|tr\left(\Pi_{j=1}^{p}\Sigma_N\left(g_j\right)\Sigma_N\left(h_j^{-1}\right)\right) - tr\left(\Pi_{j=1}^{p}\Sigma_N\left(g_j\right)\Sigma_N^{-1}\left(f_j\right)\right)\right| = O\left(N^{p(\beta-\alpha)_++\delta}\right).$$

Proof of Lemma 4: First, we consider the case $p = 1$. Let $A_1 = \Sigma_N(g)\Sigma_N(f)^{-1}$

and $B_1 = \Sigma_N(g)\Sigma_N(h^{-1})$. We have,

$$
\begin{aligned}
tr[A_1 - B_1] &= tr[\Sigma_N(g)\Sigma_N^{-1}(f)(I_N - \Sigma_N(f)\Sigma_N(h^{-1}))] \\
&= tr[\Sigma_N(g)(\Sigma_N^{-1}(f) - \Sigma_N(h^{-1}))(I_N - \Sigma_N(f)\Sigma_N(h^{-1}))] \\
&\quad + tr[\Sigma_N(g)\Sigma_N(h^{-1})(I_N - \Sigma_N(f)\Sigma_N(h^{-1}))] \\
&= a_N + b_N,
\end{aligned}
$$

say. The first term is

$$
\begin{aligned}
a_N &= tr[\Sigma_N(g)(\Sigma_N^{-1}(f) - \Sigma_N(h^{-1}))(I_N - \Sigma_N(f)\Sigma_N(h^{-1}))] \\
&\le \|\Sigma_N(g)^{1/2}\Sigma_N^{-1/2}(f)\|^2 |I_N - \Sigma_N^{1/2}(f)\Sigma_N(h^{-1})\Sigma_N^{1/2}(f)|^2.
\end{aligned}
$$

By Lemma 2, the first term above is $O(N^{(\beta-\alpha)_+ + \delta})$ and by Lemma 3, the second

term is $O(N^\delta)$. Hence, $a_N = O(N^{(\beta-\alpha)_+ + 2\delta})$. Further, using Lemma 3, in the case

$p(\beta - \alpha) < 1$ and Theorem 1(b) of Fox and Taqqu (1987) in the case $p(\beta - \alpha) \ge 1$,

$$
b_N = tr[\Sigma_N(g)\Sigma_N(h^{-1})(I_N - \Sigma_N(f)\Sigma_N(h^{-1}))] = O(N^{(\beta-\alpha)_+ + \delta}).
$$

Next, we consider the case $p > 1$. Let $A_j = \Sigma_N(g_j)\Sigma_N(f_j)^{-1}$, $\tilde{A}_j = \Sigma_N^{1/2}(g_j)\Sigma_N^{-1}(f_j)\Sigma_N^{1/2}(g_{j+1})$,

$B_j = \Sigma_N(g_j)\Sigma_N(h_j^{-1})$, $\tilde{B}_j = \Sigma_N^{1/2}(g_j)\Sigma_N(h_j^{-1})\Sigma_N^{1/2}(g_{j+1})$, $\tilde{\Delta}_j = I_N - \Sigma_N^{1/2}(f_j)\Sigma_N(h_j^{-1})\Sigma_N^{1/2}(f_j)$,

28

and $\Delta_j = I_N - \Sigma_N(f_j)\Sigma_N\left(h_j^{-1}\right)$. Note that

$$
\begin{aligned}
|\tilde{A}_j - \tilde{B}_j| &= |\Sigma_N^{1/2}(g_j)\Sigma_N^{-1/2}(f_j)\tilde{\Delta}_j\Sigma_N^{-1/2}(f_j)\Sigma_N^{1/2}(g_{j+1})| \\
&\leq |\tilde{\Delta}_j|\|\Sigma_N^{1/2}(g_j)\Sigma_N^{-1/2}(f_j)\|\|\Sigma_N^{1/2}(g_{j+1})\Sigma_N^{-1/2}(f_j)\| \\
&= O(N^{(\beta-\alpha)_+ +\delta}), \forall \delta > 0.
\end{aligned}
$$

We proceed by using complete induction. Assume that for all $k < p$

$$
|\prod_{j=1}^{k}\tilde{A}_j - \prod_{j=1}^{k}\tilde{B}_j| = O(N^{k(\beta-\alpha)_+ +\delta}), \forall \delta > 0. \tag{26}
$$

Applying the decomposition used in eq'n (13) of Dahlhaus (1989) and the inequality $|A + B| \leq |A| + |B|$, we have

$$
|\prod_{j=1}^{p}\tilde{A}_j - \prod_{j=1}^{p}\tilde{B}_j| \leq \sum_{k=1}^{p}|\prod_{j=1}^{k-1}\tilde{B}_j(\tilde{A}_k - \tilde{B}_k)\prod_{j=k+1}^{p}\tilde{A}_j| = \sum_{k=1}^{p}a_k, \tag{27}
$$

say. In (27) and elsewhere, when $k = 1$, a term such as $\prod_{j=1}^{k-1}\tilde{B}_j$ is simply equal unity. We have,

$$
\begin{aligned}
a_k &= \leq |(\prod_{j=1}^{k-1}\tilde{B}_j - \prod_{j=1}^{k-1}\tilde{A}_j)(\tilde{A}_k - \tilde{B}_k)\prod_{j=k+1}^{p}\tilde{A}_j| + |\prod_{j=1}^{k-1}\tilde{A}_j(\tilde{A}_k - \tilde{B}_k)\prod_{j=k+1}^{p}\tilde{A}_j| \\
&\leq |\prod_{j=1}^{k-1}\tilde{B}_j - \prod_{j=1}^{k-1}\tilde{A}_j|\,|\tilde{A}_k - \tilde{B}_k|\prod_{j=k+1}^{p}\left\|\tilde{A}_j\right\| \\
&\quad + \prod_{j=1}^{k-1}\left\|\tilde{A}_j\right\|\left|\tilde{A}_k - \tilde{B}_k\right|\prod_{j=k+1}^{p}\left\|\tilde{A}_j\right\|. \tag{28}
\end{aligned}
$$

29

By the induction hypothesis (26), (28) is bounded by two terms of the order $O\left(N^{p(\beta-\alpha)_++\delta}\right)$. Hence,

$$|\prod_{j=1}^{p}\tilde{A}_j - \prod_{j=1}^{p}\tilde{B}_j| = O\left(N^{p(\beta-\alpha)_++\delta}\right),\tag{29}$$

as required. Next,

$$
\begin{aligned}
tr\left[\prod_{j=1}^{p}A_j - \prod_{j=1}^{p}B_j\right] &= \sum_{k=1}^{p}tr\left[\prod_{j=1}^{k-1}B_jB_k\Delta_k\prod_{j=k+1}^{p}B_j\right]\\
&+ \sum_{k=1}^{p}tr\left[\prod_{j=1}^{k-1}B_jB_k\Delta_k(\prod_{j=k+1}^{p}A_j - \prod_{j=k+1}^{p}B_j)\right]\\
&+ \sum_{k=1}^{p}tr\left[\prod_{j=1}^{k-1}B_j(A_k-B_k)\Delta_k\prod_{j=k+1}^{p}A_j\right].
\end{aligned}\tag{30}
$$

By Lemma 3 and Theorem 1(b) of Fox and Taqqu (1987), the first term in (30) is

$$tr\left[\prod_{j=1}^{k-1}B_jB_k\Delta_k\prod_{j=k+1}^{p}B_j\right] = tr\left(\prod_{j=1}^{p}B_j - \prod_{j=1}^{k}B_j\Sigma_N\left(f_k\right)\Sigma_N\left(f_k^{-1}\right)\prod_{j=k+1}^{p}B_j\right) = O(N^{p(\beta-\alpha)_++\delta}).$$

The second term in (30) involves

$$
\begin{aligned}
|b_k| &\equiv \left|tr\left(\prod_{j=1}^{k-1}B_jB_k\Delta_k(\prod_{j=k+1}^{p}A_j - \prod_{j=k+1}^{p}B_j)\right)\right|\\
&\leq \left|\prod_{j=1}^{k-1}\tilde{B}_j\Sigma_N^{1/2}(g_k)\Sigma_N(f_k^{-1})(I_N - \Sigma_N(f_k)\Sigma_N(f_k^{-1}))\Sigma_N^{1/2}(g_{k+1})\right|\left|\prod_{j=k+1}^{p}\tilde{A}_j - \prod_{j=k+1}^{p}\tilde{B}_j\right|.
\end{aligned}
$$

30

The first term on the rhs of the above inequality can be written as the square root of

$$tr(\prod_{j=1}^{k}\Sigma_N(g_j)\Sigma_N\left(h_j^{-1}\right)(I_N-\Sigma_N(f_k)\Sigma_N(f_k^{-1}))\Sigma_N(g_{k+1})(I_N-\Sigma_N(f_k^{-1})\Sigma_N(f_k))$$
$$\times \prod_{j=k}^{2}\left(\Sigma_N\left(h_j^{-1}\right)\Sigma_N(g_j)\right)\Sigma_N(f_1^{-1}))=O(N^{2k(\beta-\alpha)_++\delta}),$$

so that, using (29), $|b_k|=O(N^{p(\beta-\alpha)_++\delta})$. Finally, the last term in (30) involves

$$
\begin{aligned}
|c_k| &\equiv \left|tr\left[\prod_{j=1}^{k-1}B_j(A_k-B_k)\Delta_k\prod_{j=k+1}^{p}A_j\right]\right|\\
&\leq |\prod_{j=1}^{k-1}\tilde{B}_j-\prod_{j=1}^{k-1}\tilde{A}_j||\Sigma_N^{1/2}(g_k)\Sigma_N^{-1/2}(f_k)\tilde{\Delta}_k^2\Sigma_N(f_k)^{-1/2}\Sigma_N^{1/2}(g_{k+1})\prod_{j=k+1}^{p}\tilde{A}_j|\quad(31)\\
&\quad+\|\Sigma_N^{1/2}(g_k)\Sigma_N^{-1/2}(f_k)\|\|\tilde{\Delta}_k\|^2\|\Sigma_N(f_k)^{-1/2}\Sigma_N^{1/2}(g_{k+1})\|\prod_{j=k+1}^{p}\left\|\tilde{A}_j\right\|\prod_{j=1}^{k-1}\left\|\tilde{A}_j\right\|\\
&= O(N^{p(\beta-\alpha)_++\delta}),
\end{aligned}
$$

where, for the first term (31), we have used (29). Hence, we have completed the proof of the Lemma. ∎

The following Theorem provides a uniform version of Theorem 5.1 of Dahlhaus (1989), allowing for the possibility of antipersistence. Part (a) of Theorem 5 was stated in Theorem 5.1 of Dahlhaus (1989) under the stronger condition $p(\beta-\alpha)<1/2$. This condition is too restrictive for our larger parameter space. Hence, we need the result under the condition $p(\beta-\alpha)<1$, as stated below.

Theorem 5 *Let Θ^* be a compact subset of $\Theta \cap \{\theta : -1 < \alpha(\theta) < 1\}$, $p \in \mathbb{N} \cup \{0\}$, and $f_{\theta,j}(\omega)$ and $g_{\theta,j}(\omega)$, $j = 1, ..., p$, be symmetric, real-valued functions on Π. Suppose that for each $j = 1, ..., p$, $f_{\theta,j}(\omega)$ satisfies Assumptions (A.1) and (A.4) on Θ^*, with exponent $\alpha(\theta)$, and that for each $\delta > 0$, $|g_{\theta,j}(\omega)| \le K(\delta) |\omega|^{-\beta(\theta)-\delta}$, as $|\omega| \to 0$, with $\beta(\theta) < 1$, continuous on Θ^*. We have the following.*

(a)   If $p(\beta(\theta) - \alpha(\theta)) < 1$ for all $\theta \in \Theta^*$,

$$\lim_{N \to \infty} \frac{1}{N} tr \left[\prod_{j=1}^{p} \left\{ \Sigma_N (f_{\theta,j})^{-1} \Sigma_N (g_{\theta,j}) \right\} \right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \prod_{j=1}^{p} \frac{g_{\theta,j}(\omega)}{f_{\theta,j}(\omega)} \right\} d\omega$$

uniformly in $\theta \in \Theta^*$.

(b)   If $p(\beta(\theta) - \alpha(\theta)) \ge 1$ for all $\theta \in \Theta^*$,

$$tr \left[\prod_{j=1}^{p} \left\{ \Sigma_N (f_{\theta,j})^{-1} \Sigma_N (g_{\theta,j}) \right\} \right] = o \left(N^{p(\beta(\theta)-\alpha(\theta))+\delta} \right), \forall \delta > 0,$$

uniformly in $\theta \in \Theta^*$.

Proof of Theorem 5: The proof is an immediate consequence of Lemmas 3 and 4. ∎

Lemma 6 *Let A be an $n \times n$ symmetric matrix satisfying $\|A\| \le K N^\delta$, for some $0 < \delta < 1/2$. If $Z \sim \mathcal{N}(0, I_N)$, then for all $\epsilon > 0$, there exists a $c > 0$, such that*

$$P \left[Z' A Z - tr(A) \ge \epsilon N \right] \le \exp \left(-c N^{1-2\delta} \epsilon^2 \right).$$

Proof of Lemma 6: We prove the lemma by applying Chernoff's inequality. For all $s \leq N^{-\delta}/4K$, with $1/2 > \delta > 0$, $I_N - sA > I_N/2$ for N large enough and

$$P_{\theta_0} \left[Z'AZ - tr\left(A \right) > \epsilon N \right] \leq \exp \left(-s\epsilon N - str\left(A \right) - \frac{1}{2} \log \left| I_N - 2sA \right| \right).$$

Using a second-order Taylor expansion of $\log \left| I_N - 2sA \right|$ around zero, bounding $I_N - 2sA$ from below by $I_N/2$, and using the fact that $tr\left(A^2 \right) \leq N\|A\|^2$,

$$
\begin{aligned}
P_{\theta_0} \left[Z'AZ - tr[A] > \epsilon N \right] &\leq \exp \left(-s\epsilon N + 2s^2 K N^{1+2\delta} \right) \\
&\leq \exp \left(-\frac{\epsilon^2 N^{1-2\delta}}{8K} \right),
\end{aligned}
$$

which completes the proof. ∎

The following lemma deals with the case in which the spectral density is of a non-invertible process.

Lemma 7 Let $g_2(\omega) = C \left| \omega \right|^{-\alpha_0 + 1 - \epsilon}$, $g_0(\omega) = C \left| \omega \right|^{-\alpha_0}$ and $\alpha_0 \leq 0$. For all $A > 0$, there exists an $\epsilon_A > 0$ and an $N_A > 0$ such that $\forall \epsilon \leq \epsilon_A$ and $\forall N \geq N_A$,

$$tr \left[\Sigma_N(g_2)^{-1} \Sigma_N(g_0) \right] \geq AN.$$

Proof of Lemma 7: Note that we cannot use directly Lemma 3 since when $\alpha_0 < 0$ $1 - \alpha_0 > 1$. For $M > 0$, let $g_M(\omega) = g_2(\omega)$ if $g_2(\omega) \geq 1/M$ and $g_M(\omega) = 1/M$ if $g_2(\omega) < 1/M$. Then g_M is bounded from below and above. For all $t > 0$, by Lemma

3,

$$|I_N - \Sigma_N^{1/2}(g_M)\Sigma_N(1/(4\pi^2 g_M))\Sigma_N^{1/2}(g_M)| \le K(M)N^t,$$

and because $\alpha_0 \le 0$,

$$\left| \frac{1}{N}tr\left[\Sigma_N(1/(4\pi^2 g_M))\Sigma_N(g_0)\right] - \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{g_0}{g_M} d\omega \right| \le K(M)N^{-1+t}.$$

Putting both relations together, and using the same argument as in Lemma 4, we get

$$\begin{aligned} tr[\Sigma_N(g_M)^{-1}\Sigma_N(g_0)] &\ge tr\left[\Sigma_N(1/(4\pi^2 g_M))\Sigma_N(g_0)\right] - K(M)N^t \\ &\ge N\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{g_0}{g_M} d\omega - K(M)N^{t-1}\right]. \end{aligned}$$

Note that $g_M(\omega) = g_2(\omega)$ iff $|\omega| \ge M^{-1/(1-\alpha_0-\epsilon)}$. Therefore

$$tr[\Sigma_N(g_M)^{-1}\Sigma_N(g_0)] \ge N\left[\frac{1}{\pi\epsilon}\left(\pi^\epsilon - M^{-\epsilon/(1-\alpha_0-\epsilon)}\right) - K(M)N^{t-1}\right].$$

Let $M > 1$ and $\epsilon > 0$ small enough. Using the inequalities $\pi^\epsilon \ge 1$ and $1 - e^{-z} \ge z/2$ where the latter applies for any $z \in (0, z_0)$ for some fixed unspecified positive z_0, we get that if $\epsilon \le \epsilon_M \equiv \min(\frac{z_0}{\log M}, 1 - \alpha_0)$, $\left(\pi^\epsilon - M^{-\epsilon/(1-\alpha_0-\epsilon)}\right)\epsilon^{-1} \ge \frac{1}{2(1-\alpha_0-\epsilon)}\log(M)$, Hence for large enough N,

$$tr[\Sigma_N(g_M)^{-1}\Sigma_N(\theta_0)] \ge \frac{N\log M}{\pi(1-\alpha_0-\epsilon)} - K(M)N^t \ge \frac{N\log M}{2(1-\alpha_0-\epsilon)}.$$

34

We remark that $g_2 \leq g_M$ and hence, $\Sigma_N(g_2)^{-1} \geq \Sigma_N(g_M)^{-1}$ for all M. Setting $\text{logM} = a_0/\epsilon$ and ϵ small enough, $N^{-1}tr[\Sigma_N(g_2)^{-1}\Sigma_N(\theta_0)]$ can be made as large as desired. ∎

The following Lemma is the analogue of Theorem 5 when the parameter is in the noninvertible region.

Lemma 8 *Let $p \geq 1$ and f_θ, $f_{j,\theta}$, $j = 1, ..., p$ be symmetric, real-valued functions on Π., where*

$$f_{j,\theta}(\omega) = O(|\omega|^{-\alpha(\theta)-\delta}), \quad f_\theta(\omega) = O(|\omega|^{-\alpha(\theta)}), \quad \alpha(\theta) < 1, \quad \forall \delta > 0,$$

Let $\theta_0 \in \Theta$ such that $\alpha(\theta_0) \leq -1 + \epsilon'$, for some $0 < \epsilon' < 1/2$. Assume that the function α is continuous on $\{|\theta - \theta_0| \leq \epsilon'\}$ and assume that $f_{j,\theta}, g_\theta$ are continuous in $\theta \in \Theta, \omega \neq 0$ and satisfy

$$\frac{\partial f_{j,\theta}(\omega)}{\partial \omega} = O\left(|\omega|^{-\alpha(\theta)-1-\delta}\right), \frac{\partial f_\theta(\omega)}{\partial \omega} = O\left(|\omega|^{-\alpha(\theta)-1-\delta}\right), \quad \forall \delta > 0.$$

Then if $\epsilon > 0$ is small enough, there exists $1 > c > 0$, such that uniformly on $\{\theta; |\theta - \theta_0| \leq \epsilon\}$,

$$\left| \frac{1}{N}tr\left[\prod_{j=1}^{p} \Sigma_N(f_\theta)^{-1}\Sigma_N(f_{j,\theta})\right] - \frac{1}{2\pi}\int_{-\pi}^{\pi} \prod_{j=1}^{p} \frac{f_{j,\theta}(\omega)}{f_\theta(\omega)}d\omega \right| \leq N^{-c}.$$

Proof of Lemma 8: For $M > 0$, consider the function $g_M(\omega) = f_\theta(\omega)$ if $f_\theta(\omega) \geq 1/M$ and $g_M(\omega) = 1/M$ if $f_\theta(\omega) < 1/M$. For all x, y such that $xy > 0$,

$$\sup_{|\theta - \theta_0| \leq \epsilon} |1/g_M(x) - 1/g_M(y)| \leq C|x - y|[M^{3+\alpha_0+\delta} + 1] \quad \forall \delta > 0.$$

Therefore there exists a $\gamma \geq 0$, such that uniformly in $\{\theta; |\theta - \theta_0| \leq \epsilon\}$,

$$\left| \frac{1}{N} tr \left[\left(\Sigma_N(f_\theta) \Sigma_N(1/(4\pi^2 g_M)) - I_N \right)^2 \right] - \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_\theta(\omega)}{g_M(\omega)} d\omega - 1 \right)^2 \right| \leq KM^\gamma N^{-1+t},$$

$$(32)$$

for any $t > 0$ and $N \geq N_0$ where N_0 is some large integer, independent of M. Because

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{f_\theta(\omega)}{g_M(\omega)} d\omega - 1 \right)^2 \leq K \int_{f_\theta(\omega) \leq 1/M} d\omega \leq KM^{1/\alpha(\theta)}, \quad \text{with } \alpha(\theta) < 0,$$

(32) implies that

$$tr \left[\left(\Sigma_N(f_\theta) \Sigma_N(1/(4\pi^2 g_M)) - I_N \right)^2 \right] \leq KN \left(M^{1/\alpha(\theta)} + M^\gamma N^{-1+t} \right).$$

Choose $M = N^r$ with $r < 1/\gamma$. Then

$$tr \left[\left(\Sigma_N(f_\theta) \Sigma_N(1/(4\pi^2 g_M)) - I_N \right)^2 \right] \leq K(N^{1+r/\alpha(\theta)} + N^{t+r\gamma}) = o(N), \quad \forall t > 0.$$

For simplicity's sake we not $f_{j,\theta}$ by f_j in the following calculations. Using the developments in the proof of Lemma 4, we have

$$
\left| tr \left[\prod_{j=1}^{p} \Sigma_N(f_\theta)^{-1} \Sigma_N(f_j) \right] - tr \left[\prod_{j=1}^{p} \Sigma_N(1/(4\pi^2 g_M)) \Sigma_N(f_j) \right] \right|
$$
$$
\leq \ \left| \Sigma_N(f_\theta)^{1/2} \Sigma_N(1/(4\pi^2 g_M)) \Sigma_N(f_\theta)^{1/2} - I_N \right|
$$
$$
\left(\sum_{j=1}^{p} \left| \prod_{i \leq j} \Sigma_N(1/(4\pi^2 g_M)) \Sigma_N(f_i) \right| \prod_{i=j+1}^{p} \| \Sigma_N(f_\theta)^{-1/2} \Sigma_N(f_i)^{1/2} \|^2 \right.
$$
$$
\left. + \sum_{j=2}^{p} \left| \prod_{i<j} \Sigma_N(1/(4\pi^2 g_M)) \Sigma_N(f_i) \right| \prod_{i=j}^{p} \| \Sigma_N(f_\theta)^{-1/2} \Sigma_N(f_i)^{1/2} \|^2 \right)
$$
$$
\leq \ CN^{\delta+1/2} \big(N^{1/2+r/(2\alpha(\theta))} + N^{\delta+r\gamma/2} \big) \big(1 + N^{r\gamma(p)} N^{-1+\delta} \big)
$$
$$
\leq \ N^{1-c}
$$

for some $0 < c < 1/2$, by choosing $r > 0$ small enough. This ends the proof. ∎