

# Edgeworth Expansion for the Whittle Maximum Likelihood Estimator of Linear Regression Processes with Long Memory Residuals

Mosisa Aga

Correspondence: Mosisa Aga, Department of Mathematics, Auburn University at Montgomery, AL, USA

Received: May 13, 2021 Accepted: June 22, 2021 Online Published: June 23, 2021

doi:10.5539/ijsp.v10n4p119 URL: <https://doi.org/10.5539/ijsp.v10n4p119>

## Abstract

We establish an Edgeworth expansion for the distribution of the Whittle maximum likelihood estimator of the parameter of a time series generated by a linear regression model with Gaussian, stationary, and long-memory residuals. This is done by imposing an extra condition on coefficients of the regression model in addition to the standard conditions imposed on the the spectral density function and the parameter values and making use of the results of Andrews et al. (2005), who provided an Edgeworth expansion for the residual component.

**Keywords:** edgeworth expansion, linear regression, long memory, Whittle maximum likelihood, spectral density function

**AMS (2000) Subject Classification Number:** Primary: 62M10

## 1. Introduction

We consider a linear regression model  $\{Y_t = X_t\beta + \epsilon_t, t \geq 1\}$  where  $\beta$  is a  $p$  vector of deterministic but unknown real numbers,  $\{X_t \in \mathbb{R}^p, t \geq 1\}$  are non-stochastic regressors, and the error terms  $\{\epsilon_t, t \geq 1\}$  are stationary, Gaussian, and strongly dependent discrete time series. The process  $\{\epsilon_t, t \geq 1\}$  is assumed to have mean zero and and spectral density  $f_\theta(\lambda)$  for  $\lambda \in (-\pi, \pi)$ , where  $\theta = (\theta_1, \theta_2, \dots, \theta_r)' \in \Theta \subseteq \mathbb{R}^r$  and  $f_\theta(\lambda) = O(|\lambda|^{-2d-\delta})$  as  $|\lambda| \rightarrow 0, \forall \delta > 0, 0 < d < 0.5$ , and  $\theta_1$  is assumed to be the long-memory parameter  $d$  of the process.

Let  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$  be an observed sample of size  $n$  and

$$\mathcal{E} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)'.$$

Then clearly the covariance matrix of  $\mathbf{Y}$  is the same as the covariance matrix of  $\mathcal{E}$ . Let  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$  be the true mean of  $\mathbf{Y}$ . Then, the least square estimate (LSE)  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$  of  $\beta$  is given by  $\hat{\beta} = V^{-1} \sum_{t=1}^n Y_t X_t$ , where  $V = \sum_{t=1}^n (X_t X_t')$  is a  $p \times p$  matrix. Thus, an estimator of  $\boldsymbol{\mu}$  is  $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_n)$ , where  $\hat{\mu}_t = X_t' \hat{\beta}$ ,  $t = 1, 2, \dots, n$ . Let  $\mathbf{X}$  denote the design matrix given by  $\mathbf{X} = (x_{ij})$  for  $i = 1, \dots, n$  and  $j = 1, \dots, p$  of our regression model. We shall assume the rank of  $\mathbf{X}$  is  $p$ . Then the matrix  $V$  is symmetric and positive definite.

The  $n \times n$  (Toeplitz) covariance matrix of  $f_\theta(\lambda)$  is denoted by  $T_n(f_\theta)$  and has  $(j, k)$  element defined by:

$$T_n(f_\theta)_{j,k} = \int_{-\pi}^{\pi} \exp(i(j-k)\lambda) f_\theta(\lambda) d\lambda \quad (1.1)$$

The log-likelihood function is

$$L_n(\theta, \boldsymbol{\mu}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(T_n(f_\theta))) - \frac{1}{2} (\mathbf{Y} - \boldsymbol{\mu})' T_n^{-1}(f_\theta) (\mathbf{Y} - \boldsymbol{\mu}). \quad (1.2)$$

Based on the fact that  $\frac{1}{n} \ln(\det(T_n(f_\theta))) \rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(f_\theta(\lambda)) d\lambda$  as  $n \rightarrow \infty$  and  $T_n((2\pi)^{-2} f_\theta^{-1})$  approximates  $T_n^{-1}(f_\theta)$  as  $n \rightarrow \infty$  (Beran, 1994),  $L_n(\theta, \boldsymbol{\mu})$  can be approximated by the Whittle log-likelihood function,  $W_n(\theta, \boldsymbol{\mu})$ , as

$$W_n(\theta, \boldsymbol{\mu}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{4\pi} \int_{-\pi}^{\pi} \ln(f_\theta(\lambda)) d\lambda - \frac{1}{2} (\mathbf{Y} - \boldsymbol{\mu})' T_n((2\pi)^{-2} f_\theta^{-1}) (\mathbf{Y} - \boldsymbol{\mu}). \quad (1.3)$$

We refer to  $W_n(\theta, \hat{\boldsymbol{\mu}})$ , where  $\hat{\boldsymbol{\mu}}$  is replaced for  $\boldsymbol{\mu}$  in (1.3) above, as the *plug-in Whittle log-likelihood* (PWLL) function. Let

$$Q_n = \mathbf{X} V^{-1} \mathbf{X}' \quad (1.4)$$

and let  $M_n = I_n - Q_n$ , where  $I_n$  is the  $n \times n$  identity matrix. It is easy to verify the following: a)  $M_n$  and  $Q_n$  are both symmetric. b)  $\mathbf{Y}' M_n = (\mathbf{Y} - \hat{\boldsymbol{\mu}})'$ . c) If  $\mathbf{U} = \mathbf{Y} - \boldsymbol{\mu}$ , then  $M_n \mathbf{Y} = M_n \mathbf{U}$ .

Making use of the above properties, the PWLL function can now be written as

$$\begin{aligned}
 W_n(\theta, \hat{\mu}) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{4\pi} \int_{-\pi}^{\pi} \ln((f_{\theta})(\lambda)) - \frac{1}{2} \mathbf{Y}' M_n T_n ((2\pi)^{-2} f_{\theta}^{-1}) M_n \mathbf{Y} \\
 &= -\frac{n}{2} \ln(2\pi) - \frac{n}{4\pi} \int_{-\pi}^{\pi} \ln(f_{\theta}(\lambda)) + f_{\theta}^{-1}(\lambda) I_n(\lambda) d\lambda,
 \end{aligned}
 \tag{1.5}$$

where  $I_n(\lambda) = |\frac{1}{2n\pi} \sum_{j=1}^n e^{ij\lambda} (Y_j - \hat{\mu}_j)|^2$ . By definition, the Whittle Maximum Likelihood estimators (WMLE),  $\hat{\theta}_n$ , of the true parameter  $\theta$  solves the equation:

$$\int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_r} (\log f_{\theta}(\lambda) + f_{\theta}^{-1}(\lambda) I_n(\lambda)) d\lambda = 0
 \tag{1.6}$$

for  $r = 1, \dots, d_{\theta}$ , where  $d_{\theta} = \dim(\theta)$ . Andrews et al. (2005) have established a valid Edgeworth expansion for the PWMLE of error component  $\{\epsilon_t, t \geq 1\}$  of our model. We shall make use of the results of Andrews et al. (2005) and acquire an Edgeworth expansion of the PWMLE of our linear regression processes by imposing an extra condition on the regression coefficients.

The remaining sections are organized as follows. Section 2 presents the background assumptions and some preliminary concepts. Section 3 states and proves the main results of the paper.

## 2. Assumptions and Preliminaries

### 2.1 Assumptions

Assumptions A1 through A8 are provided in this section and A9 will be stated in section 2.2.

A1. The parameter space  $\Theta$  is a subset of  $\mathbb{R}^r$  with non-empty interior, where  $r$  is the dimension of the true parameter.

A2. For some integer  $s \geq 3$ ,  $g(\theta) = \int_{-\pi}^{\pi} \ln f_{\theta}(\lambda) d\lambda$  and  $h(\theta) = \int_{-\pi}^{\pi} f_{\theta}^{-1}(\lambda) I_n(\lambda) d\lambda$  can be differentiated  $s + 1$  times under the integral sign.

A3.  $f_{\theta}(\lambda)$  is continuous at all  $(\lambda, \theta)$  for which  $\lambda \neq 0$ ,  $f_{\theta}^{-1}(\lambda)$  is continuous at all  $(\lambda, \theta)$ , and  $\forall \delta > 0, \exists c_1(\theta, \delta) < \infty$  such that

$$|f_{\theta}(\lambda)| \leq c_1(\theta, \delta) |\lambda|^{-2d-\delta}$$

for all  $\lambda$  in the neighborhood  $N_{\delta}$  of the origin, where  $\theta = (d, \theta_1, \dots, \theta_r)$  and  $d \in (0, 0.5)$ .

A4. For all  $(j_1, \dots, j_k)$  with  $k \leq s + 1$  and  $j_i \in \{1, \dots, d_r\}$ ,  $(\partial^k / (\partial \theta_{j_1} \dots \partial \theta_{j_k})) f_{\theta}^{-1}(\lambda)$  is continuous at all  $(\lambda, \theta)$  and  $\forall \delta > 0, \exists c_2(\theta, \delta) < \infty$  such that

$$\left| \frac{\partial^k f_{\theta}^{-1}(\lambda)}{\partial \theta_{j_1} \dots \partial \theta_{j_k}} \right| \leq c_2(\theta, \delta) |\lambda|^{2d-\delta}, \forall \lambda \in N_{\delta}$$

A5.  $(\partial / \partial \lambda) f_{\theta}(\lambda)$  is continuous at all  $(\lambda, \theta)$  for which  $\lambda \neq 0$  and  $\forall \delta > 0, \exists c_4(\theta, \delta) < \infty$  such that

$$\left| \frac{\partial f_{\theta}(\lambda)}{\partial \lambda} \right| \leq c_4(\theta, \delta) |\lambda|^{2d-1-\delta}, \forall \lambda \in N_{\delta}$$

A6. For all  $(j_1, \dots, j_k)$  with  $k \leq s + 1$  and  $j_i \in \{1, \dots, d_r\}$ ,  $(\partial^{k+1} / (\partial \lambda \partial \theta_{j_1} \dots \partial \theta_{j_k})) f_{\theta}^{-1}(\lambda)$  is continuous at all  $(\lambda, \theta)$  for which  $\lambda \neq 0$  and  $\forall \delta > 0, \exists c_5(\theta, \delta) < \infty$  such that

$$\left| \frac{\partial^{k+1} f_{\theta}^{-1}(\lambda)}{\partial \lambda \partial \theta_{j_1} \dots \partial \theta_{j_k}} \right| \leq c_5(\theta, \delta) |\lambda|^{2d-1-\delta}, \forall \lambda \in N_{\delta}$$

A7. For any compact subset  $\Theta_c$  of the parameter space there exists a constant  $C(\Theta_c, \delta) < \infty$  such that the constants  $c_i(\theta, \delta)$  for  $i = 1, \dots, 4$  are bounded by  $C(\Theta_c, \delta), \forall \theta \in \Theta_c$  and  $\forall \delta > 0$ .

A8. The design matrix  $X$  is chosen in such a way that for the matrix

$$Q_n = (q_{ij}), i, j = 1, \dots, n,
 \tag{2.1}$$

defined by (1.4) above,  $\exists N < \infty$  such that  $|q_{ij}| \leq \frac{N}{\sqrt{n}}$  for  $1 \leq i, j \leq n$ .

Most of these assumptions are standard background assumptions in asymptotic theory and have appeared in the literature including Andrews et al. (2005), Lieberman et al. (2003), and Dahlhaus (1989) among others and are needed to control the behavior of the spectral density function. This paper includes Assumption A8 which puts a restriction on the matrix  $X$ . It is this assumption that enables us to extend the results of Andrews et al. (2005) and Lieberman et al. (2003) (both do not have design matrix as our present model does) to our linear regression model. We use this assumption to establish that the  $r$ -th cumulants  $\kappa_r(\theta)$  of the WLLDs (see (3.1) below) in the Edgeworth expansion are bounded by  $O(n)$ .

### 2.2 The Whittle Log-likelihood Derivatives

Let  $\nu = (r_1, r_2, \dots, r_q)'$  denote a q-vector of positive integers each less than or equal to  $r = \dim(\theta)$ . We write the real valued q-th order partial derivative of the PWLL function indexed by  $\nu$  as

$$W_{n,\nu} = D_\nu W_n(\theta, \hat{\mu}) = \frac{\partial^q}{\partial \theta_{r_1} \dots \partial \theta_{r_q}} W_n(\theta, \hat{\mu}) = F_{n,\nu}(\theta) + Y' M_n B_{n,\nu}(\theta) M_n Y \tag{2.2}$$

where

$$F_{n,\nu}(\theta) = -\frac{n}{4\pi} \int_{-\pi}^{\pi} D_\nu \ln(f_\theta(\lambda)) d\lambda \tag{2.3}$$

and

$$B_{n,\nu}(\theta) = -\frac{1}{2} D_\nu T_n((2\pi)^{-2} f_\theta^{-1}). \tag{2.4}$$

Let

$$Z_n(\theta) = (W_{n,\nu(1)}(\theta), \dots, W_{n,\nu(r)}(\theta)), \tag{2.5}$$

where each vector  $\nu(j)$  is of the same form as  $\nu$  defined in (2.2)-(2.4) above for some  $q \leq r = \dim(Z_n(\theta))$  for  $j = 1, 2, \dots, r$  and let

$$W_n(\theta) = n^{-1/2} (Z_n(\theta) - E_\theta Z_n(\theta)). \tag{2.6}$$

Without loss of generality we may assume that  $E_\theta Z_n(\theta) = 0$ . Let

$$D_n(\theta) = E[W_n(\theta)W_n(\theta)'] \tag{2.7}$$

and let  $D(\theta) = \lim_{n \rightarrow \infty} D_n(\theta)$ .

Because  $W_n(\theta)$  is a vector of central quadratic forms in Gaussian variables plus a vector of nonrandom quantities (Anderson, 1984) we have

$$D_n(\theta)_{ij} = \text{tr}(B_{n,\nu_i} T_n(f_\theta) B_{n,\nu_j} T_n(f_\theta)) \tag{2.8}$$

We now add one more assumption:

A9.  $D_n(\theta)$  and  $D(\theta)$  are positive definite.

### 2.3 Cumulants and Edgeworth Expansion

The jth cumulant,  $\kappa_j$ , of a random variable X with a characteristic function  $\chi(t) = E(e^{itX})$ , is defined as the coefficient of  $\frac{1}{j!} (it)^j$  in a power series expansion of  $\log \chi(t) = \sum_{j \geq 1} \frac{1}{j!} \kappa_j (it)^j$ . It can be shown that  $\kappa_1 = E(X)$ ,  $\kappa_2 = E(X^2) - (EX)^2$ ,  $\kappa_3 = E(X - EX)^3$ ,  $\kappa_4 = E(X - EX)^4 - 3(E(X^2) - (EX)^2)^2$ , and so on.

Now, let  $\hat{\theta}$  be an estimate of the parameter  $\theta_0$ , constructed from a sample of size  $n$ . Under certain conditions  $n^{1/2}(\hat{\theta} - \theta_0)$  is asymptotically normally distributed with zero mean and variance  $\sigma^2$  and for many situations of practical interest the distribution function of  $n^{1/2}(\hat{\theta} - \theta_0)$  is expanded as a power series in  $n^{-1/2}$  as follows:

$$P\left(\frac{n^{1/2}(\hat{\theta} - \theta_0)}{\sigma} \leq x\right) = \Phi(x) + n^{-1/2} \pi_1(x) \phi(x) + \dots + n^{-j/2} \pi_j(x) \phi(x) + \dots, \tag{2.9}$$

where  $\phi$  and  $\Phi$  are the Standard Normal density and distribution function, respectively, and  $\pi_j$  is a polynomial in terms of cumulants and is of degree  $3j - 1$ . The expansion on the right hand side of (2.9) is termed as an Edgeworth expansion of the distribution function on the right. For example, if  $X_1, X_2, \dots, X_n$  are independent and identically distributed with mean  $\mu = \theta_0$  and finite variance  $\sigma^2$  and if  $\hat{\theta}$  represent the sample mean, then  $\pi_1$  and  $\pi_2$  are of degrees 2 and 5, respectively, and are given by

$$\begin{aligned} \pi_1(x) &= -\frac{1}{6} \kappa_3 (x^2 - 1), \text{ and} \\ \pi_2(x) &= -\frac{1}{24} \kappa_4 (x^3 - 3x) - \frac{1}{72} \kappa_3^2 (x^5 - 10x^3 + 15x). \end{aligned}$$

Details on more general cumulant and Edgeworth expansion can be found in Hall, (1993), and Barndorff-Nielsen et al., (1989).

### 3. Edgeworth Expansion of the WLLD and WMLE

A key step in establishing an Edgeworth expansion for the distribution of the Whittle Log-likelihood Derivatives (WLLDs),  $W_n(\theta)$  (see 2.6), that holds uniformly over a compact subset  $\Theta_c$  of the parameter space that satisfies Assumption A9 is showing that the cumulants of  $Z_n(\theta)$  are uniformly  $O(n)$  in these sets. To this end we shall make use of mainly Assumptions A8 and A9 in making transition from our model to the case established by Andrews et al. (2005) and obtain an Edgeworth expansion of  $W_n(\theta)$ .

Let  $\kappa_r(\theta)$  denote the  $r$ th-order joint cumulant of  $Z_n(\theta)$ . Let  $g_{\theta, \nu}(\lambda) = (2\pi)^{-2} D_{\nu} f_{\theta}^{-1}(\lambda)$ . From the theory of quadratic forms in normal variables (Searl, 1971),  $\kappa_r(\theta)$  can be written as

$$\begin{aligned} \kappa_1(\theta) &= \frac{n}{2\pi} D_{\nu_j} \int_{-\pi}^{\pi} \log(f_{\theta}(\lambda)) d\lambda + \text{tr} \left( M_n T_n(g_{\theta, \nu_j}) M_n T_n(f_{\theta}) \right) \\ \kappa_r(\theta) &= C_r \text{tr} \left[ \prod_{j=1}^r M_n T_n(g_{\theta, \nu_j}) M_n T_n(f_{\theta}) \right] \end{aligned} \tag{3.1}$$

for  $r \geq 2$  and some vectors  $\{\nu_j : j = 1, 2, \dots, r\}$  of subscripts and some constant  $C_r < \infty$ .

Note that  $\kappa_r(\theta)$  involves derivatives of  $f_{\theta}^{-1}(\lambda)$  and that  $Z_n(\theta)$  is a vector whose elements are partial derivatives to order  $\leq s - 1$  of  $W_n(\theta, \hat{\mu})$ . For example, if  $\nu_j = (1, 2, 3)$ , then the  $j$ th element of  $Z_n(\theta)$  becomes

$$D_{\nu_j} W_n(\theta, \hat{\mu}) = \frac{\partial^3}{\partial \theta_1 \partial \theta_2 \partial \theta_3} W_n(\theta, \hat{\mu}).$$

Now, given the vector  $Z_n(\theta)$  of partial derivatives, the  $r$ th order cumulant,  $\kappa_r(\theta)$ , is determined by  $r$  elements of  $Z_n(\theta)$  with repeated elements allowed. Our goal is to first show that the cumulants are  $O(n)$  and use this bound in establishing the Edgeworth expansion of the WLLDs. To that end, we substitute  $M_n = I_n - Q_n$ , where  $Q_n$  is as given in (1.4) and  $I_n$  is the identity matrix, and rewrite  $\kappa_r(\theta)$  for  $r \geq 2$  as

$$\kappa_r(\theta) = C_r \sum \text{tr} \left[ \prod_{j=1}^r ((-Q_n)^{x_j} T_n(g_{\theta, \nu_j}) (-Q_n)^{y_j} T_n(f_{\theta})) \right] \tag{3.2}$$

where  $x_j, y_j$  take on values zero or one and satisfy  $0 \leq \sum_{j=1}^r (x_j + y_j) \leq 2r$ , the summation is over all possible configurations of  $(x_1, y_1, \dots, x_r, y_r)$ , and  $(-Q_n)^0 = I_n$ . For the case where  $x_j = y_j = 0$  for all  $j = 1, \dots, r$ ,  $\kappa_r(\theta) = O(n)$  by the following Lemma which is essentially a modified version of Theorem 1 of Lieberman et al. (2003).

**Lemma 3.1.** Suppose Assumptions A1-A9 hold in the parameter space  $\Theta$ . Then, for all  $r \geq 1$ ,

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta_c} \left| \frac{1}{n} \text{tr} \left[ \prod_{j=1}^r (T_n(g_{\theta, \nu_j}) T_n(f_{\theta})) \right] - (2\pi)^{2r-1} \int_{-\pi}^{\pi} \left[ \prod_{j=1}^r (g_{\theta, \nu_j}(\lambda) f_{\theta}(\lambda)) \right] d\lambda \right| = 0$$

for any compact subset  $\Theta_c$  of the parameter space  $\Theta$ .

Next, we examine the case where at least one matrix  $Q_n$  appears in (3.2). By Assumption A8 the design matrix is chosen in such a way that every element  $q_{ij}$  of the matrix  $Q_n$  defined in (1.4) satisfies  $|q_{ij}| \leq \frac{N}{\sqrt{n}}$ ,  $i, j = 1, \dots, n$  and some  $N \leq \infty$ . Therefore,  $|\text{tr}(Q_n)| \leq \frac{N}{\sqrt{n}} |\text{tr}(P_n)|$  where  $P_n = \mathbf{1}\mathbf{1}'$  and  $\mathbf{1} = (1, \dots, 1)'$  is a vector of ones.

We shall make use of the following Lemma.

**Lemma 3.2.** Let  $P_n = \mathbf{1}\mathbf{1}'$  where  $\mathbf{1} = (1, \dots, 1)'$  is a vector of ones. For any two  $n \times n$  matrices A and B,  $\text{tr}(P_n A P_n B) = \text{tr}(P_n A) \text{tr}(P_n B)$ .

**Proof.**

$$\begin{aligned} (P_n A)(P_n B) &= \begin{pmatrix} \sum_i a_{i1} & \dots & \sum_i a_{in} \\ \dots & \dots & \dots \\ \sum_i a_{i1} & \dots & \sum_i a_{in} \end{pmatrix} \begin{pmatrix} \sum_i b_{i1} & \dots & \sum_i b_{in} \\ \dots & \dots & \dots \\ \sum_i b_{i1} & \dots & \sum_i b_{in} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{ij} a_{ij} b_{i1} & \dots & \sum_{ij} a_{ij} b_{in} \\ \dots & \dots & \dots \\ \sum_{ij} a_{ij} b_{i1} & \dots & \sum_{ij} a_{ij} b_{in} \end{pmatrix} \end{aligned} \tag{3.3}$$

Thus,

$$tr[(P_n A)(P_n B)] = \sum_{ij} a_{ij} b_{i1} + \sum_{ij} a_{ij} b_{i2} + \dots + \sum_{ij} a_{ij} b_{in} = \sum_{ij} a_{ij} b_{ij} \tag{3.4}$$

which equals to

$$tr(P_n A)tr(P_n B) = \left(\sum_{ij} a_{ij}\right)\left(\sum_{ij} b_{ij}\right) = \sum_{ij} a_{ij} b_{ij} \tag{3.5}$$

Because  $|tr(Q_n)| \leq \frac{N}{\sqrt{n}}$  and using Assumptions **A3**, **A4**, and **A8**, for any two matrices A and B of the form  $T_n(f_\theta)$  or  $T_n(g_{\theta, v_j})$  that appear in (3.2), we can find  $K < \infty$  such that

$$tr(Q_n A Q_n B) \leq \frac{(NK)^2}{n} |tr[(P_n A)(P_n B)]| = \frac{(NK)^2}{n} |tr(P_n A)tr(P_n B)| \tag{3.6}$$

Consequently, each of the summands in (3.2) for which at least one  $Q_n$  appears can be written as summands which are less than or equal to a constant multiple of one of the following sums for different values of  $p : 0 \leq p \leq r$ .

$$U_{n,p} = tr \left[ -P \prod_{j=1}^p (T_n(g_{\theta, v_j}) T_n(f_\theta)) \right],$$

$$U_{n,p}^* = tr \left[ -P \prod_{j=1}^p (T_n(g_{\theta, v_j}) T_n(f_\theta)) T_n(g_{\theta, v_{p+1}}) \right], \text{ and}$$

$$U_{n,p}^+ = tr \left[ -P \prod_{j=1}^p (T_n(f_\theta) T_n(g_{\theta, v_j})) T_n(f_\theta) \right] \tag{3.7}$$

For example, if  $r = 2$ , and if we denote  $T_n(f_\theta)$  by  $T_n$  and  $T_n(g_{\theta, v_j})$  by  $T_{n_j}$ ,  $j = 1, 2$ , then a typical term  $(x_1, y_1, x_2, y_2) = (1, 1, 1, 0)$  would look like:

$$\begin{aligned} tr((-Q_n)T_{n1}(-Q_n)T_n(-Q_n)T_{n2}T_n) &\leq \frac{N^3}{n^{3/2}} |tr((-P_n)T_{n1}(-P_n)T_n(-P_n)T_{n2}T_n)| \\ &= \frac{N^3}{n^{3/2}} |tr((-1)1'T_{n1}(-1)1'T_n(-1)1'T_{n2}T_n)| \\ &= \frac{N^3}{n^{3/2}} |tr((1'T_{n1}1)(1'T_n1)(1'T_{n2}T_n1))| \\ &= \frac{N^3}{n^{3/2}} |tr(1'T_{n1}1)tr(1'T_n1)tr(1'T_{n2}T_n1)| \\ &= \frac{N^3}{n^{3/2}} |tr(-P_n T_{n1})tr(-P_n T_n)tr(-P_n T_{n2} T_n)| \end{aligned}$$

which is of the form  $U_{n,0}^*$ ,  $U_{n,0}^+$ , and  $U_{n,1}$ .

We are now ready to give bounds on the cumulants  $\kappa_r(\theta)$  defined in (3.1) which appear in the Edgeworth expansion of the WLLD and WMLE.

**Theorem 3.3.** Suppose Assumptions A1-A9 hold. Then for all  $r \geq 1$ ,  $\kappa_r(\theta) = O(n)$  uniformly over any compact subset  $\Theta_c$  of the parameter space.

**Proof.**

We have already seen that the summand on the rhs of (3.2) is  $O(n)$  for the case where  $x_j = y_j = 0$ ,  $j = 1, \dots, r$ . For the case where at least one matrix  $Q_n$  appears in (3.2), we have shown that the summands can be bounded by sums of the forms  $U_{n,p}$ ,  $U_{n,p}^*$ , and  $U_{n,p}^+$  given in (3.7). Theorem 3 of Andrews et al. (2005) has shown that each of these summands are  $O(n^\delta)$  for all  $\delta > 0$  and therefore Theorem 3.3 holds.

We now state the main Edgeworth expansion results of WLLD,  $W_n(\theta)$ , and WMLE,  $\hat{\theta}_n$ , of the true parameter  $\theta$ .

**Theorem 3.4.** Suppose Assumptions A1-A9 hold. For  $\theta \in \Theta$ , let  $F_n(u, \theta)$  be the joint density of  $W_n(\theta)$ ,  $r = dim(\theta)$  and for any integer  $\tau \geq 3$  let  $\tilde{F}_n^{\tau-2}(u, \theta)$  be its  $(\tau - 2)$ -order Edgeworth expansion. Then

a)  $F_n(u, \theta) = \tilde{F}_n^{\tau-2}(u, \theta) + o(n^{-(\tau/2)+1})$   
 b)  $P_\theta(W_n(\theta) \in C) = \int_C \tilde{F}_n^{\tau-2}(u, \theta) du + o(n^{-(\tau/2)+1})$

uniformly over  $u \in R^r$ , all Borel sets C and  $\theta$  in any compact subset  $\Theta_c$  of the parameter space  $\Theta$ .

**Proof:**

(a) It suffices to verify that Assumptions 1-4 of Theorem 1 of Durbin (1980), which establishes the validity of an Edgeworth expansion for the density of a sequence of random vectors, are satisfied. Durbin’s Assumption 1 is satisfied by our Assumption A9. Durbin’s Assumption 4 requires that the joint cumulants of the WLLD are  $O(n)$  uniformly over a compact subset of the parameter space and this is established by our Theorem 3.3.

Next, we verify Durbin’s Assumptions 2 and 3. Let  $\phi_n(\omega, \theta) = E_\theta[\exp(i\omega'Z_n(\theta))]$  be the characteristic function of  $Z_n(\theta)$ . Durbin’s Assumption 2 states that for n large enough,  $|\phi_n(\omega, \theta)|$  is integrable over  $R^r$  and

$$\int_{\|\omega\| \geq \delta \sqrt{n}} |\phi_n(\omega / \sqrt{n}, \theta)| d\omega = o(n^{-(\tau/2)+1}). \tag{3.8}$$

When this is translated to our case, since  $Z_n(\theta) = (W_{n,v(1)}(\theta), \dots, W_{n,v(r)}(\theta))$ , using the explicit forms of  $W_{n,v(j)}$  given in (2.2) - (2.4) and from standard theory on quadratic forms in Gaussian variables (Searl, 1971), we obtain

$$\begin{aligned} \phi_n(\omega, \theta) &= E_\theta \exp \left[ \frac{i}{\sqrt{n}} \sum_{j=1}^r (\omega_j F_{n,v_j} + X_n' B_j X_n) \right] \\ &= \exp \left[ \frac{i}{\sqrt{n}} \sum_{j=1}^r \omega_j F_{n,v_j} \right] \det \left[ I_n - \frac{2i}{\sqrt{n}} \sum_{j=1}^r \omega_j B_j T_n(f_\theta) \right]^{-1/2} \end{aligned} \tag{3.9}$$

where  $F_{n,v_j}$  is as given in (2.3) and  $B_j = M_n T_n(g_{\theta,v_j}) M_n$ .

Let  $\lambda_1, \dots, \lambda_n$  be the eigenvalues of  $\sum_{j=1}^r \omega_j B_j T_n(f_\theta)$ . Then

$$\begin{aligned} \phi_n(\omega, \theta) &= \left| \exp \left[ \frac{i}{\sqrt{n}} \sum_{j=1}^r \omega_j F_{n,v_j} \right] \right| \left| \det \left[ I_n - \frac{2i}{\sqrt{n}} \sum_{j=1}^r \omega_j B_j T_n(f_\theta) \right] \right|^{-1/2} \\ &= \prod_{k=1}^n \left| 1 - \frac{2i}{\sqrt{n}} \lambda_k \right|^{-1/2} \\ &= \prod_{k=1}^n \left( 1 + \frac{4}{n} \lambda_k^2 \right)^{-1/4} \\ &= \left[ 1 + \frac{4}{n} \sum_{k=1}^n \lambda_k^2 + \left( \frac{4}{n} \right)^2 \sum_{k \neq k'} \lambda_k^2 \lambda_{k'}^2 + \dots + \left( \frac{4}{n} \right)^n \prod_{k=1}^n \lambda_k^2 \right]^{-1/4}. \end{aligned} \tag{3.10}$$

The q-th term in (3.10) is given by

$$\left( \frac{4}{n} \right)^q \sum_{I_1} \lambda_{k_1}^2 \dots \lambda_{k_q}^2 = \left( \frac{4}{n} \right)^q \left( \sum_{k=1}^n \lambda_k^2 \right)^q - \left( \frac{4}{n} \right)^q \sum_{I_2} \lambda_{k_1}^2 \dots \lambda_{k_q}^2 \tag{3.11}$$

where  $I_1$  denotes the set of indices with no two indices equal and  $I_2$  denotes the set of indices with at least two indices equal. By (2.8) we have  $D_n(\theta) i, j = tr(B_{n,v_i} T_n(f_\theta) B_{n,v_j} T_n(f_\theta))$  and therefore, the second term in (3.10) is given by

$$\frac{4}{n} \sum_{j=1}^n \lambda_j^2 = \frac{4}{n} tr \left[ \sum_{j=1}^n \omega_j B_j T_n(f_\theta) \right]^2 = \omega' D_n(\theta) \omega. \tag{3.12}$$

Thus, by (3.12) the q-th term in (3.11) is less than or equal to  $(\omega' D_n(\theta) \omega)^q$  which by Lemma 3.1 and Theorem 3.2 is less than or equal to  $O(n\|\omega\|^2)$ . Now let  $\xi$  be the smallest eigenvalue of  $D(\theta)$  over  $\theta \in \Theta_c$  and let  $k \geq 2(d + \tau - 2)$  be a fixed

integer. Then, for n large enough,

$$\begin{aligned} \int_{\|\omega\| \geq \delta \sqrt{n}} |\phi_n(\omega / \sqrt{n}, \theta)| d\omega &\leq \int_{\|\omega\| \geq \delta \sqrt{n}} (n\xi \|\omega\|^2)^{-k/4} d\omega \\ &= (n\xi)^{-k/4} \int_{\|\omega\| \geq \delta \sqrt{n}} \|\omega\|^{-k/2} d\omega. \end{aligned}$$

Now, since  $d \in (0, 0.5)$  and  $\tau \geq 3$ , we have  $k/2 \geq 1$ , and therefore, the integral on the rhs above converges and consequently Durbin’s Assumption 2 follows. Durbin’s slightly modified Assumption 3 states that the rth derivative

$\frac{\partial^r \log \phi_n(\omega, \theta)}{\partial \omega^r}$  exists for  $\omega$  in a neighbourhood of the origin and the limit of

$$\frac{1}{n} \frac{\partial^r \log \phi_n(\omega, \theta)}{\partial \omega^r} \tag{3.13}$$

as  $n \rightarrow \infty$  and  $\|\omega\| \rightarrow 0$  exists where r is some specified positive integer.

From (3.9) we have

$$\phi_n(\omega, \theta) = i \sum_{j=1}^r \omega_j F_{n, \nu_j} - \frac{1}{2} \log \det \left[ I_n - \frac{2i}{\sqrt{n}} \sum_{j=1}^r \omega_j B_j T_n(f_\theta) \right]. \tag{3.14}$$

The limit and existence of derivatives of any order of the first term of (3.14) is straight forward. From the proof of Assumption 2 above, the second term is the log of a polynomial function of  $\omega$  and is therefore infinitely differentiable in  $\omega$  provided that the determinant is not zero.

The existence of limit as  $n \rightarrow \infty$  is established using Lemma 3.1 and Theorem 3.2 and this verifies Durbin’s Assumption 3.

Part (b) of the theorem is a consequence of corollary 3.3 of Skovgaard (1986) which drives the Edgeworth expansion of a distribution function from the corresponding density function. □

We are now ready to state the Edgeworth expansion result for the Whittle Maximum Likelihood Estimator (WMLE). Let

$$\tilde{G}_n(\omega, \theta) = \phi(\omega) \left[ 1 + \sum_{r=3}^s n^{-(1/2)r+1} P_{nr\theta}(\omega) \right], \tag{3.15}$$

be the  $(s - 2)th - order$  formal Edgeworth expansion of the density of  $n^{1/2}(\hat{\theta}_n - \theta)$  for  $\theta$  in the parameter space  $\Theta$ , where  $\phi$  denotes the multivariate normal density with mean zero and covariance matrix  $\Sigma(\theta)$  and  $P_{nr\theta}$  are Edgeworth polynomials whose coefficients depend on the cumulants of the WLLDs given in (3.1).

The following Edgeworth expansion result of the WMLE of our current linear regression model is analogous to a number of similar results including Bhattacharya et al. (1978), Andrews et al. (2005), and Lieberman et. al (2003) among others under different contexts but suitably adjusted to their cases.

**Theorem 3.5.** Suppose Assumptions A1-A9 hold in a parameter space  $\Theta$  and let  $\Theta_c$  denote a compact set in  $\Theta$ . Then,

(a) there exists a sequence of estimators  $\{\hat{\theta}_n : n \geq 1\}$  and a constant  $\alpha_0 = \alpha_0(\Theta_c)$  such that

$$\inf_{\theta \in \Theta_c} P_\theta(\|\hat{\theta}_n - \theta\| < \alpha_0 n^{-1/2} (\log n)^{1/2}) = 1 - o(n^{-s/2+1}), \tag{3.16}$$

where  $\hat{\theta}_n$  is a solution to (3.15),

(b) any sequence of estimators  $\{\hat{\theta}_n : n \geq 1\}$  that satisfies (3.16) admits the Edgeworth Expansion

$$P_\theta(\sqrt{n}(\hat{\theta}_n - \theta) \in C) = \int_C \tilde{G}_n(\omega, \theta) d\omega + o(n^{-s/2+1}) \tag{3.17}$$

uniformly over  $\theta \in \Theta_c$  and over every class  $\mathcal{B}$  of Borel sets that satisfies the condition

$$\sup_{\theta \in \Theta_c} \sup_{C \in \mathcal{B}} \int_{(\partial C)^\varepsilon} \phi(\omega) d\omega = O(\varepsilon) \tag{3.18}$$

as  $\varepsilon \rightarrow 0$  where  $(\partial C)^\varepsilon$  denotes the  $\varepsilon$ -neighborhood of the boundary of C.

**Proof.**

The proof of the theorem essentially relies on making a passage from Theorem 3.4 (b) to the current result. Theorem 4 of Lieberman et al (2003) extends the argument of Theorems 2 and 3 of Bhattacharya et al. (1978) to the long memory error component and Andrews et al. (2005) establish the same result for WMLE case (without the linear regression component). The essential difficulties that our linear regression component pose in making transition to the above established results are settled in Theorems 3.1 - 3.4 above and therefore the proofs of Lieberman et al. (2003) and that of Andrews et al. (2005) go through.

**References**

- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (2nd ed.). New York: Wiley.
- Andrews, D., & Lieberman, O. (2005). Valid Edgeworth Expansions for the Whittle Maximum Likelihood Estimator for Stationary Long- Memory Gaussian Time Series. *Econometric Theory*, 21(4), 710-733.  
<https://www.jstor.org/stable/3533394>
- Barndork, E., & Cox, R. (1989). *Asymptotic Techniques for Use in Statistics*. New York: Chapman and Hall.
- Battacharya, N., & Ghosh, K. (1978). On the Validity of the Formal Edgeworth Expansion. *Annals of Statistics*, 6, 434-451. <https://doi.org/10.1214/aos/1176344134>
- Dahlhaus, R. (1989). Efficient Parameter Estimation for Self-similar Processes. *Annals of Statistics*, 17, 1749-1766.  
<https://doi.org/10.1214/aos/1176347393>
- Durbin, J. (1980). Approximations for Densities of Sufficient Estimators. *Biometrika*, 67, 311-333.  
<https://doi.org/10.1093/biomet/67.2.311>
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer. <https://doi.org/10.1007/978-1-4612-4384-7>
- Lieberman, O., Rousseau, J., & Zucker, M. (2003). Valid Edgeworth Expansions For the Maximum Likelihood Estimator of the Parameter of a Stationary, Gaussian, Strongly Dependent Processes. *Annals of Statistics*, 31, 586-612.  
<https://www.jstor.org/stable/3448407>
- Searle, S. R. (1971). *Linear Models*. New York: Wiley.
- Skovgaard, I. M. (1986). On Multivariate Edgeworth Expansions. *International Statistical Review*, 54, 169-186.  
<https://doi.org/10.2307/1403142>

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).