Econometric Theory, 2011, Page 1 of 32.
doi:10.1017/S0266466610000253

TESTING FOR WHITE NOISE UNDER UNKNOWN DEPENDENCE AND ITS APPLICATIONS TO DIAGNOSTIC CHECKING FOR TIME SERIES MODELS

XIAOFENG SHAO
University of Illinois at Urbana-Champaign

Testing for white noise has been well studied in the literature of econometrics and statistics. For most of the proposed test statistics, such as the well-known Box–Pierce test statistic with fixed lag truncation number, the asymptotic null distributions are obtained under independent and identically distributed assumptions and may not be valid for dependent white noise. Because of recent popularity of conditional heteroskedastic models (e.g., generalized autoregressive conditional heteroskedastic [GARCH] models), which imply nonlinear dependence with zero autocorrelation, there is a need to understand the asymptotic properties of the existing test statistics under unknown dependence. In this paper, we show that the asymptotic null distribution of the Box–Pierce test statistic with general weights still holds under unknown weak dependence as long as the lag truncation number grows at an appropriate rate with increasing sample size. Further applications to diagnostic checking of the autoregressive moving average (ARMA) and fractional autoregressive integrated moving average (FARIMA) models with dependent white noise errors are also addressed. Our results go beyond earlier ones by allowing non-Gaussian and conditional heteroskedastic errors in the ARMA and FARIMA models and provide theoretical support for some empirical findings reported in the literature.

1. INTRODUCTION

A fundamental problem in time series analysis is to test for white noise (or lack of serial correlation). For a zero-mean stationary process $\{u_t\}$ with finite variance $\sigma^2 = \text{var}(u_t)$, denote its covariance and correlation functions by $R_u(k) = \text{cov}(u_t, u_{t+k})$ and $\rho_u(k) = R_u(k)/\sigma^2, k \in \mathbb{Z}$, respectively. Then the null and alternative hypothesis are

$$H_0: \ \rho_u(j) = 0 \quad \text{for all } j \neq 0 \quad \text{and} \quad H_1: \ \rho_u(j) \neq 0 \quad \text{for some } j \neq 0.$$

I thank Professor Pentti Saikkonen and two referees for constructive comments that led to improvement of the paper. The work was supported in part by NSF grant DMS-0804937. Address correspondence to Xiaofeng Shao, Department of Statistics, University of Illinois at Urbana-Champaign, 725 South Wright St., Champaign, IL 61820, USA; e-mail: xshao@uiuc.edu.

© Cambridge University Press 2010

Let $f_u(\lambda) = (2\pi)^{-1} \sum_{k \in \mathbb{Z}} \rho_u(k) e^{ik\lambda}$ be the normalized spectral density function of u_t. The equivalent frequency domain expressions to H_0 and H_1 are

$$H_0 : f_u(w) = \frac{1}{2\pi}, \qquad w \in [-\pi, \pi) \quad \text{and} \quad H_1 : f_u(w) \neq \frac{1}{2\pi},$$
$$\text{for some } w \in [-\pi, \pi).$$

In statistical modeling, diagnostic checking is an integrable part of model building. A common way of testing the adequacy of the proposed model is by checking the assumption of white noise residuals. Systematic departure from this assumption implies the inadequacy of the fitted model. Thus testing for white noise is an important research topic, and it has been extensively studied in the literature of econometrics and statistics.

The methodologies can be roughly divided into two categories: time domain tests and frequency domain tests. In the time domain, the most popular test is probably the Box and Pierce (1970) (BP) portmanteau test, which admits the following form:

$$Q_n = \sum_{j=1}^{m} \hat{\rho}_u^2(h),$$

where m is the so-called lag truncation number (see Hong, 1996) and is (typically) assumed to be fixed. The empirical autocorrelation, $\hat{\rho}_u(j)$, is defined as $\hat{\rho}_u(j) = \hat{R}_u(j)/\hat{R}_u(0)$ with $\hat{R}_u(j) = n^{-1} \sum_{t=|j|+1}^{n} (u_t - \bar{u})(u_{t-|j|} - \bar{u})$, where $\bar{u} = n^{-1} \sum_{t=1}^{n} u_t$. Under the assumption that $\{u_t\}_{t \in \mathbb{Z}}$ are independent and identically distributed (i.i.d.), it can be shown that $nQ_n \to_D \chi^2(m)$, where $\to_D$ stands for convergence in distribution. If $\{u_t\}_{t=1}^{n}$ are replaced by the residuals from a well-specified model, then the limiting distribution is still χ^2 but the degree of freedom is reduced to $m - m'$, where m' is the number of parameters in the model. In the frequency domain, Bartlett (1955) proposed test statistics based on the famous U_p and T_p processes, and a rigorous theoretical treatment of their limiting distributions was provided by Grenander and Rosenblatt (1957). Other contributions to the frequency domain tests can be found in Durlauf (1991) and Deo (2000), among others.

In the literature, when deriving the asymptotic null distribution of the test statistic, most earlier works assume Gaussianity, and thus lack of correlation is equivalent to independence. Lately there has been work that stresses the distinction between lack of correlation and independence. The main reason is that the asymptotic null distributions of the previously mentioned test statistics were obtained under i.i.d. assumptions on u_t and may not hold in the presence of nonlinear dependence, such as conditional heteroskedasticity. For example, Romano and Thombs (1996) showed that the BP statistic with χ^2 approximation can lead to misleading inferences when the time series is uncorrelated but dependent. Francq, Roy, and Zakoïan (2005) also demonstrated that the BP test applied to the residuals of an autoregressive moving average (ARMA) model with uncorrelated but dependent errors performs poorly without suitable modifications. Various methods have been proposed to account for the dependence; see, e.g., Romano and Thombs (1996), Lobato, Nankervis, and Savin (2002), Francq et al. (2005), and Horowitz, Lobato, Nankervis, and Savin (2006), among others. At this point, it seems natural to ask: "Does there exist a test statistic whose asymptotic null distribution is robust to the unknown dependence of u_t?" We shall give an affirmative answer in this paper.

In a seminal paper, Hong (1996) proposed several test statistics, which measure the distance between a kernel-based spectral density estimator and the spectral density of the noise under the null hypothesis. Let

$$\hat{f}_n(w) = (2\pi)^{-1} \sum_{j=-n+1}^{n-1} K(j/m_n)\hat{\rho}_u(j)e^{ijw}$$

be the lag window estimator of the normalized spectral density function (Priestley, 1981), where $K(\cdot)$ is a nonnegative symmetric kernel function and m_n is the bandwidth that depends on the sample size. With the quadratic distance, Hong's statistic is expressed as

$$T_n = \pi \int_{-\pi}^{\pi} (\hat{f}_n(w) - (2\pi)^{-1})^2 dw$$

or, equivalently,

$$T_n = \sum_{j=1}^{n} K^2(j/m_n)\hat{\rho}_u^2(j).$$

It is worth noting that the BP statistic can be regarded as a special case of Hong's, where $K(\cdot)$ is taken to be the truncated kernel $K(x) = \mathbf{1}(|x| \leq 1)$. Under the i.i.d. assumptions on u_t and $1/m_n + m_n/n \to 0$, Hong (1996) established the asymptotic null distribution of T_n, i.e.,

$$\frac{nT_n - C_n(K)}{\sqrt{2D_n(K)}} \to_D N(0,1), \tag{1}$$

where $C_n(K) = \sum_{j=1}^{n-1}(1 - j/n)K^2(j/m_n)$, $D_n(K) = \sum_{j=1}^{n-2}(1 - j/n)(1 - (j+1)/n)K^4(j/m_n)$, and $N(0,1)$ stands for the standard normal distribution. Under some additional assumptions on $K(\cdot)$ and m_n, (1) holds with $C_n(K)$ and $D_n(K)$ replaced by $m_nC(K)$ and $m_nD(K)$, respectively, where $C(K) = \int_0^\infty K^2(x)\,dx$ and $D(K) = \int_0^\infty K^4(x)\,dx$. Later Hong and Lee (2003) established the preceding result assuming u_t to be martingale differences with conditional heteroskedasticity of unknown form. One of the major contributions of this paper is to show that Hong's test statistic is still asymptotically valid under general white noise assumption on u_t. Further, we establish that when replacing u_t by $\hat{u}_t$, the residuals from the ARMA model with uncorrelated and dependent errors, the asymptotic null distribution of T_n still holds. Our assumptions and results differ from those in Francq et al. (2005) in that m is held fixed in their asymptotic distributional theory, whereas $m = m(n)$ grows with the sample size n in our setting. From a theoretical standpoint, the fourth cumulant of u_t plays a nonnegligible role in the asymptotic distribution of Q_n when m is fixed, whereas it turns out to be asymptotically negligible in T_n when $m_n \to \infty$. So in the latter case, the asymptotic null distribution does not change under dependent white noise, i.e., the dependence is automatically accounted for if m and n both grow to infinity. The theoretical finding is also consistent with the empirical results reported in the simulation studies of Francq et al. (2005), where the empirical size of the BP test is seen to be reasonably close to the nominal one when n is large and m is relatively large compared to n.

Recently there has been considerable attention paid to testing goodness of fit for long memory time series models. Here we only mention some representative work. Extending the Hong (1996) idea, Chen and Deo (2004a) proposed a generalized portmanteau test based on the discrete spectral average estimator and obtained the asymptotic null distribution for Gaussian long memory time series. Following the early work by Bartlett (1955), Delgado, Hidalgo, and Velasco (2005) studied Bartlett's T_p process with estimated parameters, and a martingale transform approach was used to make the null distribution asymptotically distribution free. In a related work, Hidalgo and Kreiss (2006) proposed using bootstrap methods in the frequency domain to approximate the sampling distribution of Bartlett's T_p statistic with estimated parameters. In these two papers, the asymptotic distributional theory heavily relies on the assumption that the noise processes are conditionally homoskedastic martingale differences.

In the last decade, the fractional autoregressive integrated moving average (FARIMA) models with generalized autoregressive conditional heteroskedastic (GARCH) errors have been widely used in the modeling literature (see Lien and Tse, 1999; Elek and Márkus, 2004; Koopman, Oohs, and Carnero, 2007). In the modeling stage of a FARIMA-GARCH model, it is customary to fit a FARIMA model first and then fit a GARCH model to the residuals. It is crucial to specify the FARIMA model correctly because the model misspecification of the conditional mean often leads to the misspecification of the GARCH model; see Lumsdaine and Ng (1999). Thus diagnostic checking of FARIMA models with unknown GARCH errors is a very important issue. Note that Ling and Li (1997) and Li and Li (2008) have studied the BP type tests for FARIMA-GARCH models assuming a parametric form for the GARCH model. To the best of our knowledge, there seems to be no diagnostic checking methodology known or theoretically justified to work for long memory time series models with nonparametric conditionally heteroskedastic martingale difference errors. In this paper, we shall fill this gap by proving asymptotic validity of Hong's test statistic when we replace the unobserved errors by the estimated counterpart from a FARIMA model.

We now introduce some notation. For a column vector $x = (x_1, \ldots, x_q)' \in \mathbb{R}^q$, let $|x| = \left(\sum_{j=1}^{q} x_j^2\right)^{1/2}$. For a random vector ξ, write $\xi \in \mathcal{L}^p$ $(p > 0)$ if $\|\xi\|_p :=$

$[\mathbb{E}(|\xi|^p)]^{1/p} < \infty$ and let $\|\cdot\| = \|\cdot\|_2$. For $\xi \in \mathcal{L}^1$ define projection operators $\mathcal{P}_k\xi = \mathbb{E}(\xi|\mathcal{F}_k) - \mathbb{E}(\xi|\mathcal{F}_{k-1})$, $k \in \mathbb{Z}$, where $\mathcal{F}_k = (\ldots, \varepsilon_{k-1}, \varepsilon_k)$ with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ being i.i.d. random variables. Let $C > 0$ denote a generic constant that may vary; denote by $\to_p$ convergence in probability. The symbols $O_p(1)$ and $o_p(1)$ signify being bounded in probability and convergence to zero in probability, respectively. The paper is structured as follows. In Section 2 we introduce our assumptions on u_t and establish the asymptotic distributions of T_n under the null and alternative hypotheses. Section 3 discusses the case when u_t are not directly observable. Here we consider the ARMA and FARIMA models with dependent white noise errors in Sections 3.1 and 3.2, respectively. Section 4 concludes. Proofs are gathered in the Appendix.

2. WHEN u_t IS OBSERVABLE

Suitable structural assumptions on the process (u_t) are certainly needed. Throughout, we assume that (u_t) is a mean zero stationary causal process of the form

$$u_t = F(\ldots, \varepsilon_{t-1}, \varepsilon_t), \tag{2}$$

where ε_t are i.i.d. random variables and F is a measurable function for which u_t is well defined. Further we assume that u_t satisfies the geometric-moment contraction (GMC) condition (Hsing and Wu, 2004; Shao and Wu, 2007; Wu and Shao, 2004). Let $(\varepsilon'_k)_{k\in\mathbb{Z}}$ be an i.i.d. copy of $(\varepsilon_k)_{k\in\mathbb{Z}}$; let $u'_n = F(\ldots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n)$ be a coupled version of u_n. We say that u_n is GMC(α), $\alpha > 0$, if there exist $C > 0$ and $\rho = \rho(\alpha) \in (0,1)$ such that

$$\mathbb{E}(|u_n - u'_n|^\alpha) \le C\rho^n, \qquad n \in \mathbb{N}. \tag{3}$$

The property (3) indicates that the process $\{u_n\}$ forgets its past exponentially fast, and it can be verified for many nonlinear time series models, such as the threshold model, the bilinear model, and various forms of GARCH models; see Wu and Min (2005) and Shao and Wu (2007).

Besides conditional heteroskedastic models, which imply lack of correlation because of the martingale difference structure, there are a few commonly used models (see Lobato et al., 2002) that are uncorrelated but are not martingale differences. We shall show that these models satisfy the GMC property under appropriate assumptions.

Example 2.1. Bilinear model (Granger and Anderson, 1978)

$$u_t = \varepsilon_t + b\varepsilon_{t-1}u_{t-2},$$

where ε_t are i.i.d. $N(0, \sigma_\varepsilon^2)$ and $|b| < 1$. According to Example 5.3 in Shao and Wu (2007), u_t is GMC(α), $\alpha \ge 1$, if

$$\mathbb{E}\left|\begin{pmatrix} 0 & 1 \\ b\varepsilon_t & 0 \end{pmatrix}\right|_\alpha < 1,$$

where for a $p \times p$ matrix A, $|A|_\alpha = \sup_{z \neq 0} |Az|_\alpha / |z|_\alpha$, $\alpha \geq 1$, is the matrix norm induced by the vector norm $|z|_\alpha = (\sum_{j=1}^{p} |z_j|^\alpha)^{1/\alpha}$.

Example 2.2. All-pass ARMA(1,1) model (Breidt, Davis, and Trindade, 2001)

$$u_t = \phi u_{t-1} + \varepsilon_t - \phi^{-1} \varepsilon_{t-1},$$

where $|\phi| < 1$ and $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. Note that $u_t = \varepsilon_t + \sum_{j=1}^{\infty} (\phi^j - \phi^{j-2}) \varepsilon_{t-j}$. Because $|\phi| < 1$, u_t is GMC(α) if $\varepsilon_t \in \mathcal{L}^\alpha$. In view of Theorem 5.2 in Shao and Wu (2007), the all-pass ARMA(p, p) model also satisfies GMC(α) provided that $\varepsilon_t \in \mathcal{L}^\alpha$.

Example 2.3. Nonlinear moving average (MA) model (Granger and Teräsvirta, 1993)

$$u_t = \beta \varepsilon_{t-1} \varepsilon_{t-2} + \varepsilon_t,$$

where $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$ and $\beta \in \mathbb{R}$. It is easily seen that u_t is GMC(α) if $\varepsilon_t \in \mathcal{L}^\alpha$.

To obtain the asymptotic distribution of T_n, the following assumption is made on the kernel function $K(\cdot)$ and is satisfied by several commonly used kernels in spectral analysis, such as the Bartlett, Parzen, and Tukey kernels (see Priestley, 1981, pp. 446–447).

Assumption 2.1. Assume that the kernel function $K : \mathbb{R} \to [-1, 1]$ has compact support on $[-1, 1]$ and is differentiable except at a finite number of points and symmetric with $K(0) = 1$, $\max_{x \in [-1,1]} |K(x)| = K_0 < \infty$.

The assumption that $K(\cdot)$ has compact support can presumably be relaxed at the expense of a longer and more technical proof; see Chen and Deo (2004a). Here we decide to retain it to avoid more technical complications.

THEOREM 2.1. *Suppose Assumption 2.1 and (3) hold with* $\alpha = 8$. *Assume that* $\log n = o(m_n)$ *and* $m_n = o(n^{1/2})$. *Under* H_0, *we have*

$$\frac{nT_n - m_n C(K)}{\sqrt{2 m_n D(K)}} \to_D N(0, 1). \tag{4}$$

Remark 2.1. As pointed out by a referee, the eighth moment condition on u_t is fairly strong and it excludes some interesting GARCH models, such as the integrated GARCH model. In addition, the permissible parameter space for the regular GARCH(r, s) model is quite small under the eighth moment assumption. At this point, we are unable to relax this assumption as it seems necessary in our technical argument. Nevertheless, the preceding result suggests that the asymptotic null distribution of the Hong (1996) statistic is unaffected by unknown (weak) dependence. From a technical point of view, the asymptotic null distribution of the

BP statistic depends on the fourth cumulants of u_t because the number of lags m is fixed. In contrast, for Hong's statistic, as $m_n \to \infty$, the fourth cumulant effect appears to be asymptotically negligible. For a fixed m, our result in Theorem 2.1 is not applicable.

The condition on the bandwidth is less restrictive than it looks. I am not aware of any theoretical results on the optimal bandwidth choice for T_n in the hypothesis testing context. In terms of estimating the spectral density function, the optimal bandwidth is $m_n = Cn^{1/5}$ if the kernel (e.g., Parzen kernel) is quadratic around zero and $m_n = Cn^{1/3}$ if the kernel (e.g., Bartlett kernel) is linear around zero. Note that the problem of testing for white noise bears some resemblance to testing lack of fit (or specification testing) in the nonparametric regression context. The latter problem has been well studied in the literature, and the data-driven bandwidth choice for the smoothing type test has been addressed in Horowitz and Spokoiny (2001) and Guerre and Lavergne (2005), among others.

For the optimal choice of the kernel function, we refer the reader to Hong (1996) for more details. The consistency of T_n is stated in the following theorem.

THEOREM 2.2. *Suppose Assumption 2.1 and (3) hold with* $\alpha = 8$. *Assume that* $1/m_n + m_n/n \to 0$. *Under* H_1, *we have*

$$\frac{\sqrt{m_n}}{n}\left(\frac{nT_n - m_n C(K)}{\sqrt{2m_n D(K)}}\right) \to_p \frac{1}{2}\sum_{j\neq 0}\rho_u^2(j)/(2D(K))^{1/2}.$$

Proof of Theorem 2.2. It follows from the argument in the proof of Theorem 6 of Hong (1996) by noting that $R_u(j) \leq Cr^j$ for some $r \in [0,1)$ and the absolute summability of the fourth cumulants under GMC(4) (see Wu and Shao, 2004, Prop. 2). We omit the details. ■

Remark 2.2. In related work, Chen and Deo (2006) considered the variance ratio statistic to test for white noise based on the first differenced series and proved that when the horizon k satisfies $1/k + k/n = o(1)$, the asymptotic null distribution of the variance ratio statistic is also robust to conditional heteroskedasticity of unknown form. Their result is akin to ours in that the asymptotic null distribution of the test statistic is nuisance parameter free and the horizon k in the variance ratio statistic plays a similar role as our bandwidth m_n. However, in their conditions (A.1)–(A.6), the white noise process is assumed to be a sequence of martingale differences with additional regularity conditions imposed on the higher order moments (up to eighth); compare Deo (2000). Under our framework, the white noise process does not have to be martingale difference under the null. This has some practical implications because there are nonlinear time series models that are uncorrelated but are not martingale differences, as shown in Examples 2.1–2.3. From a technical point of view, the relaxation of the martingale difference assumption, which was imposed in Hong and Lee (2003) and

Chen and Deo (2006), is a very nontrivial step and is made feasible with the novel martingale approximation techniques; see the Appendix for more discussion.

Remark 2.3. For the BP test statistic, $K(x) = \mathbf{1}(|x| \le 1)$ and $C(K) = D(K) = 1$. Thus the statement (4) reduces to $\{n \sum_{j=1}^{m_n} \hat{\rho}_u^2(j) - m_n\}/\sqrt{2m_n} \to_D N(0,1)$. In the implementation of the BP test, we use the critical values based on $\chi^2(m_n)$ and compare them with the realized value of $n \sum_{j=1}^{m_n} \hat{\rho}_u^2(j)$, whereas in Hong's test, the critical values are based on the standard normal distribution. Loosely speaking, the two procedures are asymptotically equivalent, because as $m_n \to \infty$, the central limit theorem implies $\chi^2(m_n) \approx N(m_n, 2m_n)$. This suggests that the use of the BP test is valid in the presence of unknown weak dependence when m_n is relatively large compared to n.

3. WHEN u_t IS UNOBSERVABLE

In practice, the errors $\{u_t\}_{t=1,2,\ldots,n}$ are often unobservable as a part of the model but can be estimated. Hong (1996) studied the residuals from a linear dynamic model that includes both lagged dependent variables and exogenous variables. In principle, our results can be extended to the residuals from any parametric time series model with uncorrelated errors, including the setup studied by Hong (1996). Instead of pursuing full generality, we shall treat the residuals from ARMA and FARIMA models in Sections 3.1 and 3.2, respectively. This is motivated by recent interest in ARMA models with dependent white noise errors (see Francq and Zakoïan, 2005; Francq et al., 2005, and the references therein) and goodness of fit for long memory time series models (see Section 3.2 for more references).

3.1. ARMA Model

Consider a stationary ARMA time series generated by

$$(1 - \alpha_1 B - \cdots - \alpha_p B^p) X_t = (1 + \beta_1 B + \cdots + \beta_q B^q) u_t, \tag{5}$$

where B is the backward shift operator, $\{u_t\}$ is a sequence of uncorrelated random variables, and $\Lambda = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$ is an unknown parameter vector. Let $\phi_\Lambda(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$ and $\psi_\Lambda(z) = 1 + \beta_1 z + \cdots + \beta_q z^q$ be autoregressive (AR) and MA polynomials, respectively. Denote by $\Lambda_0 = (\alpha_{10}, \ldots, \alpha_{p0}, \beta_{10}, \ldots, \beta_{q0})'$ the true value of Λ and assume that Λ_0 is an interior point of the set

$$\Omega_\delta = \{\Lambda \in \mathbb{R}^{p+q};\ \text{the roots of polynomials } \phi_\Lambda(z) \text{ and } \psi_\Lambda(z) \text{ have moduli} \ge 1 + \delta\}$$

for some $\delta > 0$. Following Francq et al. (2005), we call (5) a weak ARMA model if (u_t) is only uncorrelated, a semistrong ARMA model if (u_t) is a martingale difference, and a strong ARMA model if (u_t) is an i.i.d. sequence.

Denote by $\hat{\Lambda}_n = (\hat{\alpha}_{1n}, \ldots, \hat{\alpha}_{pn}, \hat{\beta}_{1n}, \ldots, \hat{\beta}_{qn})'$ the estimator of Λ. Then the residuals $\hat{u}_t$, $t = 1, 2, \ldots, n$, are usually obtained by the following recursion:

$$\hat{u}_t = X_t - \hat{\alpha}_{1n} X_{t-1} - \cdots - \hat{\alpha}_{pn} X_{t-p} - \hat{\beta}_{1n} \hat{u}_{t-1} - \cdots - \hat{\beta}_{qn} \hat{u}_{t-q},$$
$$t = 1, 2, \ldots, n,$$

where the initial values $(X_0, X_{-1}, \ldots, X_{1-p})' = (\hat{u}_0, \ldots, \hat{u}_{1-q})' = 0$. Following Francq et al. (2005), we test

$H_0 : (X_t)$ has an ARMA(p,q) representation (5)

against the alternative

$H_1 : (X_t)$ does not admit an ARMA representation,
or admits an ARMA (p', q') representation with $p' > p$ or $q' > q$.

If p and q are correctly specified, we would expect the estimated residuals to behave like a white noise sequence under H_0. The following theorem states the asymptotic null distribution of the test statistic $T_{1n} = \sum_{j=1}^{n} K^2(j/m_n)\hat{\rho}_{\hat{u}}^2(j)$.

THEOREM 3.1. *Suppose the assumptions in Theorem 2.1 hold. Assume that* $\hat{\Lambda}_n - \Lambda_0 = O_p(n^{-1/2})$. *Then, under* H_0,

$$\frac{nT_{1n} - m_n C(K)}{(2m_n D(K))^{1/2}} \to_D N(0,1).$$

The proof of Theorem 3.1 follows the argument used in the proof of Theorem 3.2 in the Appendix and is simpler. We omit the details. Note that as a common feature of the smoothing type test, the use of the residuals $\{\hat{u}_t\}$ in place of the true unobservable errors $\{u_t\}$ has no impact on the limiting distribution.

Remark 3.1. In the simulation studies of Francq et al. (2005), it can be seen that when m is large relative to n, the level of the BP test is reasonably close to the nominal one. Here our result provides theoretical support for this phenomenon because if we let K be the truncated kernel, the resulting test statistic is exactly the same as BP's. As mentioned in Remark 2.3, the difference between the use of the χ^2-based critical values as done in the BP test, and the use of the $N(0,1)$-based critical values for Hong's test is asymptotically negligible because the number of model parameters (i.e., $p+q$) is fixed and $m_n \to \infty$. Therefore, it seems fair to say that the use of the BP test is still justified when the lag truncation number m is large, as the unknown dependence in u_t does not kick in asymptotically.

As mentioned in Francq et al. (2005), weak ARMA models can arise from various situations, such as transformation of strong ARMA processes, causal representation of noncausal ARMA processes, and nonlinear processes. In what follows, we demonstrate that the GMC condition for the noise process in the weak

ARMA respresentation can be verified for the two leading examples in Francq et al. (2005).

Example 3.1

Consider the process

$$X_t - aX_{t-1} = \varepsilon_t - b\varepsilon_{t-1}, \qquad a \neq b \in (-1, 1),$$

where ε_t are i.i.d. random variables with $\mathbb{E}(\varepsilon_t) = 0$ and $\varepsilon_t \in \mathcal{L}^\alpha, \alpha \geq 1$. Let $Y_t = X_{2t}$. Then $Y_t - a^2 Y_{t-1} = \xi_t = u_t - \theta u_{t-1}$, where $\theta \in (-1, 1)$, $\xi_t = \varepsilon_{2t} + (a-b)\varepsilon_{2t-1} - ab\varepsilon_{2t-2}$, u_t is white noise, and $u_t = R_{1t} + R_{2t} + \theta\xi_{t-1}$, where $R_{1t} = -ab\varepsilon_{2t-2} + \theta^2\varepsilon_{2t-4} + \varepsilon_{2t} + (a-b)\varepsilon_{2t-1} + \theta^2[(a-b)\varepsilon_{2t-5} - ab\varepsilon_{2t-6}]$ and $R_{2t} = \sum_{i \geq 3} \theta^i u_{t-i}$. It is easily seen that ξ_t and R_{1t} satisfy GMC(α). By Theorem 5.2 in Shao and Wu (2007), R_{2t} also satisfies GMC(α). Therefore, u_t is GMC(α).

Example 3.2

Consider the process

$$X_t = \varepsilon_t - \phi\varepsilon_{t-1}, \qquad |\phi| > 1.$$

Let $u_t = \sum_{i=0}^{\infty} \phi^{-i} X_{t-i}$. Then X_t admits the causal MA(1) representation: $X_t = u_t - \phi^{-1}u_{t-1}$. Because X_t is GMC(α), u_t is also GMC(α) by Theorem 5.2 in Shao and Wu (2007).

Remark 3.2. To study the local power of T_{1n}, we follow Hong (1996) and define the local alternative $H_{1n} : f_{un}(w) = (2\pi)^{-1} + a_n g(w)$ for $w \in [-\pi, \pi]$, where $a_n = o(1)$. The function g is symmetric and 2π-periodic and satisfies $\int_{-\pi}^{\pi} g(w)dw = 0$, which ensures that f_{un} is a valid normalized spectral density function for large n. Let $\mu(K) = 2\pi \int_{-\pi}^{\pi} g^2(w)dw/(2D(K))^{1/2}$. It can be shown that under H_{an} with $a_n = m_n^{1/4}/n^{1/2}$,

$$\frac{nT_{1n} - m_n C(K)}{(2m_n D(K))^{1/2}} \to_D N(\mu(K), 1) \tag{6}$$

provided that $\hat{\Lambda}_n - \Lambda_0 = O_p(n^{-\frac{1}{2}})$ and the assumptions in Theorem 2.1 hold. Because the proof basically repeats the argument in the proof of Theorem 4 in Hong (1996), we omit the details. It is worth mentioning that the asymptotic distribution (6) under the local alternative still holds for T_n, whereas a similar result for T_{2n} (see Section 3.2 for the definition) in the long memory case may still hold but the proof seems tedious and is thus not pursued. Compared to the BP test with a fixed m, Hong's test is locally less powerful in that the BP test has nontrivial power against the local alternative of order $n^{-1/2}$. On the other hand, the BP test only has trivial power against nonzero correlations at lags beyond m, whereas Hong's test is able to detect nonzero correlations at any nonzero lags asymptotically.

3.2. FARIMA Model

In this section, we extend our result to the goodness-of-fit problem for long memory time series. A commonly used model in the long memory time series literature is the FARIMA model:

$$(1-B)^d \phi_\Lambda(B) Y_t = \psi_\Lambda(B) u_t, \qquad \textbf{(7)}$$

where $d \in (0, \frac{1}{2})$ is the long memory parameter. Let $\theta = (d, \Lambda')'$ and denote by $\theta_0 = (d_0, \Lambda_0')'$ its true value. Assume that θ_0 lies in the interior of $\Theta_\delta = [\Delta_1, \Delta_2] \times \Omega_\delta$, where $0 < \Delta_1 < \Delta_2 < \frac{1}{2}$.

Testing goodness of fit for short/long memory time series models has attracted a lot of attention recently. Most tests were constructed in the frequency domain, and they can be roughly categorized into two types: spectral density–based test and spectral distribution function–based test. Tests developed by Hong (1996), Paparoditis (2000), and Chen and Deo (2004a) are of the first type, and they usually involve a smoothing parameter and have trivial power against $n^{-1/2}$ local alternatives. The advantage of this type of test is that the asymptotic null distributions are free of nuisance parameters. For the second type, see Beran (1992), Chen and Romano (1999), Delgado et al. (2005), and Hidalgo and Kreiss (2006), among others. Typically, the tests of this type avoid the issue of choosing the smoothing parameter, and they can distinguish the alternatives within $n^{-1/2}$ neighborhoods of the null model. However, a disadvantage associated with this kind of test is that the asymptotic null distributions often depend on the underlying data generating mechanism and are not asymptotically distribution free. The martingale transform method (see Delgado et al., 2005) and the bootstrap approach (Chen and Romano, 1999; Hidalgo and Kreiss, 2006) have been utilized to make the tests usable in practice. So far, the tests proposed by Chen and Deo (2004a), Delgado et al. (2005), and Hidalgo and Kreiss (2006) have been justified to work for long memory time series models. However, the authors assumed either Gaussian processes or linear processes with the noise processes being conditionally homoskedastic martingale differences, which excludes interesting models, such as FARIMA models with unknown GARCH errors.

Because $d_0 \in \left(0, \frac{1}{2}\right)$, the process Y_t is invertible. We have the following autoregressive representation:

$$u_t = \sum_{k=0}^{\infty} e_k(\theta_0) Y_{t-k}.$$

Given the observations $Y_t, t = 1, 2, \ldots, n$, we follow Beran (1995) and form the residuals by

$$\hat{u}_t = \sum_{j=0}^{t-1} e_j(\hat{\theta}_n) Y_{t-j}, \qquad t = 1, 2, \ldots, n, \qquad \textbf{(8)}$$

where $\hat{\theta}_n$ is an estimator of θ. Similar to the ARMA case, the null and alternative hypotheses are

$H_0 : (Y_t)$ has a FARIMA(p,d,q) representation

and

$$H_1 : (Y_t) \text{ does not admit a FARIMA representation,} \\ \text{or admits a FARIMA } (p',d,q') \text{ representation with } p' > p \text{ or } q' > q.$$

The test statistic is $T_{2n} = \sum_{j=1}^{n} K^2(j/m_n)\hat{\rho}_{\hat{u}}^2(j)$, where $\{\hat{u}_t\}_{t=1}^{n}$ are from (8).

THEOREM 3.2. *Suppose that the assumptions in Theorem 2.1 hold. Assume that $\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})$. Then under H_0 we have*

$$\frac{nT_{2n} - m_n C(K)}{(2m_n D(K))^{1/2}} \to_D N(0,1).$$

Theorem 3.2 is a new contribution to the literature, even for the model (7) with i.i.d. errors. Here we can take the Whittle pseudo-maximum likelihood estimator as $\hat{\theta}_n$. The root-n asymptotic normality of the Whittle estimator for long memory time series models with general white noise errors has been established by Hosoya (1997) and Shao (2010).

Remark 3.3. The Hong (1996) statistic has been reformulated in the discrete form by Chen and Deo (2004a), who showed asymptotic equivalence of the two statistics for Gaussian long memory time series. Note that the applicability of the Chen and Deo (2004a) test statistic has only been proved for the Gaussian case. The latter authors conjectured that their assumptions can be relaxed to allow long memory linear processes with i.i.d. innovations. The work presented here partially solves their conjecture, and our results even allow for dependent innovations.

A limitation of our theory is that we need to assume the mean of Y_t is known. In practice, if the mean is unknown, we need to modify our $\hat{u}_t$ (cf. (8)) by replacing Y_t with $Y_t - \bar{Y}_n$, where $\bar{Y}_n = n^{-1}\sum_{t=1}^{n} Y_t$. It turns out that our technical arguments are no longer valid with this modification except for the case $d_0 \in (0, \frac{1}{4})$ with additional restrictions on m_n. The main reason is that the sample mean of a long memory time series converges to the population mean relatively slowly at the rate of $n^{(1/2-d_0)}$. The larger d_0 is, the slower it becomes. When $d_0 \in [\frac{1}{4}, \frac{1}{2})$, the effect of mean adjustment becomes asymptotically nonnegligible. As pointed out by a referee, the Chen and Deo (2004a) frequency domain test statistic is mean invariant, and so no mean adjustment is needed. It might be possible to extend the theory presented in Chen and Deo (2004a) directly to the case of dependent innovations, but such an extension seems very challenging and is beyond the scope of this paper. In the short memory case, i.e., $d_0 = 0$, the mean adjustment does not affect the asymptotic null distribution of the test statistic T_{1n}. In other words, Theorem 3.1 still holds if we use the mean adjusted residuals in the calculation of T_{1n}.

Remark 3.4. It seems natural to ask if a central limit theorem for statistics based on Bartlett's T_p process can be obtained under GMC conditions on the errors. Although it might be possible to obtain a nonpivotal asymptotic null distribution under GMC conditions, the martingale transformation method used in Delgado et al. (2005) and the frequency domain bootstrap approach in Hidalgo and Kreiss (2006) may no longer be able to take care of the estimation effect for the long memory model with unknown conditional heteroskedastic errors. The main reason is that the validity of both approaches relies on the assumption that the fourth-order spectrum of the innovation sequence is a constant, which happens to be true for conditional homoskedastic martingale differences (see Shao, 2010). In the case of conditional heteroskedastic errors, I am not aware of any feasible tests based on Bartlett's T_p process. Further study along this direction certainly would be interesting.

4. CONCLUSIONS

In this paper, we showed that the Hong (1996) test is robust to conditional heteroskedasticity of unknown form in large-sample theory and is applicable to a large class of dependent white noise series. Further, when applied to the residuals from short/long memory time series models, the asymptotical null distribution is still valid. The main focus of this paper is on the theoretical aspect, although the empirical performance is also very important. The finite-sample performance of Hong's test statistic has been examined by Hong (1996) and Chen and Deo (2004b), among others, to assess the goodness of fit of time series models with i.i.d. errors. It was found that the sampling distribution of the test statistic is right skewed, and the size distortion can presumably be reduced by adopting a power transformation method (Chen and Deo, 2004b) or frequency domain bootstrap approach (Paparoditis, 2000). The performance of the aforementioned test statistics along with size-correction devices has yet to be examined for time series models with dependent errors. An in-depth study is certainly worthwhile and will be pursued in a separate work.

REFERENCES

Bartlett, M.S. (1955) *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*. Cambridge University Press.

Beran, J. (1992) A goodness-of-fit test for time series with long range dependence. *Journal of the Royal Statistical Society, Series B* 54, 749–760.

Beran, J. (1995) Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society, Series B* 57, 659–672.

Box, G. & D. Pierce (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–1526.

Breidt, F.J., R.A. Davis, & A.A. Trindade (2001) Least absolute deviation estimation for all-pass time series models. *Annals of Statistics* 29, 919–946.

Brillinger, D.R. (1975) *Time Series: Data Analysis and Theory*. Holden-Day.
Chen, W. & R.S. Deo (2004a). A generalized portmanteau goodness-of-fit test for time series models. *Econometric Theory* 20, 382–416.
Chen, W. & R.S. Deo (2004b) Power transformations to induce normality and their applications. *Journal of the Royal Statistical Society, Series B* 66, 117–130.
Chen, W. & R.S. Deo (2006) The variance ratio statistic at large horizons. *Econometric Theory* 22, 206–234.
Chen, H. & J.P. Romano (1999) Bootstrap-assisted goodness-of-fit tests in the frequency domain. *Journal of Time Series Analysis* 20, 619–654.
Delgado, M.A., J. Hidalgo, & C. Velasco (2005) Distribution free goodness-of-fit tests for linear processes. *Annals of Statistics* 33, 2568–2609.
Deo, R.S. (2000) Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *Journal of Econometrics* 99, 291–315.
Durlauf, S. (1991) Spectral based testing for the martingale hypothesis. *Journal of Econometrics* 50, 1–19.
Elek, P. & L. Márkus (2004) A long range dependent model with nonlinear innovations for simulating daily river flows. *Natural Hazards and Earth System Sciences* 4, 277–283.
Francq, C., R. Roy, & J.M. Zakoïan (2005) Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* 100, 532–544.
Francq, C. & J.M. Zakoïan (2000) Covariance matrix estimation for estimators of mixing weak ARMA-models. *Journal of Statistical Planning and Inference* 83, 369–394.
Francq, C. & J.M. Zakoïan (2005) Recent results for linear time series models with non independent innovations. In P. Duchesne & B. Rémillard (eds.), *Statistical Modeling and Analysis for Complex Data Problems*, pp. 241–267. Springer-Verlag.
Granger, C.W.J. & A.P. Anderson (1978) *An Introduction to Bilinear Time Series Models.* Vandenhoek and Ruprecht.
Granger, C.W.J. & T. Teräsvirta (1993) *Modelling Nonlinear Economic Relationships*. Oxford University Press.
Grenander, U. & M. Rosenblatt (1957) *Statistical Analysis of Stationary Time Series*. Wiley.
Guerre, E. & P. Lavergne (2005) Data-driven rate-optimal specification testing in regression models. *Annals of Statistics* 33, 840–870.
Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Applications*. Academic Press.
Hidalgo, J. & J.P. Kreiss (2006) Bootstrap specification tests for linear covariance stationary processes. *Journal of Econometrics* 133, 807–839.
Hong, Y. (1996) Consistent testing for serial correlation of unknown form. *Econometrica* 64, 837–864.
Hong, Y. & Y.J. Lee (2003) Consistent Testing for Serial Uncorrelation of Unknown Form under General Conditional Heteroscedasticity. Preprint, Cornell University, Department of Economics.
Horowitz, J.L., I.N. Lobato, J.C. Nankervis, & N.E. Savin (2006) Bootstrapping the Box-Pierce Q test: A robust test of uncorrelatedness. *Journal of Econometrics* 133, 841–862.
Horowitz, J.L. & V.G. Spokoiny (2001) An adaptive rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–631.
Hosoya, Y. (1997) A limit theory for long-range dependence and statistical inference on related models. *Annals of Statistics* 25, 105–137.
Hsing, T. & W.B. Wu (2004) On weighted U-statistics for stationary processes. *Annals of Probability* 32, 1600–1631.
Koopman, S.J., M. Oohs, & M.A. Carnero (2007) Periodic seasonal Reg-ARFIMA-GARCH models for daily electricity spot prices. *Journal of the American Statistical Association* 102, 16–27.
Li, G. & W.K. Li (2008) Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *Biometrika* 95, 399–414.
Lien, D. & Y.K. Tse (1999) Forecasting the Nikkei spot index with fractional cointegration. *Journal of Forecasting* 18, 259–273.

Ling, S. & W.K. Li (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association* 92, 1184–1194.

Lobato, I.N., J.C. Nankervis, & N.E. Savin (2002) Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory* 18, 730–743.

Lumsdaine, R.L. & S. Ng (1999) Testing for ARCH in the presence of a possibly misspecified conditional mean. *Journal of Econometrics* 93, 257–279.

Paparoditis, E. (2000) Spectral density based goodness-of-fit tests for time series models, *Scandinavian Journal of Statistics* 27, 143–176.

Priestley, M.B. (1981) *Spectral Analysis and Time Series*, vol 1. Academic Press.

Robinson, P.M. (2005) Efficiency improvements in inference on stationary and nonstationary fractional time series. *Annals of Statistics* 33, 1800–1842.

Romano, J.L. & L.A. Thombs (1996) Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association* 91, 590–600.

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

Shao, X. & W.B. Wu (2007) Asymptotic spectral theory for nonlinear time series. *Annals of Statistics* 4, 1773–1801.

Wu, W.B. (2005) Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Science* 102, 14150–14154.

Wu, W.B. (2007) Strong invariance principles for dependent random variables. *Annals of Probability* 35, 2294–2320.

Wu, W.B. & W. Min (2005) On linear processes with dependent innovations. *Stochastic Processes and Their Applications* 115, 939–958.

Wu, W.B. & X. Shao (2004) Limit theorems for iterated random functions. *Journal of Applied Probability* 41, 425–436.

Wu, W.B. & X. Shao (2007) A limit theorem for quadratic forms and its applications. *Econometric Theory* 23, 930–951.

Wu, W.B. & M. Woodroofe (2004) Martingale approximations for sums of stationary processes. *Annals of Probability* 32, 1674–1690.

TECHNICAL APPENDIX

Throughout the Appendix, u_t is assumed to be an uncorrelated stationary sequence with the representation (2). For convenience of notation, let $k_{nj} = K(j/m_n)$. Denote $Z_{jt} = u_t u_{t-j}$ and $D_{j,k} = \sum_{t=k}^{\infty} \mathcal{P}_k(Z_{jt})$. Note that for each $j \in \mathbb{N}$, $D_{j,k}$ is a sequence of stationary and ergodic martingale differences. For $a, b \in \mathbb{R}$, denote $a \vee b = \max(a,b)$ and $a \wedge b = \min(a,b)$. Let $\mathcal{F}_i^j = (\varepsilon_i, \ldots, \varepsilon_j)$ and $\mathcal{F}_t' = (\ldots, \varepsilon_{-1}', \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_t)$, $t \in \mathbb{N}$. For $X \in \mathcal{L}^1$, denote $\mathcal{P}_t' X = \mathbb{E}(X|\mathcal{F}_t') - \mathbb{E}(X|\mathcal{F}_{t-1}')$. Let $u_k^* = F(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_k)$, $k \in \mathbb{N}$. Denote by $\delta_\alpha(k) = \|u_k - u_k^*\|_\alpha$, $k \in \mathbb{N}$, $\alpha \ge 1$ the physical dependence measure introduced by Wu (2005). According to Wu (2007), we have $\|\mathcal{P}_0 Z_{jk}\|_\alpha \le C(\delta_{2\alpha}(k) + \delta_{2\alpha}(k-j)\mathbf{1}(k \ge j))$ if $u_t \in \mathcal{L}^{2\alpha}$, and $\delta_\alpha(k) \le Cr^k$ for some $r \in (0,1)$ provided that u_t is GMC(α), $\alpha \ge 1$.

One of the major technical contributions of this paper is to replace the martingale difference assumption in Hong and Lee (2003) by the GMC condition under the white noise null hypothesis. This is achieved by approximating the double array sequence $\sum_{t=j+1}^{n} Z_{jt}$ using its martingale counterpart $\sum_{t=j+1}^{n} D_{j,t}$ for $j = 1, \ldots, m_n$. Note that the martingale approximation for the single array sequence u_t has been well studied (see, among others, Hsing and Wu, 2004; Wu and Woodroofe, 2004; Wu and Shao, 2007), but the techniques there are not directly applicable. The major difficulty is that in our setting the martingale approximation error has to be bounded uniformly in $j = 1, \ldots, m_n$ and the application of the martingale central limit theorem after martingale approximation requires very delicate analysis because of the presence of dependence.

We separate the proofs of Theorem 2.1 and Theorem 3.2 along with necessary lemmas into Sections A.1 and A. 2, respectively.

A.1. Proof of Theorem 2.1. Let $\theta_{j,r,\alpha} = \|\mathcal{P}_0 Z_{jr}\|_\alpha$, $\alpha \ge 1$, and $\Theta_{j,n,\alpha} = \sum_{r=n}^{\infty} \theta_{j,r,\alpha}$. The following lemma is an extension of Theorem 1(ii) in Wu (2007). Because the proof basically repeats that in Wu (2007), we omit the details.

LEMMA A.1. *Assume that $u_t \in \mathcal{L}^{2\alpha}$, $\alpha \ge 2$. For $0 < a_n < b_n \le n$, we have*

$$\left\| \sum_{r=a_n}^{b_n} (Z_{jr} - D_{jr}) \right\|_\alpha^2 \le C \sum_{k=1}^{b_n - a_n + 1} \Theta_{j,k,\alpha}^2.$$

Part (i) of Lemma A.2, which follows, states the variance and covariances of the approximating martingale difference $D_{j,k}$ and may be of independent interest.

LEMMA A.2. *Assume that u_t is GMC*(8).

(i) For $j > 0$, we have

$$\mathbb{E}(D_{j,k}^2) = \sigma^4 + \mathrm{cov}(u_t^2, u_{t-j}^2) + \sum_{k \ne 0, k \in \mathbb{Z}} \mathrm{cum}(u_0, u_k, u_{-j}, u_{k-j}),$$

and $\mathbb{E}(D_{j,k} D_{j',k}) = (1/2)\sum_{k \in \mathbb{Z}} \{\mathrm{cum}(u_0, u_{-j}, u_k, u_{k-j'}) + \mathrm{cum}(u_0, u_{-j'}, u_k, u_{k-j})\}$ *when* $j \ne j' > 0$.

(ii) Let $D'_{j,k} = \sum_{t=k}^{\infty} \mathcal{P}'_k (u'_t u'_{t-j})$. Then $\|D_{j,k} - D'_{j,k}\|_4 \le C(\rho^{k-j} \mathbf{1}(k \ge j) + |j-k| \mathbf{1}(k < j))$.

(iii) Let $\tilde{D}_{j,k} = \mathbb{E}(D_{j,k} | (\varepsilon_k, \ldots, \varepsilon_{k-l+1}))$, $l \in \mathbb{N}$. Then $\|\tilde{D}_{j,k} - D_{j,k}\|_4 \le C(\rho^{l-j} \mathbf{1}(l \ge j) + |j-l| \mathbf{1}(l < j))$.

Here the positive constant C appeared in (ii), and (iii) is independent of j.

Proof of Lemma A.2.

(i) It follows that when $j = j' > 0$,

$$\mathbb{E}(D_{j,k}^2) = \sum_{k=-\infty}^{\infty} \mathrm{cov}(Z_{jt}, Z_{j(t+k)}) = \mathrm{var}(Z_{jt}) + \sum_{k \ne 0, k \in \mathbb{Z}} \mathrm{cov}(u_t u_{t-j}, u_{t+k} u_{t+k-j})$$

$$= \sigma^4 + \mathrm{cov}(u_t^2, u_{t-j}^2) + \sum_{k \ne 0, k \in \mathbb{Z}} \mathrm{cum}(u_0, u_k, u_{-j}, u_{k-j})$$

and when $j \ne j' > 0$,

$$\mathbb{E}(D_{j,k} D_{j',k}) = (1/4)\mathbb{E}\{(D_{j,k} + D_{j',k})^2 - (D_{j,k} - D_{j',k})^2\}$$

$$= (1/4) \sum_{k \in \mathbb{Z}} \{\mathrm{cov}(u_t u_{t-j} + u_t u_{t-j'}, u_{t+k} u_{t+k-j} + u_{t+k} u_{t+k-j'})$$

$$- \mathrm{cov}(u_t u_{t-j} - u_t u_{t-j'}, u_{t+k} u_{t+k-j} - u_{t+k} u_{t+k-j'})\}$$

$$= (1/2) \sum_{k \in \mathbb{Z}} \{\mathrm{cov}(u_t u_{t-j}, u_{t+k} u_{t+k-j'}) + \mathrm{cov}(u_t u_{t-j'}, u_{t+k} u_{t+k-j})\}$$

$$= (1/2) \sum_{k \in \mathbb{Z}} \{\mathrm{cum}(u_0, u_{-j}, u_k, u_{k-j'}) + \mathrm{cum}(u_0, u_{-j'}, u_k, u_{k-j})\}.$$

(ii) In general, for $V_t = J(\ldots, \varepsilon_{t-1}, \varepsilon_t)$, we have $\mathbb{E}(V_t|\mathcal{F}'_k) = \mathbb{E}(V'_t|\mathcal{F}_k)$ when $t \geq k$. So for $\alpha \geq 1$,

$$\begin{aligned}\|\mathbb{E}(V_t|\mathcal{F}_k) - \mathbb{E}(V'_t|\mathcal{F}'_k)\|_\alpha &\leq \|\mathbb{E}(V_t|\mathcal{F}_k) - \mathbb{E}(V'_t|\mathcal{F}_k)\|_\alpha + \|\mathbb{E}(V_t|\mathcal{F}'_k) - \mathbb{E}(V'_t|\mathcal{F}'_k)\|_\alpha \\ &\leq 2\|V_t - V'_t\|_\alpha,\end{aligned}$$

which implies that

$$\|\mathcal{P}_k V_t - \mathcal{P}'_k V_{t'}\|_\alpha \leq 4\|V_t - V'_t\|_\alpha. \quad \textbf{(A.1)}$$

Note that $D_{j,k} = \sum_{t=k}^{\infty} \mathcal{P}_k(u_t u_{t-j})$ and $D'_{j,k} = \sum_{t=k}^{\infty} \mathcal{P}'_k(u'_t u'_{t-j})$. Then when $k \leq t \leq k+j-1$, $\mathcal{P}_k(u_t u_{t-j}) = u_{t-j}\mathcal{P}_k u_t$ and $\mathcal{P}'_k(u'_t u'_{t-j}) = u'_{t-j}\mathcal{P}'_k u'_t$. So by the Cauchy–Schwarz inequality and (A.1),

$$\begin{aligned}\|D_{j,k} - D'_{j,k}\|_4 &\leq \sum_{t=k}^{k+j-1} \|u_{t-j}\mathcal{P}_k u_t - u'_{t-j}\mathcal{P}'_k u'_t\|_4 \\ &\quad + \sum_{t=k+j}^{\infty} \|\mathcal{P}_k(u_t u_{t-j}) - \mathcal{P}'_k(u'_t u'_{t-j})\|_4 \\ &\leq C \sum_{t=k}^{k+j-1} \{\|u_{t-j} - u'_{t-j}\|_8 + \|\mathcal{P}_k u_t - \mathcal{P}'_k u'_t\|_8\} \\ &\quad + C \sum_{t=k+j}^{\infty} \|u_t u_{t-j} - u'_t u'_{t-j}\|_4 \\ &\leq C \sum_{t=k}^{k+j-1} \{\rho^{t-j} + \mathbf{1}(t \leq j) + \rho^t\} + C \sum_{t=k+j}^{\infty} \{\rho^t + \rho^{t-j}\} \\ &\leq C\{\rho^{k-j}\mathbf{1}(k \geq j) + |j-k|\mathbf{1}(k < j)\}.\end{aligned}$$

As to (iii), applying the fact that $\mathbb{E}(D_{j,l}|\varepsilon_l, \ldots, \varepsilon_1) = \mathbb{E}(D'_{j,l}|\mathcal{F}_l)$, we get

$$\begin{aligned}\|\tilde{D}_{j,k} - D_{j,k}\|_4 &= \|\tilde{D}_{j,l} - D_{j,l}\|_4 = \|D_{j,l} - \mathbb{E}(D_{j,l}|\varepsilon_l, \ldots, \varepsilon_1)\|_4 \\ &= \|\mathbb{E}((D_{j,l} - D'_{j,l})|\mathcal{F}_l)\|_4 \leq \|D_{j,l} - D'_{j,l}\|_4 \\ &\leq C\{\rho^{l-j}\mathbf{1}(l \geq j) + |j-l|\mathbf{1}(l < j)\}.\end{aligned}$$

The proof is complete. ■

Proof of Theorem 2.1. Because $\hat{R}_u(0) = \sigma^2 + O_p(n^{-1/2})$, we have

$$n \sum_{j=1}^{m_n} k_{nj}^2 \hat{\rho}_u^2(j) = n\sigma^{-4} \sum_{j=1}^{m_n} k_{nj}^2 \hat{R}_u^2(j) + o_p(m_n^{1/2}).$$

Let $G_n := n\sum_{j=1}^{m_n} k_{nj}^2 \tilde{R}_u^2(j)$, where $\tilde{R}_u(j) = n^{-1}\sum_{t=|j|+1}^{n} u_t u_{t-|j|}$. Note that $\tilde{R}_u(j) - \hat{R}_u(j) = \bar{u}\{(1-j/n)\bar{u} - n^{-1}\sum_{t=1}^{n-j+1} u_t - n^{-1}\sum_{t=j+1}^{n} u_t\}$ for $j \geq 1$. Under GMC(2), $\bar{u}^2 = O_p(n^{-1})$, $\sum_{j=1}^{m_n} k_{nj}^2 \mathbb{E}(\sum_{t=j+1}^{n} u_t + \sum_{t=1}^{n-j+1} u_t)^2 = O(nm_n)$. Consequently, $n\sum_{j=1}^{m_n} k_{nj}^2(\tilde{R}_u(j) - \hat{R}_u(j))^2 = o_p(1)$. Then it suffices to show

$$\frac{G_n - \sigma^4 m_n C(K)}{(2\sigma^8 m_n D(K))^{1/2}} \to_D N(0,1). \tag{A.2}$$

We shall approximate G_n by $\tilde{G}_n = \sum_{j=1}^{m_n} k_{nj}^2 n^{-1}\left(\sum_{k=j+1}^{n} D_{j,k}\right)^2$. By the Cauchy–Schwarz inequality,

$$|G_n - \tilde{G}_n|^2 \leq \sum_{j=1}^{m_n} \frac{k_{nj}^2}{n}\left(\sum_{k=j+1}^{n}(Z_{jk} - D_{j,k})\right)^2 \times \sum_{j=1}^{m_n} \frac{k_{nj}^2}{n}\left(\sum_{k=j+1}^{n}(Z_{jk} + D_{j,k})\right)^2,$$

where the second term on the right-hand side of the inequality is easily shown to be $O_p(m_n)$ in view of the proof to be presented hereafter. As to the first term, we apply Lemma A.1 and get

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{m_n}\left\|\sum_{k=j+1}^{n}(Z_{jk} - D_{j,k})\right\|^2 &\leq \frac{C}{n}\sum_{j=1}^{m_n}\sum_{h=1}^{\infty}\left(\sum_{k=h}^{\infty}\|\mathcal{P}_0 Z_{jk}\|\right)^2 \\
&\leq \frac{C}{n}\sum_{j=1}^{m_n}\sum_{h=1}^{\infty}\left(\sum_{k=h}^{\infty}(\delta_4(k) + \delta_4(k-j)\mathbf{1}(k \geq j))\right)^2 \\
&\leq \frac{C}{n}\sum_{j=1}^{m_n}\sum_{h=1}^{\infty}\left(\sum_{k=h}^{\infty}(\delta_4(k) + \delta_4(k-j)\mathbf{1}(k \geq j))\right) \\
&\leq \frac{Cm_n}{n}\sum_{k=1}^{\infty} k\delta_4(k) + \frac{C}{n}\sum_{k=1}^{\infty}\sum_{h=1}^{k}\sum_{j=1}^{m_n \wedge k}\delta_4(k-j) \\
&\leq Cm_n^2/n = o(1).
\end{aligned}$$

So $G_n = \tilde{G}_n + o_p(m_n^{1/2})$. Write

$$\begin{aligned}
\tilde{G}_n &= n^{-1}\sum_{j=1}^{m_n} k_{nj}^2\left(\sum_{k=j+1}^{n} D_{j,k}\right)^2 \\
&= n^{-1}\sum_{j=1}^{m_n} k_{nj}^2 \sum_{k=j+1}^{n} D_{j,k}^2 + 2n^{-1}\sum_{j=1}^{m_n} k_{nj}^2 \sum_{k=j+2}^{n}\sum_{r=j+1}^{k-1} D_{j,k}D_{j,r} = \tilde{G}_{1n} + \tilde{G}_{2n}.
\end{aligned}$$

Under the assumption that u_t is GMC(8), it is easy to show that u_t^2 is GMC(4), which implies that $|\mathrm{cov}(u_t^2, u_{t-j}^2)| \leq Cr^j$ for some $r \in (0,1)$. So by Lemma A.2,

$$\begin{aligned}
\mathbb{E}(\tilde{G}_{1n}) &= n^{-1}\sum_{j=1}^{m_n} k_{nj}^2 (n-j)\left(\sigma^4 + \mathrm{cov}(u_t^2, u_{t-j}^2) + \sum_{k \neq 0}\mathrm{cum}(u_0, u_k, u_{-j}, u_{k-j})\right) \\
&= \sigma^4 \sum_{j=1}^{m_n} k_{nj}^2 + O(1) = \sigma^4 m_n C(K) + O(1),
\end{aligned}$$

where we have applied the absolute summability of the fourth joint cumulants under GMC(4) (Wu and Shao, 2004, Prop. 2). Let $\tilde{D}_{j,k} = \mathbb{E}(D_{j,k}|\varepsilon_k, \varepsilon_{k-1}, \ldots, \varepsilon_{k-l+1})$, where $l = l_n = 2m_n$. By Lemma A.2 and the assumption that $\log n = o(m_n)$,

$$\sup_{1\le j\le m_n} \|\tilde{D}_{j,k} - D_{j,k}\|_4 = O(n^{-\kappa}) \quad \text{for any } \kappa > 0. \tag{A.3}$$

Write

$$\tilde{G}_{1n} = n^{-1}\sum_{j=1}^{m_n} k_{nj}^2 \sum_{k=j+1}^{n} \tilde{D}_{j,k}^2 + n^{-1}\sum_{j=1}^{m_n} k_{nj}^2 \sum_{k=j+1}^{n} (D_{j,k}^2 - \tilde{D}_{j,k}^2) = \tilde{G}_{11n} + \tilde{G}_{12n},$$

where $\text{var}(\tilde{G}_{11n}) = O(m_n^3/n) = o(m_n)$ by the l_n-dependence of $\tilde{D}_{j,k}$ and by (A.3),

$$\|\tilde{G}_{12n}\| \le \frac{C}{n}\sum_{j=1}^{m_n} k_{nj}^2 \sum_{k=j+1}^{n} \|D_{j,k}^2 - \tilde{D}_{j,k}^2\| = o(1).$$

So (A.2) follows if we can show that $\tilde{G}_{2n}/(2\sigma^8 m_n D(K))^{1/2} \to_D N(0,1)$.

Write

$$\tilde{G}_{2n} = 2n^{-1}\sum_{j=1}^{m_n} k_{nj}^2 \times \left(\sum_{k=j+2}^{6m_n}\sum_{r=j+1}^{k-1} + \sum_{k=6m_n+1}^{n}\sum_{r=j+1}^{m_n+1} + \sum_{k=6m_n+1}^{n}\sum_{r=k-2l_n+1}^{k-1} + \sum_{k=6m_n+1}^{n}\sum_{r=m_n+2}^{k-2l_n}\right) D_{j,k}D_{j,r} = U_{1n} + U_{2n} + U_{3n} + U_{4n}. \tag{A.4}$$

We proceed to show that $U_{kn} = o_P\left(m_n^{1/2}\right)$, $k = 1,2,3$. Note that the summands in U_{1n} form martingale differences. So

$$\mathbb{E}(U_{1n}^2) = \frac{4}{n^2}\sum_{k=3}^{6m_n}\left\|\sum_{r=2}^{k-1} k_{nj}^2 \sum_{j=1}^{(r-1)\wedge m_n} D_{j,k}D_{j,r}\right\|^2 = O(m_n^5/n^2) = o(m_n).$$

Regarding U_{2n}, we let $\tilde{U}_{2n} = 2n^{-1}\sum_{j=1}^{m_n} k_{nj}^2 \sum_{k=6m_n+1}^{n}\sum_{r=j+1}^{m_n+1}\tilde{D}_{j,k}\tilde{D}_{j,r}$. It is easy to show that $U_{2n} - \tilde{U}_{2n} = o_P(1)$ in view of (A.3). Further, by Lemma A.2,

$$\begin{aligned}\mathbb{E}(\tilde{U}_{2n}^2) &= \frac{4}{n^2}\sum_{k,k'=6m_n+1}^{n}\sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2 \sum_{r=j+1}^{m_n+1}\sum_{r'=j'+1}^{m_n+1}\mathbb{E}(\tilde{D}_{j,k}\tilde{D}_{j',k'})\mathbb{E}(\tilde{D}_{j,r}\tilde{D}_{j',r'}) \\ &= \frac{4(1+o(1))}{n^2}\sum_{k=6m_n+1}^{n}\sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2 \sum_{r=(j+1)\vee(j'+1)}^{m_n+1}\mathbb{E}(\tilde{D}_{j,k}\tilde{D}_{j',k})\mathbb{E}(\tilde{D}_{j,r}\tilde{D}_{j',r}) \\ &= O(m_n^3/n) = o(m_n).\end{aligned}$$

Thus $U_{2n} = o_p\left(m_n^{1/2}\right)$. Concerning U_{3n}, because it is a martingale, we have

$$\begin{aligned}
\mathbb{E}(U_{3n}^2) &= \frac{4}{n^2}\sum_{k=6m_n+1}^{n}\left\|\sum_{j=1}^{m_n} k_{nj}^2 \sum_{r=k-2l_n+1}^{k-1} D_{j,k}D_{j,r}\right\|^2 \\
&\le \frac{C}{n^2}\sum_{k=6m_n+1}^{n}\left(\sum_{j=1}^{m_n}\left\|\sum_{r=k-2l_n+1}^{k-1} D_{j,k}D_{j,r}\right\|\right)^2 \\
&\le \frac{C}{n^2}\sum_{k=6m_n+1}^{n}\left(\sum_{j=1}^{m_n}\left\|\sum_{r=k-2l_n+1}^{k-1} D_{j,r}\right\|_4\right)^2.
\end{aligned}$$

Because $D_{j,r}$'s are martingale differences for each j, we apply Burkholder's inequality (Hall and Heyde, 1980) and get

$$\left\|\sum_{r=k-2l_n+1}^{k-1} D_{j,r}\right\|_4 \le C\left\|\sum_{r=k-2l_n+1}^{k-1} D_{j,r}^2\right\|^{1/2} \le C\left(\sum_{r=k-2l_n+1}^{k-1}\left\|D_{j,r}^2\right\|\right)^{1/2} \le C m_n^{1/2}.$$

Note that the constant C in the preceding display does not depend on j. So $\mathbb{E}(U_{3n}^2) \le Cm_n^3/n = o(m_n)$. Let $\tilde{U}_{4n} = 2n^{-1}\sum_{j=1}^{m_n} k_{nj}^2 \sum_{k=6m_n+1}^{n}\sum_{r=m_n+2}^{k-2l_n}\tilde{D}_{j,k}\tilde{D}_{j,r}$. Because $U_{4n} - \tilde{U}_{4n} = o_p(1)$ by (A.3), it remains to show $\tilde{U}_{4n}/(2\sigma^8 m_n D(K))^{1/2} \to_D N(0,1)$ in view of (A.4).

Write $\tilde{U}_{4n} = n^{-1}\sum_{k=6m_n+1}^{n} V_{nk}$, where $V_{nk} := 2\sum_{r=m_n+2}^{k-2l_n}\sum_{j=1}^{m_n} k_{nj}^2\tilde{D}_{j,k}\tilde{D}_{j,r}$. Then $\{V_{nk}\}$ forms a sequence of martingale differences with respect to $\mathcal{F}_k$. By the martingale central limit theorem, it suffices to verify the following conditions:

$$\sigma^2(n) := \mathbb{E}(\tilde{U}_{4n}^2) = 2\sigma^8 m_n D(K)(1+o(1)), \tag{A.5}$$

$$\sum_{t=6m_n+1}^{n}\mathbb{E}(V_{nt}^2\mathbf{1}(|V_{nt}| > \epsilon n\sigma(n))) = o(\sigma^2(n)n^2), \qquad \epsilon > 0, \tag{A.6}$$

$$\sigma^{-2}(n)n^{-2}\sum_{t=6m_n+1}^{n}\bar{V}_{nt}^2 \to_p 1, \quad \text{where } \bar{V}_{nt}^2 = \mathbb{E}(V_{nt}^2|\mathcal{F}_{t-1}). \tag{A.7}$$

By Lemma A.2 and (A.3), we have

$$\begin{aligned}
\sigma^2(n) &= n^{-2}\sum_{k=6m_n+1}^{n}\mathbb{E}(V_{nk}^2) \\
&= \frac{4}{n^2}\sum_{k=6m_n+1}^{n}\sum_{r,r'=m_n+2}^{k-2l_n}\sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2\mathbb{E}(\tilde{D}_{j,k}\tilde{D}_{j',k})\mathbb{E}(\tilde{D}_{j,r}\tilde{D}_{j',r'}) \qquad \text{(A.8)} \\
&= \frac{4}{n^2}\sum_{k=6m_n+1}^{n}\sum_{r=m_n+2}^{k-2l_n}\sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2\mathbb{E}(\tilde{D}_{j,k}\tilde{D}_{j',k})\mathbb{E}(\tilde{D}_{j,r}\tilde{D}_{j',r}) \\
&= \frac{4}{n^2}\sum_{k=6m_n+1}^{n}\sum_{r=m_n+2}^{k-2l_n}\sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2\mathbb{E}(D_{j,k}D_{j',k})\mathbb{E}(D_{j,r}D_{j',r}) + o(1)
\end{aligned}$$

$$= \frac{4}{n^2} \sum_{k=6m_n+1}^{n} \sum_{r=m_n+2}^{k-2l_n} \sum_{j=1}^{m_n} k_{nj}^4 \mathbb{E}(D_{j,k}^2)\mathbb{E}(D_{j,r}^2)(1+o(1))$$
$$= 2\sigma^8 m_n D(K) + o(m_n).$$

For (A.6), again by Burkholder's inequality, we get

$$\mathbb{E}(V_{nk}^4) = \mathbb{E}\left(\sum_{r=m_n+2}^{k-2l_n} \sum_{j=1}^{m_n} k_{nj}^2 \tilde{D}_{j,k}\tilde{D}_{j,r}\right)^4 \le Cm_n^3 \sum_{j=1}^{m_n} \mathbb{E}\left(\sum_{r=m_n+2}^{k-2l_n} \tilde{D}_{j,k}\tilde{D}_{j,r}\right)^4$$
$$\le Cm_n^3 \sum_{j=1}^{m_n} \mathbb{E}(\tilde{D}_{j,k}^4)\mathbb{E}\left(\sum_{r=m_n+2}^{k-2l_n} \tilde{D}_{j,r}^2\right)^2 \le Cm_n^4 k^2,$$

which implies (A.6). To show (A.7), we let $\bar{V}_n^2 = n^{-2}\sum_{t=6m_n+1}^{n} \bar{V}_{nt}^2$, where

$$\bar{V}_{nt}^2 = 4 \sum_{r,r'=m_n+2}^{t-2l_n} \sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2 \mathbb{E}\left(\tilde{D}_{j,t}\tilde{D}_{j',t}|\mathcal{F}_{t-1}\right)\tilde{D}_{j,r}\tilde{D}_{j',r'}.$$

Then we can write

$$\begin{aligned}\bar{V}_n^2 - \sigma^2(n) = \frac{4}{n^2} &\sum_{t=6m_n+1}^{n} \sum_{r,r'=m_n+2}^{t-2l_n} \sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2 \\ &\Big\{\mathbb{E}\Big([\tilde{D}_{j,t}\tilde{D}_{j',t} - D_{j,t}D_{j',t}]|\mathcal{F}_{t-1}\Big)\tilde{D}_{j,r}\tilde{D}_{j',r'} \\ &+ \Big[\mathbb{E}(D_{j,t}D_{j',t}|\mathcal{F}_{t-1}) - \mathbb{E}\Big(D_{j,t}D_{j',t}|\mathcal{F}_{t-l+1}^{t-1}\Big)\Big]\tilde{D}_{j,r}\tilde{D}_{j',r'} \\ &+ \Big[\mathbb{E}\Big(D_{j,t}D_{j',t}|\mathcal{F}_{t-l+1}^{t-1}\Big) - \mathbb{E}(D_{j,t}D_{j',t})\Big]\tilde{D}_{j,r}\tilde{D}_{j',r'} \\ &+ \mathbb{E}(D_{j,t}D_{j',t})\Big[\tilde{D}_{j,r}\tilde{D}_{j',r'} - \mathbb{E}(\tilde{D}_{j,r}\tilde{D}_{j',r'})\Big] \\ &+ \mathbb{E}(D_{j,t}D_{j',t})\mathbb{E}(\tilde{D}_{j,r}\tilde{D}_{j',r'})\Big\} - \sigma^2(n) =: \sum_{k=1}^{5} J_{kn} - \sigma^2(n).\end{aligned} \tag{A.9}$$

By a similar argument as in (A.8), $J_{5n} = \sigma^2(n)(1+o(1))$. So (A.7) follows if we can show $\sigma^{-2}(n)J_{kn} = o_p(1)$ for $k = 1,\ldots,4$. By (A.3), $J_{1n} = o_p(m_n)$. As to J_{2n}, it follows from Lemma A.2 and (A.3) that uniformly in $j, j' = 1,2,\ldots,m_n$,

$$\begin{aligned}&\|\mathbb{E}(D_{j,t}D_{j',t}|\mathcal{F}_{t-1}) - \mathbb{E}(D_{j,t}D_{j',t}|\mathcal{F}_{t-l+1}^{t-1})\| \\ &\quad= \|\mathbb{E}(D_{j,l}D_{j',l}|\mathcal{F}_{l-1}) - \mathbb{E}(D_{j,l}D_{j',l}|\mathcal{F}_1^{l-1})\| \\ &\quad\le \|\mathbb{E}((D_{j,l}D_{j',l} - D'_{j,l}D'_{j',l})|\mathcal{F}_{l-1})\| + \|\mathbb{E}((D_{j,l}D_{j',l} - D'_{j,l}D'_{j',l})|\mathcal{F}_1^{l-1})\| \\ &\quad\le 2\|D_{j,l}D_{j',l} - D'_{j,l}D'_{j',l}\| \le C\rho^{m_n} = O(n^{-\kappa}) \quad \text{for any } \kappa > 0.\end{aligned}$$

So $J_{2n} = o_p(m_n)$. Lemmas A.3 and A.4, which follow, assert that $J_{3n} = o_p(m_n)$ and $J_{4n} = o_p(m_n)$, respectively. Thus (A.7) holds, and the conclusion follows. ■

LEMMA A.3. *Under the assumptions in Theorem 2.1, the random variable* $J_{3n} = 4/n^2 \sum_{t=6m_n+1}^{n} \sum_{r,r'=m_n+2}^{t-2l_n} \sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2 [\mathbb{E}(D_{j,t} D_{j',t} | \mathcal{F}_{t-l+1}^{t-1}) - \mathbb{E}(D_{j,t} D_{j',t})] \tilde{D}_{j,r} \tilde{D}_{j',r'}$ *as defined in (A.9) is* $o_p(m_n)$.

Proof of Lemma A.3. Let $M(j,j';t) = \mathbb{E}(D_{j,t} D_{j',t} | \mathcal{F}_{t-l+1}^{t-1}) - \mathbb{E}(D_{j,t} D_{j',t})$ and

$$\tilde{J}_{3n} = \frac{4}{n^2} \sum_{t=6m_n+1}^{n} \sum_{r,r'=m_n+2}^{t-2l_n} \sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2 M(j,j';t) D_{j,r} D_{j',r'}.$$

It is easy to see that $\tilde{J}_{3n} = J_{3n} + o_p(m_n)$ in view of (A.3). For notational convenience, denote $H_D(j,t) = \sum_{r=m_n+2}^{t-2l_n} D_{j,r}$ and $H_Z(j,t) = \sum_{r=m_n+2}^{t-2l_n} Z_{jr}$. Write $\tilde{J}_{3n} = J_{31n} + J_{32n} + J_{33n}$, where

$$J_{31n} = \frac{4}{n^2} \sum_{t=6m_n+1}^{n} \sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2 M(j,j';t)(H_D(j,t) - H_Z(j,t)) H_D(j',t),$$

$$J_{32n} = \frac{4}{n^2} \sum_{t=6m_n+1}^{n} \sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2 M(j,j';t) H_Z(j,t)(H_D(j',t) - H_Z(j',t)),$$

$$J_{33n} = \frac{4}{n^2} \sum_{t=6m_n+1}^{n} \sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2 M(j,j';t) H_Z(j,t) H_Z(j',t).$$

We shall first prove $J_{31n} = o_p(m_n)$. Because $M(j,j',t)$ is l_n-dependent with respect to t, we obtain by the Cauchy–Schwarz inequality

$$\begin{aligned}\mathbb{E}(J_{31n}^2) \le \frac{C}{n^4} &\sum_{t=6m_n+1}^{n} \sum_{t'=(6m_n+1)\vee(t-l_n)}^{n\wedge(t+l_n)} \sum_{j_1,j_1',j_2,j_2'=1}^{m_n} \|H_D(j_1,t) - H_Z(j_1,t)\|_4 \\ &\times \left\|H_D(j_2,t') - H_Z(j_2,t')\right\|_4 \left\|H_D(j_1',t)\right\|_4 \left\|H_D(j_2',t')\right\|_4.\end{aligned}$$

Because the summands in $H_D(j,t)$ form martingale differences, we apply Burkholder's inequality and obtain

$$\|H_D(j,t)\|_4^4 \le C\mathbb{E}\left(\sum_{r=m_n+1}^{t-2l_n} D_{j,r}^2\right)^2 \le Ct^2, \qquad j = 1,2,\ldots,m_n. \tag{A.10}$$

Applying Lemma A.1 and the fact that $\delta_8(k) \le Cr^k$ for some $r \in (0,1)$, we get

$$\begin{aligned}&\sum_{j_1=1}^{m_n} \|H_D(j_1,t) - H_Z(j_1,t)\|_4 \\ &\quad\le C\sum_{j_1=1}^{m_n} \left(\sum_{k_1=1}^{t-5m_n-1} \Theta_{j_1,k_1,4}^2\right)^{1/2} \\ &\quad\le C\sum_{j_1=1}^{m_n} \left(\sum_{k_1=1}^{t-5m_n-1} \sum_{h=k_1}^{\infty} (\delta_8(h) + \delta_8(h-j_1)\mathbf{1}(h \ge j_1))\right)^{1/2} \le Cm_n^{3/2}. \qquad \textbf{(A.11)}\end{aligned}$$

Therefore, in view of (A.10) and (A.11), we obtain $\mathbb{E}(J_{31n}^2) \le Cm_n^6/n^2 = o(m_n^2)$. To show $J_{32n} = o_p(m_n)$, we note that

$$\begin{aligned}
\|H_Z(j,t)\|_4^4 &= \sum_{r_1,r_2,r_2,r_4=m_n+1}^{t-2l_n} \mathbb{E}(Z_{jr_1}Z_{jr_2}Z_{jr_3}Z_{jr_4}) \\
&= \sum_{r_1,r_2,r_2,r_4=m_n+1}^{t-2l_n} \{\mathrm{cov}(Z_{jr_1},Z_{jr_2})\mathrm{cov}(Z_{jr_3},Z_{jr_4}) \\
&\qquad + \mathrm{cov}(Z_{jr_1},Z_{jr_3})\mathrm{cov}(Z_{jr_2},Z_{jr_4}) \\
&\qquad + \mathrm{cov}(Z_{jr_1},Z_{jr_4})\mathrm{cov}(Z_{jr_2},Z_{jr_3}) \\
&\qquad + \mathrm{cum}(Z_{jr_1},Z_{jr_2},Z_{jr_3},Z_{jr_4})\}. \qquad \textbf{(A.12)}
\end{aligned}$$

Because $\{u_t\}$ are uncorrelated and the kth ($k = 2,3,\ldots,8$) joint cumulants are absolutely summable under GMC(8) (see Wu and Shao, 2004, Prop. 2), it is not hard to see that $\|H_Z(j,t)\|_4^4 \le Ct^2$. Following the same argument as in the derivation of $\mathbb{E}(J_{31n}^2)$, we can derive $\mathbb{E}(J_{32n}^2) = o(m_n^2)$, and so $J_{32n} = o_p(m_n)$.

It remains to show that $J_{33n} = o_p(m_n)$. Note that

$$\begin{aligned}
\mathbb{E}(J_{33n}^2) &\le \frac{C}{n^4} \sum_{t=6m_n+1}^{n} \sum_{t'=(6m_n+1)\vee(t'-l_n)}^{n\wedge(t+l_n)} \sum_{j_1,j_1',j_2,j_2'=1}^{m_n} \sum_{r_1,r_2=m_n+2}^{t-2l_n} \sum_{r_1',r_2'=m_n+2}^{t'-2l_n} \\
&\quad \times \left|\mathbb{E}(Z_{j_1r_1}Z_{j_2r_2}Z_{j_1'r_1'}Z_{j_2'r_2'})\right| \le \frac{C}{n^4} \sum_{t=6m_n+1}^{n} \sum_{t'=(6m_n+1)\vee(t'-l_n)}^{n\wedge(t+l_n)} H_n(t,t').
\end{aligned}$$

Following (A.12), we can write $\mathbb{E}\big(Z_{j_1r_1}Z_{j_2r_2}Z_{j_1'r_1'}Z_{j_2'r_2'}\big)$ as a sum of four components, which implies $H_n(t,t') = \sum_{k=1}^4 H_{kn}(t,t')$. For $H_{1n}(t,t')$, it follows from the absolute summability of the fourth cumulant that

$$\begin{aligned}
H_{1n}(t,t') = \sum_{j_1,j_1',j_2,j_2'=1}^{m_n} \sum_{r_1,r_2=m_n+2}^{t-2l_n} \sum_{r_1',r_2'=m_n+2}^{t'-2l_n} \Big| &\Big\{\mathrm{cov}\big(u_{r_1},u_{r_2}\big)\mathrm{cov}\big(u_{r_1-j_1},u_{r_2-j_2}\big) \\
&+ \mathrm{cum}\big(u_{r_1},u_{r_1-j_1},u_{r_2},u_{r_2-j_2}\big)\Big\} \\
&\times \Big\{\mathrm{cov}\big(u_{r_1'},u_{r_2'}\big)\mathrm{cov}\big(u_{r_1'-j_1'},u_{r_2'-j_2'}\big) \\
&+ \mathrm{cum}\big(u_{r_1'},u_{r_2'},u_{r_1'-j_1'},u_{r_2'-j_2'}\big)\Big\}\Big| \\
\le Cm_n^2(t\vee t')^2.
\end{aligned}$$

By the same argument, we have $H_{kn}(t,t') \le Cm_n^2(t\vee t')^2$, $k = 2,3$. Regarding $H_{4n}(t,t')$, we apply the product theorem for the joint cumulants (Brillinger, 1975) and write

$$\mathrm{cum}\big(Z_{j_1r_1},Z_{j_2r_2},Z_{j_1'r_1'},Z_{j_2'r_2'}\big) = \sum_{\upsilon} \mathrm{cum}\big(u_{i_j}, i_j \in \upsilon_1\big)\ldots\mathrm{cum}\big(u_{i_j}, i_j \in \upsilon_p\big),$$

where the summation is over all indecomposable partitions $\upsilon = \upsilon_1 \cup \cdots \cup \upsilon_p$ of the following two-way table:

$$\begin{array}{ll} r_1 & r_1 - j_1 \\ r_2 & r_2 - j_2 \\ r_1' & r_1' - j_1' \\ r_2' & r_2' - j_2'. \end{array}$$

Again by the absolute summability of kth ($k = 2, \ldots, 8$) cumulants, we get $H_{4n}(t,t') \le Cm_n^2(t \vee t')^2$. Therefore, $\mathbb{E}(J_{33n}^2) \le Cm_n^3/n = o(m_n^2)$ and $J_{33n} = o_p(m_n)$. Thus the conclusion is established. ■

LEMMA A.4. *Under the assumptions in Theorem 2.1, the random variable* $J_{4n} = 4/n^2 \sum_{t=6m_n+1}^{n} \sum_{r,r'=m_n+2}^{t-2l_n} \sum_{j,j'=1}^{m_n} k_{nj}^2 k_{nj'}^2 \mathbb{E}(D_{j,t} D_{j',t}) \big[\tilde{D}_{j,r}\tilde{D}_{j',r'} - \mathbb{E}(\tilde{D}_{j,r}\tilde{D}_{j',r'})\big]$ *as defined in (A.9) is* $o_p(m_n)$.

Proof of Lemma A.4. Write $J_{4n} = J_{41n} + J_{42n}$, where

$$J_{41n} = \frac{4}{n^2} \sum_{t=6m_n+1}^{n} \sum_{r,r'=m_n+2}^{t-2l_n} \times \sum_{j,j'=1, j\neq j'}^{m_n} k_{nj}^2 k_{nj'}^2 \mathbb{E}(D_{j,t} D_{j',t})[\tilde{D}_{j,r}\tilde{D}_{j',r'} - \mathbb{E}(\tilde{D}_{j,r}\tilde{D}_{j',r'})],$$

$$J_{42n} = \frac{4}{n^2} \sum_{t=6m_n+1}^{n} \sum_{r,r'=m_n+2}^{t-2l_n} \sum_{j=1}^{m_n} k_{nj}^4 \mathbb{E}(D_{j,t}^2)[\tilde{D}_{j,r}\tilde{D}_{j,r'} - \mathbb{E}(\tilde{D}_{j,r}\tilde{D}_{j,r'})].$$

Note that

$$\begin{aligned}\mathbb{E}(J_{41n}^2) = O(n^{-4}) &\sum_{t_1,t_2=6m_n+1}^{n} \sum_{r_1,r_1'=m_n+2}^{t_1-2l_n} \sum_{r_2,r_2'=m_n+2}^{t_2-2l_n} \sum_{j_1,j_1'=1, j_1\neq j_1'}^{m_n} \sum_{j_2,j_2'=1, j_2\neq j_2'}^{m_n} k_{nj_1}^2 k_{nj_1'}^2 \\ &\times k_{nj_2}^2 k_{nj_2'}^2 \mathbb{E}(D_{j_1,t_1} D_{j_1',t_1}) \mathbb{E}(D_{j_2,t_2} D_{j_2',t_2}) \\ &\times \{\mathrm{cov}(\tilde{D}_{j_1,r_1}, \tilde{D}_{j_2,r_2})\mathrm{cov}(\tilde{D}_{j_1',r_1'}, \tilde{D}_{j_2',r_2'}) \\ &\quad + \mathrm{cov}(\tilde{D}_{j_1,r_1}, \tilde{D}_{j_2',r_2'})\mathrm{cov}(\tilde{D}_{j_1',r_1'}, \tilde{D}_{j_2,r_2}) \\ &\quad + \mathrm{cum}(\tilde{D}_{j_1,r_1}, \tilde{D}_{j_1',r_1'}, \tilde{D}_{j_2,r_2}, \tilde{D}_{j_2',r_2'})\}.\end{aligned}$$

By Lemma A.2 and (A.3), the first two terms in braces in the preceding expression contribute $O(m_n)$. Because $\mathrm{cum}(\tilde{D}_{j_1,r_1}, \tilde{D}_{j_1',r_1'}, \tilde{D}_{j_2,r_2}, \tilde{D}_{j_2',r_2'})$ vanishes when any two neighboring indexes (say, (r_1, r_1'), (r_1', r_2), and (r_2, r_2') if $r_1 \ge r_1' \ge r_2 \ge r_2'$) are more than l_n apart, the third term is $O(l_n^3/n) = o(m_n^2)$. So $J_{41n} = o_p(m_n)$. Concerning J_{42n}, we have

$J_{42n} = J_{421n} + J_{422n}$, where

$$J_{421n} = \frac{4}{n^2} \sum_{t=6m_n+1}^{n} \sum_{r=m_n+2}^{t-2l_n} \sum_{j=1}^{m_n} k_{nj}^4 \mathbb{E}\left(D_{j,t}^2\right) \left[\tilde{D}_{j,r}^2 - \mathbb{E}(\tilde{D}_{j,r}^2)\right],$$

$$J_{422n} = \frac{8}{n^2} \sum_{t=6m_n+1}^{n} \sum_{r=m_n+3}^{t-2l_n} \sum_{r'=m_n+2}^{r-1} \sum_{j=1}^{m_n} k_{nj}^4 \mathbb{E}\left(D_{j,t}^2\right) \left[\tilde{D}_{j,r} \tilde{D}_{j,r'} - \mathbb{E}(\tilde{D}_{j,r} \tilde{D}_{j,r'})\right].$$

Because $\tilde{D}_{j,r}^2$ is l_n-dependent, we can easily derive $\mathbb{E}(J_{421n}^2) = O(m_n^3/n)$, which implies $J_{421n} = o_p(m_n)$. Let

$$\tilde{J}_{422n} = \frac{8}{n^2} \sum_{t=6m_n+1}^{n} \sum_{r=m_n+3}^{t-2l_n} \sum_{r'=m_n+2}^{r-1} \sum_{j=1}^{m_n} k_{nj}^4 \mathbb{E}(D_{j,t}^2) D_{j,r} D_{j,r'}.$$

Then by (A.3), $J_{422n} - \tilde{J}_{422n} = o_p(1)$. Because for each j, $\left\{ \sum_{r=m_n+3}^{t-2l_n} \sum_{r'=m_n+2}^{r-1} D_{j,r} D_{j,r'} \right\}$ form martingale differences with respect to $\mathcal{F}_{t-2l_n}$, we get

$$\begin{aligned}
\mathbb{E}(\tilde{J}_{422n}^2) &\le Cn^{-4} m_n \sum_{j=1}^{m_n} k_{nj}^8 \mathbb{E}\left(\sum_{t=6m_n+1}^{n} \sum_{r=m_n+3}^{t-2l_n} \sum_{r'=m_n+2}^{r-1} \mathbb{E}(D_{j,t}^2) D_{j,r} D_{j,r'} \right)^2 \\
&\le \frac{Cm_n}{n^4} \sum_{j=1}^{m_n} \sum_{t=6m_n+1}^{n} \mathbb{E}\left(\sum_{r=m_n+3}^{t-2l_n} \sum_{r'=m_n+2}^{r-1} D_{j,r} D_{j,r'} \right)^2 \\
&= \frac{Cm_n}{n^4} \sum_{j=1}^{m_n} \sum_{t=6m_n+1}^{n} \sum_{r=m_n+3}^{t-2l_n} \mathbb{E}\left[D_{j,r}^2 \left(\sum_{r'=m_n+2}^{r-1} D_{j,r'} \right)^2 \right],
\end{aligned}$$

where we have applied the fact that for each j, $\{\sum_{r'=m_n+2}^{r-1} D_{j,r} D_{j,r'}\}$ is a sequence of martingale differences with respect to $\mathcal{F}_r$. By the Cauchy–Schwarz inequality and Burkholder's inequality,

$$\mathbb{E}\left[D_{j,r}^2 \left(\sum_{r'=m_n+2}^{r-1} D_{j,r'} \right)^2 \right] \le C \left\| \sum_{r'=m_n+2}^{r-1} D_{j,r'} \right\|_4^2 \le C(r - m_n - 2).$$

Thus $\mathbb{E}(\tilde{J}_{422n}^2) \le Cm_n^2/n = o(m_n^2)$; in other words, $\tilde{J}_{422n} = o_p(m_n)$. The proof is complete. ∎

A.2. Proof of Theorem 3.2. Throughout this section, we let $u_t(\theta) = \sum_{k=0}^{\infty} e_k(\theta) Y_{t-k}$ and $\hat{u}_t = \sum_{k=0}^{t-1} e_k(\hat{\theta}_n) Y_{t-k}$, $t = 1, 2, \ldots, n$. Write $\hat{u}_t = u_t + \lambda_{nt}$, where $\lambda_{nt} = \lambda_{1t} + \lambda_{2nt}$, $\lambda_{1t} = -\sum_{k=t}^{\infty} e_k(\theta_0) Y_{t-k} = \sum_{k=0}^{\infty} \psi_{k,t} u_{-k}$, and $\lambda_{2nt} = \sum_{k=0}^{t-1} (e_k(\hat{\theta}_n) - e_k(\theta_0)) Y_{t-k}$. Denote $e_{k;m_1}(\theta) = \partial e_k(\theta)/\partial\theta_{m_1}$ and $e_{k;(m_1,m_2)}(\theta) = \partial^2 e_k(\theta)/\partial\theta_{m_1}\partial\theta_{m_2}$ for any m_1, $m_2 \in \{1, 2, \ldots, p+q+1\}$ and assume that they are the same as those expressions in Lemma A.7 later in this section without loss of generality.

LEMMA A.5. *Under the assumptions in Theorem 3.2, we have* *(i)* $n\sum_{j=1}^{m_n} k_{nj}^2 \hat{\rho}_{\hat{u}}^2(j) = n\sigma^{-4}\sum_{j=1}^{m_n} k_{nj}^2 \hat{R}_{\hat{u}}^2(j) + o_p\left(m_n^{1/2}\right)$ *and*

(ii) $n\sum_{j=1}^{m_n} k_{nj}^2 (\hat{R}_{\hat{u}}^2(j) - \tilde{R}_{\hat{u}}^2(j)) = o_p\left(m_n^{1/2}\right)$, *where* $\tilde{R}_{\hat{u}}(j) = n^{-1}\sum_{t=|j|+1}^{n} \hat{u}_t\hat{u}_{t-|j|}$.

Proof of Lemma A.5. To prove (i), it suffices to show that

$$\hat{R}_{\hat{u}}(0) = n^{-1}\sum_{t=1}^{n}\hat{u}_t^2 - \left(n^{-1}\sum_{t=1}^{n}\hat{u}_t\right)^2 = \sigma^2 + O_p(n^{-1/2}). \qquad \textbf{(A.13)}$$

To this end, let $G_{1n} = n^{-1}\sum_{t=1}^{n} u_t\lambda_{1t}$, $G_{2n} = n^{-1}\sum_{t=1}^{n}\lambda_{1t}^2$, and $G_{3n} = n^{-1}\sum_{t=1}^{n}\lambda_{2nt}^2$. Because $n^{-1}\sum_{t=1}^{n} u_t^2 - \sigma^2 = O_p(n^{-1/2})$, (A.13) follows if we can show $G_{1n} = O_p(n^{-1/2})$, $G_{2n} = O_p(n^{-1/2})$, and $G_{3n} = O_p(n^{-1})$. Note that

$$\begin{aligned}\mathbb{E}(G_{1n}^2) &= n^{-2}\sum_{t,t'=1}^{n}\sum_{k,k'=0}^{\infty}\psi_{k,t}\psi_{k',t'}\mathbb{E}(u_t u_{t'} u_{-k} u_{-k'})\\ &= n^{-2}\sum_{t=1}^{n}\sum_{k=0}^{\infty}\psi_{k,t}^2\sigma^4 + n^{-2}\sum_{t,t'=1}^{n}\sum_{k,k'=0}^{\infty}\psi_{k,t}\psi_{k',t'}\mathrm{cum}(u_t, u_{t'}, u_{-k}, u_{-k'})\\ &= O\left(\log n/n^2 + n^{-1}\right) = O\left(n^{-1}\right),\end{aligned}$$

where we have applied the fact that $\sum_{k=0}^{\infty}\psi_{k,t}^2 = O(t^{-1})$ (see Robinson, 2005) and the absolute summability of the fourth cumulants. Because $\mathbb{E}(G_{2n}) = O(\log n/n)$, $G_{2n} = O_p\left(n^{-1/2}\right)$. To show $G_{3n} = o_p(n^{-1})$, we apply the mean-value theorem and get $e_k(\hat{\theta}_n) - e_k(\theta_0) = \sum_{m_1=1}^{p+q+1} e_{k;m_1}(\bar{\theta}_{kn})\left(\hat{\theta}_n^{(m_1)} - \theta_0^{(m_1)}\right)$, where $\bar{\theta}_{kn} = \theta_0 + \beta_k(\hat{\theta}_n - \theta_0)$ for some $\beta_k \in (0,1)$. Then

$$\begin{aligned}nG_{3n} &= \sum_{t=1}^{n}\sum_{k,k'=0}^{t-1}(e_k(\hat{\theta}_n) - e_k(\theta_0))(e_{k'}(\hat{\theta}_n) - e_{k'}(\theta_0))Y_{t-k}Y_{t-k'}\\ &= \sum_{t=1}^{n}\sum_{k,k'=0}^{t-1}\sum_{m_1,m_1'=1}^{p+q+1}\left(\hat{\theta}_n^{(m_1)} - \theta_0^{(m_1)}\right)\left(\hat{\theta}_n^{(m_1')} - \theta_0^{(m_1')}\right)e_{k;m_1}(\bar{\theta}_{kn})e_{k';m_1'}\\ &\quad\times(\bar{\theta}_{k'n})Y_{t-k}Y_{t-k'}.\end{aligned}$$

When $\hat{\theta}_n \in \Theta_\delta$, by Lemma A.7, for any $(m_1, m_1') \in \{1,\ldots,p+q+1\}^2$,

$$\sum_{t=1}^{n}\sum_{k,k'=0}^{t-1}|e_{k;m_1}(\bar{\theta}_{kn})||e_{k';m_1'}(\bar{\theta}_{k'n})|\mathbb{E}|Y_{t-k}Y_{t-k'}| = O(n).$$

Because $\hat{\theta}_n - \theta_0 = O_p(n^{-\frac{1}{2}})$, we have $P(\hat{\theta}_n \notin \Theta_\delta) \to 0$. Consequently $nG_{3n} = nG_{3n}\mathbf{1}(\hat{\theta}_n \in \Theta_\delta) + nG_{3n}\mathbf{1}(\hat{\theta}_n \notin \Theta_\delta) = O_p(1)$. Therefore part (i) is proved.

As to part (ii), write $\tilde{R}_{\hat{u}}(j) - \hat{R}_{\hat{u}}(j) = -n^{-1}\bar{\hat{u}}\left(\sum_{t=1}^{n-j}\hat{u}_t + \sum_{t=j+1}^{n}\hat{u}_t\right) + (1 - j/n)\bar{\hat{u}}^2$, where $\bar{\hat{u}} = n^{-1}\sum_{t=1}^{n}\hat{u}_t$. Following the argument for part (i), it is straightforward to show that $\bar{\hat{u}} = O_p(n^{-1/2})$ and $\sum_{j=1}^{m_n} k_{nj}^2 \left(\sum_{t=1}^{n-j} \hat{u}_t + \sum_{t=j+1}^{n} \hat{u}_t\right)^2 = O_p(nm_n)$. So $n\sum_{j=1}^{m_n} k_{nj}^2 (\hat{R}_{\hat{u}}(j) - \tilde{R}_{\hat{u}}(j))^2 = o_p(1)$. Applying the Cauchy–Schwarz inequality, part (ii) follows. ■

Proof of Theorem 3.2. By Lemma A.5, we only need to show that

$$\frac{n\sum_{j=1}^{m_n} k_{nj}^2 \tilde{R}_{\hat{u}}^2(j) - \sigma^4 m_n C(K)}{(2\sigma^8 m_n D(K))^{1/2}} \to_D N(0,1).$$

Note that $\tilde{R}_{\hat{u}}^2(j) - \tilde{R}_u^2(j) = (\tilde{R}_{\hat{u}}(j) - \tilde{R}_u(j))^2 + 2\tilde{R}_u(j)(\tilde{R}_{\hat{u}}(j) - \tilde{R}_u(j))$. By Theorem 2.1, it suffices to show that

$$n\sum_{j=1}^{m_n} k_{nj}^2 (\tilde{R}_{\hat{u}}(j) - \tilde{R}_u(j))^2 = o_p(1),$$

because it implies $n\sum_{j=1}^{m_n} k_{nj}^2 \tilde{R}_u(j)(\tilde{R}_{\hat{u}}(j) - \tilde{R}_u(j)) = o_p\left(m_n^{1/2}\right)$ by the Cauchy–Schwarz inequality. To this end, we note that

$$\begin{aligned} n\sum_{j=1}^{m_n} k_{nj}^2 (\tilde{R}_{\hat{u}}(j) - \tilde{R}_u(j))^2 &\le \frac{C}{n}\sum_{j=1}^{m_n} \left\{ \left(\sum_{t=j+1}^{n} \lambda_{1t} u_{t-j}\right)^2 + \left(\sum_{t=j+1}^{n} \lambda_{2nt} u_{t-j}\right)^2 \right. \\ &\quad + \left(\sum_{t=j+1}^{n} u_t \lambda_{1(t-j)}\right)^2 + \left(\sum_{t=j+1}^{n} u_t \lambda_{2n(t-j)}\right)^2 \\ &\quad \left. + \left(\sum_{t=j+1}^{n} \lambda_{nt}\lambda_{n(t-j)}\right)^2 \right\} \\ &=: C(L_{1n} + L_{2n} + L_{3n} + L_{4n} + L_{5n}). \end{aligned}$$

We proceed to show that $L_{kn} = o_p(1), k = 1, \ldots, 5$. First,

$$\begin{aligned} \mathbb{E}(L_{1n}) &= n^{-1}\sum_{j=1}^{m_n} \mathbb{E}\left(\sum_{t=j+1}^{n}\sum_{k=0}^{\infty} \psi_{k,t} u_{-k} u_{t-j}\right)^2 \\ &= n^{-1}\sum_{j=1}^{m_n}\sum_{t,t'=j+1}^{n}\sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} \psi_{k,t}\psi_{k',t'}\mathbb{E}(u_{-k}u_{-k'}u_{t-j}u_{t'-j}) \\ &= n^{-1}\sum_{j=1}^{m_n}\sum_{t,t'=j+1}^{n}\sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} \psi_{k,t}\psi_{k',t'}\{\mathrm{cov}(u_{-k},u_{-k'})\mathrm{cov}(u_{t-j},u_{t'-j}) \\ &\qquad + \mathrm{cum}(u_{-k},u_{-k'},u_{t-j},u_{t'-j})\}, \end{aligned}$$

where the first term is $(\sigma^4/n)\sum_{j=1}^{m_n}\sum_{t=j+1}^{n}\sum_{k=0}^{\infty}\psi_{k,t}^2 = O(m_n \log n/n)$. Applying Proposition 2 in Wu and Shao (2004), we have $|\mathrm{cum}(u_{-k},u_{-k'},u_{t-j},u_{t'-j})| \le Cr^{t\vee t'-j+k\vee k'}$ for some $r \in (0,1)$. So the second term in $\mathbb{E}(L_{1n})$ is bounded by

$$Cn^{-1}\sum_{j=1}^{m_n}\sum_{t,t'=j+1}^{n}\sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} |\psi_{k,t}\psi_{k',t'}| r^{t\vee t'-j+k\vee k'} = O(m_n/n).$$

Following the same argument, we get $\mathbb{E}(L_{3n}) = O(m_n/n) = o(1)$.

To show $L_{5n} = o_p(1)$, we note that

$$L_{5n} \le \frac{C}{n}\sum_{j=1}^{m_n}\left\{\left(\sum_{t=j+1}^{n}\lambda_{1t}\lambda_{1(t-j)}\right)^2 + \left(\sum_{t=j+1}^{n}\lambda_{1t}\lambda_{2n(t-j)}\right)^2 + \left(\sum_{t=j+1}^{n}\lambda_{2nt}\lambda_{1(t-j)}\right)^2 + \left(\sum_{t=j+1}^{n}\lambda_{2nt}\lambda_{2n(t-j)}\right)^2\right\} =: C(L_{51n}+L_{52n}+L_{53n}+L_{54n}). \quad \textbf{(A.14)}$$

As to L_{51n}, we have

$$\begin{aligned}\mathbb{E}(L_{51n}) &= \frac{1}{n}\sum_{j=1}^{m_n}\sum_{t,t'=j+1}^{n}\mathbb{E}(\lambda_{1t}\lambda_{1t'}\lambda_{1(t-j)}\lambda_{1(t'-j)})\\ &= \frac{1}{n}\sum_{j=1}^{m_n}\sum_{t,t'=j+1}^{n}\sum_{k_1,k_2,k_3,k_4=0}^{\infty}\psi_{k_1,t}\psi_{k_2,t'}\psi_{k_3,t-j}\psi_{k_4,t'-j}\\ &\quad\times\mathbb{E}(u_{-k_1}u_{-k_2}u_{-k_3}u_{-k_4})\\ &= \frac{1}{n}\sum_{j=1}^{m_n}\sum_{t,t'=j+1}^{n}\sum_{k_1,k_2,k_3,k_4=0}^{\infty}\psi_{k_1,t}\psi_{k_2,t'}\psi_{k_3,t-j}\psi_{k_4,t'-j}\{\mathrm{cov}(u_{-k_1},u_{-k_2})\\ &\quad\times\mathrm{cov}(u_{-k_3},u_{-k_4})+\mathrm{cov}(u_{-k_1},u_{-k_3})\mathrm{cov}(u_{-k_2},u_{-k_4})+\mathrm{cov}(u_{-k_1},u_{-k_4})\\ &\quad\times\mathrm{cov}(u_{-k_2},u_{-k_3})+\mathrm{cum}(u_{-k_1},u_{-k_2},u_{-k_3},u_{-k_4})\}.\end{aligned}$$

Because $\sum_{k=0}^{\infty}\psi_{k,t}^2 \le Ct^{-1}$ (cf. Robinson, 2005), the first three terms in the preceding display are $O(m_n\log^2 n/n)$ under the null hypothesis. By Proposition 2 in Wu and Shao (2004), $|\mathrm{cum}(u_{-k_1},u_{-k_2},u_{-k_3},u_{-k_4})| \le Cr^{\max(k_1,k_2,k_3,k_4)-\min(k_1,k_2,k_3,k_4)}$ for some $r \in (0,1)$. Thus the fourth term in the preceding display is bounded by

$$\begin{aligned}&\frac{C}{n}\sum_{j=1}^{m_n}\sum_{t,t'=j+1}^{n}\sum_{k_1\ge k_2\ge k_3\ge k_4=0}^{\infty}|\psi_{k_1,t}\psi_{k_2,t'}\psi_{k_3,t-j}\psi_{k_4,t'-j}|r^{k_1-k_4}\\ &\quad\le\frac{C}{n}\sum_{j=1}^{m_n}\sum_{t,t'=j+1}^{n}\sum_{h_1,h_3=0}^{\infty}\sum_{k_2=0}^{\infty}|\psi_{k_2+h_1,t}\psi_{k_2,t'}|\sum_{k_4=0}^{\infty}|\psi_{k_4+h_3,t-j}\psi_{k_4,t'-j}|r^{h_1+h_3}\\ &\quad\le\frac{C}{n}\sum_{j=1}^{m_n}\sum_{t,t'=j+1}^{n}(tt'(t-j)(t'-j))^{-1/2}\sum_{h_1,h_3=0}^{\infty}r^{h_1+h_3}=o(1).\end{aligned}$$

Lemma A.6, which follows, asserts that $L_{52n} = o_p(1)$, and the same argument leads to $L_{53n} = o_p(1)$. Following the some line as in the derivation of G_{3n} (see Lemma A.5), we can derive $L_{54n} = O_p(m_n/n) = o_p(1)$. Thus $L_{5n} = o_p(1)$, and a similar and simpler argument yields $L_{kn} = o_p(1)$, $k = 2,4$. We omit the details. The conclusion is established. ■

LEMMA A.6. *Under the assumptions in Theorem 3.2, the random variable* $L_{52n} = n^{-1}\sum_{j=1}^{m_n}\left(\sum_{t=j+1}^{n}\lambda_{1t}\lambda_{2n(t-j)}\right)^2$ *as defined in (A.14) is* $o_p(1)$.

Proof of Lemma A.6. We apply a Taylor's expansion for each k and obtain

$$e_k(\hat{\theta}_n) - e_k(\theta_0) = \sum_{m_1=1}^{p+q+1} (\hat{\theta}_n^{(m_1)} - \theta_0^{(m_1)}) e_{k;m_1}(\theta_0) + \sum_{m_1,m_2=1}^{p+q+1} (\hat{\theta}_n^{(m_1)} - \theta_0^{(m_1)})(\hat{\theta}_n^{(m_2)} - \theta_0^{(m_2)}) e_{k;(m_1,m_2)}(\tilde{\theta}_{kn}),$$

where $\tilde{\theta}_{kn} = \theta_0 + \alpha_k(\hat{\theta}_n - \theta_0)$ for some $\alpha_k \in (0,1)$. By Lemma A.7, $|e_{k;m_1}(\theta_0)| \le Ck^{-1-\epsilon}$ and $\sup_{\theta\in\Theta_\delta} |e_{k;(m_1,m_2)}(\theta)| \le Ck^{-1-\epsilon}$ for some $\epsilon > 0$. Denote $e_k(\theta_0) = e_k$ and $e_{k;m_1}(\theta_0) = e_{k;m_1}$. Because $e_0(\theta) = 1$, we have

$$\begin{aligned}
L_{52n} &= \frac{1}{n} \sum_{j=1}^{m_n} \sum_{t_1,t_2=j+1}^{n} \lambda_{1t_1}\lambda_{1t_2}\lambda_{2n(t_1-j)}\lambda_{2n(t_2-j)} \\
&= \frac{1}{n} \sum_{j=1}^{m_n} \sum_{t_1,t_2=j+1}^{n} \sum_{k_1,k_2=0}^{\infty} \sum_{k_3=1}^{t_1-j-1} \sum_{k_4=1}^{t_2-j-1} \psi_{k_1,t_1}\psi_{k_2,t_2} u_{-k_1}u_{-k_2} \\
&\quad \times (e_{k_3}(\hat{\theta}_n) - e_{k_3}(\theta_0))(e_{k_4}(\hat{\theta}_n) - e_{k_4}(\theta_0)) Y_{t_1-j-k_3} Y_{t_2-j-k_4} \\
&= \frac{1}{n} \sum_{j=1}^{m_n} \sum_{t_1,t_2=j+1}^{n} \sum_{k_1,k_2=0}^{\infty} \sum_{k_3=1}^{t_1-j-1} \sum_{k_4=1}^{t_2-j-1} \psi_{k_1,t_1}\psi_{k_2,t_2} u_{-k_1}u_{-k_2} Y_{t_1-j-k_3} Y_{t_2-j-k_4} \\
&\quad \times \left(\sum_{m_1=1}^{p+q+1} \left(\hat{\theta}_n^{(m_1)} - \theta_0^{(m_1)}\right) e_{k_3;m_1} + \sum_{m_1,m_2=1}^{p+q+1} \left(\hat{\theta}_n^{(m_1)} - \theta_0^{(m_1)}\right) e_{k_3;(m_1,m_2)}(\tilde{\theta}_{k_3 n}) \right. \\
&\qquad \left. \times \left(\hat{\theta}_n^{(m_2)} - \theta_0^{(m_2)}\right) \right) \left(\sum_{m_3=1}^{p+q+1} \left(\hat{\theta}_n^{(m_3)} - \theta_0^{(m_3)}\right) e_{k_4;m_3} \right. \\
&\qquad\qquad + \sum_{m_3,m_4=1}^{p+q+1} \left(\hat{\theta}_n^{(m_3)} - \theta_0^{(m_3)}\right) e_{k_4;(m_3,m_4)}(\tilde{\theta}_{k_4 n}) \\
&\qquad\qquad \left. \times \left(\hat{\theta}_n^{(m_4)} - \theta_0^{(m_4)}\right) \right) = \sum_{h=1}^{4} L_{52hn}.
\end{aligned}$$

Write $Y_t = \sum_{k=0}^{\infty} a_k u_{t-k}$. To show $L_{521n} = o_p(1)$, it suffices to show that for any $(m_1, m_3) \in \{1, \ldots, p+q+1\}^2$,

$$\tilde{L}_{521n} = \sum_{j=1}^{m_n} \sum_{t_1,t_2=j+1}^{n} \sum_{k_1,k_2=0}^{\infty} \psi_{k_1,t_1}\psi_{k_2,t_2} \sum_{k_3=1}^{t_1-j-1} \sum_{k_4=1}^{t_2-j-1} e_{k_3;m_1} e_{k_4;m_3} u_{-k_1}u_{-k_2} \times Y_{t_1-j-k_3} Y_{t_2-j-k_4}$$

$$= \sum_{j=1}^{m_n} \sum_{t_1,t_2=j+1}^{n} \sum_{k_1,k_2=0}^{\infty} \sum_{k_3=1}^{t_1-j-1} \sum_{k_4=1}^{t_2-j-1} \sum_{h_1,h_2=0}^{\infty} \psi_{k_1,t_1}\psi_{k_2,t_2}$$
$$\times a_{h_1}a_{h_2}e_{k_3;m_1}e_{k_4;m_3}u_{-k_1}u_{-k_2}u_{t_1-j-k_3-h_1}u_{t_2-j-k_4-h_2}$$
$$= o_p(n^2).$$

Note that

$$\mathbb{E}(\tilde{L}_{521n}^2) = \sum_{j,j'=1}^{m_n} \sum_{t_1,t_2=j+1}^{n} \sum_{t_1',t_2'=j'+1}^{n} \sum_{k_1,k_2,k_1',k_2'=0}^{\infty} \sum_{k_3=1}^{t_1-j-1} \sum_{k_4=1}^{t_2-j-1} \sum_{k_3'=1}^{t_1'-j'-1}$$
$$\times \sum_{k_4'=1}^{t_2'-j'-1} \sum_{h_1,h_2,h_1',h_2'=0}^{\infty} \psi_{k_1,t_1}\psi_{k_2,t_2}\psi_{k_1',t_1'}\psi_{k_2',t_2'}a_{h_1}a_{h_2}a_{h_1'}a_{h_2'}e_{k_3;m_1}e_{k_4;m_3}$$
$$\times e_{k_3';m_1}e_{k_4';m_3}$$
$$\times \mathbb{E}\Big(u_{-k_1}u_{-k_2}u_{t_1-j-k_3-h_1}u_{t_2-j-k_4-h_2}u_{-k_1'}u_{-k_2'}u_{t_1'-j'-k_3'-h_1'}$$
$$\times u_{t_2'-j'-k_4'-h_2'}\Big)$$
$$\le C \sum_{j,j'=1}^{m_n} \sum_{t_1,t_2=j+1}^{n} \sum_{t_1',t_2'=j'+1}^{n} \sum_{k_1,k_2,k_1',k_2'=0}^{\infty} \sum_{k_3,k_4,k_3',k_4'=1}^{\infty} \sum_{h_1,h_2,h_1',h_2'=0}^{\infty}$$
$$\times \left|\psi_{k_1,t_1}\psi_{k_2,t_2}\right|\left|\psi_{k_1',t_1'}\psi_{k_2',t_2'}\right|\left|a_{h_1}a_{h_2}\right|\left|a_{h_1'}a_{h_2'}\right|\left|k_3k_3'k_4k_4'\right|^{-1-\epsilon}\mathrm{II},$$

where

$$\mathrm{II} = |\mathbb{E}(u_{-k_1}u_{-k_2}u_{t_1-j-k_3-h_1}u_{t_2-j-k_4-h_2}u_{-k_1'}u_{-k_2'}u_{t_1'-j'-k_3'-h_1'}u_{t_2'-j'-k_4'-h_2'})|$$
$$= \left|\sum_g \mathrm{cum}(u_{i_j}, i_j \in g_1)\ldots\mathrm{cum}(u_{i_j}, i_j \in g_p)\right|.$$

In the preceding equation, Σ_g is over all partitions $g = \{g_1 \cup \cdots \cup g_p\}$ of the index set $\{-k_1, t_1-j-k_3-h_1, -k_1', t_1'-j'-k_3'-h_1', -k_2, t_2-j-k_4-h_2, -k_2', t_2'-j'-k_4'-h_2'\}$. Because $\mathbb{E}(u_t) = 0$, only partitions g with $\#g_i > 1$ for all i contribute. We shall divide all contributing partitions into the following several types and treat them one by one.

1. $\#g_1 = \#g_2 = \#g_3 = \#g_4 = 2$. One such term is

$$\mathrm{cov}(u_{-k_1}, u_{t_1-j-k_3-h_1})\mathrm{cov}(u_{-k_1'}, u_{t_1'-j'-k_3'-h_1'})\mathrm{cov}(u_{-k_2}, u_{t_2-j-k_4-h_2})$$
$$\times \mathrm{cov}(u_{-k_2'}, u_{t_2'-j'-k_4'-h_2'}),$$

which is nonzero when $-k_1 = t_1 - j - k_3 - h_1$, $-k_1' = t_1' - j' - k_3' - h_1'$, $-k_2 = t_2 - j - k_4 - h_2$, and $-k_2' = t_2' - j' - k_4' - h_2'$. Define $a_h = 0$ if $h < 0$. Then for any fixed $g \in \mathbb{Z}$, $\sum_{h=0}^{\infty} |a_h a_{h+g}| \le \sum_{h=0}^{\infty} a_h^2 := S_a < \infty$. For any fixed $t_1, t_1', t_2, t_2', j, j', k_3, k_4, k_3', k_4'$, by the Cauchy–Schwarz inequality,

$$\sum_{k_1,k_2,k_1',k_2'=0}^{\infty} \left|\psi_{k_1,t_1}\psi_{k_2,t_2}\right|\left|\psi_{k_1',t_1'}\psi_{k_2',t_2'}\right|\left|a_{k_1+t_1-j-k_3}a_{k_2+t_2-j-k_4}\right|$$
$$\times\left|a_{k_1'+t_1'-j'-k_3'}a_{k_2'+t_2'-j'-k_4'}\right| \le \left(\sum_{k_1,k_2,k_1',k_2'=0}^{\infty}\psi_{k_1,t_1}^2\psi_{k_2,t_2}^2\psi_{k_1',t_1'}^2\psi_{k_2',t_2'}^2\right)^{1/2} S_a^2$$
$$= O((t_1t_2t_1't_2')^{-1/2}).$$

So this term is $O(m_n^2n^2) = o(n^4)$. Similarly, all nonvanishing terms involve four restrictions on the indexes $k_1, k_2, k_1', k_2', h_1, h_2, h_1', h_2'$ once we fix t_1, t_1', t_2, t_2', j, j', k_3, k_4, k_3', k_4'. The contributions from these terms are of order $o(n^4)$.

2. $\#g_1 = \#g_2 = 3, \#g_3 = 2$. A typical term is

$$\mathrm{cum}(u_{-k_1}, u_{t_1-j-k_3-h_1}, u_{-k_1'})\mathrm{cum}(u_{t_1'-j'-k_3'-h_1'}, u_{-k_2}, u_{t_2-j-k_4-h_2})$$
$$\times \mathrm{cov}(u_{-k_2'}, u_{t_2'-j'-k_4'-h_2'}).$$

So for any fixed $t_1, t_1', t_2, t_2', j, j', k_3, k_4, k_3', k_4'$,

$$\sum_{k_1,k_1',h_1=0}^{\infty}\left|\psi_{k_1,t_1}\psi_{k_1',t_1'}a_{h_1}\right|\left|\mathrm{cum}(u_{-k_1}, u_{t_1-j-k_3-h_1}, u_{-k_1'})\right| \tag{A.15}$$
$$\le C\sum_{k_1,k_1',h_1=0}^{\infty}\left|\psi_{k_1,t_1}\psi_{k_1',t_1'}a_{h_1}\right| r^{\max(-k_1,t_1-j-k_3-h_1,-k_1')-\min(-k_1,t_1-j-k_3-h_1,-k_1')}.$$

Consider the case $-k_1' \ge -k_1 \ge t_1 - j - k_3 - h_1$. Then the corresponding term in (A.15) is

$$C\sum_{s_1,s_2,k_1=0}^{\infty}|\psi_{k_1,t_1}\psi_{k_1-s_1,t_1'}a_{s_2+k_1+t_1-j-k_3}|r^{s_1+s_2} = O\left((t_1t_1')^{-1/2}\right),$$

where we have applied the Cauchy–Schwarz inequality and the fact that $\sum_{k=0}^{\infty}\psi_{k,t}^2 = O(t^{-1})$. Other cases can be treated in a similar fashion. So (A.15) is $O\left((t_1t_1')^{-1/2}\right)$. Similarly, we can show that

$$\sum_{h_1',h_2,k_2=0}^{\infty}\left|\psi_{k_2,t_2}a_{h_1'}a_{h_2}\mathrm{cum}(u_{t_1'-j'-k_3'-h_1'}, u_{-k_2}, u_{t_2-j-k_4-h_2})\right| = O\left(t_2^{-1/2}\right)$$

and

$$\sum_{k_2',h_2'=0}^{\infty}\left|a_{h_2'}\psi_{k_2',t_2'}\mathrm{cov}(u_{-k_2'}, u_{t_2'-j'-k_4'-h_2'})\right| = O\left((t_2')^{-1/2}\right).$$

Thus these terms contribute $O(m_n^2n^2) = o(n^4)$.

3. $\#g_1 = \#g_2 = 4$; $\#g_1 = 4, \#g_2 = \#g_3 = 2$; $\#g_1 = 5, \#g_2 = 3$; $\#g_1 = 6, \#g_2 = 2$, and $\#g_1 = 8$. Following a similar argument as the second case, it is not hard to see that the contributions of all these terms are $o(n^4)$.

So $L_{521n} = o_p(1)$. Under the assumption that u_t is GMC(8), it is not hard to show that $\mathbb{E}(Y_t^4) < \infty$ and $\sup_{t\in\mathbb{N}} \mathbb{E}\lambda_{1t}^4 < \infty$; compare the derivation of $\mathbb{E}(L_{51n})$ in the proof of Theorem 3.2. Together with Lemma A.7, we have $\mathbb{E}|L_{522n}|\mathbf{1}(\hat{\theta}_n \in \Theta_\delta) = O\left(m_n/n^{1/2}\right) = o(1)$, and so $L_{522n} = o_p(1)$. Similarly we derive $L_{52kn} = o_p(1)$, $k = 3, 4$. Now the proof is complete. ■

The following lemma is an extension of Lemma A.1 of Francq and Zakoïan (2000) to the FARIMA model.

LEMMA A.7. *For any $\theta \in \Theta_\delta$ and any $(m_1, m_2) \in \{1, \ldots, p+q+1\}^2$, there exist absolutely summable sequences $(e_k(\theta))_{k\geq 0}$, $(e_{k;m_1}(\theta))_{k\geq 1}$, and $(e_{k;(m_1,m_2)}(\theta))_{k\geq 1}$ such that almost surely*

$$u_t(\theta) = \sum_{k=0}^{\infty} e_k(\theta) Y_{t-k}, \qquad \frac{\partial u_t(\theta)}{\partial \theta_{m_1}} = \sum_{k=1}^{\infty} e_{k;m_1}(\theta) Y_{t-k}$$

and

$$\frac{\partial^2 u_t(\theta)}{\partial \theta_{m_1} \partial \theta_{m_2}} = \sum_{k=1}^{\infty} e_{k;(m_1,m_2)}(\theta) Y_{t-k}.$$

Further, there exists an $\epsilon > 0$, such that

$$\sup_{\theta\in\Theta_\delta} |e_k(\theta)| = O(k^{-1-\epsilon}), \qquad \sup_{\theta\in\Theta_\delta} |e_{k;m_1}(\theta)| = O(k^{-1-\epsilon}), \quad \text{and}$$

$$\sup_{\theta\in\Theta_\delta} |e_{k;(m_1,m_2)}(\theta)| = O(k^{-1-\epsilon}).$$

Proof of Lemma A.7. Letting $X_t = (1-B)^d Y_t$, then $\phi_\Lambda(B) X_t = \psi_\Lambda(B) u_t$. By Lemma A.1 in Francq and Zakoïan (2000), there exist sequences $(c_k(\Lambda))_{k\geq 0}$, $(c_{k;m_1}(\Lambda))_{k\geq 1}$, and $(c_{k;(m_1,m_2)}(\Lambda))_{k\geq 1}$ such that

$$u_t(\Lambda) = \sum_{j=0}^{\infty} c_j(\Lambda) X_{t-j}, \qquad \partial u_t(\Lambda)/\partial \Lambda_{m_1} = \sum_{j=1}^{\infty} c_{j;m_1}(\Lambda) X_{t-j},$$

and

$$\partial^2 u_t(\Lambda)/\partial \Lambda_{m_1} \partial \Lambda_{m_2} = \sum_{j=1}^{\infty} c_{j;(m_1,m_2)}(\Lambda) X_{t-j}.$$

Further, there exists an $r \in [0, 1)$ such that

$$\sup_{\Lambda\in\Omega_\delta} |c_j(\Lambda)| = O(r^j), \qquad \sup_{\Lambda\in\Omega_\delta} |c_{j;m_1}(\Lambda)| = O(r^j),$$

$$\sup_{\Lambda\in\Omega_\delta} |c_{j;(m_1,m_2)}(\Lambda)| = O(r^j).$$

Note that $X_t = \sum_{s=0}^{\infty} \phi_s(d) Y_{t-s}$, where $\phi_s(d) = \Gamma(s-d)/\{\Gamma(-d)\Gamma(s+1)\}$. Therefore, we get $u_t(\theta) = \sum_{k=0}^{\infty} e_k(\theta) Y_{t-k}$, where $e_k(\theta) = \sum_{j=0}^{k} c_j(\Lambda) \phi_{k-j}(d)$. The conclusion follows from the definition of Θ_δ and the fact that $d_0 \in (0, 1/2)$. ■