Econometric Theory, **18**, 2002, 278–296. Printed in the United States of America.
DOI: 10.1017.S026646660218203X

A UNIFIED APPROACH TO THE MEASUREMENT ERROR PROBLEM IN TIME SERIES MODELS

Katsuto Tanaka
Hitotsubashi University

The measurement error problem that we consider in this paper is concerned with the situation where time series data of various kinds—short memory, long memory, and random walk processes—are contaminated by white noise. We suggest a unified approach to testing for the existence of such noise. It is found that the power of our test crucially depends on the underlying process.

1. INTRODUCTION

It is sometimes the case that observations are contaminated by noise so that the true relationship between variables is somewhat obscured. This is usually called the measurement error problem, and it has been treated under various circumstances.

In this paper we focus on the time series situation and consider the model

$$y_t = x_t + u_t \qquad (t = 1, \dots, T), \tag{1}$$

where only $\{y_t\}$ is observable, $\{x_t\}$ is an underlying process or a signal, and $\{u_t\}$ is a measurement error. We assume that $\{x_t\}$ and $\{u_t\}$ are independent of each other. Moreover, $\{u_t\}$ is assumed to be independent and identically distributed with mean 0 and variance $\rho\sigma^2$, which is abbreviated as i.i.d.$(0, \rho\sigma^2)$ hereafter, where ρ is a nonnegative constant whereas σ^2 is a positive constant that is the variance of the innovation driving the signal $\{x_t\}$. The signal process $\{x_t\}$ is dependent and will be specified later.

The purpose of the present paper is to test if the measurement error really exists. To this end we consider the testing problem

$$H_0 : \rho = 0 \quad \text{vs.} \quad H_1 : \rho > 0. \tag{2}$$

Note that there exists no measurement error under H_0, whereas H_1 implies some indication of the measurement error with its degree increasing with ρ.

I am grateful to N. Katayama, E. Kurozumi, and two anonymous referees for their valuable comments and suggestions. Address correspondence to: Katsuto Tanaka, Department of Economics, Hitotsubashi University, Kunitachi, Tokyo 186, Japan; e-mail: tanaka@stat.hit-u.ac.jp.

  0266-4666/02 $9.50

To devise a test we need to specify the signal $\{x_t\}$ in (1), which will be done in Section 2, where three typical processes are considered, namely, stationary short memory, stationary long memory, and random walk processes. For these three cases we suggest the Lagrange multiplier (LM) test. The test statistics will be derived and interpreted in a systematic way. It will be shown that the statistics follow normality notwithstanding the null hypothesis being on the boundary of the parameter space.

Section 3 discusses asymptotic properties of the test by deriving limiting powers under a sequence of local alternatives of the form $\rho = c/\sqrt{T}$ with c a positive constant, whereas some simulations are conducted in Section 4 to demonstrate our methodology. It will be noticed that the identification problem emerges as the alternative deviates further from the null. This occurs in the case of both the stationary short and long memory signals, whereas it does not in the random walk signal. This is because the signals in the former are dominated by the measurement error, which tends to invalidate the estimation of the signal parameters. In particular, the identification problem turns out to be serious in the long memory signal, which may be one source of difficulties in estimating fractional autoregressive integrated moving average (denoted as ARFIMA hereafter) models. Some concluding remarks appear in Section 5, and proofs of theorems and lemmas are given in the Appendix.

2. THE LM TEST FOR THE MEASUREMENT ERROR

In this section we derive the LM test for the testing problem (2). For this purpose we specify the signal process $\{x_t\}$ in (1) as one of the following three processes:

Case 1. $\beta(L)x_t = \varepsilon_t,$ **(3)**

Case 2. $(1-L)x_t = \varepsilon_t,$ **(4)**

Case 3. $(1-L)^d x_t = \varepsilon_t,$ **(5)**

where $\{\varepsilon_t\}$ follows i.i.d.$(0, \sigma^2)$, whereas

$$\beta(L) = 1 - \beta_1 L - \cdots - \beta_p L^p$$

is a polynomial of the lag operator L. We assume that $\beta(z) = 0$ has all roots outside the unit circle. Thus $\{x_t\}$ in Case 1 is a stationary AR(p) process. The testing problem for this case was dealt with earlier in Tanaka (1983). Case 2 corresponds to the random walk process, whereas, in Case 3, we assume that the differencing parameter d is unknown and lies between 0 and $\frac{1}{2}$. Thus $\{x_t\}$ in Case 3 follows a stationary ARFIMA$(0, d, 0)$ process.

In subsequent discussions we derive the LM test for each of the preceding three cases. For this purpose we impose normality on $\{\varepsilon_t\}$ and $\{u_t\}$ so that the observable process $\{y_t\}$ is normal. We, however, note that normality is not required for asymptotic arguments (for Cases 1 and 2, see McLeod, 1978, for case 3, see Giraitis and Sargailis, 1990).

It might be thought that the LM test is easily derived in the present situation. It, however, turns out that the usual procedure for deriving the LM test cannot be applied directly. To see this let us consider the log-likelihood for $y = (y_1, \dots, y_T)'$ in (1), which is given, under normality, by

$$L(\rho, \sigma^2, \theta) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log|\Omega(\theta) + \rho I_T| - \frac{1}{2\sigma^2} y'(\Omega(\theta) + \rho I_T)^{-1} y,$$

where θ is a vector of parameters associated with the signal $\{x(t)\}$, $\sigma^2\Omega(\theta)$ is the covariance matrix of $x = (x_1, \dots, x_T)'$, and I_T is the $T \times T$ identity matrix. We now have

$$\left.\frac{\partial L(\rho, \sigma^2, \theta)}{\partial \rho}\right|_{H_0} = -\frac{1}{2}\operatorname{tr}(\Omega^{-1}(\hat{\theta})) + \frac{1}{2\hat{\sigma}^2} y'\Omega^{-2}(\hat{\theta})y,$$

where $\hat{\sigma}^2$ and $\hat{\theta}$ are the MLE's of σ^2 and θ, respectively, evaluated under H_0. Then we can devise a one-sided LM test based on $\partial L/\partial\rho|_{H_0}$, but it is not easy, in general, to compute this statistic.

One exception is Case 2, where $\Omega(\theta) = CC'$ with

$$C = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & 0 \\ 1 & 1 & \cdot & \cdot & \cdot & 1 \end{pmatrix}, \qquad C^{-1} = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ -1 & 1 & & & & \cdot \\ 0 & -1 & \cdot & & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & -1 & 1 \end{pmatrix}.$$

Then we have

$$\operatorname{tr}(\Omega^{-1}(\hat{\theta})) = 2T - 1, \qquad \hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - y_{t-1})^2,$$

$$y'\Omega^{-2}(\hat{\theta})y = \sum_{t=1}^{T-1}(y_{t+1} - 2y_t + y_{t-1})^2 + (y_T - y_{T-1})^2,$$

where $y_0 = 0$. Thus we obtain, putting $\hat{\varepsilon}_t = y_t - y_{t-1}$,

$$\left.\frac{\partial L}{\partial \rho}\right|_{H_0} = -\frac{2T-1}{2} + \frac{T}{2}\,\frac{\sum_{t=1}^{T}\hat{\varepsilon}_t^2 + \sum_{t=2}^{T}\hat{\varepsilon}_t^2 - 2\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}\hat{\varepsilon}_t}{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}$$

$$= -T\,\frac{\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}\hat{\varepsilon}_t}{\sum_{t=1}^{T}\hat{\varepsilon}_t^2} + \frac{1}{2} - \frac{T}{2}\,\frac{\hat{\varepsilon}_1^2}{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}$$

$$= -Tr_1 + O_p(1),$$

where r_1 is the first-order autocorrelation of $\{y_t - y_{t-1}\}$. We can now conduct the LM test that rejects H_0 when $\sqrt{T}r_1$ takes small values, noting that $\sqrt{T}r_1 \to N(0,1)$ under H_0.

The preceding derivation of the LM test for Case 2 is simple but exceptional. In subsequent discussions we take an alternative approach, which enables us to derive the LM test for the three cases in a unified way.

2.1. Case 1

It follows from (1) and (3) that

$$\beta(L)y_t = \varepsilon_t + \beta(L)u_t = \delta(L)a_t \qquad (t = 1, \ldots, T), \tag{6}$$

where $\delta(L) = 1 - \delta_1 L - \cdots - \delta_p L^p$ with all the roots of $\delta(z) = 0$ outside the unit circle, whereas $\{a_t\}$ follows i.i.d.$(0, \sigma_a^2)$ random variables. The parameters $\delta = (\delta_1, \ldots, \delta_p)'$ and σ_a^2 can be determined uniquely from the equation

$$\sigma^2[1 + \rho\beta(L)\beta(L^{-1})] = \sigma_a^2\delta(L)\delta(L^{-1}). \tag{7}$$

By assuming normality of $\{a_t\}$, the log-likelihood for (6) may be given by

$$L(\rho, \sigma_a^2, \beta, \delta) = -\frac{T}{2}\log(2\pi\sigma_a^2) - \frac{1}{2\sigma_a^2}\sum_{t=1}^{T}\left\{\frac{\beta(L)}{\delta(L)}\,y_t\right\}^2. \tag{8}$$

Then we have

$$\left.\frac{\partial L}{\partial \rho}\right|_{H_0} = \sum_{i=1}^{p}\left.\frac{\partial L}{\partial \delta_i}\right|_{H_0}\left.\frac{\partial \delta_i}{\partial \rho}\right|_{H_0}, \tag{9}$$

where it is not hard to see

$$\left.\frac{\partial L}{\partial \delta_i}\right|_{H_0} = -\frac{1}{\hat{\sigma}^2}\sum_{t=i+1}^{T} \hat{\varepsilon}_{t-i}\hat{\varepsilon}_t = -T\hat{r}_i. \tag{10}$$

Here $\hat{\sigma}^2 = \sum_{t=1}^{T} \hat{\varepsilon}_t^2/T$ and $\hat{\varepsilon}_t = y_t - \hat{\beta}_1 y_{t-1} - \cdots - \hat{\beta}_p y_{t-p}$ with $\hat{\beta}_i$ being the least squares estimator (LSE) of β_i in the AR(p) process $\beta(L)y_t = \varepsilon_t$. Thus $\hat{r}_i$ is the ith-order autocorrelation of the residual process $\{\hat{\varepsilon}_t\}$.

As for $\partial\delta_i/\partial\rho|_{H_0}$ in (9), we obtain the following lemma.

LEMMA 1. *For the process (6), it holds that, under $H_0: \rho = 0$,*

$$\left.\frac{\partial \delta_i}{\partial \rho}\right|_{H_0} = \hat{\beta}_i - \hat{\beta}_1\hat{\beta}_{i+1} - \cdots - \hat{\beta}_{p-i}\hat{\beta}_p = \hat{\lambda}_i \qquad (i = 1,\dots,p). \tag{11}$$

It now follows from (9)–(11) that

$$\left.\frac{\partial L}{\partial \rho}\right|_{H_0} = -T\sum_{i=1}^{p} \hat{\lambda}_i \hat{r}_i = T\hat{\alpha}'\hat{r},$$

where $\hat{\alpha} = -(\hat{\lambda}_1,\dots,\hat{\lambda}_p)'$ and $\hat{r} = (\hat{r}_1,\dots,\hat{r}_p)'$. Using the results of Box and Pierce (1970) (see also McLeod, 1978), we have, under H_0,

$$\sqrt{T}\hat{r} \to N(0, I_p - \sigma^2 J^{-1}(\beta)\Gamma_p^{-1}J^{-1}(\beta)'), \tag{12}$$

where

$$J(\beta) = \begin{pmatrix} -1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \beta_1 & -1 & & & & & \cdot \\ \beta_2 & \beta_1 & -1 & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot & 0 \\ \beta_{p-1} & \beta_{p-2} & \cdot & \cdot & \cdot & \beta_1 & -1 \end{pmatrix},$$

$$\Gamma_p = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|\beta(e^{i\omega})|^2}((e^{i(j-k)\omega}))\,d\omega : p \times p.$$

Note that Γ_p is the covariance matrix for $y_{t-1},\dots,y_{t-p}$ under H_0.

Therefore the LM statistic we suggest here takes the form

$$\begin{aligned} S_{T1} &= \sqrt{T}\hat{\alpha}'\hat{r}/(\hat{\alpha}'(I_p - \hat{\sigma}^2 J^{-1}(\hat{\beta})\hat{\Gamma}_p^{-1}J^{-1}(\hat{\beta})')\hat{\alpha})^{1/2} \\ &= \sqrt{T}\sum_{i=1}^{p}\hat{\alpha}_i\hat{r}_i \Bigg/ \left(\sum_{i=1}^{p}\hat{\alpha}_i^2 - \hat{\sigma}^2\hat{\beta}'\hat{\Gamma}_p^{-1}\hat{\beta}\right)^{1/2}, \end{aligned} \tag{13}$$

where $\hat{\Gamma}_p$ is a consistent estimator of Γ_p under H_0. It evidently holds that $S_{T1} \to N(0,1)$ under H_0, and H_0 should be rejected when S_{T1} takes large values. Tanaka (1983) derived S_{T1} via somewhat a complicated route.

The statistic S_{T1} is, apart from a normalization factor, a linear combination of the residual autocorrelations of the first p lags, where the weight $\hat{\alpha}_i = -\hat{\beta}_i + \hat{\beta}_1\hat{\beta}_{i+1} + \cdots + \hat{\beta}_{p-i}\hat{\beta}_p$ can be interpreted as follows. Put $\alpha_i = -\beta_i + \beta_1\beta_{i+1} + \cdots + \beta_{p-i}\beta_p$ and consider an auxiliary process $z_t = \beta(L)\varepsilon_t$, which is an inverse process of the signal $\{x_t\}$ satisfying $\beta(L)x_t = \varepsilon_t$. Then we have

$$\alpha_i = \mathrm{Cov}(z_t, z_{t-i})/\sigma^2 \qquad (i = 1,\ldots,p). \tag{14}$$

A similar interpretation will be given to the LM statistics derived from Cases 2 and 3.

In Section 3 we shall obtain the asymptotic distribution of S_{T1} under a sequence of local alternatives.

2.2. Case 2

Following the approach taken in the previous section, we can easily obtain the LM statistic for Case 2, whose signal is given in (4). Because we have, from (1) and (4),

$$(1 - L)y_t = \varepsilon_t + (1 - L)u_t = \delta(L)a_t \qquad (t = 1,\ldots,T), \tag{15}$$

where $\delta(L) = 1 - \delta L$ with $|\delta| < 1$, the log-likelihood for (15) is given by

$$L(\rho, \sigma_a^2, \delta) = -\frac{T}{2}\log(2\pi\sigma_a^2) - \frac{1}{2\sigma_a^2}\sum_{t=1}^{T}\left\{\frac{1-L}{\delta(L)}\,y_t\right\}^2.$$

It is now an easy matter to obtain

$$\left.\frac{\partial L}{\partial \rho}\right|_{H_0} = \left.\frac{\partial L}{\partial \delta}\right|_{H_0}\left.\frac{\partial \delta}{\partial \rho}\right|_{H_0} = -Tr_1,$$

where r_1 is the first-order autocorrelation of $(1 - L)y_t$. Because $\sqrt{T}r_1 \to N(0,1)$ under H_0, the LM test for the present case rejects H_0 when

$$S_{T2} = \sqrt{T}\alpha_1 r_1 \tag{16}$$

takes large values, where $\alpha_1 = -1$ and $S_{T2} \to N(0,1)$ under H_0.

It is noticed that the statistic S_{T2} is of a similar form to S_{T1} in (13), although the former is much simpler. In fact, S_{T2} is based only on the first-order autocorrelation of residuals. This is because the signal in the present case follows a random walk that is a special case of AR(1), whereas it follows AR(p) in Case 1. The coefficient $\alpha_1(= -1)$ in (16) also has the same interpretation as in (14), that is, $\alpha_1 = \mathrm{Cov}(z_t, z_{t-1})/\sigma^2$, where $z_t = (1 - L)\varepsilon_t$.

We note in passing that the test based on S_{T2} in (16) is asymptotically uniformly most powerful and invariant (UMPI). In fact, the testing problem in the present case is invariant under the group of scale transformations, and $y/\sqrt{y'y}$ is a maximal invariant. Then it can be shown (Tanaka, 1996, Ch. 9; Tanaka, 1999) that the MPI test of $H_0: \rho = 0$ against $H_1: \rho = c/\sqrt{T}$ with c a fixed positive constant is asymptotically the same as the test based on S_{T2}.

2.3. Case 3

This case is most complicated but can be dealt with in the same way as before. We first have, from (1) and (5),

$$(1-L)^d y_t = \varepsilon_t + (1-L)^d u_t = \delta(L) a_t \qquad (t = 1,\ldots,T), \tag{17}$$

where $0 < d < \frac{1}{2}$ and $\delta(L)$ is now a lag polynomial of infinite order determined from

$$\sigma^2[1 + \rho(1-L)^d(1-L^{-1})^d] = \sigma_a^2 \delta(L)\delta(L^{-1}).$$

Note here that

$$(1-L)^d = \frac{1}{\Gamma(-d)} \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)} L^j.$$

We then consider the log-likelihood $L(\rho, \sigma_a^2, d, \delta)$ for (17) given by

$$L(\rho, \sigma_a^2, d, \delta) = -\frac{T}{2}\log(2\pi\sigma_a^2) - \frac{1}{2\sigma_a^2}\sum_{t=1}^{T}\left\{\frac{(1-L)^d}{\delta(L)} y_t\right\}^2,$$

which leads us to obtain

$$\left.\frac{\partial L}{\partial \rho}\right|_{H_0} = T \sum_{i=1}^{T-1} \hat{\alpha}_i \hat{r}_i, \tag{18}$$

where $\hat{r}_i$ is the ith order autocorrelation of $\hat{\varepsilon}_t = (1-L)^{\hat{d}} y_t$ with $\hat{d}$ being the MLE of d under H_0. It is known that $\sqrt{T}(\hat{d} - d) \to N(0, 6/\pi^2)$ under H_0. On the other hand $\hat{\alpha}_i$ is a consistent estimator, under H_0, of

$$\begin{aligned}\alpha_i &= \mathrm{Cov}((1-L)^d \varepsilon_t, (1-L)^d \varepsilon_{t-i})/\sigma^2 \\ &= \frac{(-1)^i \Gamma(1+2d)}{\Gamma(1-i+d)\Gamma(1+i+d)} \\ &= \frac{\Gamma(i-d)\Gamma(1+2d)}{\Gamma(-d)\Gamma(1+d)\Gamma(1+i+d)}.\end{aligned} \tag{19}$$

Here the second equality is due to Adenstedt (1974), whereas the last is due to Hosking (1981).

The asymptotic null distribution of $\hat{r} = (\hat{r}_1, \ldots, \hat{r}_m)'$ for a fixed integer m is given by the following lemma, which is essentially due to Li and McLeod (1986).

LEMMA 2. *Suppose that*

$$(1-L)^d y_t = \varepsilon_t, \qquad 0 < d < \frac{1}{2} \qquad (t = 1, \ldots, T), \tag{20}$$

where $\{\varepsilon_t\} \sim i.i.d.N(0, \sigma^2)$, *and let* $\hat{d}$ *be the* MLE *of d. Define also* $\hat{\varepsilon}_t = (1-L)^{\hat{d}} y_t$ *and* $\hat{r} = (\hat{r}_1, \ldots, \hat{r}_m)'$ *with* $\hat{r}_i$ *being the ith-order autocorrelation of* $\{\hat{\varepsilon}_t\}$. *Then it holds that, for any fixed m,*

$$\sqrt{T}\,\hat{r} \to N\left(0, I_m - \frac{6}{\pi^2} g_m g_m'\right), \tag{21}$$

where $g_m = (1, \frac{1}{2}, \ldots, 1/m)'$.

We note in passing that the preceding result holds true without imposing normality on $\{\varepsilon_t\}$ (see Giraitis and Surgailis, 1990). Then we obtain the following theorem.

THEOREM 1. *The LM test for* $H_0\colon \rho = 0$ *vs.* $H_1\colon \rho > 0$ *in the model (17) rejects* H_0 *when*

$$S_{T3} = \sqrt{T} \sum_{i=1}^{T-1} \hat{\alpha}_i \hat{r}_i \Bigg/ \left(\sum_{i=1}^{T-1} \hat{\alpha}_i^2 - \frac{6}{\pi^2} \left(\sum_{i=1}^{T-1} \frac{1}{i} \hat{\alpha}_i \right)^2 \right)^{1/2} \tag{22}$$

takes large values, where $S_{T3} \to N(0,1)$ *under* H_0.

We have derived the LM tests based on S_{T1}, S_{T2}, and S_{T3} for Cases 1–3, respectively. These statistics are, apart from the normalizing factor, a linear combination of residual autocorrelations, where the weights are autocovariances of the inverse process to the signal. In the next section we examine the asymptotic properties of these tests under a sequence of local alternatives.

3. ASYMPTOTIC LOCAL POWERS OF THE LM TESTS

In this section we investigate the asymptotic properties of the LM tests derived in the last section. For this purpose we compute the limiting powers of the tests under a sequence of local alternatives, which takes the form of

$$H_1 : \rho = \frac{c}{\sqrt{T}}, \tag{23}$$

where c is a positive constant.

3.1. Case 1

We consider the asymptotic distribution of S_{T1} in (13) as $T \to \infty$ under $\rho = c/\sqrt{T}$. Let us define by $r_i(\beta, \rho)$ the ith-order autocorrelation of the true innovation $\{a_t\}$ in (6). Then we have, by the Taylor expansion,

$$\hat{r}_i = r_i(\hat{\beta},0) = r_i(\beta,\rho) + \sum_{j=1}^{p} \frac{\partial r_i(\beta,\rho)}{\partial \beta_j}(\hat{\beta}_j - \beta_j) + \frac{\partial r_i(\beta,\rho)}{\partial \rho}(-\rho) + o_p\left(\frac{1}{\sqrt{T}}\right). \tag{24}$$

The partial derivatives on the right side of (24) can be evaluated by using the following lemma.

LEMMA 3. *For the model (6) with $\rho = c/\sqrt{T}$, it holds that, as $T \to \infty$,*

$$\frac{\partial r_i(\beta,\rho)}{\partial \beta_j} \to \begin{cases} -1 & i = j \\ \psi_{i-j} & i > j \\ 0 & i < j, \end{cases}$$

in probability, where ψ_i's are the coefficients in the expansion $1/\beta(L) = 1 - \psi_1 L - \psi_2 L^2 - \dots$, and

$$\frac{\partial r_i(\beta,\rho)}{\partial \rho} \to -\alpha_i = \beta_i - \beta_1\beta_{i+1} - \cdots - \beta_{p-i}\beta_p,$$

in probability.

It follows from (24) and Lemma 3 that

$$\sqrt{T}\hat{r}_i = \sqrt{T} r_i(\beta,\rho) + \sum_{j=1}^{p} \phi_{i-j}\sqrt{T}(\hat{\beta}_j - \beta_j) + c\alpha_i + o_p(1), \tag{25}$$

where $\phi_{i-j} = -1$ for $i = j$, ψ_{i-j} for $i > j$, and 0 for $i < j$. Note also that the asymptotic distribution of $\hat{\beta}_j$ is affected by $\rho = c/\sqrt{T}$. Then we obtain the following theorem.

THEOREM 2. *For the* LM *statistic S_{T1} in (13) associated with the model in (6), it holds that, as $T \to \infty$ under $\rho = c/\sqrt{T}$,*

$$S_{T1} \to N(c\omega, 1),$$

where $\omega = (\alpha'\alpha - \sigma^2\beta'\Gamma_p^{-1}\beta)^{1/2}$ with Γ_p given in (12).

It now follows that the limiting local power of the S_{T1}-test can be computed from

$$P(S_{T1} > x) \to P(Z > x - c\omega),$$

where $Z \sim N(0,1)$.

3.2. Case 2

Let us consider the model in (15), for which the LM statistic takes the form $S_{T2} = -\sqrt{T}r_1$ in (16). We define by $r_1(\rho)$ the first-order autocorrelation of the true innovation $\{a_t\}$ in (15). Then we have

$$\begin{aligned} r_1 &= r_1(0) = r_1(\rho) + \frac{\partial r_1(\rho)}{\partial \rho}(-\rho) + o_p\left(\frac{1}{\sqrt{T}}\right) \\ &= r_1(\rho) - \frac{c}{\sqrt{T}} + o_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned} \tag{26}$$

Thus we obtain the following theorem.

THEOREM 3. *For the* LM *statistic S_{T2} in (16) associated with the model in (15), it holds that, as $T \to \infty$ under $\rho = c/\sqrt{T}$, $S_{T2} \to N(c,1)$.*

It follows that the power of the S_{T2}-test can be computed from

$$P(S_{T2} > x) \to P(Z > x - c).$$

3.3. Case 3

Let us deal with the LM statistic S_{T3} in (22) for the model in (17). Defining by $r_i(d,\rho)$ the ith-order autocorrelation of the true innovation $\{a_t\}$ in (17), we have

$$\begin{aligned} \hat{r}_i &= r_i(\hat{d},0) = r_i(d,\rho) + \frac{\partial r_i(d,\rho)}{\partial d}(\hat{d} - d) + \frac{\partial r_i(d,\rho)}{\partial \rho}(-\rho) + o_p\left(\frac{1}{\sqrt{T}}\right) \\ &= r_i(d,\rho) - \frac{1}{i}(\hat{d} - d) + \frac{c}{\sqrt{T}}\alpha_i + o_p\left(\frac{1}{\sqrt{T}}\right), \end{aligned} \tag{27}$$

where α_i is defined in (19). We note here that the asymptotic distribution of $\sqrt{T}(\hat{d} - d)$ is affected by $\rho = c/\sqrt{T}$, like that of $\sqrt{T}(\hat{\beta} - \beta)$ in Case 1.

The following theorem can be established by using (27).

THEOREM 4. *For the* LM *statistic S_{T3} in (22) associated with the model in (17), it holds that, as $T \to \infty$ under $\rho = c/\sqrt{T}$,*

$$S_{T3} \to N(c\omega, 1),$$

where $\omega = \left(\sum_{i=1}^{\infty} \alpha_i^2 - \frac{6}{\pi^2}\left(\sum_{i=1}^{\infty} \frac{1}{i}\alpha_i\right)^2\right)^{1/2}$.

It follows that the power of the S_{T3}-test can be computed from

$$P(S_{T3} > x) \to P(Z > x - c\omega).$$

In the next section we examine the finite sample powers of the LM tests, comparing them with the theoretical results obtained in this section.

4. SOME SIMULATIONS

In this section we examine, by simulations, the finite sample properties of the present test. Our main concern here is the power performance of the test when the data generating process (DGP) is made up of the signal that follows one of Cases 1–3 discussed in Section 2 and measurement error. For this purpose we take up DGP's 1–4 subsequently. We are also interested in how the test is sensitive to misspecification of the signal that follows a different process from those considered in this paper. For this purpose we take up DGP's 5–10.

Let us first consider DGP's 1–4 given by

DGP 1. $y_t = \dfrac{\varepsilon_t}{1-\beta L} + u_t, \qquad \beta = 0.5, 0.8,$

DGP 2. $y_t = \dfrac{\varepsilon_t}{1-\beta_1 L - \beta_2 L^2} + u_t \qquad (\beta_1, \beta_2) = (0.5, -0.2), (0.8, -0.5),$

DGP 3. $y_t = \dfrac{\varepsilon_t}{1-L} + u_t,$

DGP 4. $y_t = \dfrac{\varepsilon_t}{(1-L)^d} + u_t, \qquad d = 0.1, 0.3, 0.45,$

where $\{\varepsilon_t\} \sim$ i.i.d.$N(0,\ 1)$, $\{u_t\} \sim$ i.i.d.$N(0,\ \rho)$, and these two sequences are independent of each other in all DGP's. Note that DGP's 1 and 2 correspond to Case 1 dealt with before, whereas DGP 3 corresponds to Case 2 and DGP 4 to Case 3.

Table 1 is concerned with DGP 1 and reports the powers at the nominal 5% significance level, where the sample sizes examined are $T = 100$, 200, and 500 and the results are based on 5,000 replications. It is seen that the powers do not increase as ρ gets large, although they increase with T for ρ fixed. In fact, for the DGP with $\beta = 0.8$ and $T = 500$, the power under $\rho = 1$ is 99.6%, whereas

TABLE 1. Percentage powers of the LM test for DGP 1 at the 5% level

	$\rho = 0$	0.5	1	5	10	20	50
				$\beta = 0.5$			
$T = 100$	3.8	11.3	12.1	8.9	6.3	5.3	4.5
200	4.0	18.0	20.8	12.3	9.1	6.2	4.9
500	4.6	33.5	41.6	22.0	13.3	8.5	6.1
				$\beta = 0.8$			
$T = 100$	3.6	38.7	52.2	41.9	26.0	14.8	7.1
200	4.0	65.6	82.5	66.9	43.7	23.6	10.7
500	4.8	96.2	99.6	96.3	78.7	43.9	16.7

TABLE 2. Percentage powers of the LM test for DGP 2 at the 5% level

	$\rho = 0$	0.5	1	5	10	20	50
			$\beta_1 = 0.5,\ \beta_2 = -0.2$				
$T = 100$	4.7	7.0	6.8	4.1	4.2	4.3	3.5
200	4.4	9.8	9.5	4.4	4.4	4.3	4.3
500	4.9	14.7	16.5	4.6	5.0	4.8	4.6
			$\beta_1 = 0.8,\ \beta_2 = -0.5$				
$T = 100$	5.4	25.9	34.4	3.6	4.2	4.4	3.5
200	4.7	43.4	61.0	4.6	4.5	4.2	4.3
500	5.1	77.2	92.5	5.4	4.8	4.5	4.5

it is 43.9% under $\rho = 20$. The reason may be that the signal $\{\varepsilon_t/(1 - \beta L)\}$ becomes negligible and is dominated by $\{u_t\}$ as ρ gets large, which causes unidentification of β. The power performance is worse for the signal with $\beta = 0.5$ than for $\beta = 0.8$. This is because the former is closer to white noise and is consistent with the approximation formula for the local powers derived in Section 3, which reduces, in the present case, to

$$P(S_{T1} > x) \cong P(Z > x - c\beta^2),$$

where x is the upper 5% point of $N(0, 1)$, $Z \sim N(0,1)$, and $c = \rho\sqrt{T}$. Because this approximation cannot capture the nonmonotonic nature of the actual power, it is evidently very poor as an overall approximation.

Table 2 is concerned with DGP 2, where two models of the AR(2) signal are examined. The general feature is the same as in Table 1, but the situation is worse in the present case because the test is powerful only in a very small region of ρ. The approximation to the limiting local power is given by

$$P(S_{T1} > x) \cong P(Z > x - c|\beta_2|\sqrt{\beta_2^2 + 4\beta_1^2}),$$

which implies that the test is more powerful for the signal with $(\beta_1, \beta_2) = (0.8, -0.5)$ than for $(\beta_1, \beta_2) = (0.5, -0.2)$. The simulation results are in accord with this, but the approximation turns out to be quite poor.

The situation is somewhat different in the case of the random walk signal as Table 3 reports on DGP 3, where the sample sizes $T = 100$, 200 and 300 are examined at the 5% level. The powers are increasing with ρ and also with T. In the present case the signal $\{\varepsilon_t/(1 - L)\}$ is nonstationary and is not dominated by $\{u_t\}$ as ρ gets large. The approximation to the power obtained in Section 3 is $P(S_{T2} > x) \cong P(Z > x - c)$, where $c = \sqrt{T}\rho$. For example, it is 63.9% for $\rho = 0.2$ and $T = 100$, whereas the actual power is 40.6%. When $\rho = 0.2$ and $T = 300$, the approximate power is 96.6%, whereas the actual power is 81.0%. As a whole the approximation gives upward bias.

TABLE 3. Percentage powers of the LM test for DGP 3 at the 5% level

	$\rho = 0$	0.2	0.5	1	2	10	50
$T = 100$	4.7	40.6	81.0	97.0	99.8	100.0	100.0
200	4.8	64.7	97.8	99.9	100.0	100.0	100.0
300	4.9	81.0	99.7	100.0	100.0	100.0	100.0

Table 4 reports the powers for DGP 4 with the ARFIMA$(0, d, 0)$ signal at the 5% level, where $T = 100$, 200, and 500 are examined. It is seen that the powers crucially depend on the value of d and that the test is almost useless when d is small. This is because the signal $\{\varepsilon_t/(1 - L)^d\}$ looks like white noise when d is close to 0, which is one source of unidentification of d. The nonmonotonic behavior of the power is also observed even for d large. This is another source of unidentification of d. It is really difficult to correctly identify the long memory model. The approximation to the power is given by $P(S_{T3} > x) \cong P(Z > x - c\omega)$, where $\omega = 0.019$ for $d = 0.1$, $= 0.098$ for $d = 0.3$, $= 0.176$ for $d = 0.45$, although it is very poor.

We next examine how the present test is sensitive to the misspecification of the signal process. For this purpose we first consider testing AR(1) against DGP's 5 and 6:

TABLE 4. Percentage powers of the LM test for DGP 4 at the 5% level

	$\rho = 0$	0.5	1	5	10	20	50
				$d = 0.1$			
$T = 100$	2.0	1.7	1.1	0.6	0.4	0.3	0.3
200	2.4	2.6	2.3	1.1	0.6	0.7	0.6
500	3.5	4.0	3.8	1.9	1.7	1.6	1.0
				$d = 0.3$			
$T = 100$	3.6	6.3	5.8	4.8	2.3	1.2	0.7
200	3.1	8.9	10.9	8.9	5.6	3.3	1.4
500	4.0	13.2	17.9	18.4	11.1	6.1	2.7
				$d = 0.45$			
$T = 100$	3.8	10.3	12.0	15.5	10.5	6.3	2.4
200	3.5	15.3	22.7	27.7	22.6	13.2	5.4
500	4.8	27.8	42.2	54.7	45.3	30.5	12.9

DGP 5. $y_t = \dfrac{(1-\alpha L)\varepsilon_t}{1-\beta L}, \qquad \beta = 0.6, \qquad \alpha = -0.2, -0.1, 0, 0.1, 0.2,$

DGP 6. $y_t = \dfrac{\varepsilon_t}{1-\beta L-\alpha L^2}, \qquad \beta = 0.6, \qquad \alpha = -0.2, -0.1, 0, 0.1, 0.2.$

We also consider testing the random walk against DGP's 7 and 8:

DGP 7. $y_t = \dfrac{(1-\alpha L)\varepsilon_t}{1-L}, \qquad \alpha = -0.2, -0.1, 0, 0.1, 0.2,$

DGP 8. $y_t = \dfrac{\varepsilon_t}{(1-L)(1-\alpha L)}, \qquad \alpha = -0.2, -0.1, 0, 0.1, 0.2.$

Finally we consider testing the ARFIMA$(0,d,0)$ against the DGP's 9 and 10:

DGP 9. $y_t = \dfrac{(1-\alpha L)\varepsilon_t}{(1-L)^d}, \qquad \alpha = -0.2, -0.1, 0, 0.1, 0.2, \qquad d = 0.3,$

DGP 10. $y_t = \dfrac{\varepsilon_t}{(1-L)^d(1-\alpha L)}, \qquad \alpha = -0.2, -0.1, 0, 0.1, 0.2, \qquad d = 0.3.$

Tables 5–7 report the rejection probability of the present test against DGP's 5–10, where Table 5 is concerned with DGP's 5 and 6, Table 6 with DGP's 7 and 8, and Table 7 with DGP's 9 and 10. Note that the null model for DGP's 5 and 6 is AR(1), that for DGP's 7 and 8 it is the random walk, and that for DGP's 9 and 10 it is the ARFIMA$(0,d,0)$. It is seen from these tables that the

TABLE 5. Percentage powers of the LM test for DGP's 5 and 6 at the 5% level

	$\alpha = -0.2$	-0.1	0	0.1	0.2
			DGP 5		
$T = 100$	0.1	0.8	3.8	10.4	17.8
	(27.4)	(9.3)	(4.5)	(6.4)	(11.3)
200	0.0	0.5	4.4	15.8	29.2
	(49.4)	(16.7)	(4.8)	(9.7)	(18.8)
500	0.0	0.0	4.5	29.5	59.6
	(88.1)	(32.6)	(5.0)	(19.5)	(47.2)
			DGP 6		
$T = 100$	0.0	0.3	3.8	21.7	57.0
	(51.1)	(17.1)	(4.5)	(13.2)	(43.7)
200	0.0	0.1	4.4	37.5	85.2
	(80.6)	(31.0)	(4.8)	(25.7)	(77.0)
500	0.0	0.0	4.5	69.5	99.5
	(99.6)	(62.9)	(5.0)	(57.8)	(99.2)

Note: The entries in the parentheses are the percentage powers of the two-sided test.

Table 6. Percentage powers of the LM test for DGP's 7 and 8 at the 5% level

	$\alpha = -0.2$	-0.1	0	0.1	0.2
			DGP 7		
$T = 100$	0.0	0.3	5.0	24.4	61.7
	(47.3)	(14.8)	(4.5)	(15.1)	(47.8)
200	0.0	0.1	5.0	39.8	86.5
	(78.1)	(28.1)	(4.8)	(27.5)	(77.8)
500	0.0	0.0	5.0	70.8	99.6
	(99.1)	(59.9)	(5.0)	(58.5)	(99.1)
			DGP 8		
$T = 100$	62.9	24.8	5.0	0.6	0.0
	(49.8)	(16.2)	(4.5)	(15.7)	(49.6)
200	87.0	40.3	5.0	0.1	0.0
	(79.7)	(28.5)	(4.8)	(29.1)	(80.1)
500	99.8	72.6	5.0	0.0	0.0
	(99.3)	(61.4)	(5.0)	(61.7)	(99.5)

Note: The entries in the parentheses are the percentage powers of the two-sided test.

Table 7. Percentage powers of the LM test for DGP's 9 and 10 at the 5% level

	$\alpha = -0.2$	-0.1	0	0.1	0.2
			DGP 9		
$T = 100$	0.2	0.8	3.8	9.5	19.1
	(11.5)	(3.4)	(2.5)	(5.6)	(12.1)
200	0.0	0.4	4.3	15.5	36.0
	(29.0)	(7.9)	(3.0)	(10.1)	(26.6)
500	0.0	0.1	3.9	29.3	69.4
	(72.8)	(20.9)	(3.7)	(19.9)	(59.0)
			DGP 10		
$T = 100$	21.6	9.9	3.6	0.7	0.1
	(14.2)	(6.2)	(2.6)	(3.2)	(8.6)
200	40.3	16.7	4.1	0.3	0.0
	(30.7)	(10.5)	(3.0)	(7.0)	(23.1)
500	75.3	32.4	3.9	0.1	0.0
	(66.4)	(23.1)	(3.7)	(19.5)	(63.8)

Note: The entries in the parentheses are the percentage powers of the two-sided test.

rejection of the null hypothesis of no measurement error can be caused by small misspecification of the signal process as likely as by the existence of measurement error. This fact, in turn, leads us to use the present test for testing AR(p) against general ARMA, testing ARIMA(0,1,0) against general ARIMA, and testing ARFIMA$(0,d,0)$ against general ARFIMA. From this point of view, the two-sided test that rejects the null when the statistic becomes large in absolute value may be more appropriate. The entries in the parentheses in each table are percentage powers of the two-sided test at the 5% level. It is seen that the two-sided test captures a small departure from the null that the one-sided test fails to detect. Therefore it is dangerous to ascribe the rejection of the null only to the existence of measurement error.

5. CONCLUDING REMARKS

We have suggested a unified approach to testing for the existence of measurement error in time series models. The signal processes we considered were short memory, long memory, and random walk processes, for which we suggested the LM test. It was found that the power of the test crucially depends on the signal. When the signal is stationary, it is quite difficult to detect measurement error because of the poor performance of the test. This fact is closely related to unidentification, and the test loses its power when the signal is dominated by the measurement error. In particular, it emerges that the stationary long memory model is really difficult to correctly specify. Our approximation to the power turned out to be very poor for the stationary signals because it could not capture the nonmonotonic behavior of the actual power.

We also found that the present test has nonnegligible power against the misspecification of the signal process, so that it should be used with care unless we have strong prior knowledge about the signal process.

REFERENCES

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

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

Giraitis, L. & D. Surgailis (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields* 86, 87–104.

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

Li, W.K. & A.I. McLeod (1986) Fractional time series modelling. *Biometrika* 73, 217–221.

McLeod, A.I. (1978) On the distribution of residual autocorrelations in Box-Jenkins models. *Journal of the Royal Statistical Society (B)* 40, 296–302.

Tanaka, K. (1983) The one-sided Lagrange multiplier test of the AR(p) model vs the AR(p) model with measurement error. *Journal of the Royal Statistical Society, Series B* 45, 77–80.

Tanaka, K. (1996) *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. New York: Wiley.
Tanaka, K. (1999) The nonstationary fractional unit root. *Econometric Theory* 15, 549–582.

APPENDIX

Proof of Lemma 1. We have from (7) that

$$\sigma^2[1+\rho(1+\beta_1^2+\cdots+\beta_p^2)] = \sigma_a^2(1+\delta_1^2+\cdots+\delta_p^2),$$

$$\sigma^2\rho(-\beta_i+\beta_1\beta_{i+1}+\cdots+\beta_{p-i}\beta_p) = \sigma_a^2(-\delta_i+\delta_1\delta_{i+1}+\cdots+\delta_{p-i}\delta_p)$$

for $i=1,\dots,p$. Noting that $\rho=\delta_i=0$ and $\sigma_a^2=\sigma^2$ under H_0, we take the partial derivative of

$$\delta_i = \frac{\rho\sigma^2}{\sigma_a^2}(\beta_i-\beta_1\beta_{i+1}-\cdots-\beta_{p-i}\beta_p)+\delta_1\delta_{i+1}+\cdots+\delta_{p-i}\delta_p$$

with respect to ρ and evaluate it under H_0. Then we can deduce (11). ■

Proof of Lemma 2. Let $r_i(d)$ be the ith-order autocorrelation of $(1-L)^d y_t$ and put $r(d)=(r_1(d),\dots,r_m(d))'$. Then we have

$$\begin{aligned}\hat{r} = r(\hat{d}) &= r(d)+\frac{\partial r(d)}{\partial d}(\hat{d}-d)+o_p\left(\frac{1}{\sqrt{T}}\right)\\ &= r(d)-g_m(\hat{d}-d)+o_p\left(\frac{1}{\sqrt{T}}\right),\end{aligned}$$

so that

$$\sqrt{T}\hat{r} = \sqrt{T}r(d)-g_m\sqrt{T}(\hat{d}-d)+o_p(1),$$

where $\hat{d}$ is the MLE of d that minimizes

$$L(d) = -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\{(1-L)^d y_t\}^2.$$

It is not hard to see

$$\begin{aligned}\hat{d}-d &= -\left(\frac{\partial^2 L(d)}{\partial d^2}\right)^{-1}\frac{\partial L(d)}{\partial d}+o_p\left(\frac{1}{\sqrt{T}}\right)\\ &= \frac{6}{\pi^2}\sum_{i=1}^{T-1}\frac{1}{i}r_i(d)+o_p\left(\frac{1}{\sqrt{T}}\right)\end{aligned}$$

and $\sqrt{T}(\hat{d} - d) \rightarrow N(0,\ 6/\pi^2)$. Because it is known that $\sqrt{T}r(d) \rightarrow N(0,\ I_m)$, we obtain

$$\begin{pmatrix} \sqrt{T}r(d) \\ \sqrt{T}(\hat{d} - d) \end{pmatrix} \rightarrow N\left(0, \begin{pmatrix} I_m & \dfrac{6}{\pi^2}g_m \\ \dfrac{6}{\pi^2}g_m' & \dfrac{6}{\pi^2} \end{pmatrix}\right),$$

so that

$$\sqrt{T}\hat{r} = (I_m, -g_m)\begin{pmatrix} \sqrt{T}r(d) \\ \sqrt{T}(\hat{d} - d) \end{pmatrix} + o_p(1)$$

$$\rightarrow N\left(0, I_m - \frac{6}{\pi^2}g_m g_m'\right),$$

which establishes the lemma. ■

Proof of Theorem 1. It follows from (21) that

$$\sqrt{T}\sum_{i=1}^{T-1}\alpha_i\hat{r}_i \rightarrow N\left(0, \sum_{i=1}^{\infty}\alpha_i^2 - \frac{6}{\pi^2}\left(\sum_{i=1}^{\infty}\frac{1}{i}\alpha_i\right)^2\right).$$

Then we can deduce that S_{T3} in (22) tends to $N(0,1)$ because $\hat{\alpha}_i$ converges to α_i in probability, which establishes the theorem. ■

Proof of Lemma 3. Putting $r_i = r_i(\beta, \rho)$ we first have

$$\frac{\partial r_i}{\partial \beta_j} = \frac{\partial}{\partial \beta_j}\sum_{t=i+1}^{T}a_{t-i}a_t\Big/\sum_{t=1}^{T}a_t^2 = \frac{1}{T\sigma_a^2}\sum_{t=i+1}^{T}\frac{\partial a_t}{\partial \beta_j}a_{t-i} + o_p(1),$$

where

$$\frac{\partial a_t}{\partial \beta_j} = \frac{\partial}{\partial \beta_j}\frac{\beta(L)y_t}{\delta(L)} = -\frac{a_{t-j}}{\beta(L)}.$$

This gives the expression for $\partial r_i/\partial \beta_j$ in the lemma. Similarly we have

$$\frac{\partial r_i}{\partial \rho} = \frac{1}{T\sigma_a^2}\sum_{t=i+1}^{T}\frac{\partial a_t}{\partial \rho}a_{t-i} + o_p(1),$$

where

$$\frac{\partial a_t}{\partial \rho} = \sum_{j=1}^{p}\frac{\partial a_t}{\partial \delta_j}\frac{\partial \delta_j}{\partial \rho} = \sum_{j=1}^{p}\frac{a_{t-j}}{\delta(L)}\left(-\alpha_j + O\left(\frac{1}{\sqrt{T}}\right)\right).$$

Then we have the expression for $\partial r_i/\partial \rho$ given in the lemma. ■

Proof of Theorem 2. Putting $\underset{\sim}{y}_{t-1} = (y_{t-1}, \dots, y_{t-p})'$ we have

$$\hat{\beta} = \left(\sum \underset{\sim}{y}_{t-1} \underset{\sim}{y}'_{t-1}\right)^{-1} \sum \underset{\sim}{y}_{t-1} y_t = \beta + \left(\sum \underset{\sim}{y}_{t-1} \underset{\sim}{y}'_{t-1}\right)^{-1} \sum \underset{\sim}{y}_{t-1}(\varepsilon_t + \beta(L)u_t),$$

which yields, under $\rho = c/\sqrt{T}$,

$$\begin{aligned}\sqrt{T}(\hat{\beta} - \beta) &= \Gamma_p^{-1}\left(\frac{1}{\sqrt{T}} \sum \underset{\sim}{x}_{t-1} \varepsilon_t - c\sigma^2 \beta\right) + o_p(1) \\ &\to N(-c\sigma^2 \Gamma_p^{-1} \beta, \sigma^2 \Gamma_p^{-1}),\end{aligned}$$

where $\underset{\sim}{x}_{t-1} = (x_{t-1}, \dots, x_{t-p})'$. Substituting this into (25), we obtain in matrix notation

$$\sqrt{T}\hat{r} = \sqrt{T}r + J^{-1}(\beta)\sqrt{T}(\hat{\beta}(0) - \beta) + c(\alpha - \sigma^2 J^{-1}(\beta)\Gamma_p^{-1}\beta) + o_p(1),$$

where $\hat{\beta}(0)$ is the LSE of β under H_0. Then, noting that $\alpha' J^{-1}(\beta) = \beta'$ and using the arguments in McLeod (1978), we can establish the theorem. ■

Proof of Theorem 3. It follows from (26) that

$$\sqrt{T}r_1 = \sqrt{T}r_1(\rho) - c + o_p(1),$$

which evidently yields the theorem because $S_{T2} = -\sqrt{T}r_1$. ■

Proof of Theorem 4. Putting

$$L(d, \rho) = -\frac{T}{2} \log \sigma_a^2 - \frac{1}{2\sigma_a^2} \sum_{t=1}^{T} \left\{ \frac{(1-L)^d}{\delta(L)} y_t \right\}^2,$$

we have

$$0 = \frac{\partial L(\hat{d}, 0)}{\partial d} = \frac{\partial L(d, \rho)}{\partial d} + \frac{\partial^2 L(d, \rho)}{\partial d^2}(\hat{d} - d) - \rho \frac{\partial^2 L(d, \rho)}{\partial d \partial \rho} + O_p(1),$$

which yields

$$\begin{aligned}\sqrt{T}(\hat{d} - d) &\cong \left(-\frac{1}{T} \frac{\partial^2 L(d, \rho)}{\partial d^2}\right)^{-1} \frac{1}{\sqrt{T}} \left(\frac{\partial L(d, \rho)}{\partial d} - \frac{c}{\sqrt{T}} \frac{\partial^2 L(d, \rho)}{\partial d \partial \rho}\right) \\ &\cong \frac{6}{\pi^2}\left(\sqrt{T} \sum_{i=1}^{T-1} \frac{1}{i} r_i(d, \rho) + c \sum_{i=1}^{T} \frac{1}{i} \alpha_i\right).\end{aligned}$$

Substituting this into (27), we obtain

$$\sqrt{T} \sum_{i=1}^{T-1} \alpha_i \hat{r}_i \cong \sqrt{T} \sum_{i=1}^{T-1} \left(\alpha_i - \frac{6}{i\pi^2} \sum_{j=1}^{T} \frac{1}{j} \alpha_j\right) r_i(d, \rho) + c\left(\sum_{i=1}^{T} \alpha_i^2 - \frac{6}{\pi^2}\left(\sum_{i=1}^{T} \frac{1}{i} \alpha_i\right)^2\right),$$

which establishes the theorem. ■