Estimating the differencing parameter via the partial autocorrelation function

Terence Tai-Leung Chong*

Department of Economics, The Chinese University of Hong Kong, Shatin, Hong Kong

Abstract

This paper provides an explanation for the puzzling phenomenon in Tieslau et al. (1996, Journal of Econometrics 71, 249–264) that a substantial efficiency loss occurs if low-order autocorrelations are omitted when estimating the differencing parameter, d. This is because for all n strictly bigger than 1, the nth-order autocorrelation function does not depend uniquely on the differencing parameter. We construct a new estimator for the differencing parameter based on the partial autocorrelation function. Comparisons of the asymptotic and finite-sample variance of our estimator and those of TSB are made. A substantial efficiency gain is achieved by our estimator as compared to TSB's. © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: C22

Keywords: Differencing parameter; Fractionally integrated model

1. Introduction

The long-memory processes have found wide applications in various academic disciplines since 1950s, in fields as diverse as hydrology and climatology. The most familiar example has been in hydrology, originally documented by

---

* Tel.: + 852-26097045; fax: + 852-26035805.

E-mail address: b792703@mailserv.cuhk.edu.hk (T.T.-L. Chong).

Hurst (1951) who used 900 geophysical time series to study the persistence of stream-flow data and the design of reservoirs.

The theory of fractionally integrated models, stemming from the work of Granger and Joyeux (1980), was designed to fit the behavior of long-memory processes. The main economic application of such a model has been to study whether real GNP is difference stationary or trend stationary. Using quarterly post-World-War II US real GNP data in first differences, Diebold and Rudebusch (1989) found an estimated value of d, the order of integration, equal to -0.5. Sowell (1992b) found that first differenced US real GNP quarterly data for 1947–1989 can be characterized by an $ARFIMA$ (3, -0.59, 2) model. Other researchers also used fractionally integrated models in their studies of asset pricing models (Ding et al., 1993), stock returns (Lo, 1991), exchange rates (Diebold et al., 1991; Cheung, 1993; Baillie and Bollerslev, 1994), interest rates (Shea, 1991; Backus and Zin, 1993; Crato and Rothman, 1994) and inflation rates (Hassler and Wolters, 1995; Baillie et al., 1996). Baillie (1996) provides an excellent survey of the literature in this area.

When the development of ARFIMA models was still in its infancy, researchers usually assumed that the fractional differencing parameter was known in advance. Recently, there has been a tremendous amount of promising research on the estimation of d via various methods. Estimator after estimator has been proposed. Geweke and Porter-Hudak (1983) developed the well-known GPH estimator for estimating d of a stationary and invertible fractionally integrated process at low frequencies. Hurvich and Ray (1995) have extended the GPH estimator to the case of a non-stationary, non-invertible process. Maximum likelihood estimation has been proposed by Li and McLeod (1986) and Sowell (1992a). More recently, Tieslau et al. (1996) (hereafter referred to as TSB) have suggested estimating d by minimizing the difference between the sample and the population autocorrelations. A shortcoming of the TSB estimator is that a substantial efficiency loss occurs if low-order autocorrelations are omitted.

The purpose of this paper is to provide a new estimator for the differencing parameter. This estimator is constructed by equalizing the sample and population partial autocorrelations of the underlying process. The advantage of using the partial autocorrelation function is that it has a unique mapping with the differencing parameter. Thus, the problem of multiple solutions is avoided and efficiency of the estimator is increased. A difficulty encountered in using sample partial autocorrelation functions is that its variance–covariance structure is not available to us. This paper adds to the literature by deriving the theoretical variance–covariance structure of the sample partial autocorrelation function. Such a finding enables us to construct an estimator analogous to that of Tieslau et al. (1996).

We also endeavor to explain the puzzling phenomenon arising from TSB's simulation that a substantial efficiency loss occurs if low-order autocorrelations

are omitted. We argue that their results are due to the fact that for all n strictly bigger than 1, the nth-order autocorrelation function does not depend uniquely on the differencing parameter d. Thus, the problem of multiple solutions arises. This will increase the variation of the estimate of d.

The notations and structure of this paper closely follow those of TSB's paper. Section 2 constructs the model. Section 3 derives our estimator and its asymptotic distribution. Sections 4–6 compare the asymptotic and finite-sample variance of our estimator with TSB's. The conclusions are drawn in Section 7.

2. The model

A time series process $\{y_t\}$ is said to be integrated of order d if $(1 - L)^d y_t$ is stationary, where L is a lag operator such that $Ly_t = y_{t-1}$. If d is not an integer, then the process is said to be fractionally integrated. In this paper, we shall assume for ease of exposition that the process under discussion is a fractionally integrated white noise process. Consider the following model:

$$(1 - L)^d y_t = u_t \quad (t = 1, 2, \ldots, T), \tag{1}$$

where L is the lag operator and u_t is the white noise.

The fractional difference operator $(1 - L)^d$ is defined by its Maclaurin series

$$(1 - L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j - d)}{\Gamma(-d)\Gamma(j + 1)} L^j, \tag{2}$$

where $\Gamma(x)$ is the Euler gamma function defined as

$$\Gamma(x) = \int_0^{\infty} z^{x-1} \exp(-z)\, dz \quad \text{for } x > 0,$$

$$\Gamma(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(x + k)k!} + \int_1^{\infty} z^{x-1} \exp(-z)\, dz$$

$$\text{for } x < 0, \, x \neq -1, -2, -3, \ldots .$$

The jth autocorrelation of this $ARFIMA(0, d, 0)$ process is given by

$$\rho_j(d) = \prod_{i=1}^{j} \frac{d + i - 1}{i - d} \quad (j = 1, 2, \ldots, n). \tag{3}$$

The sample autocorrelations are defined as

$$\hat{\rho}_j = \frac{\sum_{t=1}^{T-j}(y_t - \bar{y})(y_{t+j} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}. \tag{4}$$

In a recent study, Tieslau et al. (1996) proposed a minimum distance estimator of d defined as

$$\hat{d} = \underset{d \in (-0.5, \ 0.25)}{\text{Argmin}} \ S(d), \tag{5}$$

where $S(d) = [\hat{\rho} - \rho(d)]' C^{-1} [\hat{\rho} - \rho(d)]$, $\rho(d)$ is a vector of dimension $(n \times 1)$ with the jth element $\rho_j(d)$, and $\hat{\rho}$ is a vector of dimension $(n \times 1)$ with the jth element $\hat{\rho}_j$. C is the asymptotic variance–covariance matrix of $\hat{\rho}$ with the (i, j)th element given by

$$c_{i,j} = \sum_{s=1}^{\infty} (\rho_{s+i} + \rho_{s-i} - 2\rho_s\rho_i)(\rho_{s+j} + \rho_{s-j} - 2\rho_s\rho_j). \tag{6}$$

In Tables 2–4 of TSB's paper, a substantial efficiency loss occurs when the first-order autocorrelation is not used in the estimation of d. This implies that the first-order autocorrelation carries most of the information needed for the estimation of d. We argue that their findings are due to the fact that the mapping between $\rho_n(d)$ and d is not one to one for all $n \geqslant 2$.

Note that when $n = 1$,

$$\rho_1(d) = \frac{d}{1-d}. \tag{7}$$

In this instance, the mapping between d and $\rho_1(d)$ is unique. However, for $n \geqslant 2$, different values of d may generate the same $\rho_n(d)$. Consider the values of d used in Table 2 of TSB's paper. Table 1 below shows all other values of d which share the same nth-order autocorrelation for $n = 2, 3$.

Table 1
Values of d which share the same nth-order autocorrelation for $n = 2, 3$

d	$n = 2$	$n = 3$
-0.49	-0.2576	$-0.1392, -2.6410$
-0.45	-0.2895	$-0.1608, -2.6801$
-0.4	-0.3333	$-0.1917, -2.7213$
-0.3	-0.4375	$-0.2714, -2.7665$
-0.2	-0.5714	$-0.3879, -2.7298$
-0.1	-0.75	$-0.5772, -2.5430$
0	-1	$-1.0000, -2.0000$
0.1	-1.375	$-1.3497 \pm 0.9213i$
0.2	-2	$-1.0789 \pm 1.5197i$
0.24	-2.3846	$-0.9317 \pm 1.7192i$

For example, considering the case where $n = 2$, we have

$$\rho_2(d) = \prod_{i=1}^{2} \frac{d + i - 1}{i - d}.$$

In this case, $\rho_2(-0.4) = \rho_2(-0.3333) = -0.0714$.

Thus if the true d is $d_0 = -0.4$, and if we estimate d_0 by using the second-order autocorrelation only, the estimator converges to the set $\{-0.4, -0.3333\}$.

Obviously, for $n = 1$, the criterion function is U-shaped and thus it has a unique minimum. However, for $n \geqslant 2$, the shape changes with the true value of d. Figs. 1 and 2 show the shape of the criterion function against the differencing parameter d, for $n = 2, 3$, when the true d is $d_0 = -0.4$.

It should be noted that the minimum of the function is not well located. In Fig. 1, $S(d)$ has two minima, one at $d = -0.4$ and the other at $d = -0.33$. The function is relatively insensitive to d for $d < 0$. Similarly, in Fig. 2, the minima of $S(d)$ occur at $d = -0.4$ and -0.1917.

The existence of multiple solutions widens the variation of $\hat{d}$. This will make the variance of $\sqrt{T}(\hat{d} - d)$ diverge to infinity as the sample size increases.
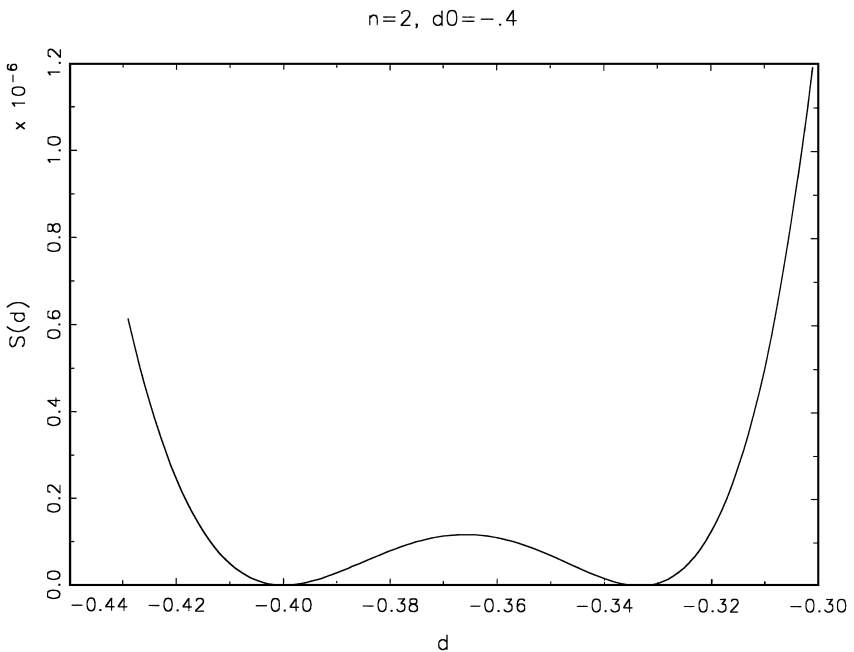

n=2, d0=−.4

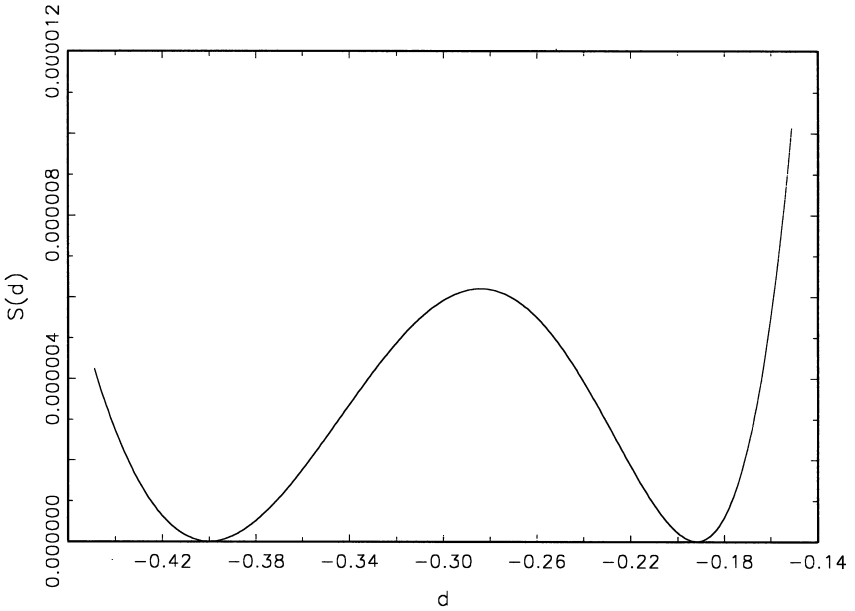Fig. 1.

$$n=3, \quad d0=-.4$$

Fig. 2.

3. The new estimator

Our estimator differs from the TSB estimator in that we use the sample partial correlation function to form the moment conditions. The nth-order partial autocorrelation function of a fractional white noise process is

$$\alpha_n(d) = \frac{d}{n-d}. \tag{8}$$

This expression originally was given by Hosking (1981). For $n = 1$, the partial correlation function and the autocorrelation are identical. A salient feature of $\alpha_n(d)$ is that its relationship with d is unique for all n. Hence, we can either use a single $\alpha_n(d)$ or a combination of them to form an estimator of d.

Next, we discuss how to obtain an estimate of $\alpha_n(d)$. Let

$$Y = (y_1, \, y_2, \dots, y_T)',$$

$$\boldsymbol{\rho}(n) = (\rho_1, \, \rho_2, \dots, \rho_{n-1}, \, \rho_n)',$$

$$X_n = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ y_1 & 0 & \cdots & 0 \\ y_2 & y_1 & \cdots & 0 \\ \vdots & y_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & y_1 \\ \vdots & \vdots & \cdots & \vdots \\ y_{T-1} & y_{T-2} & \cdots & y_{T-n} \end{pmatrix},$$

$$\hat{\boldsymbol{\beta}}(n) = (\hat{\beta}_{n,1}, \hat{\beta}_{n,2}, \ldots, \hat{\beta}_{n,n-1}, \hat{\beta}_{n,n})' = (X_n'X_n)^{-1}X_n'Y$$

$$= \begin{pmatrix} \sum_{t=2}^{T} y_{t-1}^2 & \sum_{t=3}^{T} y_{t-1}y_{t-2} & \cdots & \sum_{t=n+1}^{T} y_{t-1}y_{t-n} \\ \sum_{t=3}^{T} y_{t-1}y_{t-2} & \sum_{t=3}^{T} y_{t-2}^2 & \cdots & \sum_{t=n+1}^{T} y_{t-2}y_{t-n} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=n+1}^{T} y_{t-1}y_{t-n} & \sum_{t=n+1}^{T} y_{t-2}y_{t-n} & \cdots & \sum_{t=n+1}^{T} y_{t-n}^2 \end{pmatrix}^{-1}$$

$$\begin{pmatrix} \sum_{t=2}^{T} y_t y_{t-1} \\ \sum_{t=3}^{T} y_t y_{t-2} \\ \vdots \\ \sum_{t=n+1}^{T} y_t y_{t-n} \end{pmatrix}.$$

Since $E(y_t)$ is assumed to be 0 for all t, if we divide each element in the above matrix by $\sum_{t=2}^{T} y_{t-1}^2$ and take probability limit, we have

$$\hat{\boldsymbol{\beta}}(n) \overset{p}{\to} \Phi(n-1)^{-1}\rho(n) \overset{\text{def}}{=} \boldsymbol{\beta}(n),$$

where $\Phi(n-1)$ is a Toeplitz matrix of dimension $(n \times n)$ defined as

$$\Phi(n-1) = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{pmatrix}. \tag{9}$$

The elements of $\hat{\boldsymbol{\beta}}(n)$ will converge in probability to a function of d, in particular,

$$\hat{\boldsymbol{\beta}}(2) \overset{\text{p}}{\to} \Phi(1)^{-1}\boldsymbol{\rho}(2) = \left(\rho_1 \frac{1-\rho_2}{1-\rho_1^2} \quad \frac{\rho_1^2 - \rho_2}{\rho_1^2 - 1}\right)' = \left(\frac{2d}{2-d} \quad \frac{d}{2-d}\right)',$$

$$\hat{\boldsymbol{\beta}}(3) \overset{\text{p}}{\to} \Phi(2)^{-1}\boldsymbol{\rho}(3) = \left(\frac{3d}{3-d} \quad \frac{3d(1-d)}{(3-d)(2-d)} \quad \frac{d}{3-d}\right)',$$

$$\hat{\boldsymbol{\beta}}(4) \overset{\text{p}}{\to} \Phi(3)^{-1}\boldsymbol{\rho}(4) = \left(\frac{4d}{4-d} \quad \frac{6d(1-d)}{(4-d)(3-d)} \quad \frac{4d(1-d)}{(4-d)(3-d)} \quad \frac{d}{4-d}\right)'.$$

One can easily generalize that the probability limit of the nth element of $\hat{\boldsymbol{\beta}}(n)$ is given by

$$\hat{\beta}_{n,n} \overset{\text{p}}{\to} \frac{d}{n-d} = \alpha_n(d).$$

Hence, the nth-order sample partial autocorrelation can be obtained from the estimated coefficient of y_{t-n} in the regression of y_t on $y_{t-1}, y_{t-2}, \ldots, y_{t-n}$.

Our estimator of d is defined as

$$\hat{d} = \underset{d \in (-0.5, 0.25)}{\text{Argmin}} \ S(d), \tag{10}$$

where

$$S(d) = [\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}(d)]' W [\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}(d)], \tag{11}$$

$\boldsymbol{\alpha}(d)$ is a vector of dimension $(n \times 1)$ with the jth element $d/(j-d)$, $\hat{\boldsymbol{\alpha}}$ is a vector of dimension $(n \times 1)$ with the jth element $\hat{\beta}_{j,j}$, and W is a symmetric, positive-definite weighting matrix.

Let

$$D = \frac{\partial \boldsymbol{\alpha}(d)}{\partial d} = \left(\frac{1}{(1-d)^2}, \frac{2}{(2-d)^2}, \ldots, \frac{n}{(n-d)^2}\right)'. \tag{12}$$

Note that

$$\frac{\partial S(d)}{\partial d} = -2D' W [\hat{\alpha} - \alpha(d)], \tag{13}$$

$$\frac{\partial^2 S(d)}{\partial d^2} = 2D' W^{-1} D + \text{o}_{\text{p}}(1). \tag{14}$$

Since we use $\hat{\beta}_{n,n}$ as the estimator for $\alpha_n(d)$, the vector $\hat{\boldsymbol{\alpha}}$ is

$$\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1(d), \ \hat{\alpha}_2(d), \ldots, \hat{\alpha}_n(d))' = (\hat{\beta}_{1,1}, \ \hat{\beta}_{2,2}, \ldots, \hat{\beta}_{n,n})'.$$

For stationary process y_t, the estimator $\hat{\beta}_{j,j}$ $(j = 1, 2, \ldots, n)$ is asymptotically normal. Thus we have

$$\sqrt{T}[\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}(d)] \overset{\text{d}}{\to} \text{N}(0, \Omega), \tag{15}$$

where

$$\Omega = \lim_{T \to \infty} T \begin{pmatrix} \mathrm{Var}(\hat{\beta}_{1,1}) & \mathrm{Cov}(\hat{\beta}_{1,1}, \hat{\beta}_{2,2}) & \cdots & \mathrm{Cov}(\hat{\beta}_{1,1}, \hat{\beta}_{n,n}) \\ \mathrm{Cov}(\hat{\beta}_{2,2}, \hat{\beta}_{1,1}) & \mathrm{Var}(\hat{\beta}_{2,2}) & \cdots & \mathrm{Cov}(\hat{\beta}_{2,2}, \hat{\beta}_{n,n}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\hat{\beta}_{n,n}, \hat{\beta}_{1,1}) & \mathrm{Cov}(\hat{\beta}_{n,n}, \hat{\beta}_{2,2}) & \cdots & \mathrm{Var}(\hat{\beta}_{n,n}) \end{pmatrix}. \quad (16)$$

Note also that

$$\frac{\partial S(d)}{\partial \hat{d}} = \frac{\partial S(d)}{\partial d} + \frac{\partial^2 S(d)}{\partial d_*^2}(\hat{d} - d) = 0, \quad (17)$$

where d_* lies between d and $\hat{d}$. Solving for $(\hat{d} - d)$ gives

$$\hat{d} - d = -\left[\frac{\partial^2 S(d)}{\partial d_*^2}\right]^{-1} \frac{\partial S(d)}{\partial d}, \quad (18)$$

$$\sqrt{T}\, \frac{\partial S(d)}{\partial d} \xrightarrow{\mathrm{d}} \mathrm{N}(0,\, 4D'W\Omega W D), \quad (19)$$

$$\sqrt{T}(\hat{d} - d) \xrightarrow{\mathrm{d}} \mathrm{N}(0,\, [D'WD]^{-1}D'W\Omega W D[D'WD]^{-1}). \quad (20)$$

Thus, the optimal weighting matrix is

$$W = \Omega^{-1} \quad (21)$$

and therefore, under the optimal case, we have

$$\sqrt{T}(\hat{d} - d) \xrightarrow{\mathrm{d}} \mathrm{N}(0,\, [D'\Omega^{-1}D]^{-1}). \quad (22)$$

Note that the (l, m)th element of the variance–covariance matrix Ω is given by

$$\Omega_{l,m} = \lim_{T \to \infty} T[L(l)\mathrm{E}(\hat{\boldsymbol{\beta}}(l) - \boldsymbol{\beta}(l))(\hat{\boldsymbol{\beta}}(m) - \boldsymbol{\beta}(m))'L(m)'],$$

where

$$L(i) = \underbrace{(0\ 0 \ldots 0\ 1)}_{i\ \text{terms}}, \quad (23)$$

The remaining conundrum is to find $\mathrm{E}(\hat{\boldsymbol{\beta}}(l) - \boldsymbol{\beta}(l))(\hat{\boldsymbol{\beta}}(m) - \boldsymbol{\beta}(m))'$. Since

$$\begin{aligned} \hat{\rho}(n) - \rho(n) &= \hat{\Phi}(n-1)\hat{\boldsymbol{\beta}}(n) - \Phi(n-1)\boldsymbol{\beta}(n) \\ &= \Phi(n-1)(\hat{\boldsymbol{\beta}}(n) - \boldsymbol{\beta}(n)) + (\hat{\Phi}(n-1) - \Phi(n-1))\boldsymbol{\beta}(n) \\ &\quad + (\hat{\Phi}(n-1) - \Phi(n-1))(\hat{\boldsymbol{\beta}}(n) - \boldsymbol{\beta}(n)) \end{aligned}$$

$$= \Phi(n-1)(\hat{\boldsymbol{\beta}}(n) - \boldsymbol{\beta}(n)) + (\hat{\Phi}(n-1) - \Phi(n-1))\boldsymbol{\beta}(n)$$
$$+ \; O_p(T^{-1}),$$

we have

$$\hat{\boldsymbol{\beta}}(n) - \boldsymbol{\beta}(n) = \Phi(n-1)^{-1}\Delta(n), \tag{24}$$

where

$$\Delta(n) = (\hat{\boldsymbol{\rho}}(n) - \boldsymbol{\rho}(n)) - (\hat{\Phi}(n-1) - \Phi(n-1))\boldsymbol{\beta}(n) + O_p(T^{-1}). \tag{25}$$

Hence, $\Omega_{l,m}$ is reduced to

$$\lim_{T\to\infty} T[L(l)\Phi(l-1)^{-1} E(\Delta(l)\Delta(m)')\Phi(m-1)^{-1}L(m)'].$$

Let

$$\Psi(l,m) = \lim_{T\to\infty} T E(\Delta(l)\Delta(m)').$$

Note that

$$\Psi(l,m) = C(l,m) - \lim_{T\to\infty} E(\hat{\Phi}(l-1) - \Phi(l-1))\boldsymbol{\beta}(l)(\hat{\boldsymbol{\rho}}(m) - \boldsymbol{\rho}(m))'$$

$$- \lim_{T\to\infty} E(\hat{\boldsymbol{\rho}}(l) - \boldsymbol{\rho}(l))\boldsymbol{\beta}(m)'(\hat{\Phi}(m-1) - \Phi(m-1))'$$

$$+ \lim_{T\to\infty} E(\hat{\Phi}(l-1) - \Phi(l-1))\boldsymbol{\beta}(l)\boldsymbol{\beta}(m)'(\hat{\Phi}(m-1) - \Phi(m-1)),$$

where $C(l,m)$ is an $l \times m$ matrix with the (i,j)th element $c_{i,j}$ defined in (6).

Lastly, the (i,j)th element of the $\Psi(l,m)$ is given by

$$\Psi(l,m)_{i,j} = c_{i,j} - \sum_{h=1,h\neq i}^{l} \beta_{l,h}c_{|i-h|,j} - \sum_{k=1,k\neq j}^{m} \beta_{m,k}c_{|j-k|,i}$$

$$+ \sum_{h=1,h\neq i}^{l} \sum_{k=1,k\neq j}^{m} E(\hat{\Phi}(l-1) - \Phi(l-1))_{i,h}$$

$$\times \quad (\hat{\Phi}(m-1) - \Phi(m-1))_{k,j}\beta_{l,h}\beta_{m,k}$$

$$= c_{i,j} - \sum_{h=1,h\neq i}^{l} \beta_{l,h}c_{|i-h|,j} - \sum_{k=1,k\neq j}^{m} \beta_{m,k}c_{|j-k|,i}$$

$$+ \sum_{h=1,h\neq i}^{l} \sum_{k=1,k\neq j}^{m} c_{|i-h|,|k-j|}\beta_{l,h}\beta_{m,k}. \tag{26}$$

Thus, an equivalent of Bartlett's formula for the partial autocorrelation function is given by

$$\text{cov}(\hat{\alpha}_l(d),\ \hat{\alpha}_m(d)) = \text{cov}(\hat{\beta}_{l,l},\ \hat{\beta}_{m,m})$$
$$= L(l)\Phi(l-1)^{-1}\Psi(l,\ m)\Phi(l-1)^{-1}L(m)', \quad (27)$$

where L, Φ and Ψ are defined in (23) (9) and (26), respectively.

Since all the elements of the variance–covariance matrix Ω are uncovered, the matrix $W(=\Omega^{-1})$ can be constructed and we can now evaluate $S(d)$ at various values of d.

4. Comparison of asymptotic variance of $\sqrt{T}(\hat{d}-d)$

In this section, we compare the efficiency of our estimator to that of TSB's. In all the tables below, $c_{i,j}$ is approximated by

$$c_{i,j} = \sum_{s=1}^{x} (\rho_{s+i} + \rho_{s-i} - 2\rho_s\rho_i)(\rho_{s+j} + \rho_{s-j} - 2\rho_s\rho_j),$$

where the number of items in the summation is $x = 5 \times 10^6$.

In TSB's paper, they did not report the value of x used to calculate $c_{i,j}$. Therefore, their results may be slightly different from ours, especially for $d > 0$. In addition, there are some slight misprints in TSB's tables. In their Table 1, for $d = 0$ and $n = 3$, the value should be 0.7347 instead of 0.7437. In the same table, for $d = -0.3$ and $n = 20$, the correct value should be 0.7426 instead of 0.8625. In Table 3, for $d = 0$ and $n = 2$–6, $n = 3$–7, it is fairly clear that the true values should be 2.035 and 3.8198 respectively instead of 1.5866 and 2.4957. This is because all the values reported in their Table 3 should be bigger than those reported in Table 4.

Table 2
Asymptotic variance of $\sqrt{T}(\hat{d}-d)$, using 5 partial autocorrelations

d	1–5	2–6	3–7	5–9	10–14
−0.49	1.2305	2.6125	4.4028	9.2274	28.378
−0.45	1.2063	2.5530	4.3376	9.1503	28.270
−0.4	1.1399	2.4806	4.2583	9.0569	28.139
−0.3	1.0122	2.3425	4.1088	8.8829	27.897
−0.2	0.8917	2.2159	3.9756	8.7328	27.694
−0.1	0.7803	2.1075	3.8693	8.6247	27.564
0	0.6832	2.0350	3.8198	8.6091	27.603
0.1	0.6187	2.0669	3.9412	8.8813	28.192
0.2	0.7060	2.7407	5.1597	11.044	32.426
0.24	0.9373	4.2363	7.8733	15.785	41.386

Table 3
Asymptotic variance of $\sqrt{T}(\hat{d} - d)$, using 10 partial autocorrelations

d	1–10	2–11	3–12	5–14	10–19
− 0.49	0.9776	1.9036	3.0681	6.0495	17.134
− 0.45	0.9481	1.8827	3.0563	6.0566	17.190
− 0.4	0.9115	1.8581	3.0440	6.0696	17.266
− 0.3	0.8395	1.8146	3.0294	6.1128	17.448
− 0.2	0.7697	1.7818	3.0337	6.1880	17.685
− 0.1	0.7034	1.7674	3.0704	6.3177	18.016
0	0.6453	1.7920	3.1748	6.5623	18.554
0.1	0.6108	1.9309	3.1436	6.3321	19.733
0.2	0.6908	2.7321	5.0047	9.9176	25.045
0.24	0.8372	4.1292	7.8532	15.264	35.234

Tables 2 and 3 below are counterparts of Tables 3 and 4 in TSB's paper, respectively. We compare the performance of the asymptotic variance of $\sqrt{T}(\hat{d} - d)$ of our estimator, which uses the partial autocorrelation, with that of TSB, which uses the pure autocorrelation.

Note from Table 2 that for $n \geqslant 2$, an increase in d leads to a decrease in the asymptotic variance, up to the point where $d = 0$. For $d < 0$, the use of the partial autocorrelation function yields a smaller asymptotic variance than the use of the autocorrelation function. For example, for $d = -0.2$ and $n = 5$–9, the asymptotic variance of the TSB estimator is 7417.6, while ours is only 8.7328.

Table 3 also shows that, for $d < 0$, there is a sharp reduction in the asymptotic variance of our estimator as compared to the TSB estimator. For example, for $d = -0.2$ and $n = 5$–14, the asymptotic variance of the TSB estimator is 1398, while ours is only 6.188.

A general remark from Tables 2 and 3 is that for $d < 0$, our estimator has an asymptotic variance much smaller than that of TSB's. For $d > 0$, our estimator has a moderately larger asymptotic variance. Thus if the sign of d is known a priori, our results, together with TSB's, provide a useful guiding principle for the selection of estimation method of d.

5. Short-run dynamics

In the previous sections, we only consider a fractional white noise process. We would like to examine how robust our estimator is to serial correlation. We will discuss the asymptotic bias in our estimator of d obtained from the $ARFIMA(0, d, 0)$ model caused by ignoring short-run dynamics. For comparison, we will consider estimators of d based on a single partial autocorrelation α_k.

d^* is the value of d that generates the same value of α_k for $ARFIMA(0, d, 0)$ model.

Suppose the true model is an $ARFIMA(1, d, 0)$ process of the form

$$(1 - L)^d(1 - \phi L)y_t = \varepsilon_t,$$

then we have

$$y_t = (1 - \phi L)^{-1}(1 - L)^{-d}\varepsilon_t = \sum_{j=0}^{\infty} \mu(j)L^j\varepsilon_t,$$

where

$$\mu(j) = \sum_{i=0}^{j} \left(\frac{\Gamma(i + d)}{\Gamma(d)\Gamma(i + 1)}\phi^{j-i} \right).$$

Hence

$$\hat{\rho}_n \xrightarrow{\text{p}} \frac{\sum_{j=n}^{\infty}(\mu(j)\mu(j - n))}{\sum_{j=0}^{\infty}(\mu(j))^2}.$$

If the true model is an $ARFIMA(0, d, 1)$ process of the form

$$(1 - L)^d y_t = (1 + \theta L)\varepsilon_t,$$

then we have

$$y_t = \sum_{j=0}^{\infty} \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)}L^j(1 + \theta L)\varepsilon_t = \sum_{j=0}^{\infty} \lambda(j)L^j\varepsilon_t,$$

where

$$\lambda(0) = 1,$$

$$\lambda(j) = \frac{\Gamma(j - 1 + d)}{\Gamma(d)\Gamma(j)}\theta + \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)} \quad \text{for } j \geqslant 1.$$

$$
\begin{aligned}
\hat{\rho}_n \xrightarrow{\text{p}} & \frac{\sum_{j=n}^{\infty}(\lambda(j)\lambda(j - n))}{\sum_{j=0}^{\infty}(\lambda(j))^2} \\
= & \frac{\dfrac{(1 + \theta^2)\Gamma(n + d)}{\Gamma(n + 1 - d)} + \dfrac{\theta\Gamma(n + 1 + d)}{\Gamma(n + 2 - d)} + \dfrac{\theta\Gamma(n - 1 + d)}{\Gamma(n - d)}}{\dfrac{(1 + \theta^2)\Gamma(d)}{\Gamma(1 - d)} + \dfrac{2\theta\Gamma(1 + d)}{\Gamma(2 - d)}}.
\end{aligned}
$$

For the partial autocorrelations in the case of the $ARFIMA(0, d, 1)$ and the $ARFIMA(1, d, 0)$ processes, one is referred to Hosking (1981, Lemmas 1 and 2).

Table 4
Asymptotic bias $|d^* - d|$ for the MDE from the $(0, d, 0)$ model

d	ϕ	Lag 1	Lag 2	Lag 5	Lag 10	Lag 20	Lag 50
True model is ARFIMA (1, *d*, 0) *model*							
0.1	0.4	0.238	0.068	0.024	0.013	0.009	0.008
0.2	0.4	0.185	0.145	0.052	0.033	0.027	0.034
0.24	0.4	0.162	0.180	0.065	0.044	0.039	0.052
0.1	0.8	0.364	0.186	0.082	0.052	0.032	0.019
0.2	0.8	0.278	0.412	0.168	0.108	0.073	0.056
0.24	0.8	0.242	0.517	0.203	0.133	0.094	0.078
True model is ARFIMA (0, *d*, 1) *model*							
0.1	0.4	0.208	0.343	0.035	0.006	0.003	0.001
0.2	0.4	0.158	0.374	0.027	0.012	0.006	0.002
0.24	0.4	0.137	0.387	0.024	0.014	0.007	0.003
0.1	0.8	0.263	0.915	0.541	0.415	0.089	0.002
0.2	0.8	0.197	0.922	0.521	0.426	0.094	0.004
0.24	0.8	0.171	0.925	0.512	0.430	0.096	0.004

Table 4 gives values of d^* for $d = 0.1$, 0.2, and 0.24, for the $(1, d, 0)$ and $(0, d, 1)$ models. For the $(1, d, 0)$ model we consider $\phi = 0.4$ and 0.8, whereas for the $(0, d, 1)$ model we consider $\theta = 0.4$ and 0.8.

Comparing our Table 4 to Table 5 of TSB's paper, if the true model is an $ARFIMA(1, d, 0)$ model, our estimator yields a smaller bias in most cases. If the true model is $ARFIMA(0, d, 1)$ model and the number of lags more than 2, our estimator has a smaller bias for $\theta = 0.4$. Therefore, we argue that our estimator is more robust to misspecification in the autoregressive component.

6. Finite sample comparisons

In the previous sections, we only consider the asymptotic performances of our estimator. In fact, if the first-order autocorrelation is included, both our estimator and the TSB estimator contain the same amount of information, so the asymptotic variance should be the same. Thus, a more interesting question is to compare their small-sample behavior. To investigate this, we simulate an $ARFIMA(0, d, 0)$ process of sample size T. The values of T and d are set at $T = 100$, 200, and $d = -0.4$, -0.3, -0.2, -0.1, 0, 0.1 and 0.2.

The simulation method is similar to that of Chung and Baillie (1993). We make use of the Cholesky decomposition of the covariance matrix assuming Gaussian noise. In all the simulations, 1000 replications are performed.

Table 5
Variance of $\sqrt{T}(\hat{d} - d)$, using 2nd autocorrelation/partial autocorrelation only

d	$T = 100$ Auto	$T = 100$ Partial	$T = 200$ Auto	$T = 200$ Partial
− 0.4	4.83	2.56	7.43	3.15
− 0.3	4.70	3.25	7.02	4.14
− 0.2	4.55	3.72	7.22	4.69
− 0.1	4.85	4.17	7.22	4.73
0	4.55	3.56	6.50	4.25
0.1	3.98	3.19	3.84	3.41
0.2	2.54	2.44	1.56	2.05

Table 6

$d \backslash n$	$T = 100$			$T = 200$		
	1–5	2–6	3–7	1–5	2–6	3–7
(a) *Variance of $\sqrt{T}(\hat{d} - d)$, using five autocorrelations*						
− 0.4	0.75	2.94	3.35	0.86	5.09	6.38
− 0.3	0.94	2.86	3.28	1.02	5.09	5.73
− 0.2	1.00	3.33	3.53	0.87	5.59	5.89
− 0.1	0.96	3.66	3.99	0.87	6.36	6.86
0	0.97	3.42	4.20	0.76	5.52	7.24
0.1	0.82	2.70	3.49	0.80	3.50	5.81
0.2	0.59	2.20	2.90	0.58	1.53	3.53
(b) *Variance of $\sqrt{T}(\hat{d} - d)$, using five partial autocorrelations*						
− 0.4	0.66	1.45	2.93	0.51	1.15	2.45
− 0.3	0.73	1.61	3.04	0.64	1.45	2.93
− 0.2	0.83	1.94	3.56	0.68	1.74	3.30
− 0.1	0.83	2.18	4.10	0.76	2.12	3.93
0	0.95	2.48	4.13	0.65	2.32	4.53
0.1	0.81	2.27	4.06	0.75	2.31	3.96
0.2	0.53	1.84	2.93	0.48	1.41	2.36

We use the T realized values of y_t to construct the sample autocorrelations and sample partial autocorrelations. The search for d is over the range from − 0.5 to 0.25 at an increment of 0.01. Table 5 gives the variance of $\sqrt{T}(\hat{d} - d)$, using respectively the second-order autocorrelation and partial autocorrelation only. Table 6 gives the variance of $\sqrt{T}(\hat{d} - d)$, using five autocorrelations and partial autocorrelations.

From Tables 5 and 6, it is clear that for $T = 100$ and 200, our estimator performs better than the TSB estimator in most cases.

7. Conclusion

In this paper, we derive the theoretical variance–covariance structure of the sample partial autocorrelation function for the $ARFIMA(0, d, 0)$ model. The result is then used to construct an estimator analogous to that of TSB. It was shown that the asymptotic variance of our estimator is more robust to the differencing parameter as compared to the TSB estimator, when the first partial autocorrelation (or autocorrelation) is removed. Thus, the statistical inference based on partial autocorrelation function estimator should be more reliable.

We also uncover the reason behind the puzzling phenomenon of the TSB's study. We argue that their results are due to two primary reasons: first, for all n strictly bigger than 1, the nth-order autocorrelation function does not depend uniquely on the differencing parameter d. Thus, the problem of multiple solutions arises. This will increase the variation of the estimate of d. Our second reason has to do with the fact that, for $-0.5 < d < 0$, the asymptotic criterion function is rather insensitive to d for $n > 1$. This phenomenon together with the problem of multiple solutions increases the variance of the estimator for d.

Regarding the short-run dynamics when an $ARFIMA(0, d, 0)$ is specified, our estimator yields a smaller asymptotic variance if the true model is $ARFIMA(1, d, 0)$.

Finally, we compare the behavior of the two estimators in finite samples. It is found that our estimator performs better in most cases for samples of size 100 and 200.

For comparison purposes, throughout the paper, we assume the error term is i.i.d. with zero mean. Expressions in Eqs. (3) and (8) would be more cumbersome were this not assumed. Further comparison of our estimator with TSB's can be made by relaxing this assumption.

References

Backus, D.K., Zin, S.E., 1993. Long-memory inflation uncertainty: evidence from the term structure of interest rate. Journal of Money, Credit and Banking 25 (3), 681–700.

Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5–59.

Baillie, R.T., Bollerslev, T., 1994. Cointegration, fractional cointegration and exchange rate dynamic. Journal of Finance 49, 737–745.

Baillie, R.T., Chung, C.F., Tieslau, M.A., 1996. Analyzing inflation by the fractionally integrated ARFIMA-GARCH model. Journal of Applied Econometrics 11 (1), 23–40.

Cheung, Y.W., 1993. Tests for fractionally integration: a monte carlo investigation. Journal of Time Series Analysis 14, 331–345.

Chung, C.F., Baillie, R.T., 1993. Small sample bias in conditional sum-of-squares estimators of fractionally integrated ARMA models. Empirical Economics 18, 791–806.

Crato, N., Rothman, P., 1994. Fractional integration analysis of long-run behavior for US macroeconomic time series. Economics Letters 45, 287–291.

Diebold, F.X., Husted, S., Rush, M., 1991. Real exchange rates under the Gold Standard. Journal of Political Economy 99, 1252–1271.

Diebold, F.X., Rudebusch, G.D., 1989. Long memory and persistence in aggregate output. Journal of Monetary Economics 24, 189–209.

Ding, Z., Granger, C.W.J., Engle, R.F., 1993. A long memory property of stock returns and a new model. Journal of Empirical Finance 1, 83–106.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221–238.

Granger, C.W.J., Joyeux, R., 1980. An introduction to the long-memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15–29.

Hassler, U., Wolters, J., 1995. Long memory in inflation rates: international evidence. Journal of Business and Economic Statistics 13, 37–45.

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

Hurst, H.E., 1951. Long-term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers 116, 770–799.

Hurvich, C.M., Ray, B.K., 1995. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. Journal of Time Series Analysis 16, 17–41.

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

Lo, A.W., 1991. Long term memory in stock market prices. Econometrica 59, 1279–1313.

Shea, G.S., 1991. Uncertainty and implied variance bounds in long memory models of the interest rate term structure. Empirical Economics 16, 287–312.

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

Sowell, F., 1992b. Modeling long run behavior with the fractional ARIMA model. Journal of Monetary Economics 29, 277–302.

Tieslau, M.A., Schmidt, P., Baillie, R.T., 1996. A minimum distance estimator for long-memory processes. Journal of Econometrics 71, 249–264.