Maximum Likelihood Estimation of the Fractional Di®erencing Parameter in an ARFIMA Model Using Wavelets

Y. K. Tse
Department of Economics
National University of Singapore
Singapore 119260

V.V. Anh
Centre in Statistical Science and Industrial Mathematics
Queensland University of Technology
GPO Box 2434
Brisbane QLD 4001 Australia

Q. Tieng
Centre in Statistical Science and Industrial Mathematics
Queensland University of Technology
GPO Box 2434
Brisbane QLD 4001 Australia

Abstract In this paper we examine the ¯nite-sample properties of the approximate maximum likelihood estimate (MLE) of the fractional di®erencing parameter d in an ARFIMA(p, d, q) model based on the wavelet coe±cients. Ignoring wavelet coe±cients of higher order of resolution, the remaining wavelet coe±cients approximate a sample of independently and identically distributed normal variates with homogeneous variance within each level. The approximate MLE performs satisfactorily and provides a robust estimate for which the short memory component need not be speci¯ed.

Keywords ARFIMA model, fractional di®erencing parameter, maximum likelihood estimation, wavelet coe±cient

1 Introduction

In a recent paper Johnstone and Silverman (1997) showed that the wavelet coe±cients of a serially correlated noise process have the following properties. First, within each level the distribution of the wavelet coe±cients is stationary and the variance of the wavelet coe±cients is homogeneous within each level. Second, the autocorrelation of the wavelet coe±cients within each level dies away rapidly. Third, there is little or no correlation between the wavelet coe±cients at di®erent levels. Fourth, the logarithmic variance of the wavelet coe±cients decreases approximately linearly as the level increases.

McCoy and Walden (1996) suggested an approximate maximum likelihood estimation (MLE) method for estimating the fractional di®erencing parameter of a fractional white noise process. While McCoy and Walden considered only pure fractional white noise processes, the arguments of Johnstone and Silverman suggested that the method can be extended to an autoregressive fractionally integrated moving average (ARFIMA) process. In this paper we examine the ¯nite sample properties of the approximate MLE of the fractional di®erencing parameter d in an ARFIMA(p, d, q) model based on the wavelet coe±cients.

The most widely used method in the empirical literature for estimating the fractional di®erencing parameter in an ARFIMA model is perhaps the semiparametric procedure due to Geweke and Porter-Hudak (1983, hereafter GPH). An advantage of the GPH procedure is that the structure of the short memory component in the model need not be speci¯ed. Tse, Anh and Tieng (1999) examined the use of the t-statistic in the GPH regression for the estimation of the fractional di®erencing parameter as a test for cointegration. Sowell (1992) proposed the more e±cient MLE method to estimate the fractional di®erencing parameter. This method, however, is computationally very demanding. Besides, the full structure of the short memory component has to be speci¯ed. The nonparametric and semiparametric methods have been found to be more satisfactory than the MLE methods when the model (of the short memory component) is misspecified. Cheung (1993), however, found that under correct model specification the various MLE methods are superior. For a survey of the estimation of the ARFIMA models see Beran (1994).

The remaining of this paper is organised as follows. In Section 2 we briefly summarize the notations and definitions of the wavelet transform and the wavelet coefficients. Section 3 outlines the approximate likelihood function of the Gaussian ARFIMA process. Results of a Monte Carlo experiment are reported in Section 4. Section 5 concludes the paper.

2 Wavelet Coefficients and Discrete Wavelet Transform

Traditional time series analyses rely on methods involving either the time or the frequency domain. In contrast, wavelet transforms permit an analysis that combines both time and frequency information. Applications of wavelet analysis have been rapidly evolving in the areas of mathematics, quantum physics and image processing. Recently, the technique has been adopted in the areas of computational economics and econometrics (see, for example, Davidson, Labys, and Lesourd (1998) and Pan and Wang (1998) In this section we briefly review the theory of wavelet transformations.

Let $Y = (Y_1, Y_2, \ldots, Y_N)'$ be an N-vector of observations. Usually wavelet transformations are used with equally spaced observations with a sample size equal to an integer power of two. Thus, we assume $N = 2^J$ for some positive integer J.[1] If we apply a discrete wavelet transform (DWT) on Y, we obtain two 2^{J-1}-vectors of the smoothed and detailed components at the $(J-1)$th level of resolution, denoted by c_{J-1} and d_{J-1}, respectively. The DWT consists of low-pass (to generate c_{J-1}) and high-pass (to generate d_{J-1}) filters generated from a mother wavelet.[2] If we apply the DWT to c_{J-1} again,

[1] When this is not the case, the technique of padding may be used.

[2] In this paper we use the Daubechies filters of order 4. The Daubechies wavelet has many desirable

we obtain the smoothed and detailed components at the $(J-2)$th level of resolution, denoted by c_{J-2} and d_{J-2}, respectively. Repeating this process recursively, we obtain the vector of wavelet coefficients defined by the N-vector $d = (c_0; d_0; \ldots; d'_{J-2}; d'_{J-1})'$ after J applications of the DWT. Here both c_0 and d_0 are scalars, denoting the smoothed and detailed components of the lowest level of resolution.

Note that $d_j = \{d_{jk}\}$, $k = 1; \ldots; 2^j$, is an 2^j-vector, for $j = 0; \ldots; J-1$. Thus, there are 2^j wavelet coefficients at the jth level of resolution. In the next section we present the approximate likelihood function of the wavelet coefficients when Y is generated from an ARFIMA process.

3 The Approximate Likelihood Function of an ARFIMA Process

We assume $\{Y_t\}$ follows an ARFIMA(p, d, q) process given by

$$(1 - \phi_1 L - \cdots - \phi_p L^p)(1-L)^d Y_t = (1 - \theta_1 L - \cdots - \theta_q L^q)\epsilon_t$$

where $\epsilon_t \sim IIDN(0, \sigma_\epsilon^2)$ and L is the lag operator: The roots of $1 - \phi_1 L - \cdots - \phi_p L^p = 0$ and $1 - \theta_1 L - \cdots - \theta_q L^q = 0$ are assumed to lie outside the unit circle. The parameter d is not necessarily an integer so that fractional differencing is permitted. When $|d| < 0.5$, the process $\{Y_t\}$ is stationary. However, when $d \geq 0.5$, the process has an infinite variance and is thus not covariance stationary. Note that Y_t has an infinite autoregressive representation as a result of the following expansion

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k L^k;$$

where

$$\binom{d}{k} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$$

and $\Gamma(x)$ is the gamma function.

properties. In particular, it possesses the smallest support for a given number of vanishing moments.

Following the arguments of Johnstone and Silverman (1997), we assume that $d_{jk} \sim IIDN(0, \sigma_j^2)$, for $j = 0, \ldots, J-1$ and $k = 1, \ldots, 2^j$. This assumption has incorporated the noncorrelation of wavelet coefficients within the same level as well as across different levels.[3]

Next we adopt the argument of McCoy and Walden (1996) that for sufficiently small j, $\log \sigma_j^2$ is approximately linear in j. In particular, we have, for sufficiently small j, $\sigma_j^2 \approx 2^{2(J-j)d}\sigma^2$, where σ^2 depends on other parameters of the model, but does not vary with j.

Ignoring c_0 and the wavelet coefficients for $j > J-K$, the approximate likelihood function of $(d_0', \ldots, d_{J-K}')'$ is given by (after dropping the constant term)

$$\ell(d;\sigma^2) = -\frac{1}{2}\sum_{j=0}^{J-K}\left[2^j \log\left(2^{2(J-j)d}\sigma^2\right) + \sum_{k=1}^{2^j}\frac{d_{jk}^2}{2^{2(J-j)d}\sigma^2}\right].$$

The approximate MLE of d and σ^2 can be obtained by maximizing $\ell(d;\sigma^2)$ with respect to d and σ^2. In the next section we report the results of a Monte Carlo experiment on the finite sample distribution of the approximate MLE of d.

4 Monte Carlo Results

We generated $\{Y_t\}$ from an ARFIMA(p, d, q) process with (p, q) = (0, 0); (1, 0) and (0, 1). To generate the series, we used the truncated moving average representation. This is a convenient method when the process is stationary. N is taken to be 1024 and 2048 (i.e., J is 10 and 11, respectively). The approximate maximum likelihood method requires the wavelet coefficients of high resolution to be discarded. If the number of truncated levels $K-1$ is large, there will be a great reduction in the number of observations retained for estimation.[4] However, if the number of truncated levels is small, serious bias may be induced. In McCoy and Walden (1996) only the case of $K = 2$ was considered. To

[3] See Appendix A of Johnstone and Silverman (1997).

[4] Recall that there are 2^j wavelet coefficients in the jth level of resolution.

examine the e®ects of truncating the wavelet coe±cients, we let K = 2; 3 and 4. The approximate MLE of d is then denoted as W(2), W(3) and W(4), respectively. As a comparison, we also calculated the frequency-domain MLE (denoted as MLE for short) of d. Estimates of the bias and root mean squared errors (RMSE) of each estimate are summarized in Tables 1, 2 and 3. The results are based on Monte Carlo samples of 1000 each.[5]

Table 1 summarises the case when there is no short-memory component, i.e., (p, q) = (0, 0).[6] We can see that, in terms of the RMSE, W(2) performs quite well relative to the MLE, which provides the best results. W(3) provides lower bias than W(2) in all cases, but this is achieved at the expense of increasing the RMSE.

From Table 2 (for (p, q) = (1, 0)) we can see that W(2) performs quite badly for Á = 0:4 or 0.6. Indeed for Á = 0:6, even W(3) gives rise to rather large RMSE. W(4), however, appears to be satisfactory compared to the MLE. Indeed, the RMSE of W(4) is quite close to those of the MLE. The same conclusion applies to the results in Table 3 for (p, q) = (0; 1):s

5 Conclusions

Overall the results suggest that while W(4) has the largest RMSE for pure fractional processes, it performs favourably against the MLE and other W estimates with fewer truncations when the serial correlation in the short-memory component is moderate. As the W estimate has an advantage over the MLE in that the structure of the short-memory component need not be speci¯ed, it provides a robust estimate for d.

[5] All computations performed in this paper were coded in GAUSS with the application modules TSM and MAXLIK. TSM contains procedures for the generation of observations following an ARFIMA process and the computation of the frequency-domain MLE of the fractional di®erencing parameter d.

[6] This was the case considered by McCoy and Walden (1996).

References

[1] J. Beran, Statistics for Long-Memory Processes (Chapman & Hall, New York, 1994).

[2] Y.W. Cheung, Test for Fractional Integration: A Monte Carlo Investigation, Journal of Time Series Analysis 14 (1993) 331-345.

[3] R. Davidson, W. Labys and J.-B. Lesourd, Wavelet Analysis of Commodity Price Behaviour, Computational Economics 11 (1998) 103-128.

[4] J. Geweke and S. Porter-Hudak, The Estimation and Application of Long Memory Time Series Models, Journal of Time Series Analysis 4 (1983) 221{238.

[5] J.M. Johnstone and B.W. Silverman, Wavelet Threshold Estimators for Data with Correlated Noise, Journal of the Royal Statistical Society Series B 59 (1997) 319{351.

[6] E.J. McCoy and A.T. Walden, Wavelet Analysis and Synthesis of Stationary Long-Memory Processes, Journal of Computational and Graphical Statistics 5 (1996) 26{56.

[7] Z. Pan and X. Wang, A Stochastic Nonlinear Regression Estimator Using Wavelets, Computational Economics 11 (1998) 89-102.

[8] F.B. Sowell, Modelling Long-Run Behaviour with Fractional ARFIMA Model, Journal of Monetary Economics 29 (1992) 277-302.

[9] Y.K. Tse, V.V. Anh and Q. Tieng, No-Cointegration Test Based on Fractional Di®erencing: Some Monte Carlo Results, Journal of Statistical Planning and Inference 80 (1999) 257 - 267.

Table 1: Monte Carlo Results of the Wavelet Approximate MLE and the Frequency-Domain MLE for ARFIMA(0, d, 0) Model

	Estimation	Sample Size	
d	Method	1024	2048
0.1	W(2)	{0.0069	{0.0047
		0.0339	0.0249
	W(3)	{0.0063	{0.0035
		0.0484	0.0336
	W(4)	{0.0125	{0.0058
		0.0732	0.0513
	MLE	{0.0005	0.0007
		0.0245	0.0173
0.2	W(2)	{0.0106	{0.0095
		0.0348	0.0243
	W(3)	{0.0061	{0.0058
		0.0523	0.0322
	W(4)	{0.0097	{0.0058
		0.0801	0.0513
	MLE	0.0023	0.0009
		0.0246	0.0173
0.3	W(2)	{0.0108	{0.0104
		0.0349	0.0266
	W(3)	{0.0027	{0.0023
		0.0470	0.0332
	W(4)	{0.0035	0.0017
		0.0722	0.0510
	MLE	0.0062	0.0021
		0.0272	0.0174
0.4	W(2)	{0.0118	{0.0090
		0.0379	0.0261
	W(3)	{0.0015	0.0017
		0.0548	0.0361
	W(4)	{0.0053	0.0040
		0.0795	0.0511
	MLE	0.0096	0.0048
		0.0281	0.0180

Notes: For each d and estimation method, the ¯rst ¯gure under \Sample Size" is the estimated bias and the second ¯gure is the estimated root mean squared error (RMSE). For example, {0.0069 is the estimated bias, while 0.0339 is the estimated RMSE for d = 0:1, W(2) and N = 1024.

Table 2: Monte Carlo Results of the Wavelet Approximate MLE and the Frequency-Domain MLE for ARFIMA(1, d, 0) Model

		Estimation	Sample Size	
Á	d	Method	1024	2048
0.2	0.2	W(2)	0.0512	0.0524
			0.0626	0.0570
		W(3)	0.0194	0.0195
			0.0536	0.0372
		W(4)	0.0006	0.0051
			0.0755	0.0461
		MLE	0.0017	0.0016
			0.0529	0.0332
0.4	0.2	W(2)	0.1445	0.1450
			0.1485	0.1471
		W(3)	0.0612	0.0630
			0.0797	0.0719
		W(4)	0.0171	0.0175
			0.0754	0.0539
		MLE	{0.0020	0.0040
			0.0742	0.0471
0.6	0.2	W(2)	0.2932	0.2894
			0.2957	0.2906
		W(3)	0.1480	0.1469
			0.1562	0.1509
		W(4)	0.0580	0.0568
			0.0957	0.0750
		MLE	0.0009	{0.0004
			0.0969	0.0632
0.2	0.4	W(2)	0.0494	0.0468
			0.0620	0.0528
		W(3)	0.0215	0.0209
			0.0605	0.0405
		W(4)	0.0064	0.0123
			0.0821	0.0552
		MLE	0.0211	0.0096
			0.0595	0.0359
0.4	0.4	W(2)	0.1350	0.1315
			0.0585	0.0564
		W(3)	0.0217	0.0207
			0.0795	0.0662
		W(4)	0.0217	0.0207
			0.0817	0.0541
		MLE	0.0215	0.0105
			0.0752	0.0459
0.6	0.4	W(2)	0.2732	0.2713
			0.2756	0.2726
		W(3)	0.1425	0.1372
			0.1513	0.1419
		W(4)	0.0553	0.0562
			0.0973	0.0759
		MLE	0.0304	0.0147
			0.0976	0.0672

Notes: The data generation process is (1 ¡ ÁL) (1 ¡ L)^{d} Y_t = ²$_t$. See Notes to Table 1 for more details.

Table 3: Monte Carlo Results of the Wavelet Approximate MLE and the Frequency-Domain MLE for ARFIMA(0, d, 1) Model

µ	d	Estimation Method	Sample Size 1024	Sample Size 2048
0.2	0.2	W(2)	{0.0738	{0.0673
			0.0809	0.0716
		W(3)	{0.0288	{0.0241
			0.0568	0.0410
		W(4)	{0.0169	{0.0064
			0.0754	0.0494
		MLE	0.0040	0.0060
			0.0511	0.0337
0.4	0.2	W(2)	{0.1664	{0.1595
			0.1703	0.1617
		W(3)	{0.0809	{0.0773
			0.0940	0.0847
		W(4)	{0.0386	{0.0263
			0.0824	0.0547
		MLE	0.0123	0.0057
			0.0738	0.0462
0.6	0.2	W(2)	{0.2952	{0.2822
			0.2979	0.2838
		W(3)	{0.1817	{0.1722
			0.1887	0.1759
		W(4)	{0.0835	{0.0776
			0.1094	0.0923
		MLE	0.0228	0.0062
			0.1021	0.0651
0.2	0.4	W(2)	-0.0704	-0.0674
			0.0791	0.0724
		W(3)	-0.0237	-0.0216
			0.0562	0.0420
		W(4)	-0.0075	-0.0024
			0.0785	0.0539
		MLE	0.0275	0.0110
			0.0637	0.0390
0.4	0.4	W(2)	{0.1617	{0.1537
			0.1666	0.1563
		W(3)	{0.0689	{0.0640
			0.0869	0.0741
		W(4)	{0.0222	{0.0187
			0.0793	0.0543
		MLE	0.0382	0.0181
			0.0957	0.0513
0.6	0.4	W(2)	{0.2915	{0.2774
			0.2942	0.2792
		W(3)	{0.1683	{0.1565
			0.1761	0.1612
		W(4)	{0.0758	{0.0619
			0.1070	0.0827
		MLE	0.0350	0.0237
			0.1055	0.0759

Notes: The data generation process is (1 ¡ L)^d Y_t = (1 ¡ µL) ²$_t$. See notes to Table 1 for more details.