Turk J Math
35 (2011) , 151 – 158.

doi:10.3906/mat-0906-31

Jackknife and bootstrap with cycling blocks for the estimation of fractional parameter in ARFIMA model

Lorenc Ekonomi and Argjir Butka

Abstract

One of most important problems concerning the ARFIMA time series model is the estimation of fractional parameter d. Various methods have been used to solve this problem, such as the log-periodogram regression of a process. In this article we propose two jackknife and bootstrap methods, which aid in the estimation of fractional parameter d. These methods involve non-overlapping blocks and moving blocks with random starting point and length. We have conducted several simulations and the results show that the estimations obtained are very close to the real parameter value.

Key Words: Jackknife, bootstrap, fractionally parameter, ARFIMA model, moving blocks, non overlapping blocks.

1. Introduction

Over the last couple decades, there has been much special interest in the application of long memory time series to various fields such as economy, finance, hydrology and geology. Of particular focus has been estimation of ARFIMA model parameters, especially that of fractional parameter d. For this, Granger and Joueux [10] approximated the ARFIMA model by a high-order autoregressive process and estimated d by comparing variances for different choices of d. Geweke and Porter-Hudak [9] and Kashyap and Eom [12] used a regression procedure for the logarithm of the periodogram to estimate d. Hassler [11] considered tests based on the asymptotic results of Kashyap and Eom [12], of the periodogram regression estimator of d. He showed that this test is valid only if the series are generated from a fractional white noise process.

In Section 2 we give the classical method [9] for the estimation of unknown parameter d, and in Section 3 given is a short expose of the contribution to jackknife and bootstrap for time series.

In Section 4 we propose two jackknife and bootstrap methods for the estimation of parameter d. We subtract the arithmetic mean from the data corresponding the time series, obtaining thus a numeric sequence having groups made entirely of non negative terms and similar groups made of negative terms. The non negative and negative groups alternate with each other. Each non negative group followed by a negative group consists of what we call a cycle. Also, we define a block constructed of a fixed number of consecutive cycles. Such blocks will be the units of future samples. In the first jackknife and bootstrap method, blocks do not overlap one as other, while in the later methods we allow blocks to initiate with any term which is the start of a cycle. We call these moving blocks.

In Section 5 we have constructed a simulation with ARFIMA models in order to estimate the fractional parameter for different values by methods proposed in Section 4. The corresponding results are given in Tables 1 and 2 at the end of the paper. The results of the relative bias for the estimations show a surprising fit of the estimation with the real values. In Section 6 we give some conclusions.

Table 1. Jackknife and bootstrap with non overlapping blocks.

d	T	g	No. of cycles	$\hat{d}_J$	Relative bias
0.4	300	40	10	0.4263	0.06
0.3	300	40	10	0.2774	-0.07
0.2	300	40	10	0.2196	0.09
0.1	300	40	10	0.1017	0.01
-0.1	300	40	10	-0.0953	-0.04
-0.2	300	40	10	-0.1987	0.00
-0.3	300	40	10	-0.3067	0.02
-0.4	300	40	10	-0.4073	0.01

d	T	g	No. of cycles	$\hat{d}_B$	Relative bias
0.4	300	60	10	0.3627	-0.09
0.3	300	60	10	0.2954	-0.01
0.2	300	60	10	0.1966	-0.01
0.1	300	60	10	0.0807	-0.19
-0.1	300	60	10	-0.1074	0.07
-0.2	300	60	10	-0.1980	-0.01
-0.3	300	60	10	-0.2654	-0.11
-0.4	300	60	10	-0.3469	-0.13

2. Regression estimators of d

The autoregressive integrated moving average process ARFIMA(p,d,q) is defined as

$$\phi(B)(1-B)^d Y_t = \theta(B) Z_t, \tag{1}$$

where B is the back-shift operator, Z_t is a white process with mean zero and variance σ^2 and $\phi(B) = 1-\phi_1 B - ... - \phi_p B^p$ and $\theta(B) = 1-\theta_1 B - ... - \theta_q B^q$ are stationary autoregressive and invertible moving operators of order p and q, respectively, and d takes fractional values in the range $]-0.5, 0.5[$.

Let us show the estimation of the periodogram of d by [9]. The spectral density of ARFIMA(p,d,q) process is

$$f(w) = f_u(w)\left[2\sin\left(\frac{w}{2}\right)\right]^{-2d}, w \in]-\pi,\pi[, \tag{2}$$

Table 2. Jackknife and bootstrap with moving blocks.

d	T	g	No. of cycles	$\hat{d}_J$	Relative bias
0.4	300	23	2	0.4213	0.05
0.3	300	23	2	0.3201	0.07
0.2	300	23	2	0.2066	0.03
0.1	300	23	2	0.1059	0.06
-0.1	300	23	2	-0.1022	0.02
-0.2	300	23	2	-0.2127	0.06
-0.3	300	23	2	-0.3199	0.06
-0.4	300	23	2	-0.4259	0.06

d	T	g	No. of cycles	$\hat{d}_J$	Relative bias
0.4	300	50	8	0.4156	0.03
0.3	300	50	8	0.2970	-0.01
0.2	300	50	8	0.1984	0.00
0.1	300	50	8	0.0916	-0.08
-0.1	300	50	8	–0.1001	0.00
-0.2	300	50	8	-0.2100	0.05
-0.3	300	50	8	-0.3095	0.03
-0.4	300	50	8	-0.4198	0.05

d	T	g	No. of cycles	$\hat{d}_J$	Relative bias
0.4	300	50	9	0.4126	0.03
0.3	300	50	9	0.3014	0.00
0.2	300	50	9	0.1859	-0.07
0.1	300	50	9	0.0944	-0.05
-0.1	300	50	9	-0.1066	0.06
-0.2	300	50	9	-0.2069	0.03
-0.3	300	50	9	-0.3233	0.07
-0.4	300	50	9	-0.4293	0.07

d	T	g	No. of cycles	$\hat{d}_B$	Relative bias
0.4	300	50	10	0.4169	0.04
0.3	300	50	10	0.3058	0.02
0.2	300	50	10	0.1970	-0.01
0.1	300	50	10	0.1021	0.02
-0.1	300	50	10	-0.0903	-0.09
-0.2	300	50	10	-0.1834	-0.08
-0.3	300	50	10	-0.2666	-0.11
-0.4	300	50	10	-0.3341	-0.16

where $f_u(w)$ is the spectral density of the ARMA(p,q) process. If the time series has T terms, than $w_j = \frac{2\pi j}{T}, j = 1, 2, ..., \frac{T}{2}$ is a set of harmonic frequencies. After some transformations this may be written as

$$\ln f(w_j) = \ln f_u(0) - d \ln \left[2 \sin\left(\frac{w_j}{2}\right)\right]^2 + \ln \frac{f_u(w_j)}{f_u(0)}. \tag{3}$$

For a given series $y_1, y_2, ..., y_T$ the periodogram is given by

$$I(w) = \frac{1}{2\pi}\left[R(0) + 2\sum_{k=1}^{T-1} R(k)\cos(kw)\right], w \in]-\pi, \pi[. \quad (4)$$

Adding $\ln I(w_j)$ to both sides of (3), equation (3) can be written as a simple regression equation,

$$u_j = a + bv_j + e_j, j = 1, 2, ..., g, \quad (5)$$

where $u_j = \ln I(w_j), v_j = 2\ln\left[2\sin\left(\frac{w_j}{2}\right)\right], e_j = \ln\frac{I(w_j)}{f_u(w_j)} + c, b = -d, a = \ln f_u(0) - c, c = E\left[-\ln\frac{I(w_j)}{f_u(w_j)}\right]$ and g is chosen so that $\frac{g}{T} \to 0$ when $T \to \infty$.

The estimator of d is then $-\hat{b}$, where $\hat{b}$ is the least squares estimate of the slope of the regression equation (5), that is

$$\hat{d} = -\hat{b} = -\frac{\sum_{j=1}^{g}(v_j - \overline{v})u_j}{\sum_{j=1}^{g}(v_j - \overline{v})^2}, \quad (6)$$

where $\overline{v} = \frac{1}{g}\sum_{j=1}^{g} v_j$.

It is shown [9, 12] that the periodogram regression estimator of d is asymptotically normally distributed with mean d.

3. Jackknife and bootstrap in time series

The jackknife [23] and the bootstrap [5] have become well established as nonparametric estimators of the variance of a statistic. However, the assumption of independence of the observations is crucial. It is easily seen that they give incorrect answers if dependence is neglected [22]. Recently the two methods have been extended to ARMA models by reducing to innovations that which are i.i.d. [4, 8, 6]. Still, ARMA processes are not able to model essential features of many observed time series [20]. Fitting models which go beyond ARMA is, however, an extremely difficult task and it seems impossible to take into account effects of parameter estimation or incorrect specification of the model. Moreover, a variance estimator can be unreliable even if the true distribution differs only slightly from the model and the statistic is robust [13]. For these reasons there has been proposed an extension to the standard jackknife and bootstrap which does not require one to first fit a parametric or semi parametric model [13]. It works for arbitrary stationary processes with short-range dependence expressed, for instance, with mixing conditions. For the jackknife we deleted each block of l consecutive observations once and calculate the sample variance of the values of the statistic obtained in this way. Moreover we make a smooth transition between observations left out and observations with full weight, similar to tapering in time series analysis. For the bootstrap we choose $\frac{n}{l}$ blocks of length l with replacement from the $n - l + 1$ blocks of observed data. If the statistic is not a symmetric function of the observations, leaving out observations in the middle or joining randomly selected blocks causes problems.

Carlstein [2] proposed a variance estimator which selects non-overlapping blocks. So, there were proposed bootstrap methods with moving blocks [15] which randomly chose blocks with fixed length. This bootstrap was biased in the finite sets, but asymptotically unbiased for the mean. Another version was stationary bootstrap [19] where the blocks were chosen from a starting point and with randomly lengths. This estimation was unbiased for the mean, even in the finite sets. Furthermore, even the variance estimate for the mean was consistent.

But, the problem was that the extremes of blocks, during the formation of the bootstrap samples, destroyed the structure of the dependency among the observations of time series. In order to eliminate this, we can use bootstrap with linked blocks [14] and with matched blocks [3].

Phillips and Yu [17] proposed the use of jackknife techniques to reduce the bias for pricing bond options in a financial time series. Suppose a sample of n observations is decomposed into m consecutive sub-samples, each with of l observations $(n = m \times l)$. The jackknife estimator of a parameter θ in the model is defined by $\hat{\theta}_J = \frac{m}{m-1}\hat{\theta}_n - \frac{1}{m^2-m}\sum_{i=1}^{n}\hat{\theta}_{li}$ where $\hat{\theta}_n$ and $\hat{\theta}_{li}$ are the extreme estimates of θ based on the entire sample and the i-s sub-sample, respectively, and θ can be a coefficient in the diffusion process, such as the mean reversion parameter, or a much more complex function of the parameters of the diffusion process and the data, such as an asset price or derivative price [18].

Bootstrap with blocks was used for the estimation of parameters of ARFIMA model [7, 16], but local bootstrap was used for estimation of parameter d in ARFIMA model [21].

In the next section we will propose a jackknife and a bootstrap method which estimates parameter d in ARFIMA model. This will be bootstrap with blocks that have no fixed length, the starting point of every block is random and the block with its starting point is defined according to a certain criteria. Let us take in consideration two cases. In the first case jackknife and the bootstrap will be constructed with non-overlapping blocks, while in the second case the blocks will be moving with a starting point at the beginning of a cycle.

4. Jackknife and bootstrap estimation with blocks compounded by cycle

Let us suppose an ARFIMA(p, d, q) time series process with terms $y_1, y_2, ..., y_T$. We shall show how the blocks will be formed on the basis of which the bootstrap sample will be constructed.

If we subtract from the series terms their mean, the time series sequence will exhibit consecutive groups with continuously alternating positive and negative signs. Let us write this adapted series as $y'_1, y'_2, ..., y'_T$. Every two consecutive groups of positive and then negative values, we will call a cycle. Let $C_1, C_2, ..., C_k$ denote the created cycles and that each cycle C_i has n_i values for $i = 1, 2, ..., k$. We have $\sum_{i=1}^{k} n_i = T$. The form of the cycles will be $C_1 = \left\{y'_1, ..., y'_{n_1}\right\}, C_2 = \left\{y'_{n_1+1}, ..., y'_{n_1+n_2}\right\}, ..., C_k = \left\{y'_{n_1+\cdots+n_{k-1}+1}, ..., y'_T\right\}$, with $y'_{n_1+\cdots+n_i} < 0,\ y'_{n_1+\cdots+n_i+1} > 0$ for $i = 1, 2, ..., k$. It is clear enough that the length of each cycle and their number are random.

We describe two different ways of forming of blocks. In the first case the blocks are not overlapping and, if the block is compounded by s consecutive cycles, the starting point will be the observations $y_1, y_{n_1+\cdots+n_s+1}, \ldots$. For the jackknife estimation we delete consecutively each of the blocks and with the remaining observations we estimate parameter d by (6). If $\hat{d}_{(i)}$ is the estimation when the block of i-s is deleted, and we have deleted in total m blocks, then the jackknife estimation for parameter d will be

$$\hat{d}_J = m\hat{d} - \frac{m-1}{m}\sum_{i=1}^{m}\hat{d}_{(i)}. \tag{7}$$

For bootstrap estimation we randomly choose blocks until we have a subset of T values from the time series. Let us denote the chosen values by $y_1^*, y_2^*, ..., y_T^*$. Using the procedures described in Section 2, we take the estimation for the parameter d, that we call bootstrap estimation, and sign it with $\hat{d}^*$. We repeat this process M times forming a set $\hat{d}_1^*, \hat{d}_2^*, ..., \hat{d}_M^*$ of bootstrap estimations. Then, the estimation for parameter d will be

$$\hat{d}_B = \frac{1}{M}\sum_{i=1}^{M}\hat{d}_i^*. \tag{8}$$

In the second case we have moving blocks and the starting point of a block will be the observations $y_1, y_{n_1+1}, ..., y_{n_1+\cdots+n_{k-1}+1}$. The estimation procedures for bootstrap and jackknife for d parameter will be the same as in the first case.

5. Simulations

We construct a programme for the simulation of the ARFIMA process and calculate an estimation of parameter d. At first we simulate the ARFIMA $(0, d, 0)$ process for different known values of $d \in]-0.5, 0.5[$, generating the terms of the time series for $t = 1, 2, ..., T$. For this, as it appears from [9], expression (1) is transformed into

$$y_t = \sum_{k=0}^{\infty}\beta_k\varepsilon_{t-k}, \tag{9}$$

where ε_{t-k} is the white noise term with mean zero, and coefficient β_k computed via the recurrent relation

$$\beta_{k+1} = \beta_k\frac{d+k}{k+1}. \tag{10}$$

To determine of the values of T, g and the number of the cycles that constitute the block, we carry out different calculations until we find a suitable combination. This is a problem that is mentioned in most of the cases in the literature of this field. There are different ideas for the determination of these numbers, but in all cases the suitable values are determined by simulations [3, 7, 14, 19, 21].

We estimate $R(k)$ for $k = 0, 1, ..., T-1$, which appears in (4), applying the formula $R(k) = \frac{1}{T}\sum\limits_{i=k+1}^{T}(y_i - \overline{y})(y_{i-k} - \overline{y})$ where $\overline{y} = \frac{1}{T}\sum\limits_{i=1}^{T}y_i$, then $I(w_j), u_j = \ln I(w_j)$ and $v_j = 2\ln\left[2\sin\left(\frac{w_j}{2}\right)\right]$ for $j = 1, 2, ..., g$.

In Tables 1 and 2 one find some of the best results achieved. In the tables, is presented relative bias, which is given in many cases in the research articles [24], as a ratio of the estimated difference from the true value to the absolute value of the true parameter value. It is worth mentioning that the number of the cycles which form the block is nearly to 10. Judging from the simulation results, it is generally noticed a better estimation of the jackknife than that of the bootstrap. On the other hand, taking into account the calculations, the jackknife and bootstrap exhibit better behavior for moving blocks, than with non overlapping b locks. For the jackknife, this is linked with the greater number of the realized estimations, something that influences richness of the estimation histogram.

6. Conclusions

The simulation results show clearly that the jackknife and bootstrap are very good alternatives for the estimation of time series, in particular for the estimation of fractional parameter via the ARFIMA model. It is important to stress that the units of samples have to be block-constructed in such a way that they preserve the structure and dependence between terms of the time series. From the simulation it is observed that estimation with moving blocks give smaller values of relative bias. On the other hand, the optimal number of cycles which appear in a block is determined from the result of the simulation. We do not have yet any theoretical argument about what is the correlation between the number of terms of the series and the number of cycles which appear in the block in order to get optimal results from the estimation. We hope this will be a topic of future study.

References

[1] Bickel, P.J. and Freedman, D.A.: Some asymptotic theory for the bootstrap, Annals of Statistics 9, 1196–1217, (1981).

[2] Carlstein, E.: The use of subseries values for estimating the variance of a general statistic from a stationary sequence, Ann. Statist. 14, 1171–1179, (1986).

[3] Carlstein, E., Do, K., Hall, P., Hesterberg, T. and Kunsch, H.: Matched-block bootstrap for dependent data, Bernoulli 4, 305–328, (1998).

[4] Davis, W. W.: Robust interval estimation of the innovation variance of an ARMA model, Ann. Statist. 5, 700–708, (1977).

[5] Efron, B.: Bootstrap methods: another look at the jackknife, Annals of Statistics 7, 1–26, (1979).

[6] Efron, B. and Tibshirani, R.J.: Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy (with discussion), Statist. Sci. 1, 54–77, (1986).

[7] Franco, G.C. and Reisen, V.A.: Bootstrap techniques in semiparametric estimation methods for ARFIMA models: a comparison study, Computational Statistics & Data Analysis 19, 243–259, (2004).

[8] Freedman, D.A.: On bootstrapping two-stage least squares estimates in stationary linear models, Ann. Statist. 12, 827–842, (1984).

[9] Geweke, J. and Porter-Hudak, S.: The estimation and application of long memory time series model, Journal of Time Series Analysis 4, 221–238, (1983).

[10] Granger, C.W.J. and Joyeux, R.: An introduction to long memory time series models and fractional differencing, Journal of Time Series Analysis 1, 15–39, (1980).

[11] Hassler, U.: Regression of spectral estimators with fractionally integrated time series, Journal of Time Series Analysis 14, 369–379, (1993).

[12] Kashyap, R.L. and Eom, K.B.: Estimation in long-memory time series models, Journal of Time Series Analysis 9, 35–41, (1988).

[13] Kunsch, H.: The jackknife and bootstrap for general stationary observations, Annals of Statistics 17, 1217–1241, (1989).

[14] Kunsch, H. and Carlstein, E.: The linked blockwise bootstrap for serially dependent observations, Technical Report, Dept. of Statistics, Univ. of North Carolina, Chapel Hill. (1990).

[15] Liu, R. and Singh, K.: Moving blocks jackknife and bootstrap capture weak dependence, Exploring the Limits of Bootstrap, Wiley, New York, 225–248, (1992).

[16] Maharaj, A.E.: A test for the difference parameter of the ARFIMA model using the moving blocks bootstrap, Working Paper, Dept. of Economics and Business Statistics, Monash University Australia, (1999).

[17] Phillips, P.C.B. and Yu, J.: Jackknifing bond option prices, Review of Financial Studies 18, 707–742, (2005).

[18] Phillips, P.C.B. and Yu, J.: Comment: A selective overview of nonparametric methods in financial econometrics, Statistical science, Vol. 20, No.4, 338–343, (2005).

[19] Politis, D. and Romano, J.: The stationary bootstrap, JASA, 89, 1303–1313, (1994).

[20] Priestley, M.B.: Spectral analysis and time series 1, 2, Academic New York, (1981).

[21] Silva, E.M., Franco, G.C., Reisen, V.A. and Cruz F.R.B.: Local bootstrap approaches for fractional differential parameter estimation in ARFIMA models, Computational Statistics & Data Analysis 51, 1002–1011, (2006).

[22] Singh, K.: On the asymptotic accuracy of Efron's bootstrap, Annals of Statistics 9, 1187–1195, (1981).

[23] Tukey, J.: Bias and confidence in not quite large samples (abstract), Ann. Math. Statist. 29, 614, (1958).

[24] Wu, J.: Jackknife, bootstrap and other resampling methods in regression analysis, Annals of Statistics 14, 1261–1295, (1986).

Lorenc EKONOMI
Natural Sciences Department
University of Korca, Korce, ALBANIA
e-mail: lorencekonomi@yahoo.co.uk

Received 10.06.2009

Argjir BUTKA
Natural Sciences Department
University of Korca, Korce, ALBANIA
e-mail: argjirbutka@yahoo.com