Key words: **Aggregate time series**; **Long memory**; **Infinite MA operator**; **Power loss**.

1. Introduction

Many time series encountered in practice are generated by a long-memory process and consequently, before modeling any time series, it may be necessary to test it for long memory. The Autocorrelation Function of a (second-order) stationary and invertible ARMA process is geometrically bounded, i.e., $|\rho_\tau| \leq Kr^\tau,\ \tau = 1, 2, \ldots$, where K and r are constants such that $K > 0$ and $0 < r < 1$. A long-memory process is characterized by a much slower decay of its autocorrelations, which follow an asymptotically hyperbolic decay. A second order stationary process $\{Z_t\}$ is a long-memory process if, for some $K > 0$ and $\alpha < 0$, its autocorrelation function has the following asymptotic behavior, $|\rho_\tau| \sim K|\tau|^\alpha$ mbox $\tau \to \infty$, where the symbol ~ is used in the usual way, i.e., $a_N \sim kb_N$ if and only if $a_N/b_N \to k$ with $k \neq 0$.

Let B define the backshift operator, $BZ_t = Z_{t-1}$, and $\nabla^d Z_t = (1-B)^d Z_t$ the differencing operator. The process $\{Z_t\}$ is called an Autoregressive Fractionally Integrated Moving Average process of order (p, q), or an ARFIMA(p, d, q) process, with $d \in IR$, if $\nabla^d Z_t$ is a stationary solution of the difference equations

$$\phi(B)\nabla^d Z_t = \theta(B)a_t \qquad \#$$

where the polynomials $\phi(B) = (1 - \phi_1 B - \ldots - \phi_p B^p)$ and $\theta(B) = (1 - \theta_1 B - \ldots - \theta_q B^q)$ have their roots outside the unit circle and no roots in common, and a_t is a white noise process with zero mean and mbox $(a_t) = \sigma_a^2$. We will assume that $d \in (-0.5, 0.5)$, for which the ARFIMA model (ref: lm3) is stationary and invertible. In order to estimate the parameter d, we consider a spectral regression method proposed by Geweke and Porter-Hudak (1983), henceforth GPH. According to this procedure, the least squares estimator of d is

$$\hat{d}^{GPH} = -0.5 \sum_{j=1}^{n} (x_j - \bar{x}) \log I(\omega_j) \Bigg/ \sum_{j=1}^{n} (x_j - \bar{x})^2 \qquad \#$$

where n is the number of frequencies used, i.e., $\omega_j \leq \omega_n$ with ω_n small, $x_j = \log\left|1 - e^{-i\omega_j}\right| = \log|2\sin(\omega_j/2)| = \log|2\sin(\pi j/N)|$ and $I(\omega_j)$ is the periodogram of the time series $Z_1, \ldots, Z_N$. Based on simulation results, Geweke and Porter-Hudak (1983) suggest using a truncation point $n = [N^\alpha]$, where α is a value around 0.5 or 0.6 and $[s]$ represents the integer part of s. Hurvich, Deo and Brodsky (1998) proved that, under general conditions, $\hat{d}^{GPH}$ is asymptotically normal:

$$\sqrt{n}\left(\hat{d}^{GPH} - d\right) \xrightarrow{L} N(0, \pi^2/24) \qquad \#$$

where $\xrightarrow{L}$ means convergence in distribution. In order to test a time series for long memory, the null hypothesis is $d = 0$, i.e., short memory or inexistence of long memory. Rejection of this hypothesis leads to the conclusion of long memory in the time series, i.e., $d \neq 0$, and thus to the need of an ARFIMA process to model it. Based on result (ref: lm5), these hypotheses can easily be tested with the GPH estimator.

Because of the process of data collection or the practice of researchers, time series are frequently obtained through temporal aggregation. For example, the series commonly used in analysis and modeling are quarterly or annual totals. As a result, the series used in testing for long memory are often time series aggregates. In this paper, we study the effects of the use of aggregate time series on the GPH test.

2. Temporal Aggregation of Long-Memory Models

Suppose that the analyzed time series Y_T is the m-period nonoverlapping aggregates of Z_t, $Y_T = \sum_{t=m(T-1)+1}^{mT} Z_t = \sum_{j=0}^{m-1} B^j Z_{mT}$,where m is fixed and is called the order of aggregation and T is the aggregate time unit. In general, the time series Z_t will be called a basic series and Y_T an aggregate series. The number of observations of Y_T is $N_A = N/m$, where N is the length of the basic series. Tschernig (1995) derived the aggregate model of an ARFIMA$(2, d, 0)$ and for the general ARFIMA(p, d, q) model (ref: lm3), we conclude that the aggregate process $Y_T = (1 + B + \ldots + B^{m-1})Z_{mT}$ is an ARFIMA(p, d, ∞):

$$\Phi(\mathsf{B})(1 - \mathsf{B})^d Y_T = \Theta(\mathsf{B})\varepsilon_T \qquad \#$$

where B is the backshift operator on the aggregate time unit, $\mathsf{B}Y_T = Y_{T-1}$, $\Phi(\mathsf{B}) = 1 - \Phi_1\mathsf{B} - \ldots - \Phi_p\mathsf{B}^p = \prod_{j=1}^{p}\left(1 - \delta_j^m\mathsf{B}\right)$ is the autoregressive operator and δ_j^{-1} are the p roots of the AR polynomial $\phi(B)$ of model (ref: lm3), $\Theta(\mathsf{B}) = 1 - \Theta_1\mathsf{B} - \ldots$ is the infinite moving average operator and ε_T is a sequence of i.i.d. random variables with zero mean and variance σ_ε^2. The parameters Θ_j $(j = 1, \ldots)$ and σ_ε^2 are functions of the parameters ϕ_j $(j = 1, \ldots, p)$, θ_j $(j = 1, \ldots, q)$, σ_a^2 and d (a thorough discussion of aggregation of ARIMA models can be found in Wei, 1990). Model (ref: lm6) also shows that the fractional differencing parameter d remains the same.

3. Testing with Aggregate Series

Since $\Theta(\mathsf{B})$ in model (ref: lm6) is a polynomial of infinite order, we first need to check whether $\Theta(\mathsf{B})\varepsilon_T$ converges in mean square (the convergence of infinite sums of random variables is to be understood in the mean square sense). Testing for long memory with aggregate time series is crucially dependent on this convergence.

The binomial expansion of $(1 + B)^d$ is, for d any real number, $(1 + B)^d = \sum_{j=0}^{\infty} \eta_j B^j$, where

$$\binom{\mathrm{d}}{\mathrm{j}} = \frac{d(d-1)(d-2)\ldots(d-j+1)}{j(j-1)(j-2)\ldots1} = \frac{\Gamma(d+1)}{\Gamma(j+1)\Gamma(d-j+1)} = \eta_j \qquad \#$$

and $\Gamma(x)$ is the Gamma function. We now give the asymptotic representation of the coefficients η_j above.

Proposition *Let displaystyle* $\eta_j = \frac{\Gamma(d+1)}{\Gamma(j+1)\Gamma(d-j+1)}$ *for* $j = 0,1,2\ldots$ *We have displaystyle* $\eta_j \sim \frac{\Gamma(d+1)}{2\pi}(-1)^{j-d-1/2}j^{-(d+1)}$ *mbox* $j \to \infty$.

The above asymptotic representation of the η_j is useful to study the convergence in mean square of the expansion of $(1+B+\ldots+B^{m-1})^d a_t$, which is done in the following theorem.

Theorem *Let* $(1+B+\ldots+B^{m-1})^d$ *be a polynomial in the backshift operator B, where m is a positive integer. Then, we have*

$$(1+B+\ldots+B^{m-1})^d = \prod_{j=1}^{m-1}(1+\zeta_j B)^d = \prod_{j=1}^{m-1}\left[\sum_{k=0}^{\infty}\binom{d}{k}\zeta_j^k B^k\right] = \prod_{j=1}^{m-1}\left(\sum_{k=0}^{\infty}\eta_k\zeta_j^k B^k\right)$$

where the coefficients ζ_j *are such that* $\zeta_j = -\varphi_j^{-1}$ *and the* φ_j *are the roots of the polynomial* $(1+B+\ldots+B^{m-1})$, *and displaystyle* $\eta_k = \binom{d}{k}$ *is defined in (ref: lm7). Consequently,* $\prod_{j=1}^{m-1}\left(\sum_{k=0}^{\infty}\left|\eta_k\zeta_j^k\right|^2\right) < \infty$ *mbox* $d > -0.5$, *where* $|x|$ *represents the absolute value of* x.

Thus, we conclude that the expansion of $(1+B+\ldots+B^{m-1})^d$ is square summable. Consequently, the same applies to the moving average operator of the aggregate model (ref: lm6), i.e., $\sum_{j=0}^{\infty}\Theta_j^2 < \infty$ mbox $d > -0.5$ and therefore, since we assume that the roots of the autoregressive operator of model (ref: lm6) are outside the unit circle, this model is stationary.

Since the parameter d remains unchanged by aggregation, Theorem ref: lmagth1 implies that the test statistic retains its asymptotic properties for aggregate series and therefore testing for long memory can also be based on aggregate data. However, we can expect aggregation to produce significant effects on the test for finite samples and especially on its power. To assess this effect, we conducted a simulation experiment where $1,000$ time series of size $N = 3,600$ were generated from several ARFIMA models with $d = \pm 0.45, \pm 0.4, \pm 0.3, \pm 0.2, \pm 0.1$. The generated series were subsequently aggregated with orders $m = 2,3,4,6,8,12$. Following common practice, the number of frequencies used for the GPH estimator was selected as $n = [N^{0.5}]$ for the basic series and $n_A = [N_A^{0.5}]$ in the aggregate case, where $[x]$ represents the integer part of x. As an example, the power obtained from the simulation for a 5% significance level ! is shown in Table ref: tbllmag1 for $d = \pm 0.3$ and for the models AR FIMA$(1,d,0)$ model with $\phi = 0.6$ and ARFIMA$(0,d,1)$ with $\theta = 0.6$ and for $m = 1,2,3,12$, where $m = 1$ represents the basic series.

We conclude that aggregation can have very serious consequences on the power of the test, as can be seen by the low power obtained for the higher values of m. Therefore, testing for long memory with an aggregate series may lead to the wrong conclusion that the data was generated by a short-memory process and consequently to the consideration of an ARMA process instead of an ARFIMA process to model the series. Although not shown here, we obtained the same type of results with another commonly used long-memory test, namely, the R/S test (Hurst, 1951; Lo, 1991).

Aggregation

Table

REFERENCES

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

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

Hurvich, C. M., Deo, R. and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak' s estimator of the memory parameter of a long-memory time series, *Journal of Time Series Analysis*, **19**, 19-46.

Lo, A. W. (1991). Long-term memory in stock market prices, *Econometrica*, **59**, 1279-1313.

Tschernig, R. (1995). Long memory in foreign exchange rates revisited, *Journal of International Financial Markets, Institutions and Money*, **5**, 53-78.

Wei, W. W. S. (1990). *Time Series Analysis - Univariate and Multivariate Methods*, Addison-Wesley, New York.

RÉSUMÉ

Il est fréquemment necessaire de tester une série temporelle pour mémoire longue avec des séries temporelles aggrégées. Donc, il est important de connaitre les consequences de ce probléme. Pour cet effet, on utilize le teste de Geweke et Porter-Hudak (1983). La componente de moyennes mobiles du modèle aggrégé est infinite, mais elle est convergente en moyenne quadratique, ce que permet de faire le test avec des séries aggrégées. Pourtant, l' aggrégation temporélle a un effet négatif sur la puissance du teste.