View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Facultad de Ciencias Económicas y Empresariales

Working Paper nº 06/09

Time series modelling of sunspot numbers using long range cyclical dependence

Luis A. Gil-Alana
Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Time series modelling of sunspot numbers using long range cyclical dependence
Luis A. Gil-Alana

Working Paper No.06/09
October 2009

ABSTRACT

This paper deals with the analysis of the monthly structure of sunspot numbers using a new technique based on cyclical long range dependence. The results show that sunspot numbers have a periodicity of 130 months, but more importantly, that the series is highly persistent, with an order of cyclical fractional integration slightly above 0.30. That means that the series displays long memory, with a large degree of dependence between the observations that tends to disappear very slowly in time.

Luis A. Gil-Alana
Universidad de Navarra
Depto. Economía
Campus Universitario
31080 Pamplona
alana@unav.es

1. Introduction

Solar magnetic disturbances manifest as dark spots on the surface of the sun which are visible from Earth. These disturbances, usually called sunspots have been observed since ancient times. It was in the XIXth century when Schwabe (1843) collected them in terms of a time series. He collected 17 years of sunspot observations and his results revealed a 10-year periodicity in the data. In 1848, Rudolf Wolf devised a daily method of estimating solar activity by counting the number of individual spots and groups of spots on the face of the sun. He chose to compute his sunspot number by adding 10 times the number of groups to the total count of individual spots, because neither quantity alone completely captured the level of activity.[1] Today, the sunspot number is more a smoothed number based on the weighted average of measurements made from a network of observatories. This ensures that the differences in observations due to location, weather, observer and other factors do not affect the sunspot number.

Determining the sunspot cycle period is important among other things in order to compare the period estimate with weather cycles. Thus, cooling and warming of the Earth might be due to the changes in the number of observed sunspots (Linström et al., 1996; Ballester and Oliver, 1999; Olvera, 2007; etc.). From the historical data available to Wolf, he estimated a cycle period of above 11.1 years/cycle. This result was also confirmed by Schuster (1906) who employed techniques based on the periodogram and found an estimate of 11.125 years/cycle. Nowadays, it is widely accepted that the number of sunspots fluctuate with apparently regular intervals, period length averaging 10-11 years (Waldmeier, 1961). Nevertheless, the modelling of the time series of sunspots is still an open issue. According to Aguirre et al. (2008), three are two main practical

[1] See Morris (1977) for the definition and more information about the Wolf sunspot numbers.

difficulties concerning this series: one, the apparent nonstationary nature of the series, and two, the complex dynamics underlying the fluctuations in the cycle amplitude. In this paper we present a new time series approach that deals with the two above-mentioned issues by using fractionally integrated models in the context of cyclical structures.

The outline of the article is as follows: Section 2 presents the methodology employed in the paper that is based on long memory processes. Section 3 describes the data and reports the empirical results, while Section 4 contains some concluding comments.

2. Methodology

From a time series viewpoint, a process is said to be covariance or second order stationary if the mean and the variance do not depend on time and the covariance between any two observations depends on the distance between them but not on their specific locations in time. Then, given a covariance stationary process $\{x_t, t = 0, \pm 1, \dots\}$, with autocovariance function $E[(x_t - Ex_t)(x_{t-j} - Ex_t)] = \gamma_j$, according to McLeod and Hipel (1978), x_t displays the property of long memory if

$$\lim_{T \to \infty} \sum_{j=-T}^{T} |\gamma_j|$$

is infinite. An alternative definition, based on the frequency domain is as follows. Suppose that x_t has an absolutely continuous spectral distribution, so that it has a spectral density function, denoted by $f(\lambda)$, and defined as

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j \cos \lambda j, \qquad -\pi < \lambda \leq \pi.$$

Then, x_t displays long memory if the spectral density function has a pole at some frequency λ in the interval $[0, \pi]$. Most of the empirical literature has concentrated on the case where the singularity or pole in the spectrum occurs at the zero frequency. This is the case of the standard fractionally integrated or I(d) models of the form:

$$(1 - L)^d x_t = u_t, \quad t = 0, \pm 1, ..., \qquad (1)$$

with $x_t = 0$, $t \leq 0$, and where L is the lag operator ($Lx_t = x_{t-1}$), d is a positive real value, and u_t is an I(0) process defined as a covariance stationary process with spectral density function that is positive and bounded at all frequencies.[2]

The I(d) model in (1) has been widely employed in the analysis of meteorological time series, in particular, examining the warming in temperatures under the assumption that the residuals from a linear time trend follow an I(d) process (see, e.g., Bloomfield, 1992; Smith, 1993; Lewis and Ray, 1997; Pethkar and Selvam, 1997; Koscielny-Bunde et al., 1998, Pelletier and Turcotte, 1999; Percival et al., 2004; Maraun et al., 2004, and Gil-Alana, 2003, 2005). All these works were based on the observation that most of the time series examined by these authors presented a typical shape with the spectral density increasing dramatically as the frequency approaches zero and that differencing the data leads to overdifferencing at the zero frequency.

However, a process may also display a pole or singularity in the spectrum at a frequency away from zero. In this case, the process may still display the property of long memory but the autocorrelations present a cyclical structure that is decaying very slowly. This is the case of the Gegenbauer processes defined as:

$$(1 - 2\cos w_r L + L^2)^d x_t = u_t, \quad t = 1, 2, ..., \qquad (2)$$

[2] The I(0) class of models include the classical white noise process but also other structures allowing a weak dependence structure like the stationary autoregressive moving average (ARMA) models.

where w_r and d are real values, and u_t is I(0). For practical purposes we define $w_r = 2\pi r/T$, with r = T/s and thus, s will indicate the number of time periods per cycle, while r refers to the frequency that present a pole or singularity in the spectrum of x_t. Note that if r = 0 (or s = 1), the fractional polynomial in (2) becomes $(1 - L)^{2d}$, which is the polynomial associated to the common case of fractional integration at the long run or zero frequency. This type of process was introduced by Andel (1986) and subsequently analyzed by Gray, Zhang and Woodward (1989, 1994), Chung (1996a,b) and Dalla and Hidalgo (2005) among others.

Gray et al. (1989, 1994) showed that the polynomial in (2) can be expressed in terms of the Gegenbauer polynomial, such that, calling $\mu = \cos w_r$, for all $d \neq 0$,

$$(1 - 2\mu L + L^2)^{-d} = \sum_{j=0}^{\infty} C_{j,d}(\mu) L^j, \qquad (3)$$

where $C_{j,d}(\mu)$ are orthogonal Gegenbauer polynomial coefficients recursively defined as:

$$C_{0,d}(\mu) = 1,$$

$$C_{1,d}(\mu) = 2\mu d,$$

$$C_{j,d}(\mu) = 2\mu\left(\frac{d-1}{j}+1\right)C_{j-1,d}(\mu) - \left(2\frac{d-1}{j}+1\right)C_{j-2,d}(\mu), \quad j = 2, 3,....$$

(See, for instance, Magnus et al., 1966, Rainville, 1960; etc. for further details on Gegenbauer polynomials). Gray et al. (1989) showed that x_t in (2) is (covariance) stationary if $d < 0.5$ for $|\mu = \cos w_r| < 1$ and if $d < 0.25$ for $|\mu| = 1$.[3] This model has

[3] Note that if $|\mu| < 1$ and d in (2) increases beyond 0.5, the process becomes "more nonstationary" in the sense, for example, that the variance of the partial sums increases in magnitude.

been employed to analyze the sunspots number time series by several authors, including Gray et al. (1989), Hsu and Tsai (2009) and others.

The model just presented can be generalized to the case of more than one cyclical structure and we can consider processes of the form:

$$\prod_{j=1}^{k}(1 - 2\cos w_r^{(j)}L + L^2)^{d_j} y_t = u_t, \quad t = 1, 2, ..., \qquad (4)$$

where k is a finite integer indicating the maximum number of cyclical structures; and $w_r^{(j)} = 2\pi / s_{(j)}$ where $s_{(j)}$ indicates the number of time periods per cycle corresponding to the j^{th} cyclical structure. Empirical works based on multiple cyclical structures of this form (also named k-factor Gegenbauer processes) can be found in Ferrara and Guegan (2001), Sadek and Khotanzad (2004) and Gil-Alana (2007).

3. Data and empirical results

In this section we employ monthly data of sunspot numbers obtained from the Solar Influences Data Analysis Center (SIDC: http://www.sidc.be/sunspot-data). The data run from 1749m1 to 2008m12 and they are displayed in Figure 1.

[Insert Figures 1 – 3 about here]

We notice a cyclical structure in the data that is clearer when we observe the first 500 sample autocorrelation values, displayed in the correlogram in Figure 2. We note a significant cyclical pattern that is decaying very slowly, which might be consistent with a cyclical fractional model of the form represented by equation (2). Moreover, the periodogram, displayed in Figure 3 presents a large peak at a frequency away from zero, giving further support in favour of a cyclical I(d) model. Also, the periodogram, though it is not a consistent estimator of the spectral density function, it is an asymptotically

unbiased estimator of it and, evaluated at the discrete Fourier frequencies $\lambda_j = 2\pi j/T$, j = 1, 2, …, T/2, can give us an indication about the length of the cycles. In our case, we observe in Figure 3 that the highest peak takes place at $\lambda_j = 2\pi 24/T$, which implies that the cycles have a periodicity of T/24 = 3120/24 = 130 periods (=months)/cycle.

Based on this evidence, we estimate the following model,

$$y_t = \mu + x_t, \qquad t = 1, 2, ..., \tag{4}$$

$$(1 - 2\cos w_r L + L^2)^d x_t = u_t, \qquad t = 1, 2, ..., \tag{5}$$

where y_t is the observable time series (sunspot numbers), μ is an intercept, and u_t is an I(0) process that is specified in terms of a white noise process and using an AR(1) structure. Higher AR orders were also considered though several Likelihood Ratio (LR) tests conducted in previous versions of this work concluded that the AR(1) specification was sufficient to capture the short run dynamics underlying the series.

We employ here a procedure developed by Robinson (1994) that essentially tests the null hypothesis:

$$H_o : d = d_o, \tag{6}$$

in (4) and (5) for any given real value d_o, assuming that w_r is known and a given parametric structure for u_t. This method is briefly described in the Appendix and present several advantages compared with other procedures. First, it allows us to test any real value d_o, encompassing thus models with different orders of integration. Second, the limit distribution is standard N(0,1), and this standard limit behaviour holds independently of the inclusion or not of deterministic terms (like an intercept) and of the way of modeling the I(0) error term. Moreover, assuming Gaussianity on u_t, this method is found to be the most efficient one in the context of fractional integration.

[Insert Table 1 about here]

The approach employed here tests H_o (6) in (4) and (5) for d_o-values equal to 0 to 2 with 0.001 increments, assuming that w_r in (4) is equal to $2\pi j/T$, with $j = T/s$, and $s = 110, 111, \ldots, 150$.[4] The estimated values are then chosen as the values that produce the lowest statistics across w_r and d_o. In Table 1 we present the results for the two cases of $\mu = 0$ a priori in (4) and μ unknown, assuming that u_t is white noise and AR(1). The first thing to note is that for the four cases examined the lowest statistics take place at $r = T/130$, which is consistent with the plot of the periodogram in Figure 3. If we focus on the estimators, we observe that the results are very similar for the four cases, with the values of d ranging from 0.311 (white noise u_t with an intercept) to 0.336 (AR(1) with u_t with no intercept). Looking at the 95% confidence interval for the non-rejection values we observe that all them are in the range (0, 0.5) implying a low degree of cyclical long range dependence behaviour. Performing several tests on the estimated residuals we found evidence in favour of the case of AR(1) disturbances.[5]

Next we wonder if the cyclical fractional differencing parameter has remained constant across the sample period. For this purpose we estimate the model in (4) and (5) for different subsamples of 720 observations corresponding each, to 60 complete years, starting from the subsample 1749M1 – 1808M12, and adding estimates moving forward the sample five consecutive years. The results for the two cases of white noise and AR(1) disturbances are displayed in Figure 4.

[Insert Figure 4 about here]

[4] We choose these values noting that the highest peak in the periodogram takes place at $j = 24$, implying cycles of periodicity equal to 130 periods.
[5] We use here Box-Pierce and Ljung-Box-Pierce statistics (Box and Pierce, 1970; Ljung and Box, 1978).

In both cases we observe the same pattern: the estimates are relatively stable in the first subsamples, with values above 0.3; then, there is a reduction in the degree of integration in the mid-subsamples, increasing again in the final part of the sample. In any case, all values (including the confidence bands) are in the range (0.2, 0.4) implying a certain degree of stability in the results.

Next we focus on yearly data. First, we present the results for the yearly average data based on the monthly observations. A plot of the time series and its corresponding correlogram and periodogram are displayed respectively in Figures 5 – 7. The periodogram presents its highest value at 24, and given that now T = 260, then 260/24 = 10.83 periods (= years)/cycle. Because of this, we report in Tables 2 and 3 respectively the estimated coefficients with r = T/10 and T/11.

[Insert Tables 2 and 3 about here]

The results are rather similar in the two cases. The series presents an order of integration above 0.5 if the disturbances are white noise, while it is slightly below 0.4 if the they are autocorrelated. Thus, the results seem to be very sensitive to the choice of the short run dynamics underlying the series. The larger values observed in the uncorrelated case may be due to the fact that the fractional differencing parameter is in this case the only parameter used to describe the dependence across the data. Also, the values reported in these two tables are also higher than those given in Table 1 and based on the monthly observations. This may be explained by the fact that the annual data are averaged values of the monthly observations, and aggregation has been the usual argument claimed to justify fractional integration (Robinson, 1978; Granger, 1980). Using LR tests and other statistics the evidence point out in favour of the autocorrelated cases.

Finally, in Tables 4 and 5 we want to examine if the degree of dependence may be affected by the monthly structure of the data. Therefore, we compute the estimates of d, yearly, for each month of the year, again for the two frequencies and for the two cases of white noise and AR(1) u_t.

[Insert Tables 4 and 5 about here]

The first thing we observe in these two tables is that the estimated values of d are in all cases smaller than those reported in Tables 2 and 3 for the yearly averaged data, which is consistent with the argument of aggregation mentioned in the preceding paragraph. Starting with the white noise case (in Table 4) we see that all the values are in the range (0.39, 0.53) and we also note consistently higher values at r = T/10 than at T/11. If r = T/10, the highest estimate of d corresponds to April (0.532) followed by October (0.524) and November (0.503), while the lowest value is obtained in February (0.396). The same pattern follows with r = T/11. If we focus now on the case of AR(1) u_t (in Table 5) the values are substantially smaller, ranging now between 0.206 (June, r = T/10) and 0.265 (July, r = T/10). Similarly to the previous cases, we also perform various statistics to determine which of the two specifications (white noise and AR(1)) was more appropriate to describe the short run dynamics and the evidence pointed out in favour of the autocorrelated cases.

4. Concluding comments

In this paper we have examined the time series properties of the sunspot numbers, monthly, from 1749m1 to 2008m12. For this purpose we have employed a technique based on cyclical long range dependence. Using this technique we have shown that sunspot numbers have a periodicity of 130 months, and more importantly, that the series

is highly persistent, with an order of cyclical fractional integration slightly above 0.30. That means that the series displays long memory, with a degree of dependence between the observations that tends to disappear very slowly in time.

Prediction of the sunspot numbers is a natural following-up step in this work. Though various numerical prediction techniques have been used for the sunspot number time series (curve fitting, artificial intelligence, neural networks, EMD analysis, etc.), these approaches although very accurate in short-term predictions are rather unreliable in long term (Gholipour et al.,2003, Xu et al., 2008). In this context, the use of cyclical long range dependence techniques can provide better predictions in the long run.

The results presented in this work can also be used as a first step in the specification of a multivariate model including other variables such as temperatures which might be well described in terms of a cyclically fractionally integrated model (see, e.g., Gil-Alana, 2009). In this context, the concept and ideas of cyclical fractional cointegration can be a fruitful avenue for further research in the future.

Acknowledgments

The author gratefully acknowledges financial support from the Ministerio de Ciencia y Tecnologia (ECO2008-03035, Spain). Comments of an anonymous referee are also gratefully acknowledged. The usual disclaimer applies.

Appendix

Robinson (1994) proposes a Lagrange Multiplier (LM) test of the null hypothesis H_o (6) in (4) and (5) for any real value d_o. The test statistic is given by:

$$\hat{s} = \left(\frac{T}{\hat{A}}\right)^{1/2} \frac{\hat{a}}{\hat{\sigma}^2},$$

where T is the sample size, and

$$\hat{a} = \frac{-2\pi}{T} \sum_{j=1}^{*} \psi(\lambda_j)\, g(\lambda_j;\hat{\tau})^{-1} I(\lambda_j); \qquad \hat{\sigma}^2 = \sigma^2(\hat{\tau}) = \frac{2\pi}{T} \sum_{j=1}^{T-1} g(\lambda_j;\hat{\tau})^{-1} I(\lambda_j);$$

$$\hat{A} = \frac{2}{T}\left(\sum_{j=1}^{*} \psi(\lambda_j)^2 - \sum_{j=1}^{*} \psi(\lambda_j)\hat{\varepsilon}(\lambda_j)' \times \left(\sum_{j=1}^{*} \hat{\varepsilon}(\lambda_j)\hat{\varepsilon}(\lambda_j)' \right)^{-1} \times \sum_{j=1}^{*} \hat{\varepsilon}(\lambda_j)\psi(\lambda_j) \right)$$

$$\psi(\lambda_j) = \log\left| 2\left(\cos\lambda_j - \cos w_r\right)\right|; \qquad \hat{\varepsilon}(\lambda_j) = \frac{\partial}{\partial \tau} \log g(\lambda_j;\hat{\tau}),$$

$I(\lambda_j)$ is the periodogram of $\hat{u}_t = (1 - 2\cos w_r L + L^2)^{d_o} y_t - \hat{\mu}'\bar{z}_t$, with

$$\hat{\mu} = \left(\sum_{t=1}^{T} \bar{z}_t\, \bar{z}_t' \right)^{-1} \sum_{t=1}^{T} \bar{z}_t (1 - 2\cos w_r L + L^2)^{d_o} y_t; \quad \bar{z}_t = (1 - 2\cos w_r L + L^2)^{d_o} 1_t,$$

evaluated at $\lambda_j = 2\pi j/T$ and g is a known function coming from the spectral density function of $\hat{u}_t$: $f(\lambda;\tau) = \frac{\sigma^2}{2\pi} g(\lambda;\tau)$, with $\hat{\tau}$ obtained by minimizing $\sigma^2(\tau)$. Note that the test is purely parametric and, therefore, it requires specific modelling assumptions regarding the short memory specification of u_t. Thus, for example, if u_t is white noise, g $\equiv$ 1 and, if u_t is an AR process of form: $\phi(L)u_t = \varepsilon_t$, $g = |\phi(e^{i\lambda})|^{-2}$, so that the AR coefficients are a function of τ. Finally, the summation on * in the above expressions are over $\lambda \in M$, where $M = \{\lambda: -\pi < \lambda < \pi, \lambda \notin (\rho_u - \lambda_u, \rho_u + \lambda_u)$, such that ρ_u is the distinct pole of $\psi(\lambda)$ on $(-\pi, \pi]$.

References

Aguirre, L.A., C. Letellier and J. Maquet, 2008, Forecasting the time series of sunspot numbers, Solar Physics 249, 103-120.

Andel, J., 1986, Long memory time series models, Kybernetika 22, 105-123.

Ballester, J.L. and R. Oliver, 1999, Discovery of the near 158 day periodicity in group sunspot numbers during the eighteenth century, The Astrophysical Journal 522, 153-156.

Bloomfield, P., 1992, Trends in global temperatures, Climate Change 21, 275-287.

Chung, C.-F., 1996a, A generalized fractionally integrated autoregressive moving-average process, Journal of Time Series Analysis 17, 111-140.

Chung, C.-F., 1996b, Estimating a generalized long memory process, Journal of Econometrics 73, 237-259.

Dalla, V. and J. Hidalgo, 2005, A parametric bootstrap test for cycles, Journal of Econometrics 129, 219-261.

Ferrara, L. and D. Guegan, 2001, Forecasting with k-factor Gegenbauer processes. Theory and Applications. Journal of Forecasting 20, 581-601.

Gholipour, A., C. Lucas, N.A. Babak and M. Shafiee, 2005, Journal of Atmospheric and Solar Terrestrial Physics 67, 595-603.

Gil-Alana, L.A., 2003, An application of fractional integration to a long temperature time series, International Journal of Climatology 23, 1699-1710.

Gil-Alana, L.A., 2005, Statistical model of the temperatures in the northern hemisphere using fractional integration techniques, Journal of Climate 18, 5357-5369.

Gil-Alana, L.A., 2007, Testing the existence of multiple cycles in financial and economic time series. Annals of Economics and Finance 1, 1-20.

Gil-Alana, L.A., 2009, A bivariate fractional cointegration relationship in the context of cyclical structures, Journal of Statistical Computation and Simulation, forthcoming.

Granger, C.W.J., 1980, Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14, 227-238.

Gray, H.L., Zhang, N. and Woodward, W.A., 1989, On generalized fractional processes, Journal of Time Series Analysis 10, 233-257.

Gray, H.L., Zhang, N. and Woodward, W.A., 1994, On generalized fractional processes. A correction, Journal of Time Series Analysis 15, 561-562.

Hsu, N.-J- and H. Tsai, 2009, Semiparametric estimation for seasonal long memory time series using generalized exponential models, Journal of Statistical Planning and Inference, forthcoming.

Koscielny-Bunde, E., A. Bunde, S. Havlin, H.E. Roman, Y. Goldreich and H.J. Schellnhuber, 1998, Indication of a universal persistence law governing atmospheric variability, Physics Review Letters 81, 729-732.

Lewis, P.A.W. and B.K. Ray, 1997, Modelling long-range dependence, nonlinearity and periodic phenomena in sea surface temperatures using TSMARS, Journal of the American Statistical Association 92, 881-893.

Lindstrom, J., H. Kokko and E. Ranta, 1996, There is something new under the sunspots, Oikos 77, 3, 565-586.

Magnus, W., Oberhettinger, F. and R.P. Soni, 1966, Formulas and theorema for the special functions of mathematical physics, Springer, Berlin.

Maraun, D., H.W. Rust and H. Timmer, 2004, Tempting long memory on the interpretation of DFA results, Nonlinear Processes in Geophysics 11, 495-503.

McLeod, A.I. and K.W. Hipel, 1978 Preservation of the rescaled adjusted range. A reassessment of the Jurst phenomenon, Water Resources Research 14, 491-507.

Morris, M.J., 1977, Forecasting the sunspot cycle (with discussion), Journal of the Royal Statistical Society, Series A 140, 437-468.

Olvera, F.E., 2007, A spectral analysis of the sunspot time series using the periodogram, Statistical Signal Processing ECE538, 1-4.

Pelletier, J. and D. Turcotte, 1999, Self-affine time series II. Aplications and Models, Advances in Geophysics 40, 91-166.

Pethkar, J.S. and A.M. Selvam, 1997, Nonlinear dynamics and chaos. Applications for prediction of weather and climate, Proc. TROPMET 97, Bangalare, India.

Percival, D.B., J.E. Overland and H.O. Mofjeld, 2004, Modelling North Pacific climate time series, to appear in Time Series Analysis and Applications to Geophysical Systems, edited by D.R. Brillinger, E.A. Robinson and F.P. Schoenberg, Springer-Verlag.

Rainville, E.D., 1960, Special functions, MacMillan, New York.

Robinson, P.M., 1978, Statistical inference for a random coefficient autoregressive model. Scandinavian Journal of Statistics 5, 163-168.

Robinson, P.M., 1994. Efficient tests of nonstationary hypotheses. Journal of the American Statistical Association 89, 1420-1437.

Sadek, N. and A. Khotanzad, 2004, K-factor Gegenbauer ARMA process for network traffic simulation. Computers and Communications 2, 963-968.

Schuster, A., 1906, On the periodicities of sunspots, Philosophical Transactions of the Royal Society 206, 69-100.

Schwabe, S.H., 1843, Solar observations during 1843, Astronomische Nachrichten 20, 234-235.

Smith, R.L., 1993, Long-range dependence and global warming, Statistics for the Environment (ed. Barnett, V. And Turkman, K.F., pub. J. Wiley), 141-161.

Waldmeier, M., 1961, The sunspot activity in the years 1610-1960, Swiss Federal Observatory, Zurich.

Xu, T., J. Wu, Z.-S. Wu and Q. Li, 2007, Long-term sunspot number prediction based on EMD analysis and AR model, Chinese Journal of Astronomy and Astrophysics 3, 337-342.

Figure 1: Monthly sunspot numbers: 1749M1 – 2008M12

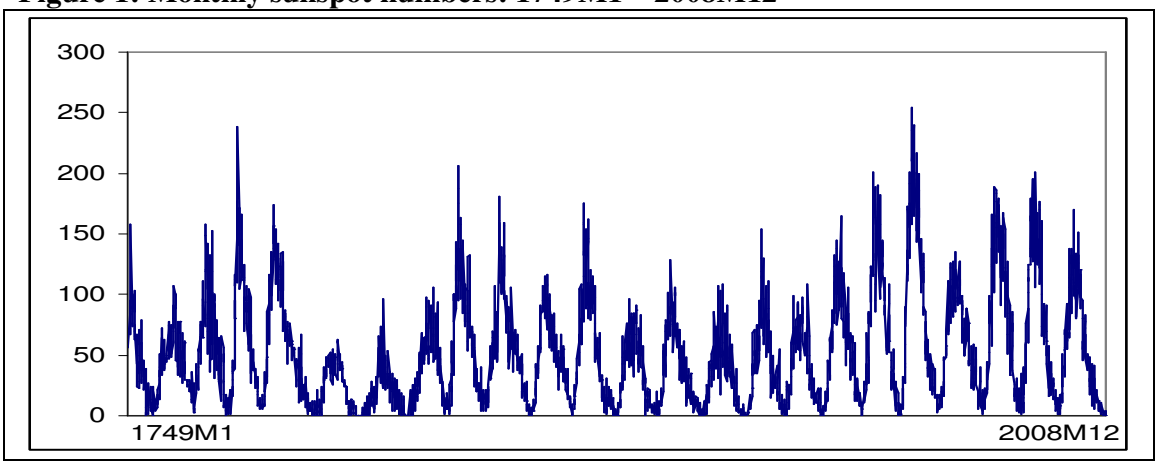
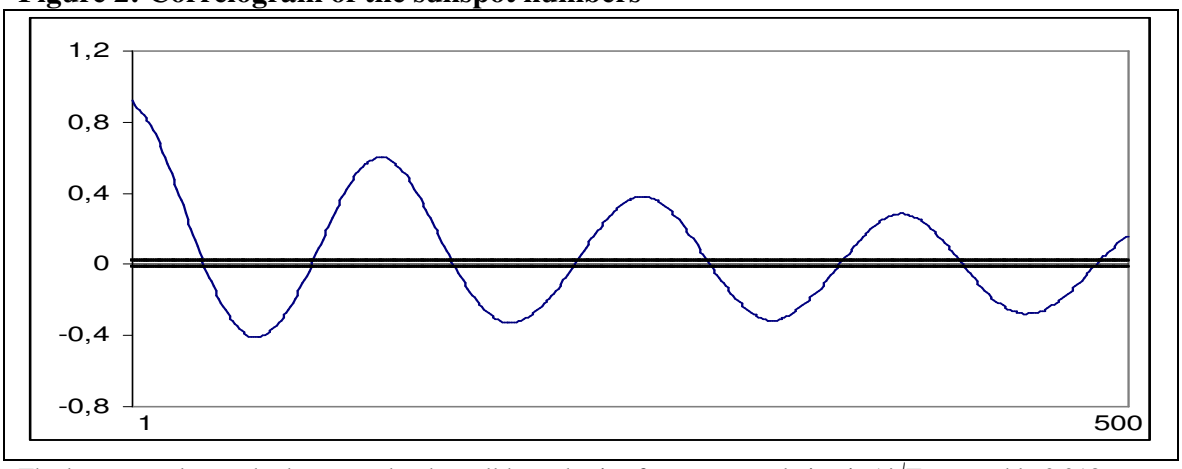
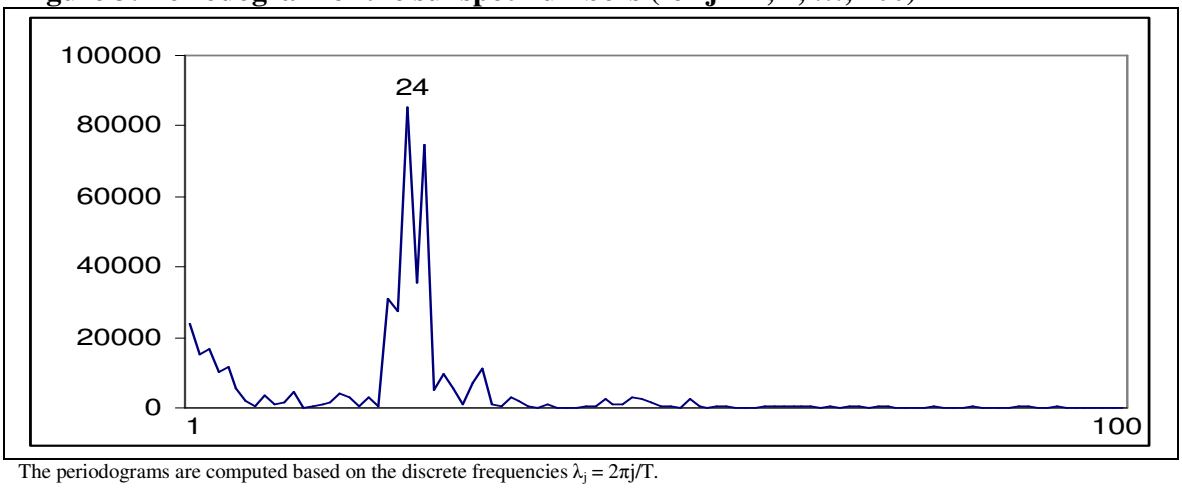


Figure 2: Correlogram of the sunspot numbers



The large sample standard error under the null hypothesis of no autocorrelation is $1/\sqrt{T}$ or roughly 0.018 for the series used in this application.

Figure 3: Periodogram of the sunspot numbers (for j = 1, 2, …, 100)



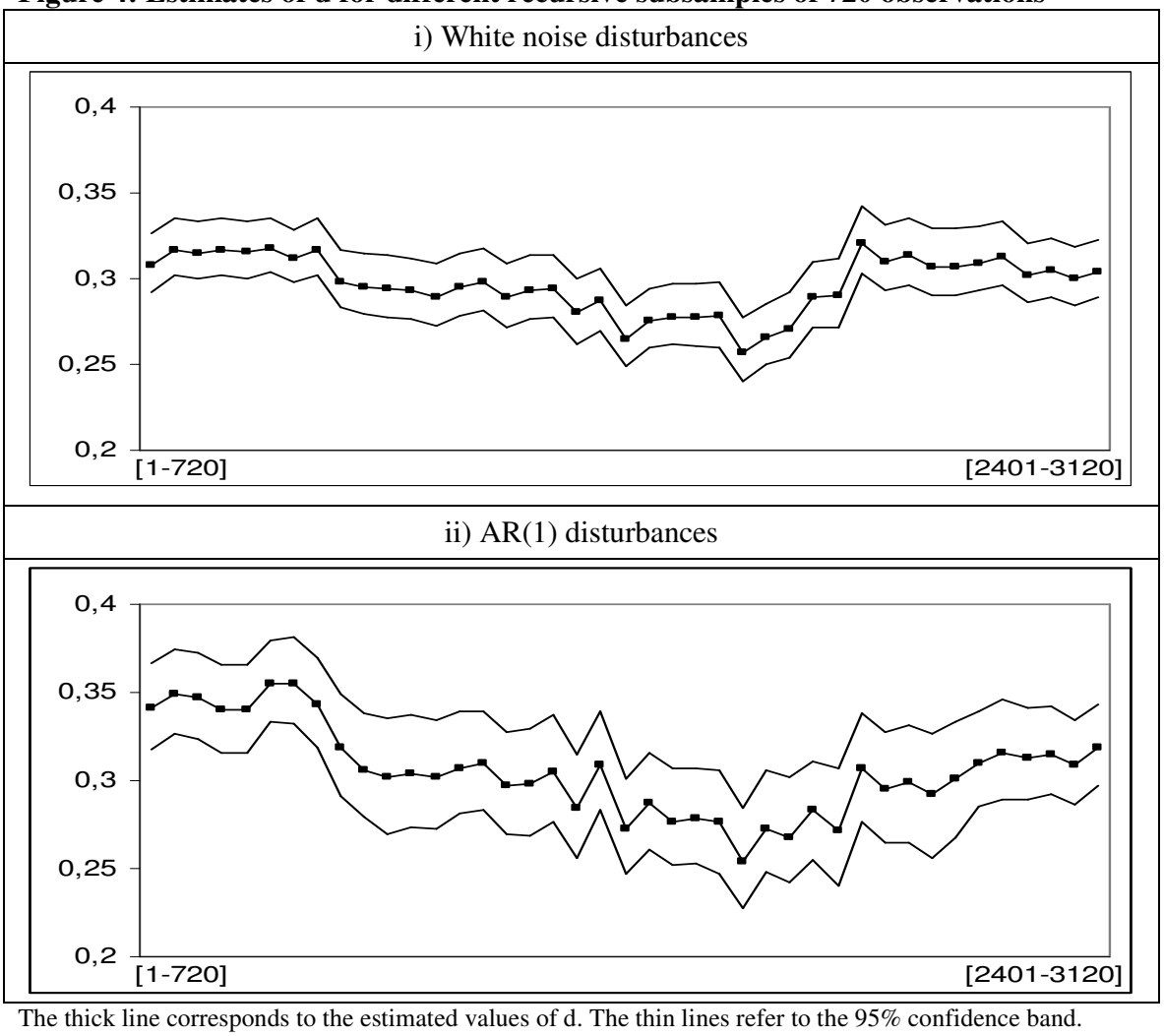
The periodograms are computed based on the discrete frequencies $\lambda_j = 2\pi j/T$.

Table 1: Estimates of the parameters in fractional cyclical models (w_r = T/130)

	No regressors		With an intercept		
	d	AR	d	Intercept	AR
White noise u_t	0.312 (0.304, 0.320)	---	0.311 (0.303, 0.320)	52.38081 (27.987)	---
AR (1) u_t	0.336 (0.325, 0.347)	-0.097	0.335 (0.324, 0.347)	52.37599 (24.277)	-0.095

The values in parenthesis behind the estimates of d refer to the 95% confidence bands. Those in parenthesis behind the estimate of the intercept refer to the t-value.

Figure 4: Estimates of d for different recursive subsamples of 720 observations



The thick line corresponds to the estimated values of d. The thin lines refer to the 95% confidence band.

Figure 4: Yearly average sunspot numbers: 1749 – 2008

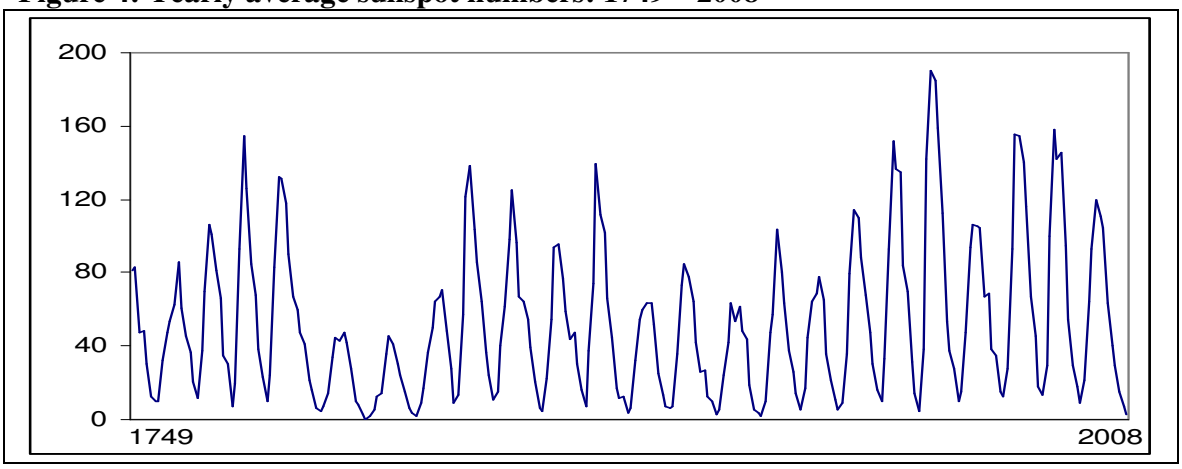
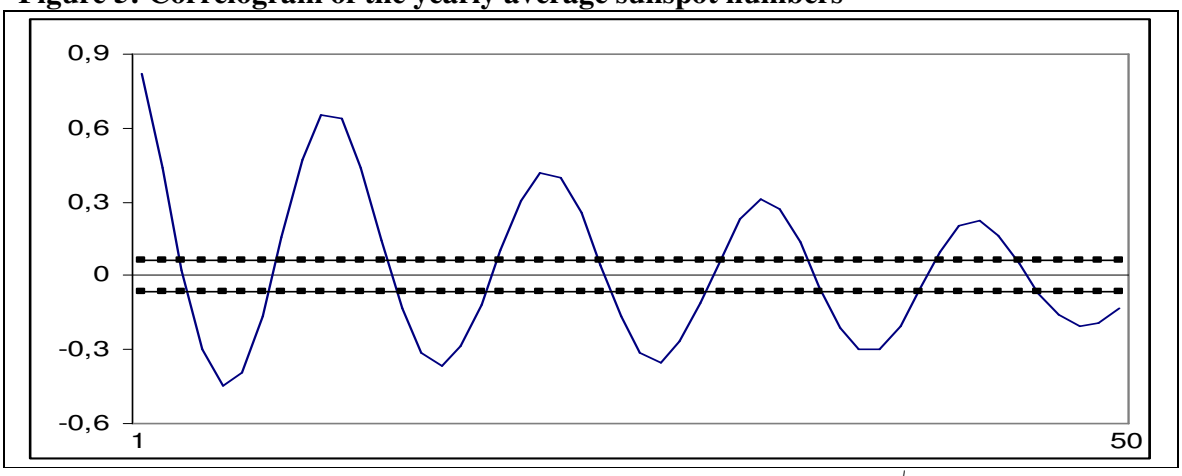
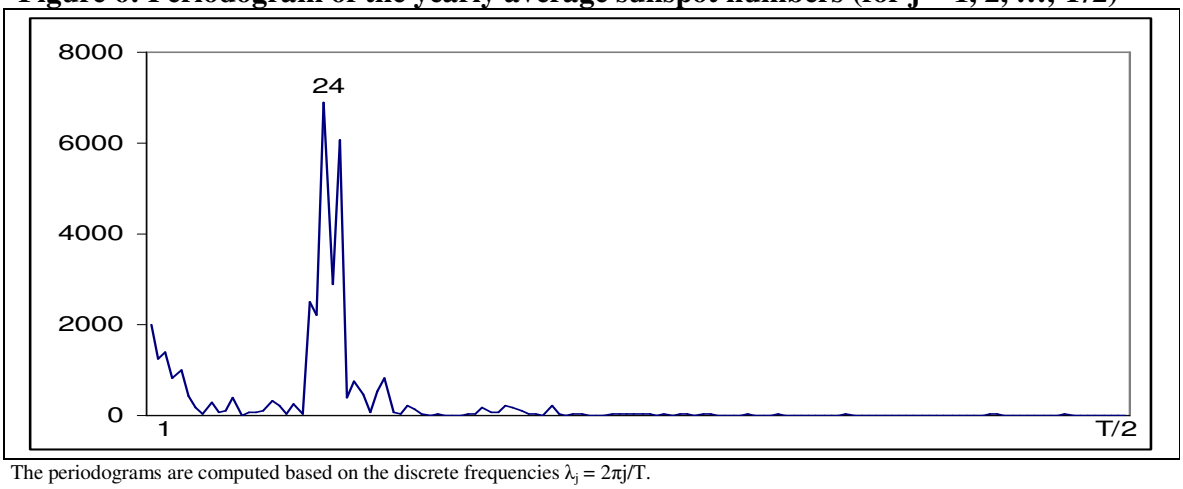


Figure 5: Correlogram of the yearly average sunspot numbers



The large sample standard error under the null hypothesis of no autocorrelation is $1/\sqrt{T}$ or roughly 0.062 for the series used in this application.

Figure 6: Periodogram of the yearly average sunspot numbers (for j = 1, 2, …, T/2)



The periodograms are computed based on the discrete frequencies $\lambda_j = 2\pi j/T$.

Table 2: Estimates of the parameters in fractional cyclical models (r = T/10)

Yearly average	No regressors		With an intercept		
	d	AR	D	Intercept	AR
White noise u_t	0.763 (0.692, 0.839)	---	0.745 (0.676, 0.821)	52.24232 (23.596)	---
AR (1) u_t	0.399 (0.338, 0.476)	0.618	0.395 (0.335, 0.467)	52.15954 (34.342)	0.611

The values in parenthesis behind the estimates of d refer to the 95% confidence bands. Those in parenthesis behind the estimate of the intercept refer to the t-value.

Table 3: Estimates of the parameters in fractional cyclical models (r = T/11)

Yearly average	No regressors		With an intercept		
	d	AR	D	Intercept	AR
White noise u_t	0.719 (0.648, 0.795)	---	0.699 (0.629, 0.775)	52.29603 (22.151)	---
AR (1) u_t	0.385 (0.323, 0.462)	0.582	0.376 (0.319, 0.445)	52.22245 (33.068)	0.585

The values in parenthesis behind the estimates of d refer to the 95% confidence bands. Those in parenthesis behind the estimate of the intercept refer to the t-value.

Table 4: Monthly estimates for the case of white noise disturbances

	r = T/10		r = T/11	
	d	Intercept	d	Intercept
JANUARY	0.499 (0.442, 0.564)	50.16569 (19.213)	0.461 (0.406, 0.525)	50.19367 (18.679)
FEBRUARY	0.396 (0.352, 0.444)	52.13392 (20.237)	0.359 (0.318, 0.406)	52.14836 (20.214)
MARCH	0.476 (0.423, 0.535)	51.01546 (20.825)	0.441 (0.390, 0.499)	51.04373 (20.281)
APRIL	0.532 (0.476, 0.594)	51.76727 (20.537)	0.492 (0.435, 0.555)	51.80709 (19.938)
MAY	0.467 (0.410, 0.531)	52.94759 (19.531)	0.436 (0.381, 0.499)	52.99719 (18.870)
JUNE	0.445 (0.396, 0.500)	52.58568 (20.542)	0.415 (0.367, 0.469)	52.64809 (19.992)
JULY	0.484 (0.434, 0.541)	52.61536 (20.624)	0.444 (0.394, 0.501)	52.66602 (20.197)
AUGUST	0.488 (0.434, 0.547)	52.25256 (19.866)	0.459 (0.407, 0.519)	53.33359 (19.104)
SEPTEMBER	0.499 (0.443, 0.562)	52.77756 (19.555)	0.478 (0.424, 0.538)	52.88548 (18.560)
OCTOBER	0.524 (0.469, 0.584)	53.04521 (20.180)	0.487 (0.431, 0.548)	53.13985 (19.495)
NOVEMBER	0.503 (0.448, 0.565)	51.52662 (19.983)	0.464 (0.409, 0.526)	51.61503 (19.423)
DECEMBER	0.484 (0.427, 0.549)	52.39317 (18.756)	0.445 (0.389, 0.509)	52.50388 (18.350)

Table 5: Monthly estimates for the case of AR(1) disturbances

	r = T/10		r = T/11	
	d	Intercept	d	Intercept
JANUARY	0.249 (0.177, 0.345)	50.14200 (24.497)	0.251 (0.188, 0.334)	50.14523 (23.721)
FEBRUARY	0.242 (0.113, 0.414)	52.09100 (23.062)	0.256 (0.141, 0.342)	52.10237 (22.326)
MARCH	0.234 (0.161, 0.364)	50.93877 (26.176)	0.248 (0.180, 0.403)	50.95422 (25.106)
APRIL	0.259 (0.195, 0.345)	51.72897 (26.903)	0.242 (0.186, 0.317)	51.74057 (26.576)
MAY	0.212 (0.142, 0.300)	52.88657 (24.965)	0.219 (0.159, 0.299)	52.92057 (24.113)
JUNE	0.206 (0.137, 0.363)	52.51120 (25.482)	0.223 (0.158, 0.373)	52.56005 (24.500)
JULY	0.265 (0.191, 0.413)	52.55519 (25.223)	0.247 (0.185, 0.354)	52.59073 (24.943)
AUGUST	0.229 (0.164, 0.321)	53.19266 (25.376)	0.240 (0.181, 0.330)	53.25668 (24.364)
SEPTEMBER	0.211 (0.150, 0.292)	52.659668 (24.364)	0.245 (0.186, 0.330)	52.76300 (24.147)
OCTOBER	0.256 (0.190, 0.352)	52.93225 (26.084)	0.247 (0.189, 0.330)	53.01456 (25.523)
NOVEMBER	0.241 (0.173, 0.342)	51.32525 (25.576)	0.234 (0.175, 0.320)	51.41100 (25.023)
DECEMBER	0.237 (0.163, 0.339)	52.32092 (23.826)	0.239 (0.175, 0.325)	52.4229 (23.177)