应用概率统计 第三十一卷
第二期 2015年4月

Chinese Journal of Applied Probability
and Statistics Vol.31 No.2 Apr. 2015

Analysis of Approximately Periodic Time Series *

WU SHUJIN* ZHU XIAOYU YANG XIAO
(*Department of Statistics and Actuarial Science, School of Finance and Statistics, East China Normal University, Shanghai, 200241*)

Abstract

In this paper, we first define the concept of approximately periodic time series, that is, the length of their periods is not any constant. For example, the sunspot series has a period about 11 years, but the length of its periods is not just 11 years, which is approximately periodic. Then we give a method to extract approximately periodic trend and bring forward a generalized difference operator, which can eliminate not only time trend and periodicity but also approximately periodicity of time series. At last, we take the sunspot data as an example to show the application of generalized difference operator.

Keywords: Time series, approximate periodicity, generalized difference operator.
AMS Subject Classification: 91B30, 91B28.

§1. Introduction

Time series has been used in statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, and communications engineering (see Fan and Yao, 2006). When a nonstationary time series is analyzed, difference operators are very important because it can eliminate both the time trend and periodicity of time series. Usually, n-order difference operator is used to eliminate time trend of time series and k-step difference operator is used to eliminate periodicity of time series (see Tsay, 2002).

In practice, a lot of time series looks like having periodicity. However, the length of their periods is not any constant. For example, Figure 1 is the sequential chart of sunspot data during twentieth century. From Figure 1, the sunspot data during twentieth century looks like having periodicity and its period is about 11 years, but the distance of adjacent two peaks is not always 11 years. In fact, during the twentieth century the sunspot peak

*The research was supported by the National Natural Science Foundation of China (11471120) and the 111 Project (B14019).
*Corresponding author, E-mail: sjwu@stat.ecnu.edu.cn.
Received November 5, 2014.
doi: 10.3969/j.issn.1001-4268.2015.02.009

years include 1905, 1917, 1928, 1937, 1947, 1957, 1968, 1979, 1989 and 2000. The distances of adjacent two peaks are 12, 11, 9, 10, 10, 11, 11, 10 and 11. We call such time series as a time series with approximate periodicity. The concept of approximate periodicity will be given in the following section.

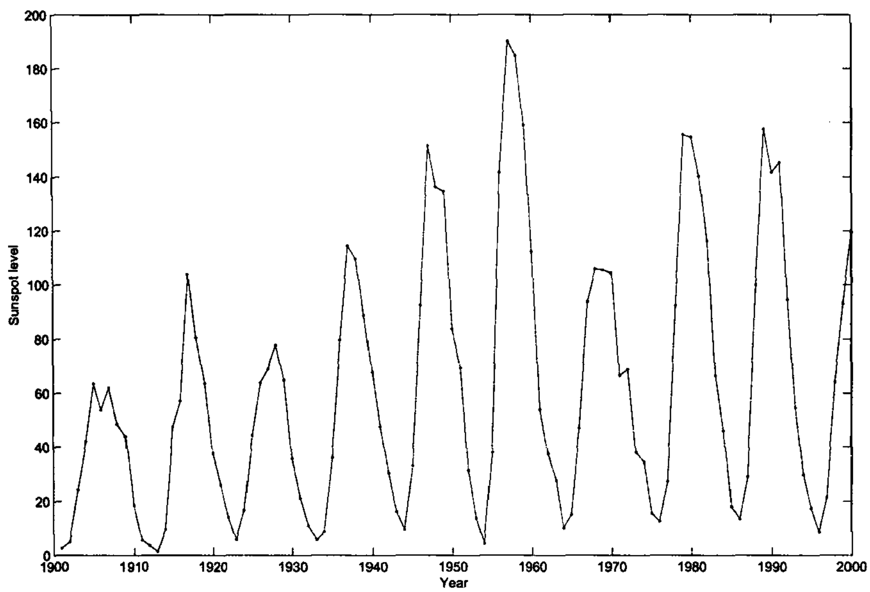

Figure 1 Sequential chart of sunspot data during twentieth century

We know difference operator can eliminate periodicity of time series, but it cannot eliminate approximate periodicity. In this paper, we first define the concept of approximately periodic time series, then give a method to extract approximately periodic trend and bring forward a generalized difference operator, which can eliminate not only time trend and periodicity but also approximately periodicity of time series. At last, take sunspot data as an example to show the application of generalized difference operator.

§2. Approximately Periodic Time Series

In this section, we mainly define the concept of approximately periodic time series. To be brief, approximately periodic time series is a time series which seasonal trend is an approximately periodic function. Before presenting the concept of approximately periodic time series, we give the concept of approximately periodic function.

2.1 Approximately Periodic Function

In this subsection, we first depict the concept of approximately periodic function, then give a proposition to show the relation of approximately periodic function and periodic function.

Definition 2.1 Let $\{f(t), t \geq 0\}$ be a real-valued function. If there exit a strictly increasing sequence $\{T_k | T_0 = 0, \lim_{k\to+\infty} T_k = +\infty\}$ and a strictly increasing continuous function $\{g(t), t \geq T_1\}$ satisfying $g(T_k) = T_{k-1}$ for all $k = 1, 2, \ldots$, such that for any $t \geq T_1$ it follows that

$$f(t) = f(g(t)),$$

then $f(t)$ is called an approximately periodic function with scale transformation g.

Remark 1 Any periodic function is obviously an approximately periodic function.

In fact, for any periodic function f with period T, denote $T_k = kT$ for all $k = 0, 1, 2, \ldots$ and $g(t) = t - T$ for all $t \geq T$, then it is easy to check that f is an approximately periodic function.

Proposition 2.1 $\{f(t), t \geq 0\}$ is an approximately periodic function if and only if there exists a strictly increasing continuous function $\{u(t), t \geq 0\}$ satisfying $u(0) = 0$ and $\lim_{t\to+\infty} u(t) = +\infty$ such that $\{f(u(t)), t \geq 0\}$ is a periodic function.

Proof (*Sufficiency*) If $\{f(u(t)), t \geq 0\}$ is a periodic function, denote its period by T, then $T > 0$. Denote

$$g(t) = u(u^{-1}(t) - T)$$

and

$$T_0 = 0, \qquad T_k = u(T + u^{-1}(T_{k-1})), \quad k = 1, 2, \ldots,$$

then it is not difficult to prove that $\{g(t), t \geq T_1\}$ and $\{T_k, k \geq 0\}$ are strictly increasing and continuous, where $T_1 = u(T)$, and $\lim_{k\to+\infty} T_k = +\infty$. Furthermore,

$$\begin{aligned} g(T_k) &= u(u^{-1}(T_k) - T) \\ &= u(u^{-1}(u(T + u^{-1}(T_{k-1}))) - T) = u(u^{-1}(T_{k-1})) \\ &= T_{k-1}, \qquad k = 1, 2, \ldots \end{aligned}$$

and

$$f(t) = f(u(u^{-1}(t))) = f(u(u^{-1}(t) - T)) = f(g(t)).$$

Thus, $\{f(t), t \geq 0\}$ is an approximately periodic function.

(*Necessary*) If $\{f(t), t \geq 0\}$ is an approximately periodic function, then there exist a strictly increasing sequence $\{T_k | T_0 = 0, \lim_{k\to+\infty} T_k = +\infty\}$ and a strictly increasing continuous function $\{g(t), t \geq T_1\}$ satisfying $g(T_k) = T_{k-1}$ for all $k = 1, 2, \ldots$, such that for any $t \geq T_1$ it follows that

$$f(t) = f(g(t)).$$

Denote

$$u(t)=\frac{T_1}{T}t, \qquad t\in[0,T]$$

and

$$u(t)=g^{-1}(u(t-T)), \qquad t\geq T,$$

where T is any given positive number, then

$$f(u(t))=f(g(u(t)))=f(g(g^{-1}(u(t-T))))=f(u(t-T)),$$

so $f(u(t))$ is a periodic function with period T. It yields from mathematical induction that $u(t)$ is strictly increasing and continuous.

Furthermore, we can prove $\lim\limits_{t\to+\infty}u(t)=+\infty$. In fact, owning to the increasing property of $u(t)$, if $\lim\limits_{t\to+\infty}u(t)=+\infty$ does not hold, then there is $M>0$ such that

$$\lim_{t\to+\infty}u(t)=M.$$

Noting that $u(t)=g^{-1}(u(t-T))$ for all $t\geq T$ and $g^{-1}(\cdot)$ is continuous, we obtain that

$$M=g^{-1}(M),$$

that is, $g(M)=M$, which contradicts that $\{g(t),t\geq T_1\}$ is strictly increasing and satisfies $g(T_k)=T_{k-1}$ for all $k=1,2,\ldots$. The proof is completed. □

Remark 2 Proposition 2.1 shows that approximately periodic function comes from a periodic function by scale transformation of its argument.

2.2 Approximately Periodic Time Series

Briefly, approximately periodic time series is a time series which seasonal trend is an approximately periodic function. In this subsection, we will give the definition of approximately periodic time series.

Definition 2.2 Let $\{f(t),t\geq 0\}$ be an approximately periodic function. Denoting $t_0\geq 0$ and $t_k=t_0+k\Delta t$, where $\Delta t>0$ and $k=1,2,\ldots$, if there exists a natural number N such that $\min\limits_{k\geq 1}\{T_k-T_{k-1}\}/\Delta t\geq N$, where T_k-T_{k-1} is the length of k^{th} approximate period, $k=1,2,\ldots$, then $\{f(t_k),k\geq 0\}$ is called an N-order skeleton of $f(t)$.

Remark 3 Generally, there are at least five points to depict a period of an approximately periodic function, so $N\geq 5$ in Definition 2.2.

Definition 2.3 If the seasonal trend of a time series is a skeleton of some approximately periodic function with scale transformation g, then we call the time series has approximate periodicity. That is, the time series is an approximately periodic time series with scale transformation g.

Bezandry and Diagana (2011) wrote a book on almost periodic stochastic processes, which mainly brought forward the concept of p-th mean almost periodic processes and studied existence of p-th mean almost periodic processes. However, approximately periodic time series is different from p-th mean almost periodic processes.

In the following, we will present a simple example to show the character of approximately periodic time series.

Example 1 $\{X_t, t = 1, 2, \ldots\}$ in Table 1 is a sample of time series draw its sequential chart in Figure 2. It is obvious from Figure 2 that $\{X_t, t = 1, 2, \ldots\}$ is an approximately periodic time series with $\{T_k, k = 1, 2, \ldots\}$ as follows

k	0	1	2	3	4	5	6	7
T_k	0	8	17	27	38	50	63	77
$T_k - T_{k-1}$	–	8	9	10	11	12	13	14

From the above table we know that average periodic length is 11, so we draw the sequential chart of the 11-step difference of $\{X_t\}$ in Figure 3. Figure 3 shows that the 11-step difference operator cannot eliminate approximate periodicity of $\{X_t\}$. Thus, we will find new method to extract or eliminate approximate periodicity of approximately periodic time series.

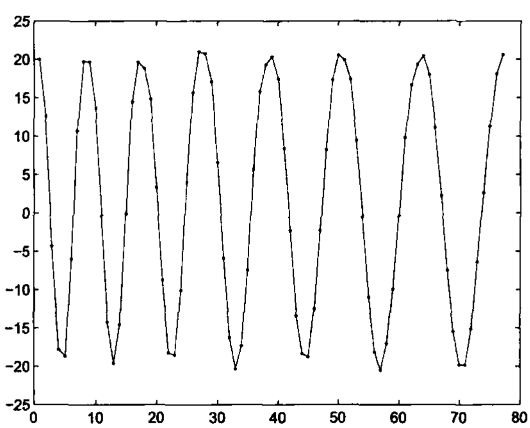

Figure 2 The sequential chart of $\{X_t\}$

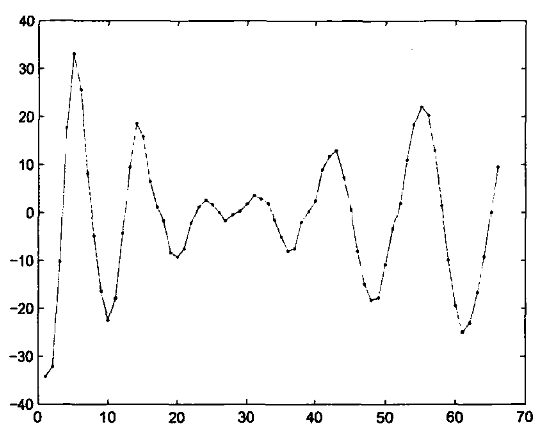

Figure 3 The sequential chart of $\{\nabla_{11} X_t\}$

§3. Extraction of Approximately Periodic Trend

Example 1 shows that approximate periodicity cannot be eliminated by multi-step difference operator. In this section, we will bring forward a method to extract approximately periodic trend of approximately periodic time series. The method includes three steps.

Table 1 Data of Example 1

t	1	2	3	4	5	6	7	8	9	10	11
X_t	10.00	6.34	-2.08	-8.79	-9.67	-3.74	4.46	9.72	9.68	6.59	-0.35
$v(t)$	1	2	3	4	5	6	7	8	0.89	1.78	2.67
ε_t	0.159	0.829	0.824	0.372	-1.031	-2.033	-2.026	-0.150	-0.200	-0.421	-0.335
t	12	13	14	15	16	17	18	19	20	21	22
X_t	-7.16	-9.59	-7.48	-0.14	7.40	9.63	8.81	7.19	1.58	-3.78	-8.81
$v(t)$	3.56	4.44	5.33	6.22	7.11	8	0.8	1.6	2.4	3.2	4
ε_t	-0.119	0.285	-0.628	-0.383	0.211	-0.240	-1.029	-0.836	-0.734	0.752	0.353
t	23	24	25	26	27	28	29	30	31	32	33
X_t	-9.11	-5.12	2.28	7.99	11.01	10.73	8.99	3.48	-2.80	-8.17	-10.35
$v(t)$	4.8	5.6	6.4	7.2	8	0.73	1.45	2.18	2.91	3.64	4.36
ε_t	0.237	-0.125	0.490	0.284	1.137	0.971	0.290	-0.642	-0.673	-0.657	-0.489
t	34	35	36	37	38	39	40	41	42	43	44
X_t	-9.23	-4.37	2.58	7.66	9.29	10.33	9.01	4.24	-0.90	-6.86	-8.72
$v(t)$	5.09	5.82	6.55	7.27	8	0.67	1.33	2	2.67	3.33	4
ε_t	-1.009	-1.121	-0.449	-0.433	-0.583	0.660	-0.144	-1.273	-0.880	-1.320	0.440
t	45	46	47	48	49	50	51	52	53	54	55
X_t	-9.16	-5.96	-0.86	4.09	8.92	10.57	9.91	8.73	4.43	-0.44	-5.99
$v(t)$	4.67	5.33	6	6.67	7.33	8	0.62	1.23	1.85	2.46	3.08
ε_t	0.506	0.900	0.849	0.065	0.540	0.696	0.342	-0.722	-2.145	-2.216	-2.446
t	56	57	58	59	60	61	62	63	64	65	66
X_t	-9.46	-10.51	-8.37	-4.94	-0.38	4.81	7.98	9.34	10.46	9.15	5.42
$v(t)$	3.69	4.31	4.92	5.54	6.15	6.77	7.38	8	0.57	1.14	1.71
ε_t	-1.638	-0.688	0.570	0.508	-0.015	-0.023	-0.629	-0.536	0.995	-0.503	-1.972
t	67	68	69	70	71	72	73	74	75	76	77
X_t	1.03	-3.97	-8.05	-10.10	-10.13	-7.63	-2.83	1.42	5.61	9.26	10.61
$v(t)$	2.29	2.86	3.43	4	4.57	5.14	5.71	6.29	6.86	7.43	8
ε_t	-2.246	-2.288	-1.841	-0.936	-0.329	0.332	1.274	0.620	0.120	0.461	0.738

3.1 Determination of $\{T_k, k=1,2,\ldots\}$

Generally, determination of $\{T_k, k=1,2,\ldots\}$ is a simple job from its background. For example, consider the sale quantity data of moon cake in China. We know moon cake is a festal food on mid-autumn festival for Chinese. Thus, the sale quantity is very large on mid-autumn festival, but less on other days. Mid-autumn festival is a conventional festival for Chinese, which is scheduled by the lunar calendar, so the time length between two adjacent mid-autumn festival is not constant. Thus, the sale quantity data of moon cake in China has approximate periodicity. We can take the dates of mid-autumn festival as $\{T_k, k=1,2,\ldots\}$.

If we can't determine $\{T_k, k=1,2,\ldots\}$ of some approximately periodic time series from its background, such as sunspot data, we can estimate $\{T_k, k=1,2,\ldots\}$ as follows: Let $x_1, x_2, \ldots, x_t, \ldots$ be a sample of approximately periodic time series $\{X_t, t=1,2,\ldots\}$ and t_0 be the first time that $\{X_t, t=1,2,\ldots\}$ reaches its maximum (or minimum) for this sample. Then we take a partial sample $x_{t_0}, x_{t_0+1}, \ldots, x_t, \ldots$ from the sample $x_1, x_2, \ldots, x_t, \ldots$, and newly denote them by $x_0, x_1, \ldots, x_{t-t_0}, \ldots$. Denote $T_0=0$ and T_k the k^{th} time that time series $\{x_1, x_2, \ldots, x_{t-t_0}, \ldots\}$ reaches its maximum (or minimum), $k=1,2,\ldots$, then $\{T_k, k=0,1,2,\ldots\}$ is just what we need. In fact, because approximately periodic function is continuous, so for any interval of an approximate period it can reach its maximum and minimum, so using its maximum (or minimum) to estimate $\{T_k, k=1,2,\ldots\}$ is always effective.

3.2 Estimation of Scale Transformation g

Assume that $\{T_k, k=1,2,\ldots\}$ has been determined. We can estimate g step by step. In order to estimate $\{g(t), t\in[T_k, T_{k+1}]\}$, $k=1,2,\ldots$, we can use the partial samples $\{x_t, t=T_{k-1}, T_{k-1}+1, \ldots, T_{k+1}\}$. According to Definition 2.1, $\{x_{g(t)}, t=T_k, T_k+1, \ldots, T_{k+1}\}$ has the same trend as $\{x_t, t=T_{k-1}, T_{k-1}+1, \ldots, T_k\}$, so using the times of diagnostic points we can estimate $\{g(t), t\in[T_k, T_{k+1}]\}$.

Assume that the form of g has been known and it has m unknown parameters, $\theta_1, \theta_2, \ldots, \theta_m$. Because $g(T_k)=T_{k-1}$ and $g(T_{k+1})=T_k$, it yields $m\geq 2$. Assume there are $m-2$ diagnostic points of the time series for time interval $[T_k, T_{k+1}]$ and $[T_{k-1}, T_k]$ as $t_1^{(k)}, t_2^{(k)}, \ldots, t_{m-2}^{(k)}$ and $t_1^{(k-1)}, t_2^{(k-1)}, \ldots, t_{m-2}^{(k-1)}$, respectively. Then

$$\begin{cases} g(T_k; \theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_m^{(k)}) = T_{k-1}; \\ g(t_1^{(k)}; \theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_m^{(k)}) = t_1^{(k-1)}; \\ g(t_2^{(k)}; \theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_m^{(k)}) = t_2^{(k-1)}; \\ \quad\vdots \\ g(t_{m-2}^{(k)}; \theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_m^{(k)}) = t_{m-2}^{(k-1)}; \\ g(T_{k+1}; \theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_m^{(k)}) = T_k. \end{cases} \tag{3.1}$$

Solving Equation (3.1) yields that there are m functions, $h_1^{(k)}, h_2^{(k)}, \ldots, h_m^{(k)}$, such that

$$\begin{cases} \widehat{\theta}_1^{(k)} = h_1^{(k)}(T_{k-1}, T_k, T_{k+1}; t_1^{(k)}, t_2^{(k)}, \ldots, t_{m-2}^{(k)}; t_1^{(k-1)}, t_2^{(k-1)}, \ldots, t_{m-2}^{(k-1)}); \\ \widehat{\theta}_2^{(k)} = h_2^{(k)}(T_{k-1}, T_k, T_{k+1}; t_1^{(k)}, t_2^{(k)}, \ldots, t_{m-2}^{(k)}; t_1^{(k-1)}, t_2^{(k-1)}, \ldots, t_{m-2}^{(k-1)}); \\ \quad\vdots \\ \widehat{\theta}_m^{(k)} = h_m^{(k)}(T_{k-1}, T_k, T_{k+1}; t_1^{(k)}, t_2^{(k)}, \ldots, t_{m-2}^{(k)}; t_1^{(k-1)}, t_2^{(k-1)}, \ldots, t_{m-2}^{(k-1)}). \end{cases} \tag{3.2}$$

Then

$$g(t) = g(t; \widehat{\theta}_1^{(k)}, \widehat{\theta}_2^{(k)}, \ldots, \widehat{\theta}_m^{(k)}), \qquad t \in [T_k, T_{k+1}],\ k = 1, 2, \ldots \tag{3.3}$$

is just what we need.

For example, if g is a linear function, then it follows from $g(T_k) = T_{k-1}$ and $g(T_{k+1}) = T_k$ that

$$g(t) = T_{k-1} + \frac{T_k - T_{k-1}}{T_{k+1} - T_k}(t - T_k), \qquad t \in [T_k, T_{k+1}],\ k = 1, 2, \ldots. \tag{3.4}$$

If g is a quadric function, the maximum (or minimum) of $\{X_t, t = 1, 2, \ldots\}$ at the times $\{T_k, T_k+1, \ldots, T_{k+1}\}$ and $\{T_{k-1}, T_{k-1}+1, \ldots, T_k\}$ as $X_{t_1^{(k)}}$ and $X_{t_1^{(k-1)}}$, respectively. Then it follows from $g(T_k) = T_{k-1}$, $g(T_{k+1}) = T_k$ and $g(t_1^{(k)}) = t_1^{(k-1)}$ that

$$\begin{cases} a_k + b_k T_k + c_k T_k^2 = T_{k-1}; \\ a_k + b_k t_1^{(k)} + c_k (t_1^{(k)})^2 = t_1^{(k-1)}; \\ a_k + b_k T_{k+1} + c_k T_{k+1}^2 = T_k. \end{cases}$$

Solving the above equation we obtain that

$$\begin{bmatrix} a_k \\ b_k \\ c_k \end{bmatrix} = \begin{bmatrix} 1 & T_k & T_k^2 \\ 1 & t_1^{(k)} & (t_1^{(k)})^2 \\ 1 & T_{k+1} & T_{k+1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T_{k-1} \\ t_1^{(k-1)} \\ T_k \end{bmatrix},$$

so

$$g(t) = \begin{bmatrix} 1 & t & t^2 \end{bmatrix} \begin{bmatrix} 1 & T_k & T_k^2 \\ 1 & t_1^{(k)} & (t_1^{(k)})^2 \\ 1 & T_{k+1} & T_{k+1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T_{k-1} \\ t_1^{(k-1)} \\ T_k \end{bmatrix}, \qquad t \in [T_k, T_{k+1}],\ k = 1, 2, \ldots.$$

3.3 Estimation of Approximately Periodic Function f

Assume $\{T_k, k = 1, 2, \ldots\}$ and g have been estimated, then we will estimate f in this subsection. Let

$$v(t) = t, \qquad t \in [0, T_1]$$

and

$$v(t) = g^{(k)}(t), \qquad t \in [T_k, T_{k+1}],\ k = 1, 2, \ldots,$$

where $g^{(k)}(\cdot)$ is the k-time composite function of $g(\cdot)$, $k = 1, 2, \ldots$.

Using the mathematical induction it is not difficult to prove the following property:

Property 3.1 $f(t) = f(v(t))$ and $v(t) \in [0, T_1]$ hold for all $t \geq 0$.

Remark 4 Property 3.1 shows that, for any $t \geq 0$, $f(t) \in \{f(s), s \in [0, T_1]\}$.

For any $t \geq 0$, denote

$$Y_{v(t)} = f(t),$$

then

$$\{(v(t), Y_{v(t)}), t \geq 0\} \subseteq \{(t, f(t)), t \in [0, T_1]\}.$$

Thus, we can estimate $\{f(t), t \in [0, T_1]\}$ by fitting $\{Y_s, s \in [0, T_1]\}$.

3.4 Example

In this subsection, we will continue Example 1 to estimate its g and f.

Assume g is a linear function. It yields from Equation (3.4) that

$$g(t) = T_{k-1} + \frac{T_k - T_{k-1}}{T_{k+1} - T_k}(t - T_k), \qquad t \in [T_k, T_{k+1}],\ k = 1, 2, \ldots,$$

where $T_k = k(k+15)/2$, $k = 0, 1, 2, \ldots$. It follows from the mathematical induction that

$$g^{(k)}(t) = T_0 + \frac{T_1 - T_0}{T_{k+1} - T_k}(t - T_k), \qquad t \in [T_k, T_{k+1}],\ k = 1, 2, \ldots.$$

Thus

$$\begin{aligned} v(t) &= T_0 + \frac{T_1 - T_0}{T_{k+1} - T_k}(t - T_k) \\ &= \frac{8}{k+8}\left(t - \frac{k(k+15)}{2}\right), \qquad t \in [T_k, T_{k+1}],\ k = 1, 2, \ldots. \end{aligned}$$

Noting that $v(t) = t$ for all $t \in [0, T_1]$ we have

$$v(t) = \frac{8}{k+8}\left(t - \frac{k(k+15)}{2}\right), \qquad t \in [T_k, T_{k+1}],\ k = 0, 1, 2, \ldots, \tag{3.5}$$

where $T_k = k(k+15)/2$, $k = 0, 1, 2, \ldots$.

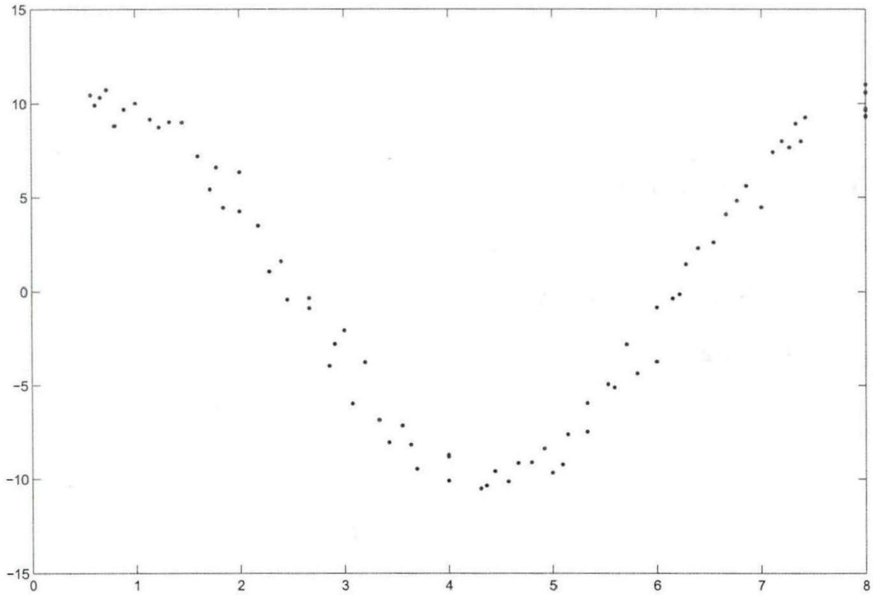

Figure 4 Sequential chart of $\{Y_{v(t)}, t \geq 0\}$

Denote $Y_{v(t)} = f(t)$, $t = 0, 1, 2, \ldots$, and draw its scatter figure in Figure 4. We judge from Figure 4 that $\{f(t), t = 0, 1, \ldots, 8\}$ may be a function with the form $a\cos(c(t-b))$, where a approximately equals 10, b approximately equals 1 and c approximately equals $2\pi/7$. Thus, we assume that

$$f(t) = a\cos\Big(\frac{\pi(t-b)}{c}\Big) + \varepsilon_t, \qquad t \in [0, 8].$$

Using the least square method we obtain $a = 9.88$, $b = 0.9$ and $c = 3.53$. That is,

$$f(t) = 9.88\cos\Big(\frac{\pi(t-0.9)}{3.53}\Big) + \varepsilon_t, \qquad t \in [0, 8]. \tag{3.6}$$

It yields from Equation (3.5), Equation (3.6) and Property 3.1 that

$$f(t) = 9.88\cos\Big(\frac{8\pi}{3.53(k+8)}\Big[t - \frac{k(k+15)}{2}\Big] - \frac{0.9\pi}{3.53}\Big) + \varepsilon_t,$$
$$t \in [T_k, T_{k+1}],\ k = 0, 1, 2, \ldots, \tag{3.7}$$

where $T_k = k(k+15)/2$, $k = 0, 1, 2, \ldots$.

Using Equation (3.7) we can estimate $\{\varepsilon_t\}$, which shown in Table 1 and draw its sequential chart in Figure 5. From Figure 5 we know $\{\varepsilon_t\}$ has no approximately periodic trend, that is, our method can extract the trend of approximate period.

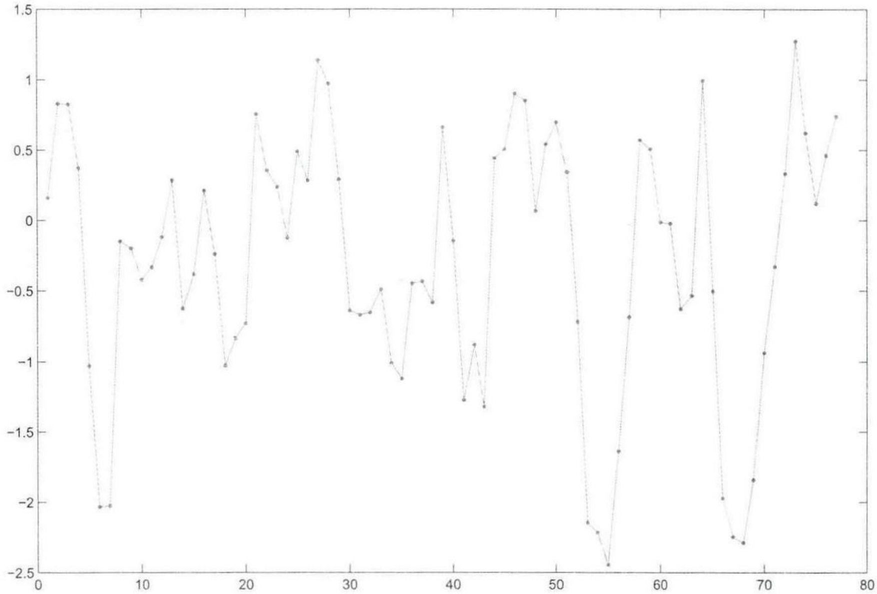

Figure 5 Sequential chart of $\{\varepsilon_t, t \geq 0\}$

§4. Elimination of Approximately Periodic Trend

In Section 3, we have given a method to extract the approximately periodic trend, which is very effective. However, the method is a little troublesome, so we hope bring forward a little simpler method to deal with the approximately periodic trend in this section, which is so-called the method of generalized difference operator.

4.1 Generalized Difference Operator

Let $\{X_{t_k}, k = 1, 2, \ldots\}$ be an approximately periodic time series with scale transformation g, where $t_1 < t_2 < \cdots < t_k < \cdots$ are the sampling times. We denote its peak/ bottom point time at each approximate period by T_k, at this time X_{T_k} is its peak point or bottom point at the approximate period, then $T_1 < T_2 < \cdots < T_k < \cdots$. To eliminate its approximate periodicity at time interval $[T_k, T_{k+1}]$, $k \geq 2$, we consider to use the data at time interval $[T_{k-1}, T_k]$ and bring forward a generalized difference operator ∇_g as follows

$$\nabla_g X_{t_i} = X_{t_i} - \widehat{X}_{g(t_i)}, \tag{4.1}$$

where

$$\begin{aligned}
&\widehat{X}_{g(t_i)} \overset{\text{def.}}{=} q_i X_{\widehat{t}_i} + (1 - q_i) X_{\widehat{t}_{i+1}}, \\
&\widehat{t}_i = \max\{t_j : t_j \leq g(t_i)\}, \\
&\widehat{t}_{i+1} = \min\{t_j : t_j > g(t_i)\}
\end{aligned}$$

and

$$q_i = \frac{\widehat{t}_{i+1} - g(t_i)}{\widehat{t}_{i+1} - \widehat{t}_i}.$$

Remark 5 The generalized difference operator includes the ordinary difference operator.

In fact, if time series is periodic, then $T_k - T_{k-1} \equiv T$ and $g(t) = t - T$, where $T > 0$ is a constant, $k = 1, 2, \ldots$. Thus,

$$\widehat{t}_i = g(t_i) = t_i - T,$$
$$q_i = \frac{\widehat{t}_{i+1} - g(t_i)}{\widehat{t}_{i+1} - \widehat{t}_i} = 1$$

and

$$\nabla_g X_{t_i} = X_{t_i} - X_{t_i - T} = \nabla_T X_{t_i}.$$

That is, for this case, ∇_g just equals the T-step difference operator ∇_T.

If time series is nonperiodic, i.e., $T_k - T_{k-1} \equiv 1$. Analyzing analogically, we can obtain that

$$\nabla_g X_{t_i} = X_{t_i} - X_{t_i - 1} = \nabla X_{t_i}.$$

That is, for this case, ∇_g just equals the 1-order difference operator ∇.

Theorem 4.1 Assume that $\{X_{t_n}, n \geq 0\}$ is a time series from a continuous time stochastic process $\{X_t, t \geq 0\}$, where $t_n = t_{n-1} + \Delta t_n$ for $n = 1, 2, 3, \ldots$, then

$$\lim_{\max\limits_n\{\Delta t_n\} \downarrow 0} \nabla_g X_{t_n} = X_{t_n} - X_{g(t_n)} \qquad \text{for all } t_n \in T.$$

That is, ∇_g can eliminate approximately periodicity as $\max\limits_n\{\Delta t_n\}$ tends to zero.

Proof Noting that X_t is continuous with respect to t, we have

$$\lim_{\Delta t_n \downarrow 0} X_{\widehat{t}_n} = X_{g(t_n)}$$

and

$$\lim_{\Delta t_{n+1} \downarrow 0} X_{\widehat{t}_{n+1}} = X_{g(t_n)}.$$

Thus, it follows from (4.1) that

$$\begin{aligned}
\lim_{\max\limits_n\{\Delta t_n\} \downarrow 0} \nabla_g X_{t_n} &= \lim_{\max\limits_n\{\Delta t_n\} \downarrow 0} \{X_{t_n} - \widehat{X}_{g(t_n)}\} \\
&= \lim_{\max\limits_n\{\Delta t_n\} \downarrow 0} \{X_{t_n} - q_n X_{\widehat{t}_n} - (1 - q_n) X_{\widehat{t}_{n+1}}\} \\
&= X_{t_n} - X_{g(t_n)}.
\end{aligned}$$

The proof is completed. □

4.2 Analysis of Sunspot Series

In Section 1, we have pointed out the sunspot data during the twentieth century is a time series with approximate periodicity, which sequential chart is drawn in Figure 1. In this subsection, we try to eliminate its approximate period using the generalized difference operator. During the twentieth century the sunspot peak years include 1905, 1917, 1928, 1937, 1947, 1957, 1968, 1979, 1989 and 2000.

To show the effect of the generalized difference operator with linear function g, we only consider the sunspot data from 1905 to 2000. In order to eliminate the approximately periodic effect on $t_i = 1918$, we compute as follows

$$g(1918) = 1905 + \frac{1917-1905}{1928-1917}(1918-1917) = 1906\frac{1}{11},$$

$$\widehat{t}_i = \max\left\{t_j : t_j \le 1906\frac{1}{11}\right\} = 1906$$

and

$$\widehat{t}_{i+1} = \min\left\{t_j : t_j > 1906\frac{1}{11}\right\} = 1907,$$

so

$$q_i = \frac{1907 - 1906\frac{1}{11}}{1907-1906} = \frac{10}{11}$$

and

$$\begin{aligned} X_{g(1918)} &= \frac{10}{11} * X_{1906} + \left(1 - \frac{10}{11}\right) * X_{1907} \\ &= \frac{10}{11} * X_{1906} + \frac{1}{11} * X_{1907} \\ &= \frac{10}{11} * 53.8 + \frac{1}{11} * 62 \\ &\approx 54.55. \end{aligned}$$

then

$$\begin{aligned} \nabla_g X_{1918} &= X_{1918} - X_{g(1918)} \\ &\approx 80.6 - 54.55 \\ &= 26.05. \end{aligned}$$

Similarly, we can compute all generalized differences from 1917 to 2000, and draw their sequential chart in Figure 6.

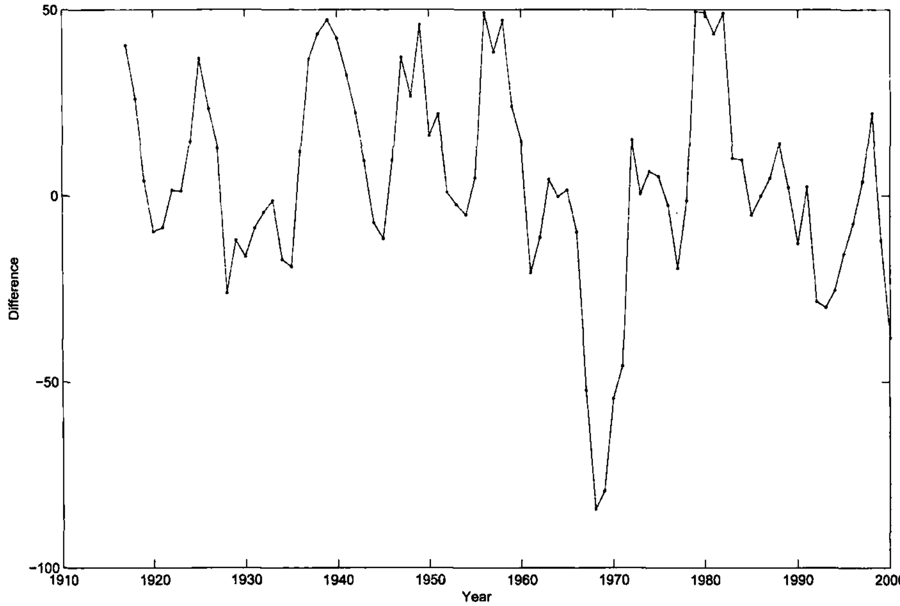

Figure 6 Sequential chart of difference on sunspot data during twentieth century

Comparing Figure 1 and Figure 6, it is obvious that the generalized difference operator can eliminate approximate periodicity for sunspot data.

References

[1] Fan, J.Q. and Yao, Q.W., *Nonlinear Time Series: Nonparametric and Parametric Methods*, Science Press, Beijing, 2006.

[2] Tsay, R.S., *Analysis of Financial Time Series*, John Wiley & Sons, Inc., 2002.

[3] Bezandry, P.H. and Diagana, T., *Almost Periodic Stochastic Processes*, Springer, 2011.

近似周期时间序列分析

吴述金 朱晓雨 杨 筱

(华东师范大学金融与统计学院统计与精算学系, 上海, 200241)

本文首先给出了近似周期时间序列概念, 即: 具有周期特征但是周期长度变化的时间序列. 比如, 太阳黑子序列具有11年左右的周期, 但是其周期并不是11, 而是在11左右变化, 这就是一个近似周期序列. 然后给出了提取近似周期趋势方法, 并且提出了广义差分算子, 这里提出的广义差分算子不仅可以消除时间序列的长期趋势和周期性, 而且还可以消除近似周期性. 最后, 以太阳黑子序列为例说明了广义差分算子的应用.

关键词: 时间序列, 近似周期性, 广义差分算子.

学科分类号: O211.6.