Chapter Two

Review and Implementation of Seasonal Adjustment Methods

2.1 Introduction

The interest in seasonally adjusted series by government agencies has brought about the development and improvement of many practical seasonal adjustment techniques.

The use of electronic computers by Shiskin and Eisenpress (1957) in the deseasoning procedures used in the U. S. Bureau of the Census, the massive manual computations involved in seasonal adjustment procedures have been greatly reduced. As a result, further accelerates the development and application of seasonal adjustment techniques by economists, geologists, scientists etc. and subsequently stimulated the interest to develop more sophisticated methods.

Techniques in seasonal adjustment, ranging from the simple moving average filters to the newer model-based approach, are mainly used in analysis of economic and business time series. A detailed review on the current development of seasonal adjustment methods for economic data is given by Bell and Hillmer (1984).

Comparatively, little effort has been made in developing techniques to decompose hydrologic time series. At present, the common approach for extracting seasonal effects from hydrologic time series is to standardize the original data. In the following section, we shall discuss harmonic analysis which is employed in the studies of hydrologic time series.

In view of the resemblance of hydrologic data to commercial data in its basic structure and composition, we shall try to investigate whether techniques that have been designed for deseasoning economic and business time series can also be used for deseasoning hydrologic time series.

In Section 2.3 and 2.4, we shall discuss and implement two seasonal adjustment techniques. They are the Census X-11 procedure of Shiskin and Eisenpress (1957) and the *ARIMA*-model-based approach of Hillmer and Tiao (1982). We shall also appraise and compare the applicability of these techniques in deseasoning the observed river flows used in this study.

2.2 Harmonic Analysis

Harmonic analysis is basically concerned with approximating a periodic discrete series by a sum of sine and cosine terms. Its overall effect is to partition the variability of a series of length N into components at frequencies $2\pi/N$, $4\pi/N$, ..., π. The component at angular frequency $\omega_j = 2\pi j/N$ is called the jth harmonic. The jth harmonic is defined as $a_j \cos \omega_j t + b_j \sin \omega_j t$, which can be rewritten as $R_j \cos(\omega_j t + \phi_j)$ where

$$R_j = \text{amplitude of the } j\text{th harmonic}$$
$$= \sqrt{(a_j^2 + b_j^2)} \ ,$$

$$\phi_j = \text{phase of the } j\text{th harmonic}$$
$$= \tan^{-1}\left(\frac{-b_j}{a_j}\right).$$

By the well known Parserval's Theorem, $R_j^2/2$ is the contribution of the jth harmonic to the total variance of the series. Let s^2 be the total variance of the series, then the ratio $R_j^2/2s^2$ represents the part of the variation explained by the jth harmonic.

In harmonic analysis, hydrologic time series are assumed to be composed by deterministic periodic components and a stationary stochastic component. The periodic components are present in parameters such as the monthly means and standard deviations of the series. The stochastic component is superposed on these periodic components.

From past investigations, monthly time series of river flows are usually non-stationary and both the monthly mean and standard deviation follow a periodic movement with a fundamental period of 12-month. Therefore, observations for each of the 12 calendar months in a year can be assumed to be drawings from different populations, each with its own population mean m_τ and population standard deviation s_τ, where $\tau = 1, 2, \ldots, 12$.

The sample estimates of m_τ and s_τ are respectively

$$\hat{m}_\tau = \frac{1}{n} \sum_{i=1}^{n} x_{,\tau}$$

and

$$\hat{s}_\tau = \left[\frac{1}{n} \sum_{i=1}^{n} (x_{i,\tau} - m_\tau)\right]^{1/2}$$

where $x_{i,\tau}$ represents the observation for the month τ of the year i and n is the total number of years.

The classical method for removing the periodicity in m_τ and s_τ from $\{x_{i,\tau}\}$ is to apply the simple transformation

$$y_{i,\tau} = \frac{x_{i,\tau} - m_\tau}{s_\tau}. \tag{2.1}$$

The transformed series $\{y_{i,\tau}\}$ will possess mean zero and standard deviation unity. The total number of statistics required in this transformation is $(\nu\omega)$, where ν is the number of periodic parameters each with period ω. For example, the above transformation needs (2x12) statistics, 12 m_τ and 12 s_τ.

From the viewpoint of sampling theory, when ν and ω are large, the $(\nu\omega)$ statistics cannot be estimated accurately. As an illustration, a daily time process (period = 365) with periodicity in two parameters will require (2x365) statistics. This is an unnecessary large number and in addition, will entail large sampling variation. In order to reduce the number of statistics needed in the transformation, an alternative is to approximate the periodic parameters m_τ and s_τ by the periodic functions

$$m_\tau = A_0 + \sum_{k=1}^{M} (A_k \cos\frac{2\pi k}{12}\tau + B_k \sin\frac{2\pi k}{12}\tau) \tag{2.2}$$

and

$$s_\tau = {}_sA_0 + \sum_{k=1}^{M'} ({}_sA_k \cos\frac{2\pi k}{12}\tau + {}_sB_k \sin\frac{2\pi k}{12}\tau) \tag{2.3}$$

where M is the number of significant harmonics for the periodic parameter m_τ and M' is the number of significant harmonics for the periodic parameter s_τ. Note that the maximum number of harmonics for monthly series is $M = M' = 6$. The Fourier coefficients A_k , B_k , ${}_sA_k$ and ${}_sB_k$ are given respectively as

$$A_k = \begin{cases} \frac{1}{6} \sum_{\tau=1}^{12} m_\tau \cos\frac{2\pi k}{12}\tau & k = 1, 2, \ldots, 5 \\ \frac{1}{12} \sum_{\tau=1}^{12} m_\tau \cos\frac{2\pi k}{12}\tau & k = 0, 6 \end{cases}$$

$$B_k = \begin{cases} \frac{1}{6} \sum_{\tau=1}^{12} m_\tau \sin\frac{2\pi k}{12}\tau & k = 1, 2, \ldots, 5 \\ 0 & k = 0, 6 \end{cases}$$

$${}_sA_k = \begin{cases} \frac{1}{6} \sum_{\tau=1}^{12} s_\tau \cos\frac{2\pi k}{12}\tau & k = 1, 2, \ldots, 5 \\ \frac{1}{12} \sum_{\tau=1}^{12} s_\tau \cos\frac{2\pi k}{12}\tau & k = 0, 6 \end{cases}$$

$${}_sB_k = \begin{cases} \frac{1}{6} \sum_{\tau=1}^{12} s_\tau \sin\frac{2\pi k}{12}\tau & k = 1, 2, \ldots, 5 \\ 0 & k = 0, 6 \end{cases}$$

Note that $A_0 = \frac{1}{12} \sum_{\tau=1}^{12} m_\tau$ and ${}_sB_0 = \frac{1}{12} \sum_{\tau=1}^{12} s_\tau$ are the mean values of m_τ and s_τ respectively.

The above approximation of periodic parameters m_τ and s_τ by harmonic analysis requires a total of 24 statistics. However, not all of the six harmonics are needed, the first few harmonics are normally sufficient for generating complex seasonal pattern.

Several methods can be used to determine the significant harmonics. We shall briefly discuss the method of Fisher and Whittle (Priestley, 1981, pg. 406 - 411). For testing the significance of the harmonic with the largest amplitude, Fisher uses the test statistics

$$g = \frac{\left(\max\limits_{1 \le j \le m} R_j^2\right)}{\sum\limits_{j=1}^{m} R_j^2}$$

where m is the total number of harmonics. The harmonic with the largest amplitude is said to be significant at 100α% level if the calculated value of g exceeds g_α, where the critical value g_α is chosen such that

$$\Pr(g > g_\alpha) = m(1-g_\alpha)^{m-1} - \frac{m(m-1)}{2}(1-2g_\alpha)^{m-1} + \ldots + (-1)^{k-1} \frac{m!}{k!\,(m-k)!}(1-kg_\alpha)^{m-1} \qquad (2.4)$$

and k is the largest integer which is less than $1/g_\alpha$.

The values of g_α can be obtained from the table below,

Table 2.1 *Fisher's g_α-critical Values For α=0.01, 0.05, m=2, 3, ..., 6*

m	6	5	4	3	2
α=0.01	0.72179	0.78874	0.78925	0.94226	0.995
α=0.05	0.61615	0.68377	0.68938	0.87090	0.975

Suppose $\max\limits_{1 \le j \le m} (R_j^2)$ occurs at $j = j_1$ (corresponding to the frequency $\frac{2\pi j_1}{12}$ for $m = 6$) and $R_{j_1}^2$ turns out to be significant for a selected probability level α, the second largest amplitude $R_{j_2}^2$ is then tested by using the statistic,

$$g_2 = \frac{R_{j_2}^2}{\left[\sum\limits_{j=1}^{m} R_j^2 - R_{j_1}^2\right]}$$

and referring to the Fisher's distribution (equation (2.4)) with m replaced by $m-1$. If the second largest harmonic also turns out to be significant, we then continue with the procedure to test the third largest harmonic and so on until we fail to obtain a significant result. In this way, we obtain an estimate of M (or M') for the periodic function (2.2) (or (2.3)).

By using the method of Fisher and Whittle (1981), the first five harmonics are found significant in the observed series. Figure 2.1 displays the time sequence plot of the seasonally adjusted series and Figure 2.2 displays the periodogram of the adjusted series. From Figure 2.2, we observe that the seasonal cycles of the original series have been removed. The sample *ACF* and *PACF* of the seasonally adjusted series, which are represented by Figure 2.3 and 2.4 respectively, show that the seasonally adjusted series has no obvious peaks at seasonal lags. The two dotted lines on the graphs are the upper and lower standard error of the estimations.

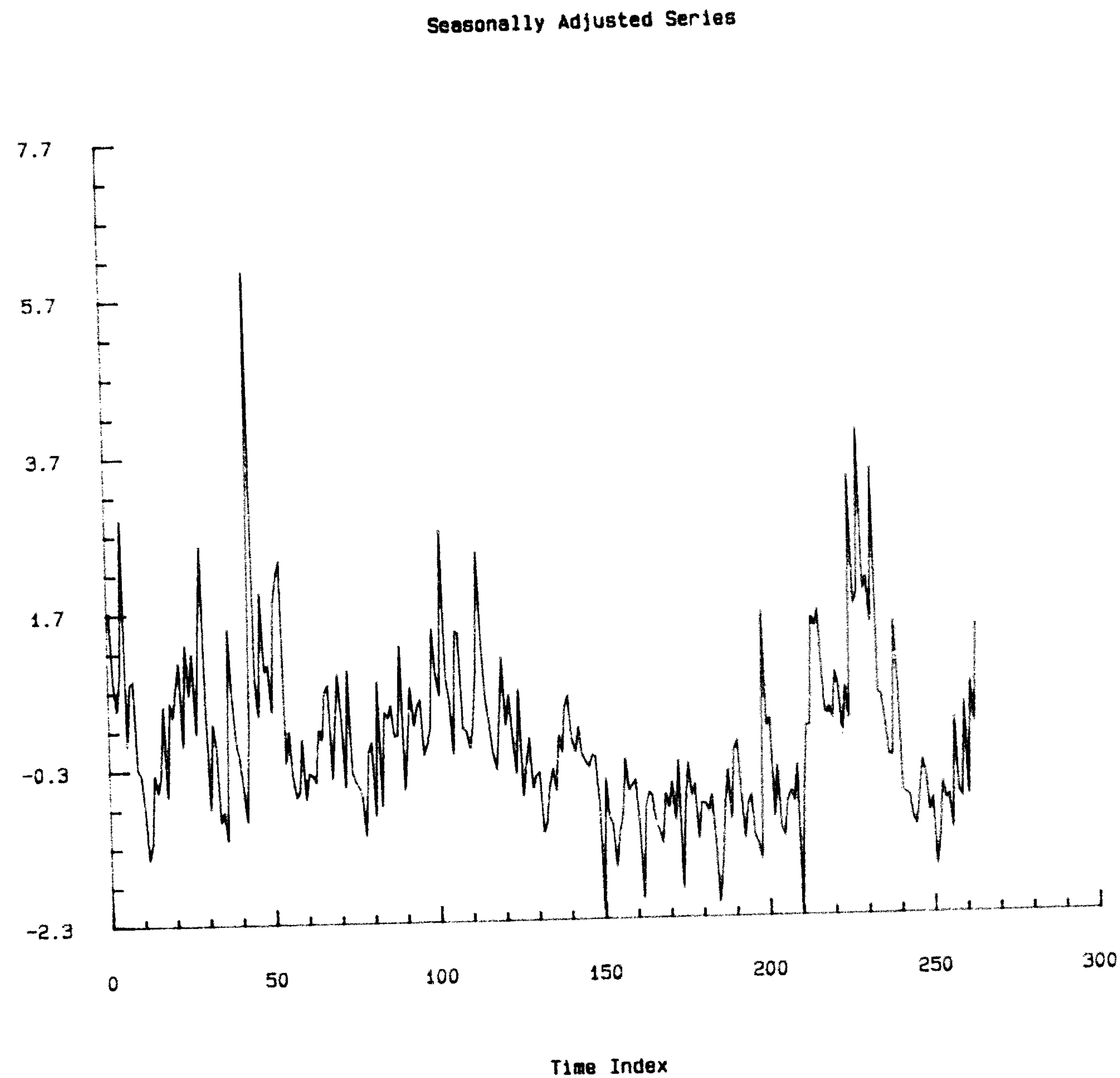

Figure 2.1 Time Series Plot Of The Seasonally Adjusted Series Of Perak River Flows By Harmonic Analysis

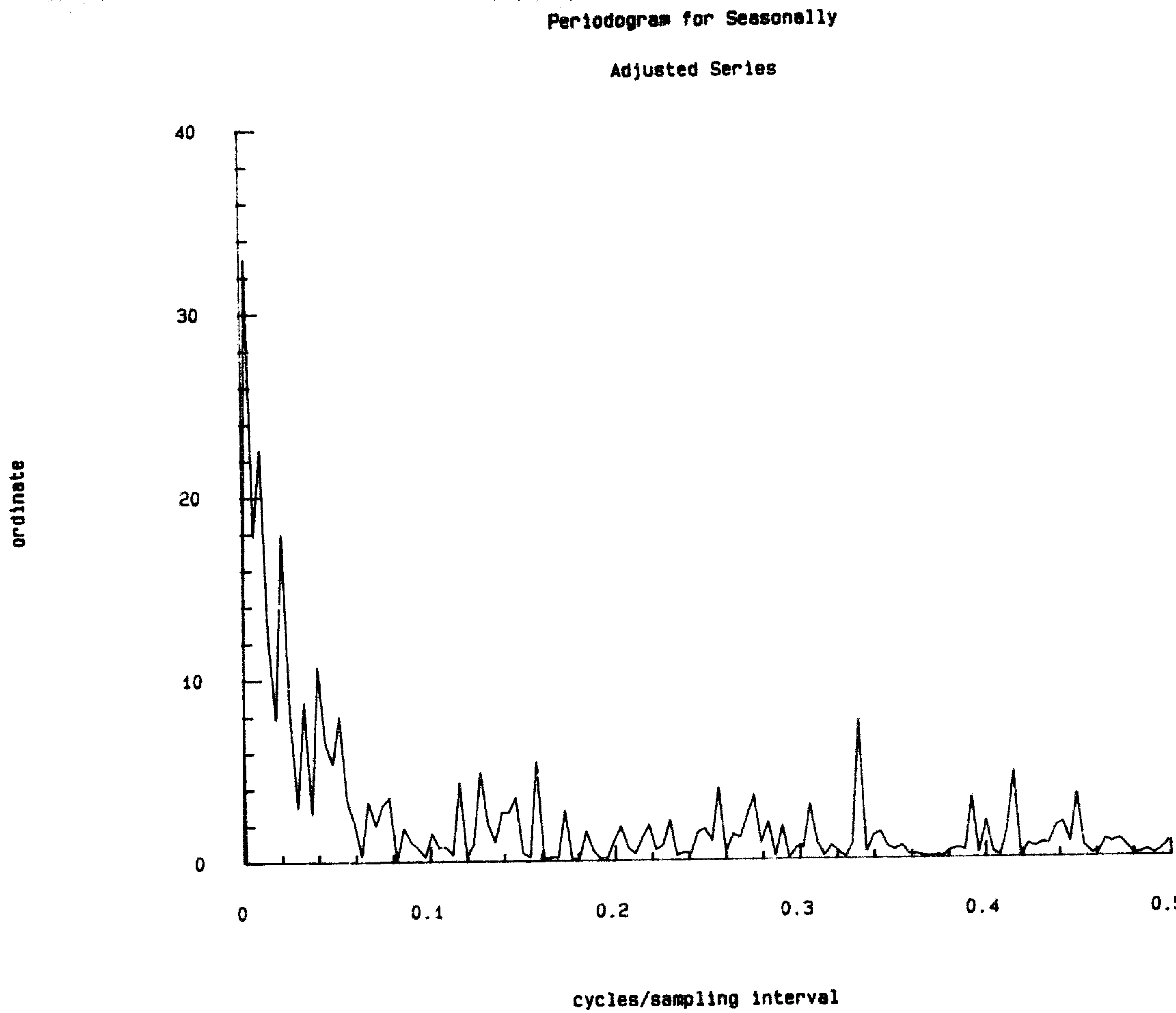

Figure 2.2 Periodogram Of The Seasonally Adjusted Series Of Perak River Flows By Harmonic Analysis

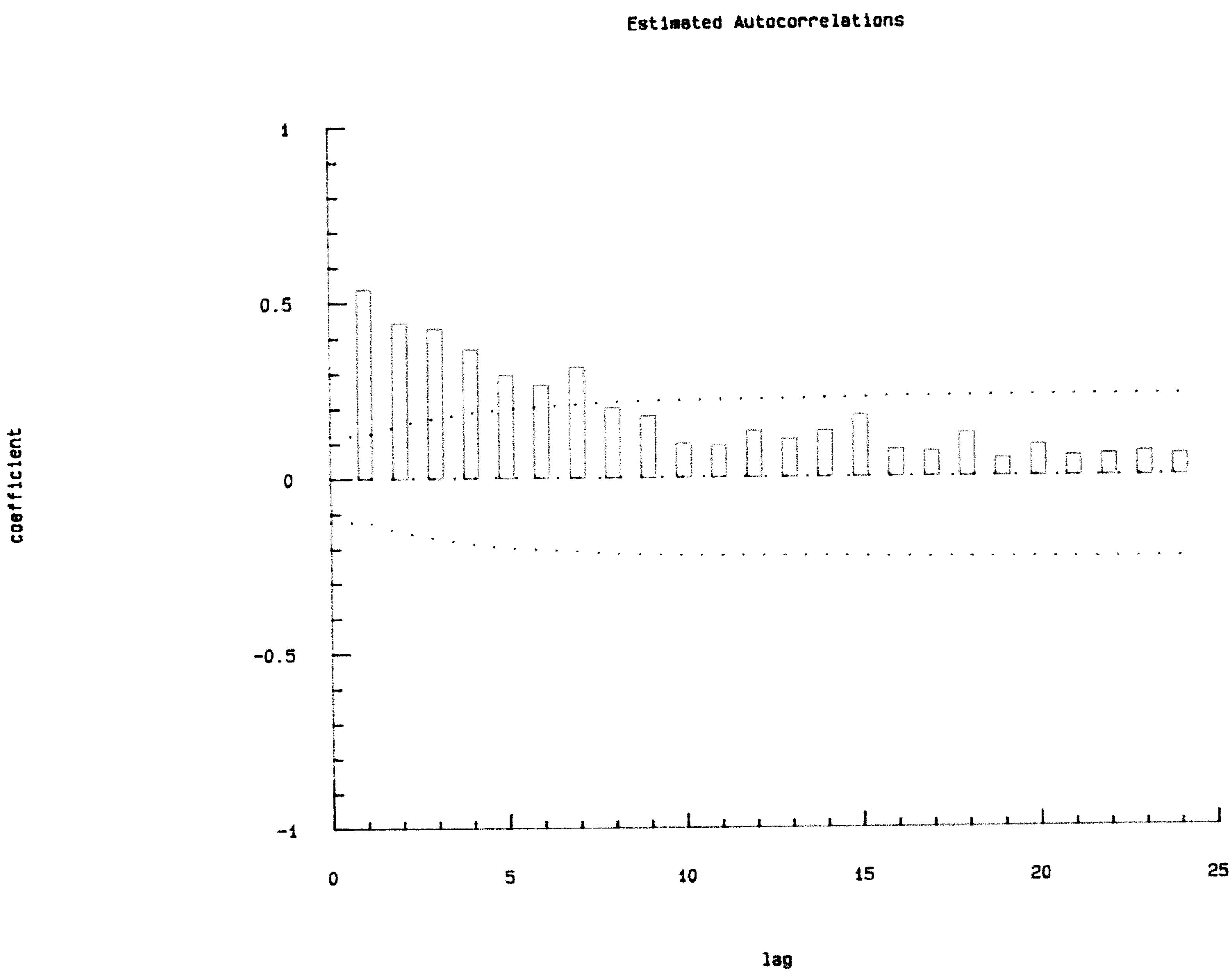

Figure 2.3 Sample ACF Of The Seasonally Adjusted Series Of Perak River Flows By Harmonic Analysis

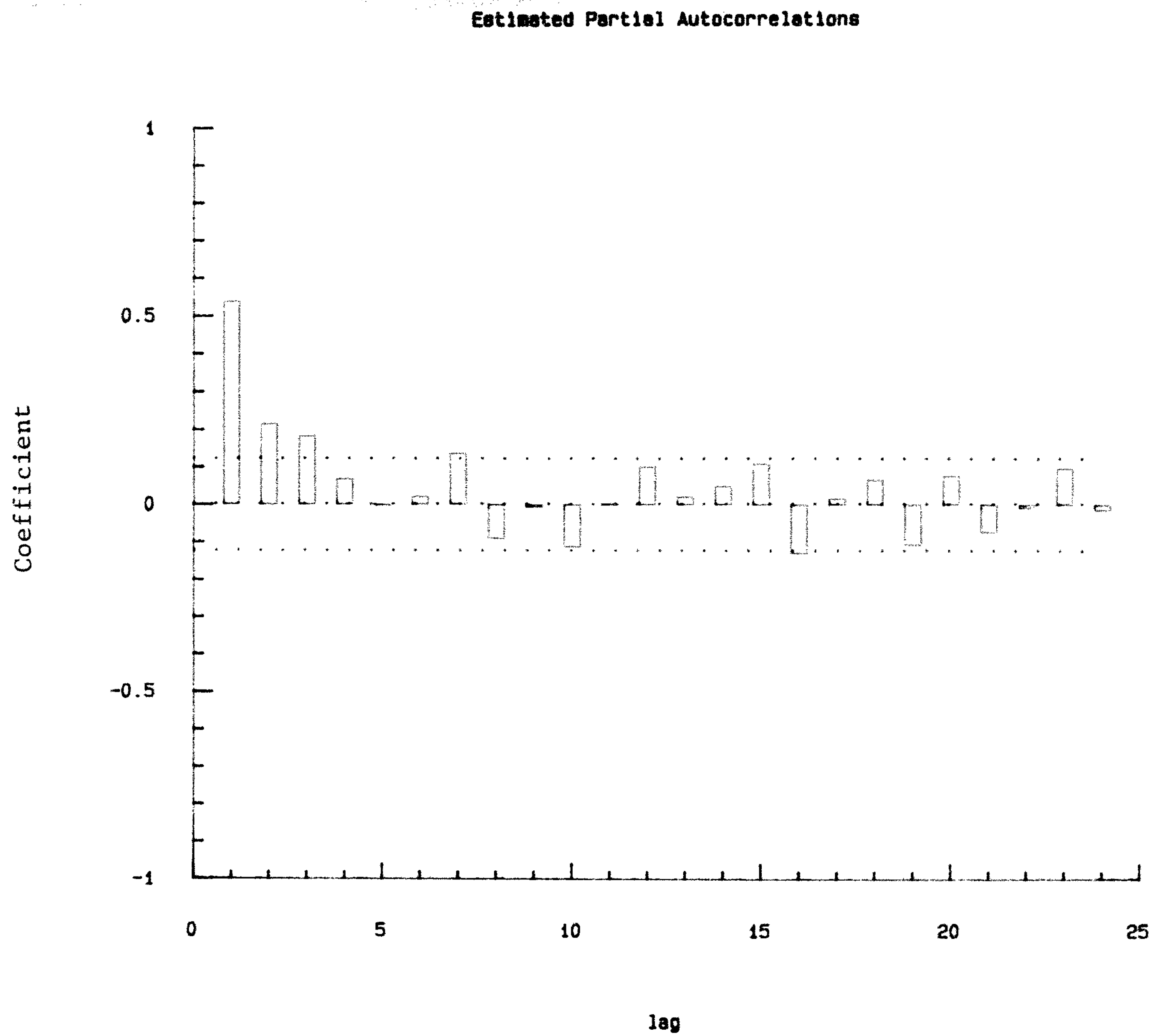

Figure 2.4 Sample PACF Of The Seasonally Adjusted Series Of Perak River Flows By Harmonic Analysis

2.3 Symmetric Moving Average Filters

Among the various types of available seasonal adjustment methods, the most widely used technique is the Bureau of the Census X-11 procedure described by Shiskin and et. al. (1967). It was introduced by the U. S. Bureau of the Census in the 1950's. The basic feature of the procedure is that it uses a sequence of moving average filters to decompose a time series into a seasonal component, a trend component and a noise component.

A Symmetric (2k+1)-term Weighted Moving Average (WMA) Filter is defined as

$$w(B) = \sum_{j=-k}^{k} w_j B^j,$$

where

(a) k is a non-negative integer,

(b) w_j's are weight constants such that $w_{-j} = w_j$ and $\sum_{j=-k}^{k} w_j = 1$,

(c) B is the back shift operator such that $BY_t = Y_{t-1}$.

If $w_j = \frac{1}{2k+1}$ for all $j = -k, \ldots, 0, 1, \ldots, k$, then it is known as *Symmetric (2k+1)-term Moving Average (MA) Filter*. As an example, the value of k in a five-term *WMA* filter is 2, and $w_0 = \frac{1}{3}$, $w_1 = w_{-1} = \frac{1}{9}$, $w_2 = w_{-2} = \frac{2}{9}$ is a possible set of weight constants.

The main purpose of applying these filters is to damp out the influence of the irregular component in order to obtain a 'smooth ' delineation of the trend component. The Census X-11 procedure is constructed based on these symmetric *WMA* and *MA* filters.

Generally, the filters used in the Census X-11 procedure are the 3-term *MA*; 5-term *MA*; 12-term *MA*; (3*3)-term *MA*, which is obtained by taking a 3-term *MA* of the 3-term *MA* already generated; (3*5)-term *MA*, which is obtained by taking a 5-term *MA* of the 3-term *MA* already generated; centered 12-term *MA*, which is obtained by taking a 2-term *MA* of the 12-term *MA* already generated and etc.

Before we proceed further to discuss the procedure, we shall denote the seasonal component of a time series, which is the intra-year variation that repeats constantly or in an evolving fashion from year to year by S. We shall also denote the trend-cycle that includes variation due to long-term trend and underlying cycle by T, and the irregular component by I. If we denote the original series by Y, then experience indicates that many time series can be represented by the multiplicative formula

$$Y = S \times T \times I. \tag{2.5}$$

Shiskin and Eisenpress (1957) have described two versions of the Census X-11 procedure and compared their applicability. Both of them are ratio-to-moving average methods and the second method is an improved version of the first method.

We shall briefly describe the principal features of Method I and Method II.

A. Method I

The principal features of Method I are as follows :

(a) Estimate the trend-cycle by applying a 12-month *MA* filter to the original series. The trend-cycle is then eliminated by dividing the 12-month *MA* into the original data leaving the seasonal-irregular component.

(b) Fit a 5-term *MA* to the time series representing the seasonal-irregular component for each of the 12 calendar months in successive years in order to obtain 12 sets of seasonal factors. The largest and the smallest seasonal-irregular ratios in each set of the 5 terms are dropped from computation before the remaining three are averaged. The purpose of doing this is to remove the extreme ratios.

(c) The seasonal factors for each year are then adjusted so that their sum is equal to 1200. These adjusted seasonal factors are later divided into the original observations to obtain a preliminary seasonal adjusted series, which represents the trend-cycle-irregular component. Note that the sum of the adjusted seasonal factors for each month is equal to 1200 because the seasonal-irregular components obtained above were multiplied by 100 prior to the adjustment.

(d) This series is in turn smoothed by a 5-month *MA* filter to provide the second approximation of the trend-cycle. This curve is more flexible than the one produced by the 12-month *MA* filter in (a). In other words, the 5-month *MA* filter can change direction over a shorter interval and produces smoother peaks.

(e) Divide the second approximation of the trend-cycle into the original data in order to obtain the second approximation of the seasonal-irregular component.

(f) Repeat (b)-(c) to get the final seasonally adjusted series.

B. Method II

Method II uses more complicated *WMA* filters in place of the simple *MA* filters in Method I. Instead of using a 5-month *MA* filter to obtain the second approximation of the trend component in (d), Method II uses the Spencer 15-month *WMA* filter. The weights of the Spencer's 15-month *WMA* are as follows :

Table 2.2 *Weights w_j Used In The Spencer's 15-Month Moving Average Filter.*

j	0	1	2	3	4	5	6	7
w_j	$\frac{-3}{320}$	$\frac{-6}{320}$	$\frac{-5}{320}$	$\frac{3}{320}$	$\frac{21}{320}$	$\frac{46}{320}$	$\frac{67}{320}$	$\frac{74}{320}$

The Spencer 15-month *WMA* filter has been proven to give smoother representation of the trend-cycle component than the irregular curve produced by the 5-month *MA* filter, so the use of Spencer 15-month *WMA* filter is more justified in terms of producing smooth trend curve.

This *WMA* filter entails the loss of seven values at both the beginning and the end of the series. In order to fill in the seven missing values at the end of the Spencer's 15-month *WMA*, which are of considerable importance, Shiskin and Eisenpress (1957) used the average of the last four months of the preliminary seasonally adjusted series as the estimate of each of the seven months immediately after the last month of this series, i.e.,

$$Y_{n+1} = Y_{n+2} = \ldots = Y_{n+7} = \frac{Y_{n-3} + Y_{n-2} + Y_{n-1} + Y_n}{4},$$

where Y_n is the last record of the preliminary seasonal adjusted series. Similarly, seven values are added to the beginning of the series. This augmented series is then smoothed by the Spencer 15-month *WMA* filter. On the other hand, Method I uses the last and the first record of the preliminary seasonal adjusted series to fill the missing values at the end and the beginning of the series respectively, which assumes that there is no trend in the seasonal factors at both ends of the series.

Method II also reduces the weights of extreme values more selectively by using a control chart to detect and improve extreme seasonal-irregular ratios. Two standard errors are selected as the upper and lower control limits of each of the twelve calendar months (the square of the standard error is defined as the average of the squared deviations of the ratios from their corresponding 5-term *MA* values for each calendar month). If a ratio falls outside these limits, it will be taken as "extreme" and is replaced by the average of the "extreme" ratio and the ratios immediately preceding and following. The extreme value is replaced by the average of the first or the last three values of the series if it is the first or the last ratio respectively.

It has been pointed out by Shiskin and Eisenpress (1957) that the 5-month *MA* filter applied in Method I, which is used to compute seasonal adjustment factors, causes erratic year-to-year changes in directions. Method II uses a (3*3)-term *MA* filter, which is equivalent to a 5-term *WMA* filter to overcome the above problem. Moreover, the (3*3)-term *MA* filter is substituted by a

(3*5)-term *MA* filter if there is a strongly marked irregular movement, i.e., more smoothings are needed to be carried out. Unlike Method I, Method II smooths the ratios after they are centered (i.e., adjusted so that their sum is equal to 1200 for each calendar year) to avoid distortions of the smoothed series.

The following steps describe Method II of the Census X-11 procedure for deseasoning a monthly time series by using multiplicative adjustment, this algorithm uses the Spencer 15-month *WMA* filter.

(a) Apply a 12-month *MA* filter to the original data. Center the resulting data by taking a 2-month *MA* filter of the smoothed series. Six missing ratios at each end of the seasonal-irregular series are extrapolated by repeating the first or last available ratios for these months.

(b) Divide the centered 12-month *MA* into the original series to get the seasonal-irregular ratios.

(c) Fit a 5-month *MA* filter to the seasonal-irregular ratios for each of the twelve calendar months. Determine two control limits and modify the extreme values. These series are prior extended by adding two values at both ends, i.e.,

$$Y_{-1} = Y_0 = \frac{Y_1 + Y_2}{2} \text{ and } Y_{n+1} = Y_{n+2} = \frac{Y_n + Y_{n-1}}{2}$$

where Y's are the seasonal-irregular ratios and Y_n is the last record of the series for each calendar month.

(d) Apply a (3*3)-term *MA* filter to the modified series for each calendar month to obtain twelve sets of seasonal factors. Divide the original series by these seasonal factors to obtain the preliminary seasonal adjusted series.

(e) Apply the Spencer 15-month *WMA* filter to this seasonally adjusted series to obtain the second approximation of the trend curve.

(f) Divide the original series by the 15-month *WMA* to obtain the second estimate of the seasonal-irregular series.

(g) Divide the preliminary adjusted series by the 15-month *WMA* to obtain the irregular ratios. Calculate d, the mean of the absolute first differences of these ratios.

(h) Repeat (c)-(d) to get the second approximation of the seasonal adjustment factors.

(i) Divide these seasonal factors into the original series to get the stable-seasonal adjusted series.

(j) Center the ratios $\frac{\text{original}}{\text{15-month WMA}}$ for all the 12 calendar months in each year.

(k) Apply a (3*5)-term *MA* filter or a (3*3)-term *MA* filter to the previously centered ratios for each of the twelve calendar months, depending on whether $d \geq 2$ percent or $d < 2$ percent . This is to obtain the final seasonal adjustment factor.

(l) Divide the final seasonal adjustment factors into the original series to obtain the final seasonally adjusted series.

Some shortcomings of Method II have been pointed out by Y. S. Leong (1962). First of all, the 12-month *MA* filter tends to cut corners at turning points of the cycles, especially where the peaks are steep and valleys are deep, erring in the direction of the curvature. Furthermore, Method II of Shiskin repeats the processes of *MA* and *WMA* filtering for a number of times, some of them are complicated graduation formulas, for example, the Spencer 15-month *WMA* filter. He commented that Method II is tedious and excessive computation is required. To overcome the above mentioned difficulties, Leong introduced the 6-month iterated-*MA* filter with triangular weights, he showed that the trend-cycle curve produced resembled the trend-cycle movements more closely than the 12-month *MA* filter of Method II.

Burman (1965) also listed some serious criticisms about Method II of the Census X-11 procedure. He pointed out that the more logical approach is to perform logarithmic transformation on (2.5). The model is then of additive seasonality,

$$\log Y = \log T + \log S + \log I$$

with the condition that the seasonal factors for the 12 calendar months in a year sum to zero.

In order to compare the applicability of the multiplicative and the additive approaches in our study, a computer program has been written for the additive model also. The algorithm used is based on the discussion given in Bovas Abraham and Johannes Ledolter (1983). The divisions in Method II of the Census X-11 procedure are replaced by subtractions in the additive model, and the Henderson 5, 9, 13 or 23-term *WMA* filters are used to obtain the second estimate of the trend series.

Table 2.3 below exhibits the weights w_j used in the Henderson's moving average filters.

Table 2.3 *Weights w_j Used In The Henderson's Moving Average Filters.*

j	5-term (m=2)	9-term (m=4)	13-term (m=6)	23-term (m=11)
0	0.558	0.330	0.240	0.148
±1	0.294	0.267	0.214	0.138
±2	-0.073	0.119	0.147	0.122
±3		-0.010	0.066	0.097
±4		-0.041	0.000	0.068
±5			-0.028	0.039
±6			-0.019	0.013
±7				-0.005
±8				-0.015
±9				-0.016
±10				-0.011
±11				-0.004

From Figure 2.5, the periodogram of the irregular series produced by the multiplicative adjustment, we observe that there is a significant peak at the low frequency, which indicates that the irregular series follows a low order autoregressive process; whereas in Figure 2.6, the periodogram of the irregular series produced by the additive adjustment, there is a peak at the frequency of one cycle every twelve months (the fundamental frequency), this indicates that there is a 12-month seasonal cycle in the irregular series. Therefore, we conclude that the additive model does not seem to remove the seasonal component of the observed series as effectively as the multiplicative model.

We next perform an logarithmic transformation on the original series and apply additive adjustment. From the periodogram of the irregular series as shown in Figure 2.7, we observe that the seasonal peaks have been reduced significantly as compared to that of Figure 2.6, which its original series was not logarithmic transformed. Figure 2.7 also shows that the irregular series follows a low-order autoregressive model. Therefore, we conclude that the multiplicative model describes the river flows used in this study better than the additive model, and the model of the series must be identified correctly before applying the appropriate adjustment.

For additive adjustment, we used the Henderson 13-term *WMA* filter to estimate the trend series of our data. In our investigation, we found that the 9-term and 23-term filters did not capture the movements of our data as closely to the original series as the 13-term filter. Figure 2.8 exhibits the trend series obtained by using the Henderson 13-term *WMA* filter. The movement of the estimated trend series is smooth and seems to follow the movement of the original series quite closely.

All of these linear smoothing filters have often been criticized for not having an explicit model for the original series. Moreover, the estimates of the observations at both ends of the series do not have the same degree of reliability as those of the central observations. This is because for a moving average filter of length $2m+1$, the first and the last m observations cannot be smoothed with the same set of symmetric weights as are applied to the central of the series (Dagum, 1978). As a consequence, the Department of Statistics, Canada has developed a composite method, *X-11-ARIMA*, which uses *ARIMA* model to forecast and backcast the data before the extended series is to be seasonally adjusted by the filters of Method II in the Census *X-11* procedure.

Dagum (1978) pointed out that the X-11-*ARIMA* performs better than the Census X-11 procedure except for very irregular series. However, the selected model must fit the data well and generate "reasonable" forecasts for the last three years of the observed data, which according to Dagum are forecasts with a mean absolute error smaller than 5 per cent for well behaved series, and smaller than 10 per cent for very noisy series. The portmanteau test is used to test the goodness of fit of the model. The following three *ARIMA* models were incorporated into the X-11-*ARIMA* program,

(a) $(1-B)(1-B^s)(1-\phi_1 B-\phi_2 B^2)Y_t = (1-\theta_1 B-\theta_2 B^2)(1-\theta_{s1}B^s)a_t$

(b) $(1-B^s)(1-\phi_1 B-\phi_2 B^2)Y_t = (1-\theta_1 B)(1-\theta_{s1}B^s-\theta_{s2}B^{2s})a_t$

(c) $(1-B)(1-B^s)(1-\phi_1 B-\phi_2 B^2)\log Y_t = (1-\theta_1 B)(1-\theta_{s1}B^s-\theta_{s2}B^{2s})a_t$

where s is the length of seasonality in the series.

Finally, Cleveland and Tiao (1976) have shown that the use of linear filters in the Census-X-11 procedure for time series decomposition is justified if the series follows the *ARIMA* model

$$
\begin{aligned}
&(1-B)(1-B^{12})Y_t \\
&\quad = \frac{(1-B^{12})}{(1-B)}(1+0.49B-0.49B^2)b_{1t} + (1-B)(1+0.64B^{12}+0.83B^{24})b_{2t} \\
&\quad\quad + (1-B)(1-B^{12})e_t
\end{aligned}
\tag{2.6}
$$

where b_{1t}, b_{2t} and e_t are three independent white noise processes which are Normally distributed as $N(0,\ \sigma^2_{b1})$, $N(0,\ \sigma^2_{b2})$ and $N(0,\ \sigma^2_e)$ respectively with $\sigma^2_{b2}/\sigma^2_{b1} = 1.3$ and $\sigma^2_e/\sigma^2_{b1} = 14.4$. They have shown that the conditional expectations of the trend and seasonal components of this model have the same weights as those of the filters that are used in the Census X-11 procedure. In other words, this model represents the underlying stochastic

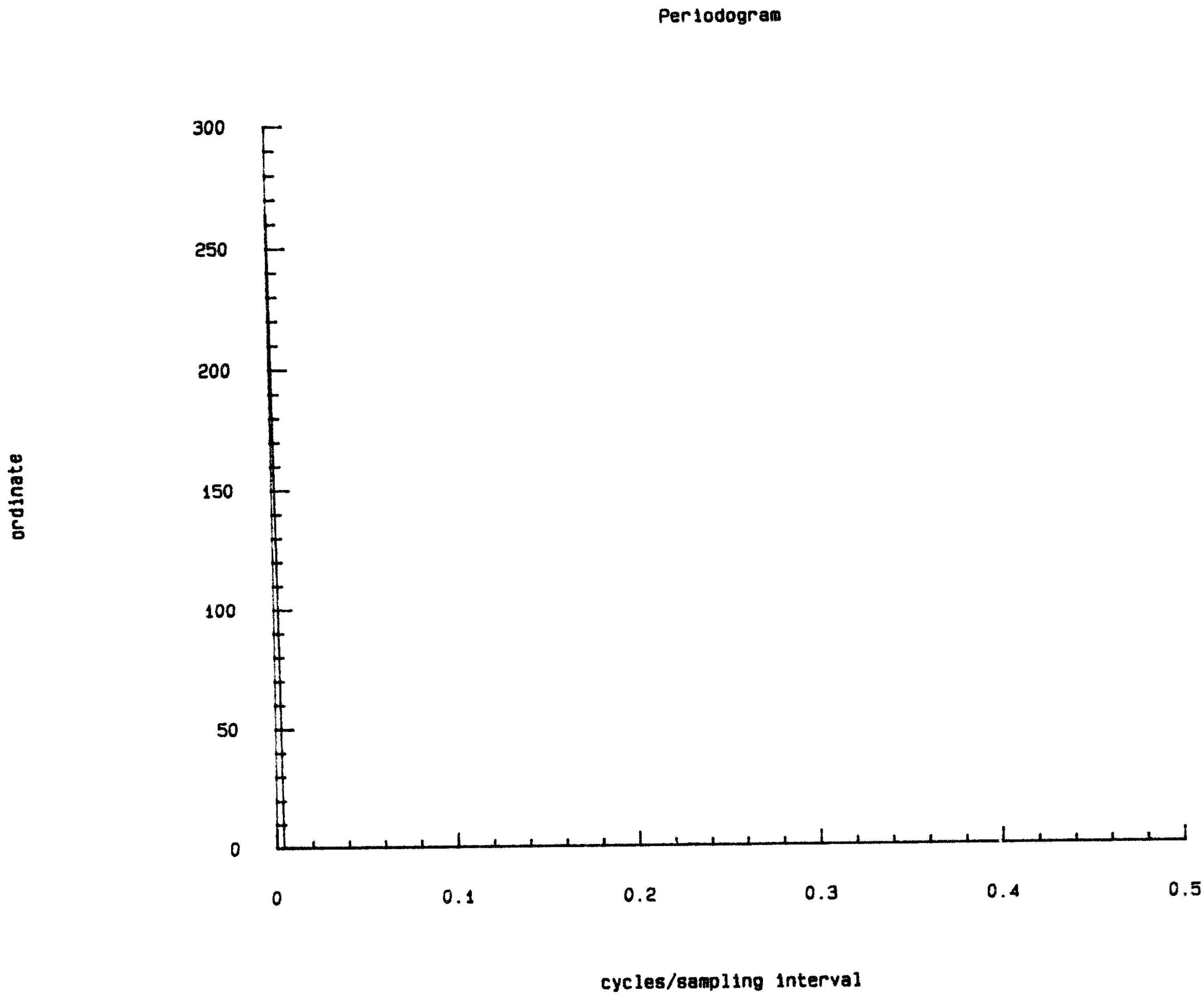

Figure 2.5 Periodogram Of The Irregular Component Of Perak River Flows Obtained By Multiplicative Model

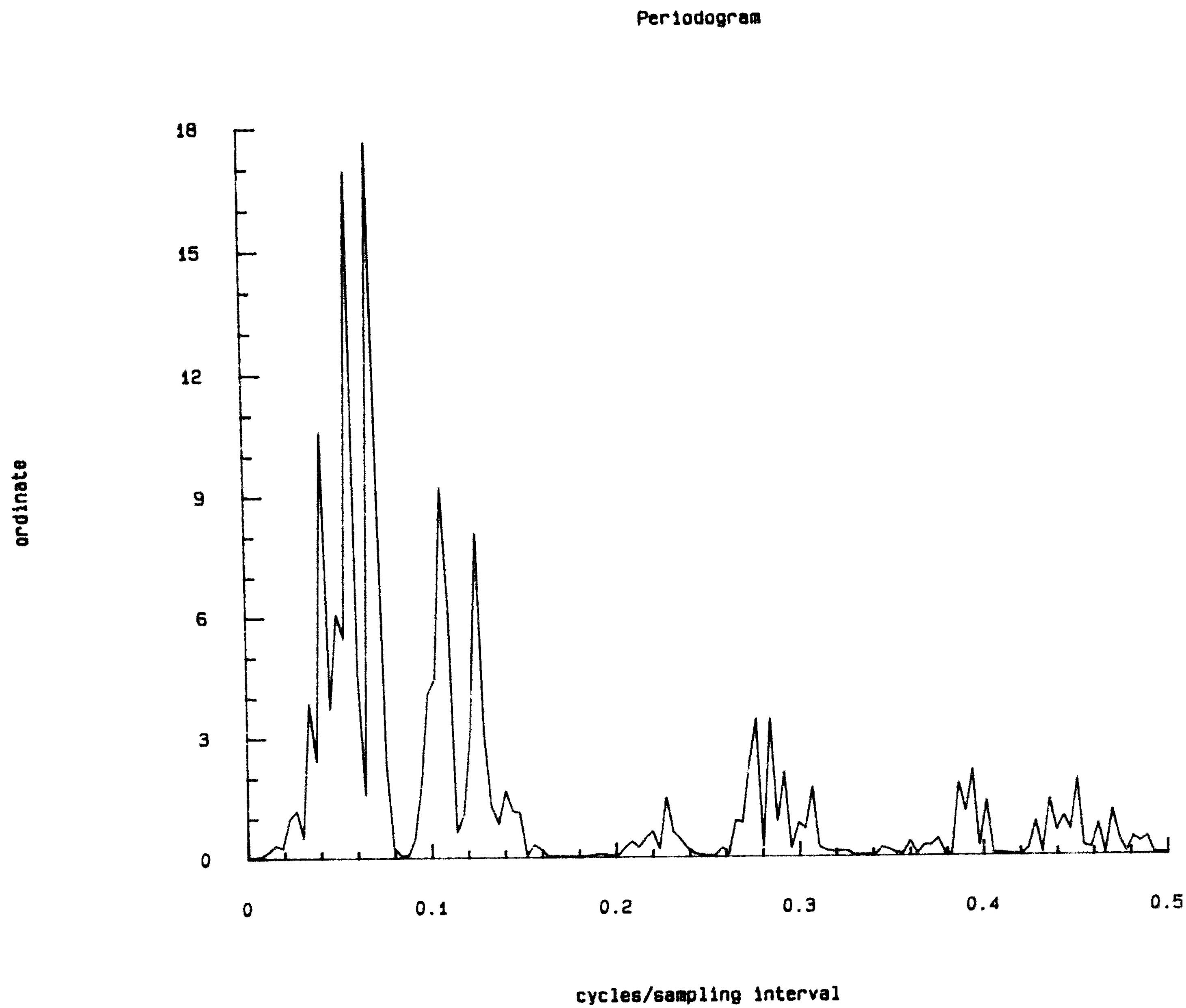

Figure 2.6 Periodogram Of The Irregular Component Of Perak River Flows Obtained By Additive Model

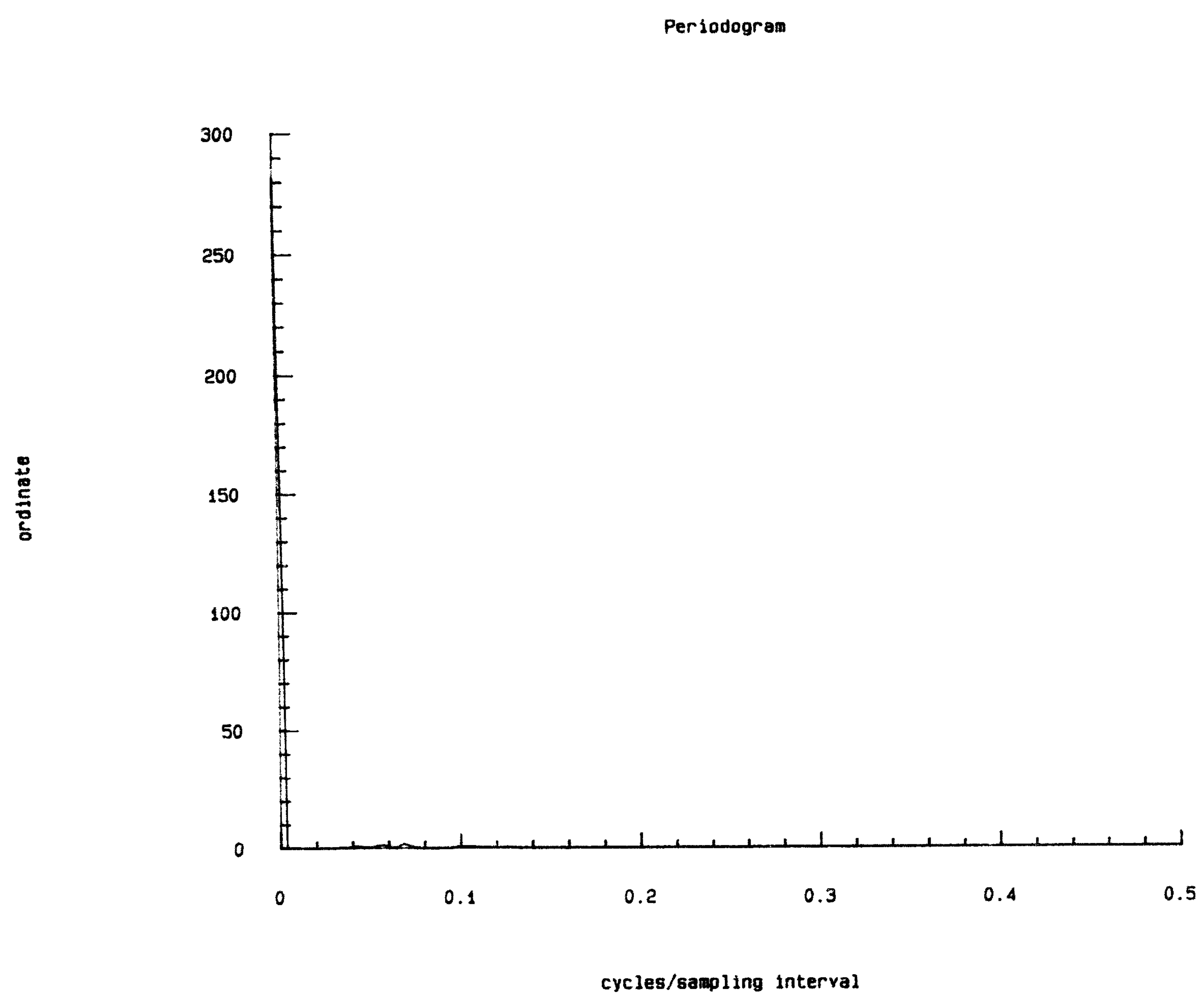

Figure 2.7 Periodogram Of The Irregular Component Of The Logarithm Transformation Of Perak River Flows Obtained By Additive Model

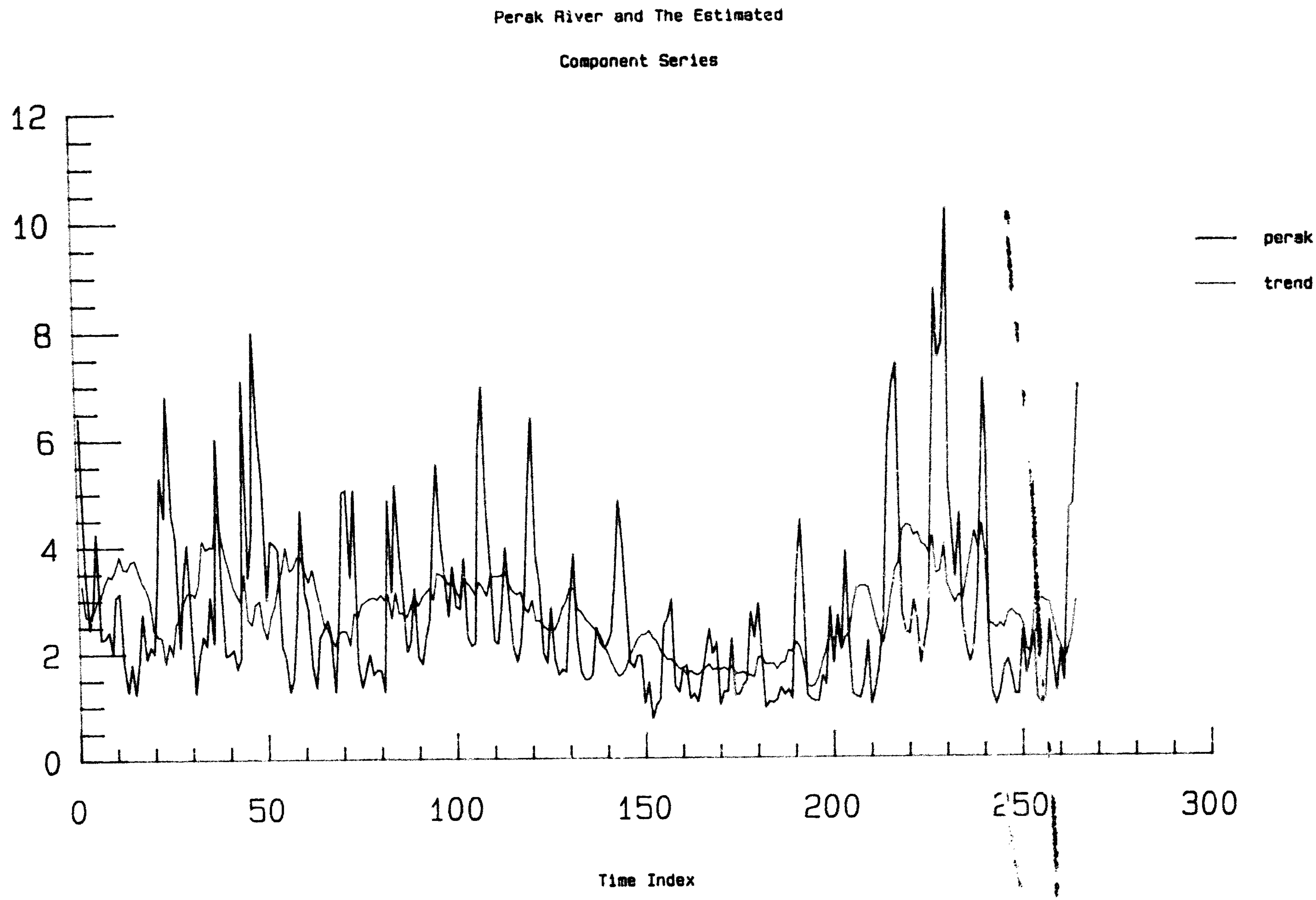

Figure 2.8 The Estimated Trend Series Obtained By Henderson 13-Term Moving Average

mechanism of those filters. However, if the model of the concerned series differs markedly from model (2.6), then the decomposition method of the Census X-11 procedure is not reliable. Therefore the usefulness of the Census X-11 procedure is limited, because it applies the same filters to all observed series.

2.4 Model-Based Approach

In the model-based approach, it is assumed that the original data Y_t can be decomposed additively as

$$Y_t = S_t + T_t + I_t$$

where S_t, T_t and I_t are respectively the seasonal, trend and irregular components of the data.

The most primitive model-based approach was introduced in the 1930's. It was the regression method where the trend and seasonal components were fitted to functions that depend linearly on some parameters. Normally, the trend component was described by polynomials in time, the seasonal component was represented by seasonal means, and the irregular component was assumed to be white noise. The seasonal component was then subtracted from the original series to yield a seasonally adjusted series. A drawback of this technique was that it required explicit specifications of the mathematical form for the trend and seasonal components.

Recently, many researchers have used the Gaussian *ARIMA* models (Box and Jenkin, 1970) in the model-based approach. An *ARIMA* stochastic model is first fitted to the original series and then the three components S_t, T_t and I_t are estimated by using the theory of signal extraction. A drawback of this technique is that the covariance structures of the seasonal and non-seasonal components of the original series must be known priorly even though these components are not observable. Therefore, any *ARIMA*-model-based approach must make assumptions that uniquely specify the covariance structures of both the seasonal and non-seasonal components of the series.

Some of the basic assumptions are listed below :

(a) S_t follows an unknown *ARIMA* model $\phi_S(B)S_t = \eta_S(B)b_t$ where $\phi_S(B) = 1 + B + \ldots + B^{s-1}$ with s denoting the length of seasonality and b_t is a white noise series with distribution $N(0, \sigma_b^2)$. The order of $\eta_S(B)$ is at most s-1.

(b) T_t follows an unknown *ARIMA* model $\phi_T(B)T_t = \eta_T(B)c_t$ where $\phi_T(B) = (1-B)^d$ with $\eta_T(B)$ of degree at most d and c_t is a white noise series with distribution $N(0, \sigma_c^2)$.

(c) I_t follows an unknown *ARIMA* model $\phi_I(B)I_t = \eta_I(B)d_t$ with d_t distributed as $N(0, \sigma_d^2)$.

(d) b_t, c_t and d_t are mutually independent.

(e) Each pair of the autoregressive and moving average polynomials of the three component models have their zeros lying on or outside the unit circle and have no common zeros.

With the above assumptions, it can then be readily shown that the overall model for Y_t is the *ARIMA* model

$$\Phi(B)Y_t = \Theta(B)a_t \tag{2.7}$$

where

$$\Phi(B) = (1-B)^d\phi_S(B)\phi_I(B),$$

and the moving average polynomial $\theta(B)$ and innovation variance σ_a^2 satisfy the constraint

$$\frac{\theta(B)\theta(F)}{\Phi(B)\Phi(F)}\sigma_a^2 = \frac{\eta_S(B)\eta_S(F)}{\phi_S(B)\phi_S(F)}\sigma_b^2 + \frac{\eta_T(B)\eta_T(F)}{\phi_T(B)\phi_T(F)}\sigma_c^2 + \frac{\eta_I(B)\eta_I(F)}{\phi_I(B)\phi_I(F)}\sigma_d^2 \tag{2.8}$$

where $F = B^{-1}$, that is $FY_t = Y_{t+1}$.

Equation (2.8) can be easily verified by noting that as S_t, T_t and I_t are assumed to be mutually uncorrelated, the covariance generating function of Y_t can be written as

$$\gamma_y(B, F) = \gamma_S(B, F) + \gamma_T(B, F) + \gamma_I(B, F)$$

where

$$\gamma_S(B, F) = \frac{\eta_S(B)\eta_S(F)}{\phi_S(B)\phi_S(F)}\sigma_b^2,$$

$$\gamma_T(B, F) = \frac{\eta_T(B)\eta_T(F)}{\phi_T(B)\phi_T(F)}\sigma_c^2,$$

and

$$\gamma_I(B, F) = \frac{\eta_I(B)\eta_I(F)}{\phi_I(B)\phi_I(F)}\sigma_d^2$$

are respectively the covariance generating functions of S_t, T_t and I_t. The covariance generating function of a process $\{X_t\}$ is defined as $\gamma_X(Z) = \sum_{-\infty}^{\infty} \gamma_k Z^k$ where $\{\gamma_k\}$ is the autocovariance function of the process $\{X_t\}$.

Clearly, given the model for Y_t, the models for S_t, T_t and I_t are not unique. In other words, given σ_a^2, $\theta(B)$ and $\Phi(B)$ ($\phi_S(B)$, $\phi_T(B)$ and $\phi_I(B)$), there are more than one ways of choosing $\eta_S(B)$, $\eta_T(B)$, $\eta_I(B)$ and the innovation variances σ_b^2, σ_c^2 and σ_d^2. Any choice of the above three moving average polynomials and innovation variances that satisfies (2.8) is called an acceptable decomposition. Hillmer and Tiao (1982) have proved the result that among acceptable decompositions, the decomposition that maximizes σ_d^2 subject to the constraints in (2.8) is unique and it is known as canonical decomposition.

The canonical decomposition can also be defined as the decomposition that minimizes the innovation variances σ_b^2 and σ_c^2, i.e., makes the seasonal and trend components as deterministic as possible while remains consistent with the information contained in the observable series Y_t.

The partial fraction of the left-hand side of equation (2.8) is an unique decomposition, i.e.,

$$\frac{\theta(B)\theta(F)}{\Phi(B)\Phi(F)}\,\sigma_a^2 = \frac{Q_S(B,\ F)}{\phi_S(B)\phi_S(F)} + \frac{Q_T(B,\ F)}{\phi_T(B)\phi_T(F)} + \frac{Q_I(B,\ F)}{\phi_I(B)\phi_I(F)} \tag{2.9}$$

where

$$Q_S(B,\ F) = q_{0S} + \sum_{j=1}^{s-2} q_{jS}\ (B^j + F^j),$$

$$Q_T(B,\ F) = q_{0T} + \sum_{j=1}^{d-1} q_{jT}\ (B^j + F^j)$$

and $Q_I(B,\ F)$ can be obtained by subtraction. This partial fraction is unique because the degrees of the numerators are lower than the degrees of the corresponding denominators.

By writing $B = e^{-i\omega}$ and $F = e^{i\omega}$ with $0 \le \omega \le \pi$, the functions $Q_S(B,\ F)$ and $Q_T(B,\ F)$ can be obtained by converting $\theta(B)\theta(F)$ and $\Phi(B)\Phi(F)$ into functions of x, where $x = \frac{1}{2}(e^{-i\omega}+e^{i\omega}) = \cos\omega$, and by applying the result that if a, b, ..., m are roots of the equation $f(x) = 0$ and α, β, ..., μ are their corresponding multiplicities so that $f(x) = (x-a)^{\alpha}(x-b)^{\beta}\ldots(x-m)^{\mu}$, then $\frac{\varphi(x)}{f(x)}$ can be decomposed as

$$\frac{\varphi(x)}{f(x)} = \frac{A_{\alpha}}{(x-a)^{\alpha}} + \frac{A_{\alpha-1}}{(x-a)^{\alpha-1}} + \ldots + \frac{A_1}{(x-a)} + \ldots$$

$$+ \frac{M_{\mu}}{(x-m)^{\mu}} + \frac{M_{\mu-1}}{(x-m)^{\mu-1}} + \ldots + \frac{M_1}{(x-m)}$$

where $A_{\alpha-k+1} = \dfrac{\psi_1^{k-1}(a)}{(k-1)!}$, ..., $M_{\alpha-k+1} = \dfrac{\psi_m^{k-1}(m)}{(k-1)!}$

with $\psi_1(x) = \dfrac{\varphi(x)(x-a)^{\alpha}}{f(x)}$, ..., $\psi_m(x) = \dfrac{\varphi(x)(x-m)^{\mu}}{f(x)}$.

We shall use the model $(1-B^S)Y_t = (1-\theta_2 B^S)a_t$ to illustrate how to obtain the function $Q_T(B, F)$. In this case, $\Phi(B) = (1-B^S) = (1-B)\phi_S(B)$. Hence

$$f(\cos\omega) = \Phi(e^{-i\omega})\Phi(e^{i\omega})$$

$$f(x) = (1-e^{-i\omega})(1-e^{i\omega})\phi_S(e^{-i\omega})\phi_S(e^{i\omega})$$

$$= -2(x-1)\phi_S(e^{-i\omega})\phi_S(e^{i\omega})$$

and $x = \cos\omega = 1$ is one of the roots of $f(x) = 0$. The function $Q_T(B, F)$ can then be evaluated as

$$Q_T(B, F) = \frac{-2\varphi(x)}{f'(x)}\Big|_{x=1} = \frac{(1-\theta_2)^2}{s^2}$$

where $\varphi(x) = \varphi(\cos\omega) = (1-\theta_2 e^{-i\omega})(1-\theta_2 e^{i\omega}) = 1 - 2\theta_2 x + \theta_2^2$.

Note that $\phi_S(e^{-i\omega})\phi_S(e^{i\omega}) = s^2$ when $\cos\omega = 1$.

Hillmer and Tiao (1982) proved that this unique decomposition is acceptable if $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 \geq 0$ where

$$\varepsilon_1 = \min_{\omega} \frac{Q_S(e^{-i\omega}, e^{i\omega})}{|\phi_S(e^{-i\omega})|^2} ,$$

$$\varepsilon_2 = \min_{\omega} \frac{Q_T(e^{-i\omega}, e^{i\omega})}{|\phi_T(e^{-i\omega})|^{2d}}$$

and

$$\varepsilon_3 = \min_{\omega} \frac{Q_I(e^{-i\omega}, e^{i\omega})}{|\phi_I(e^{-i\omega})|^2} .$$

By comparing (2.8) and (2.9), the three fractions on the right hand side of (2.8) can be written respectively as

$$\frac{|\eta_S(e^{-i\omega})|^2\sigma_b^2}{|\phi_S(e^{-i\omega})|^2} = \frac{Q_S(e^{-i\omega}, e^{i\omega})}{|\phi_S(e^{-i\omega})|^2} + \Gamma_1 \geq 0 ,$$

$$\frac{|\eta_T(e^{-i\omega})|^2\sigma_c^2}{|\phi_T(e^{-i\omega})|^{2d}} = \frac{Q_T(e^{-i\omega}, e^{i\omega})}{|\phi_T(e^{-i\omega})|^{2d}} + \Gamma_2 \geq 0$$

and

$$\frac{|\eta_I(e^{-i\omega})|^2\sigma_d^2}{|\phi_I(e^{-i\omega})|^2} = \frac{Q_I(e^{-i\omega}, e^{i\omega})}{|\phi_I(e^{-i\omega})|^2} + \Gamma_3 \geq 0$$

where Γ_1, Γ_2 and Γ_3 are three constants such that $\Gamma_1 + \Gamma_2 + \Gamma_3 = 0$ and they provide a means to change from the initial partial fractions decomposition (2.9) to an acceptable decomposition. By using the definition that canonical decomposition is a decomposition which minimizes the innovation variances, an unique canonical decomposition is then implied by the fact that $\Gamma_i + \varepsilon_i = 0$ for $i = 1, 2, 3$. Therefore (2.9) can be rearranged as

$$\frac{\theta(B)\theta(F)}{\Phi(B)\Phi(F)} = \frac{Q_S^*(B, F)}{\phi_S(B)\phi_S(F)} + \frac{Q_T^*(B, F)}{\phi_T(B)\phi_T(F)} + \frac{Q_I^*(B, F)}{\phi_I(B)\phi_I(F)} \qquad (2.10)$$

where

$$Q_S^*(B, F) = Q_S(B, F) - \phi_S(B)\phi_S(F)\varepsilon_1, \tag{2.11}$$

$$Q_T^*(B, F) = Q_T(B, F) - \phi_T(B)\phi_T(F)\varepsilon_2 \tag{2.12}$$

and

$$Q_I^*(B, F) = Q_I(B, F) + \phi_I(B)\phi_I(F)\varepsilon_3.$$

When both the seasonal and trend components and the overall series are *ARIMA* models, the minimum mean squared error (*MSE*) estimates of the seasonal and trend components S_t and T_t are respectively $\hat{S} = W_S(B, F)Y_t$ and $\hat{T} = W_T(B, F)Y_t$ (Tiao and Hillmer, 1978), where

$$W_S(B, F) = \frac{\gamma_S(B, F)}{\gamma_Y(B, F)} = \frac{\sigma_b^2\ \Phi(B)\Phi(F)\eta_S(B)\eta_S(F)}{\sigma_a^2\ \theta(B)\theta(B)\phi_S(B)\phi_S(F)} \tag{2.13}$$

$$W_T(B, F) = \frac{\gamma_T(B, F)}{\gamma_Y(B, F)} = \frac{\sigma_c^2\ \Phi(B)\Phi(F)\eta_T(B)\eta_T(F)}{\sigma_a^2\ \theta(B)\theta(F)\phi_T(B)\phi_T(F)} \tag{2.14}$$

and $\gamma_Y(B, F)$, $\gamma_S(B, F)$, $\gamma_T(B, F)$ are the covariance generating functions of Y_t, S_t and T_t respectively. The estimate of the irregular component is obtained by subtracting the seasonal and trend components from the original series, namely, $\hat{I}_t = Y_t - \hat{T}_t - \hat{S}_t$.

By comparing the corresponding terms in (2.8) and (2.10), and then substitute them in (2.13) and (2.14), we have

$$W_S(B, F) = \frac{\Phi(B)\Phi(F)Q_S^*(B, F)}{\theta(B)\theta(F)\phi_S(B)\phi_S(F)}$$

and

$$W_T(B, F) = \frac{\Phi(B)\Phi(F)Q_T^*(B, F)}{\theta(B)\theta(F)\phi_T(B)\phi_T(F)}.$$

where Q_S^* , Q_T^* can be evaluated from (2.11) and (2.12), and $\Phi(B)$, $\theta(B)$ can be estimated from the *ARIMA* model fitted to $\{Y_t\}$ in (2.7).

Hillmer and Tiao (1982) have considered the seasonal decomposition of the following three seasonal models,

(a) $(1-B^s)Y_t = (1-\theta_2 B^s)a_t$,

(b) $(1-B)(1-B^s)Y_t = (1-\theta_1 B)(1-\theta_2 B^s)a_t$ and

(c) $(1-B^s)Y_t = (1-\theta_1 B)(1-\theta_2 B^s)a_t$.

By comparing the autoregressive polynomial of the overall model with $(1-B)^d\phi_S(B)\phi_I(B)$, the value of $\phi_I(B)$ is found to be unity for all the three cases. They derived the corresponding canonical decompositions for all the three models and their study showed that for the time series of Monthly U. S. Unemployed Males From Age 16 to 19 (January, 1971 - August 1979), the *ARIMA*-model-based approach has intuitively pleasing results.

The model $(1-B)(1-B^{12})Y_t = (1-0.33059B)(1-0.82850B^{12})a_t$ is fitted to the monthly flows of Perak river with the standard errors attached to 0.33059 and 0.82850 being, respectively 0.05902 and 0.03555, and $\hat{\sigma}_a^2 = 1.23443$. We shall use this model to illustrate the performance of the *ARIMA*-model-base approach. Since the Perak river flows are monthly data, we shall let $s = 12$ for the following discussions.

From the results obtained in Hillmer and Tiao (1982), the canonical decomposition of the above model is

$$\frac{(1-\theta_1 B)(1-\theta_1 F)(1-\theta_2 B^s)(1-\theta_2 F^s)}{(1-B^s)(1-F^s)(1-B)(1-F)}$$

$$= \frac{Q_S^*(B,\ F)}{\phi_S(B)\phi_S(F)} + \frac{Q_T^*(B,\ F)}{(1-B)^2(1-F)^2} + \theta_2 \frac{(1+\theta_1)^2}{4}$$

where

$$Q_S^*(B,\ F) = Q_S(B,\ F) + \phi_S(B)\phi_S(F)\left\{C - \frac{\theta_2(1-\theta_1)^2}{4}\right\}$$

$$Q_T^*(B,\ F) = Q_T(B,\ F) - C(1-B)^2(1-F)^2$$

with

$$Q_T(B,\ F) = \frac{(1-\theta_1)^2(1-\theta_2)^2}{s^2}$$

$$\cdot\left\{1 + \left[\frac{\theta_2 s^2}{(1-\theta_2)^2} + \frac{(1+\theta_1)^2}{4(1-\theta_1)^2} + \frac{(s^2-4)}{12}\right](1-B)(1-F)\right\},$$

$$Q_S(B,\ F) = \frac{(1-\theta_1^2)\ (1-\theta_2^2)}{(1-B)^2(1-F)^2}\left\{1-\frac{1}{s^2}\ \phi_S(B)\phi_S(F)\right\}$$

$$+ \theta_1(1-\theta_2)^2 \frac{(1-B)(1-F)}{(1-B)^2(1-F)^2}$$

$$- (1-\theta_2)^2\ \frac{(1-B^s)(1-F^s)}{(1-B)^2(1-F)^2}\left\{\frac{s^2-4}{12s^2}(1-\theta_1)^2 + \frac{(1+\theta_1)^2}{4s^2}\right\}$$

and

$$C = \frac{(1-\theta_2)^2}{48s^2}\left\{(1-\theta_1)^2(s^2-1) + 3(1+\theta_1)^2\right\} + \frac{\theta_2(1-\theta_1)^2}{4}\ .$$

From Figure 2.9, the estimated trend component of Perak river by using the *ARIMA*-model-based approach, we observe that the estimated trend component follows the movements of its original series quite closely. Figure 2.10 exhibits the plot of the estimated seasonal component of Perak river.

By just looking at Figure 2.9, we would think that the model-based approach gives pleasing results and most of the authors also did not go further in the study of the characteristics of the estimated trend component. However, from our investigation, we notice that there is a significant peak at the 12th lag of the sample *ACF* of the estimated trend component $\hat{T}_t$ (as shown in Figure 2.11). From Figure 2.12, the sample *ACF* of the postulated trend series $(1-B)^2\hat{T}_t$ also exhibits peaks at seasonal lags. Hence the estimated trend component does not exhibit the behavior that is close to that expected by the postulated model $\phi_T(B)T_t = \eta_T(B)c_t$ where $\phi_T(B) = (1-B)^2$ and $\eta_T(B)$ is a polynomial in B with degree at most 2.

To clarify the above unexpected results, we have written to Hillmer and based on his reply (1989), we note that the covariance generating functions of T_t and $\hat{T}_t$ can be obtained respectively as

$$\gamma_T(B,\ F) = \frac{\sigma_c^2\ \eta_T(B)\eta_T(F)}{\phi_T(B)\phi_T(F)}$$

$$\gamma_{\hat{T}}(B,\ F) = \gamma_T(B,\ F)\frac{\sigma_c^2\ \eta_T(B)\eta_T(F)\phi_S(B)\phi_S(F)}{\sigma_a^2\ \theta(B)\theta(F)}\ .$$

Hence, it is clear that the covariance generating functions of the estimated trend component $\gamma_{\hat{T}}(B,\ F)$ has seasonal lags in the operators $\theta(B)$, $\theta(F)$ and $\phi_S(B)$, $\phi_S(F)$ that appear in $\gamma_T(B, F)$. As a result, the sample *ACF* of the estimated trend has an obvious seasonal peak at the 12th lag and the phenomenon is then explained.

From the above discussion, it is clear that the model-based approach is not suitable for deseasoning the observed Perak river flows because the estimated trend component does not seem to have the expected characteristics.

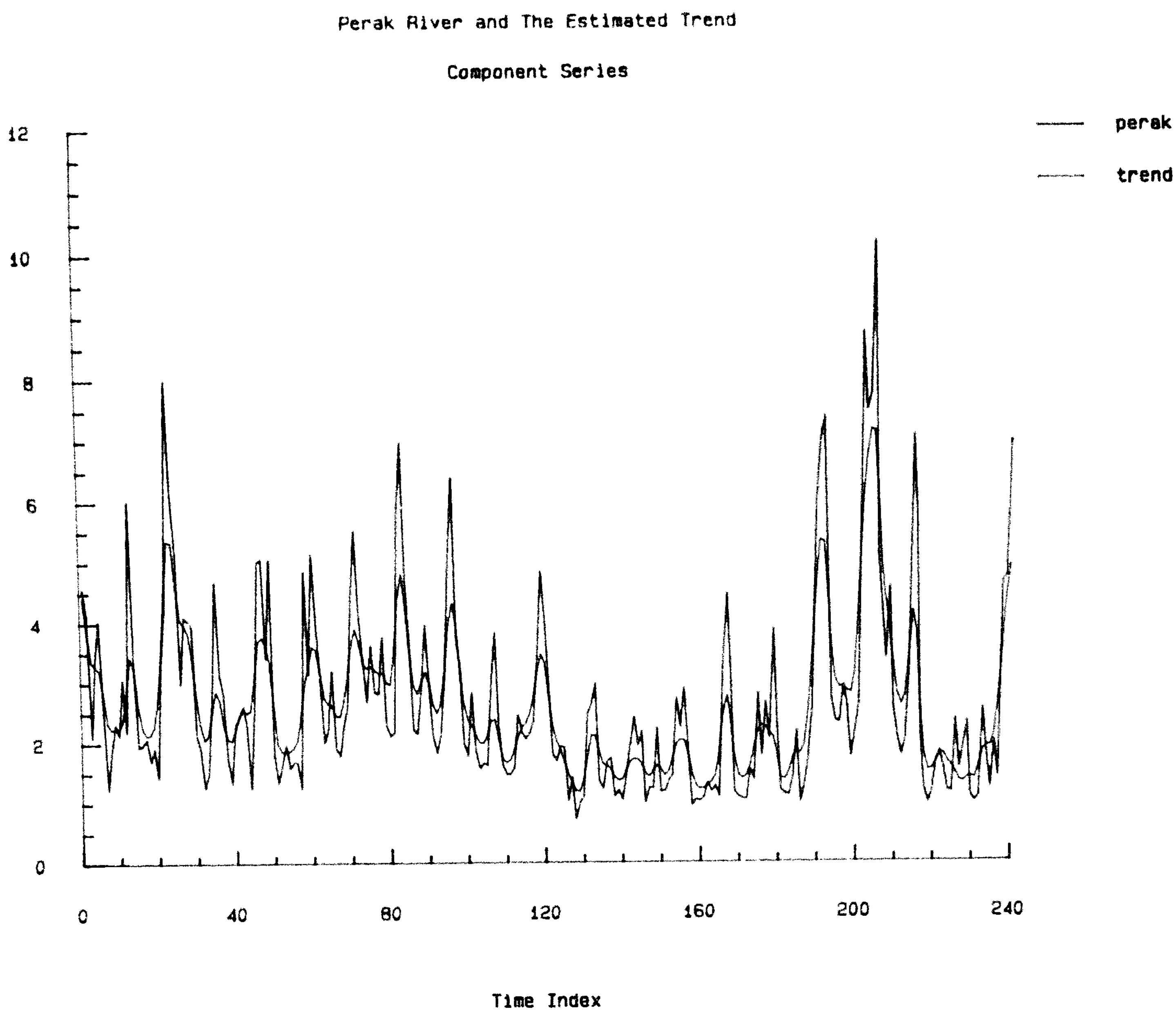

Figure 2.9 The Estimated Trend Series Obtained By Using The ARIMA-Model-Based Approach

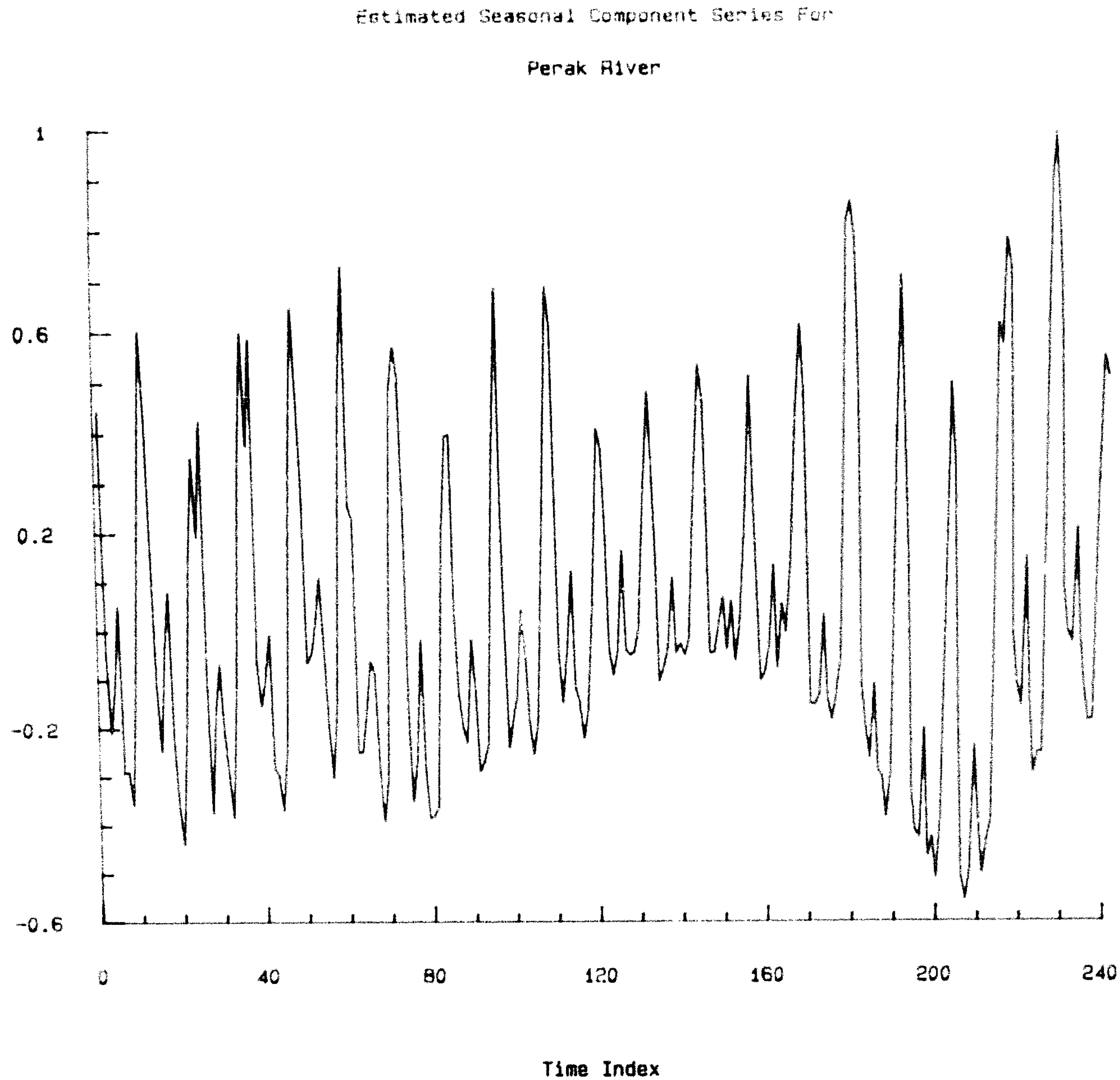

Figure 2.10 The Estimated Seasonal Series Obtained By Using The ARIMA-Model-Based Approach

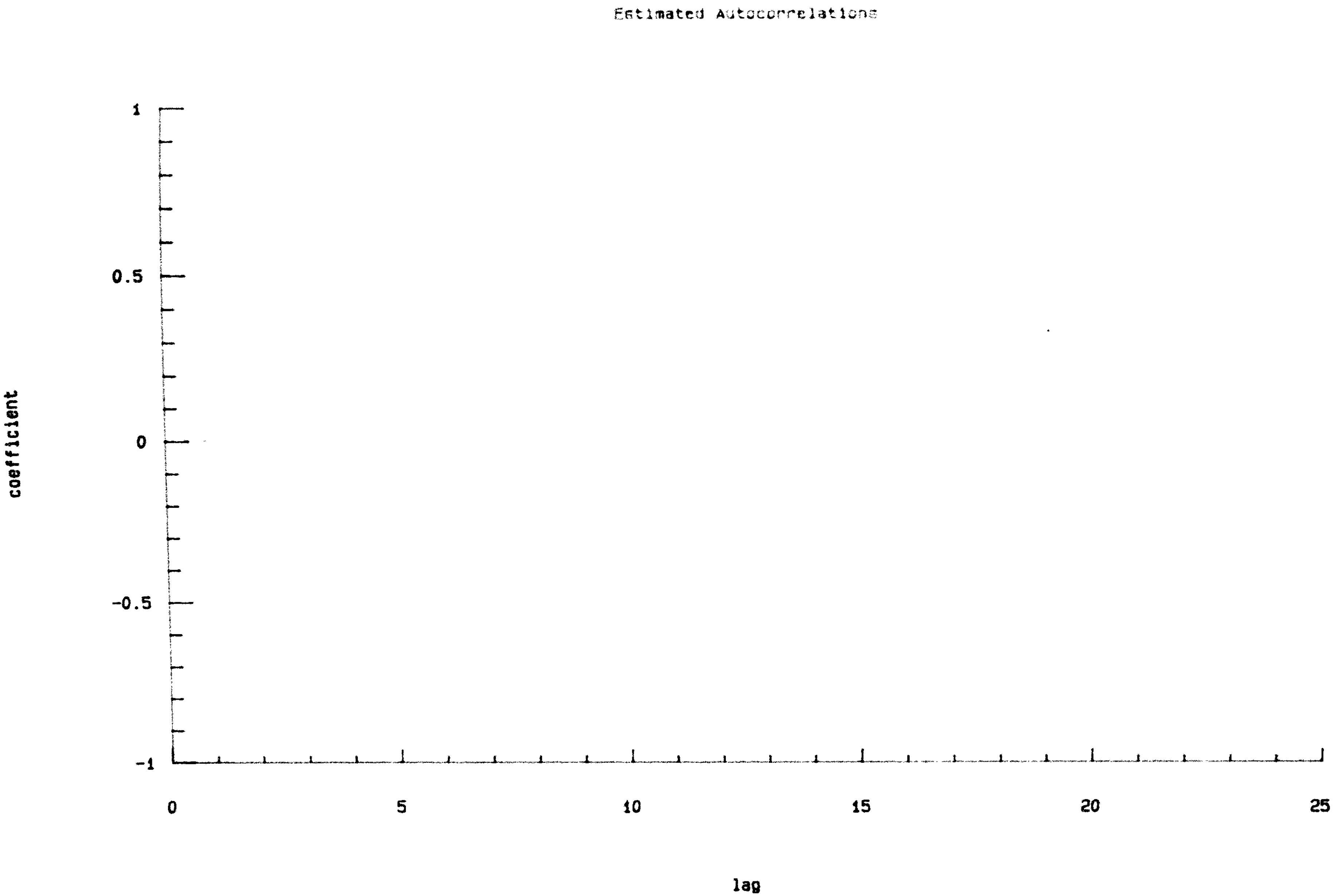

Figure 2.11 Sample ACF Of The Trend Component Obtained By Using The ARIMA-Model-Based Approach

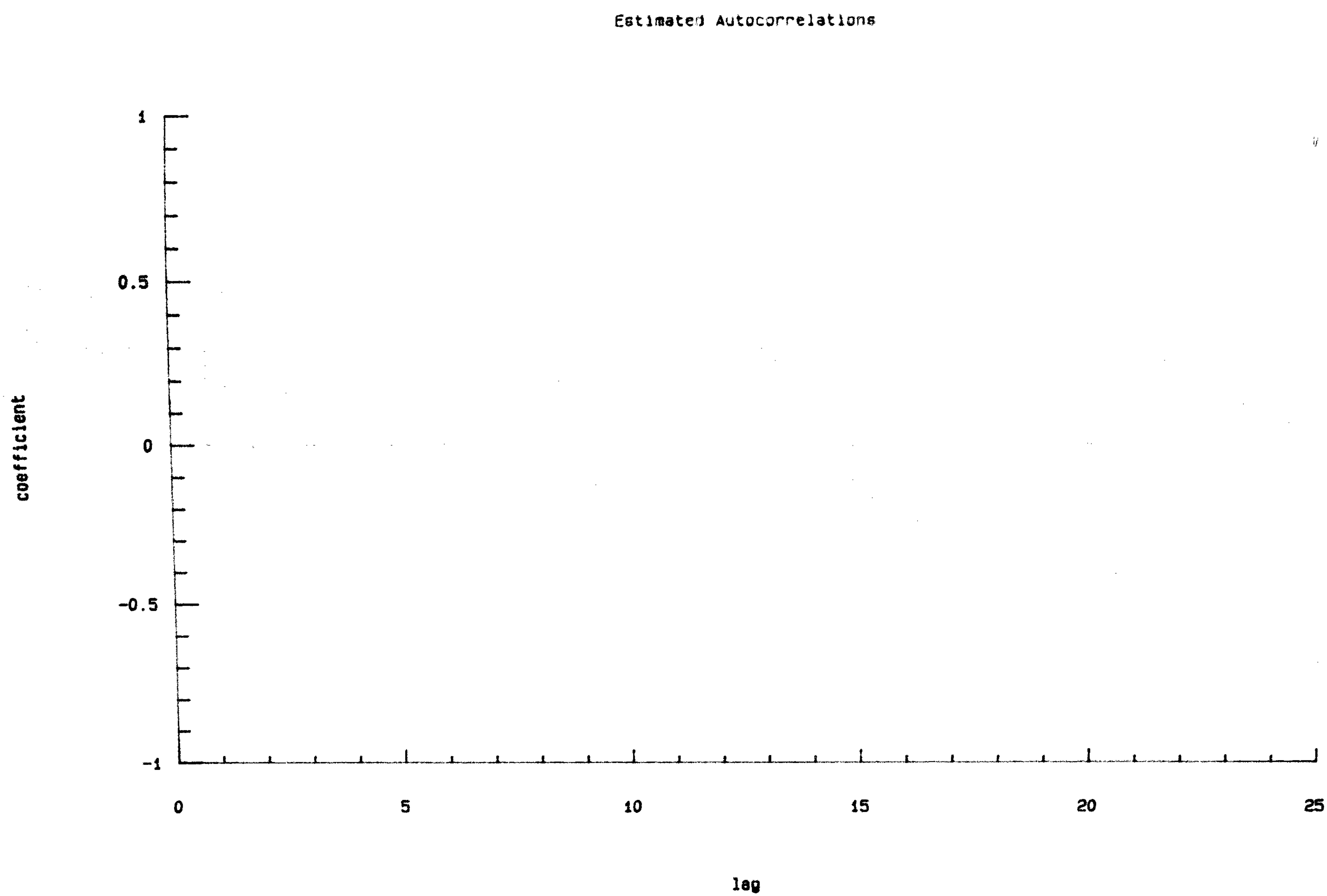

Figure 2.12 Sample ACF Of $(1-B)^2\hat{T}_t$, Where $\hat{T}_t$'s Are Obtained By Using The ARIMA-Model-Based Approach