PUEY UNGPHAKORN INSTITUTE
FOR ECONOMIC RESEARCH

Simplified Spectral Analysis and Linear Filters for Analysis of Economic Time Series.

by

Supachoke Thawornkaiwong

April 2016

Discussion Paper

No. 25

The opinions expressed in this discussion paper are those of the author(s) and should not be attributed to the Puey Ungphakorn Institute for Economic Research.

Simplified Spectral Analysis and Linear Filters for Analysis of Economic Time Series.

Supachoke Thawornkaiwong*
Bank of Thailand

December 21, 2015

Abstract

We develop and simplify spectral analysis of time series. The main focus is on the spectral representation theorem, Bochner's theorem, and some key results concerning time-invariant linear filters. We then show how to apply these key results to shed some light on various applications including Yule-Slutsky effects, seasonal adjustment and trend estimation. We also show how spectral analysis can indicate appropriateness of certain statistical models when applied with some economic time series.

Keywords: Spectral Analysis; Linear Filters; Exploratory Data Analysis; Yule-Slutzky Effect; Seasonality; Trend Estimation; HP Filters.

1 Introduction

Most economic time series, particularly macroeconomic ones, exhibit various forms of fluctuation. First time readers of these time series often find such fluctuation striking as well as puzzling. Time series analysts have long decomposed such fluctuation into three main categories, namely, a slowly changing component known as a trend component, a component with more or less stable period known as a seasonal component, and the remaining part with rather random movement known as an irregular component. From an economic point of view, a trend component is usually thought to reflect the long-run dynamics of a time series due to some fundamental factors, such as structural or institutional transformation of an economy or a company. Many factors, such as weather, harvesting seasons and calendar or fiscal year, can result in seemingly predictable fluctuation of a seasonal component. An irregular component seems to be the least predictable part and is often associated with effects of business cycles.

A trend component is often the main object of interest of policy makers responsible for the long-run performance of an economy, or of company directors concerning with the long-term success of their business. An irregular component is also of an interest to policy

E-mail address: s.thawornkaiwong@gmail.com

makers monitoring stability of an economy and the business cycles. Given the importance of these two components, a great deal of effort have been made by both economists and policy makers over a century to obtain a desirable decomposition of economic time series.

A typical procedure to get such decomposition is to first remove a seasonal component from the data by employing seasonal adjustment programmes. Then a trend component is estimated by some low-pass filters and the remaining part is an irregular component. Various smoothing techniques are employed in each of the three steps. From a mathematical point of view, all popular smoothing techniques can be regarded as time-invariant linear filters. The effects of these smoothing techniques can be understood given a working knowledge of the well-known theory of time-invariant linear filters. However, the main barrier to popularity of this elegant theory is the presumed knowledge of spectral analysis of time series.

Classical time series analysis can be divided into two main approaches, namely, time-domain and spectral-domain analysis. It can be shown, and will be demonstrated in this paper, that the two approaches are two sides of the same coin. If one wishes to understand a particular time series, one can either employ the time-domain or spectral-domain analysis in the study. These two approaches complement rather than competing against each other. The early development of time series analysis relied heavily on spectral analysis but the time-domain approach has gained its popularity particularly since the publication of the Box and Jenkins (1970) methodology. The success of the Box-Jenkins methodology is a result of the focus on relatively simple autoregressive-moving average (ARMA) models. With the ARMA models, the time-domain analysis has proven to be so flexible that, in most applications, spectral analysis can be avoided. The main requirement of the time-domain analysis is just familiarity with correlation theory. On the other hand, the main drawback of the spectral analysis is its relatively demanding mathematical requirement. To get a working knowledge of spectral analysis, a working knowledge of Fourier analysis is necessary. Relatively advanced measure theory and functional analysis are also essential to understanding its theoretical component.

As spectral analysis is the most natural and effective approach to understand behaviours of time-invariant linear filters, it will be the central issue of this paper. This paper is aimed to popularize spectral analysis to the readers in Thailand with central bankers as the main target group. Our aim is rather ambitious since we wish to introduce the spectral representation theorem to the readers. It is our belief that some understanding of the spectral representation theorem will help the readers understand the theory of time-invariant linear filters. A typical discussion of the spectral representation theorem requires a working knowledge of measure theory and functional analysis. However, advanced mathematics will be avoid at all cost in our simplified version of the theorem. The possibility of introducing the readers a glimpse of the spectral representation theorem in a relatively less technical way is a result of simple and lucid insights from Hannan (1960). An in-depth investigation of a special class of stationary processes known as circularly defined processes can give

insights into the core of the spectral representation theorem. Given a relatively demanding mathematical requirement in Hannan (1960), we will show detailed discussion of how results in Hannan (1960) can be derived. The minimum mathematical requirements for this paper are a typical first course in linear algebra with matrix diagonalization and a typical first course in probability and statistics. Familiarity with complex numbers at pre-university levels is also assumed. All results concerning circularly defined processes will be shown in details. The generalization of the representation theorem to stationary processes can be regarded as the limiting cases of the circularly defined processes. Generalization of this seemingly restricted class of stationary processes to any stationary processes can be made rigourous by the elegant and abstract idea of Toeplitz forms presented in Grenander and Szego (1958). To make the paper more accessible, all proofs are omitted for the main text and can be found in the Appendix.

Sinusoids plus noise and harmonic processes are discussed in Section 2. Detailed discussion of spectral representation of circularly defined processes is given in Section 3. In Section 4, we explain the spectral representation theorem of stationary processes with reference to the known results in Section 3. Theory of time-invariant linear filters is discussed in Section 5 where various applications of such theory will be discussed in Section 6. The content in Section 6 can be considered as our attempt to dispel myth concerning effects of various classes of linear filters mechanically and routinely employed by policy makers on original data. We show how the usual 3-month moving average routine can lead to spurious cycles known as the Yule-Slutsky effect. We explain why the simple exponential moving average smoothing is a rather inappropriate tool employed to monitor short and medium-run dynamics of volatility of certain time series. We also discuss the issue of seasonal adjustment and give a partial explanation why Gomez and Maravall, the main contributor of modern seasonal adjustment techniques, tried to caution their readers in Gomez and Maravall (2001) that "data used in econometric models should not be, as a rule, seasonally adjusted". Band-pass and low-pass filters are also discussed while a popular low-pass filter known as the Hodrick-Prescott (HP) filter is discussed in details. The inclusion of the topic of sinusoids plus noise in the main text is rather unusual compared with other papers popularizing spectral analysis. The reason of its presence in the paper is to dispel the confusion we learnt from policy makers that it is the core of spectral analysis. The last section is the concluding part containing our suggestions how policy makers could modify their mechanical treatment of the data in the light of spectral analysis, and some lessons learnt from the myth and mistakes.

Before starting serious discussion of the subject, we define certain jargons and notations employed in the paper. Policy makers often think of a time series as a set of observations from the real world, such as the quarterly GDP of Thailand. Symbolically, a time series is just a finite set of numbers denoted by x_1, ..., x_N, where N is the total number of observations policy makers have. However, from a mathematical point of view, this view

is rather restrictive. In this paper, we adopt the convention employed in the literature to assume that a time series can be traced back and forward indefinitely into the past and future. Hence, we can regard a time series as a doubly-infinite sequence of random variables

$$..., \ X_{-N}, \ ..., \ X_{-1}, \ X_0, \ X_1, \ ..., \ X_N, \ ...,$$

which can be abbreviated to $\{X_t : t \in \mathbb{Z}\}$, where $\mathbb{Z}$ is the set of integers. We often shorten this lengthy symbol by $\{X_t\}$ wherever the abbrevaiation will not cause any confusion. The data $\{x_1, \ ..., \ x_N\}$ policy makers have is just a realization of a finite section of the time series $\{X_t\}$. The notation $\mathbb{R}$ represents the set of real numbers.

2 Harmonic and Stationary Processes

2.1 Sinusoid Plus Noise

Analysis of time series data has a very long history. It can be dated back as early as in the ancient Egypt era. The notable example is the long record of the water levels in the Nile River. However, a proper and systematic treatment, with serious mathematical foundation, had not started until the second half of the 19th Century. Due to the salient up-and-down movement of most time series, the early time series analysts paid a great deal of attention to the cyclical behaviour and tried to fit time series data with sinusoidal functions, i.e. sine and cosine functions. From a statistical point of view, these early time series analysts had a vision that, in the absence of any upward or downward trend, a time series $\{X_t\}$ have a representation of the form

$$X_t = \mu + R\cos(\lambda t + \phi) + \varepsilon_t, \tag{1}$$

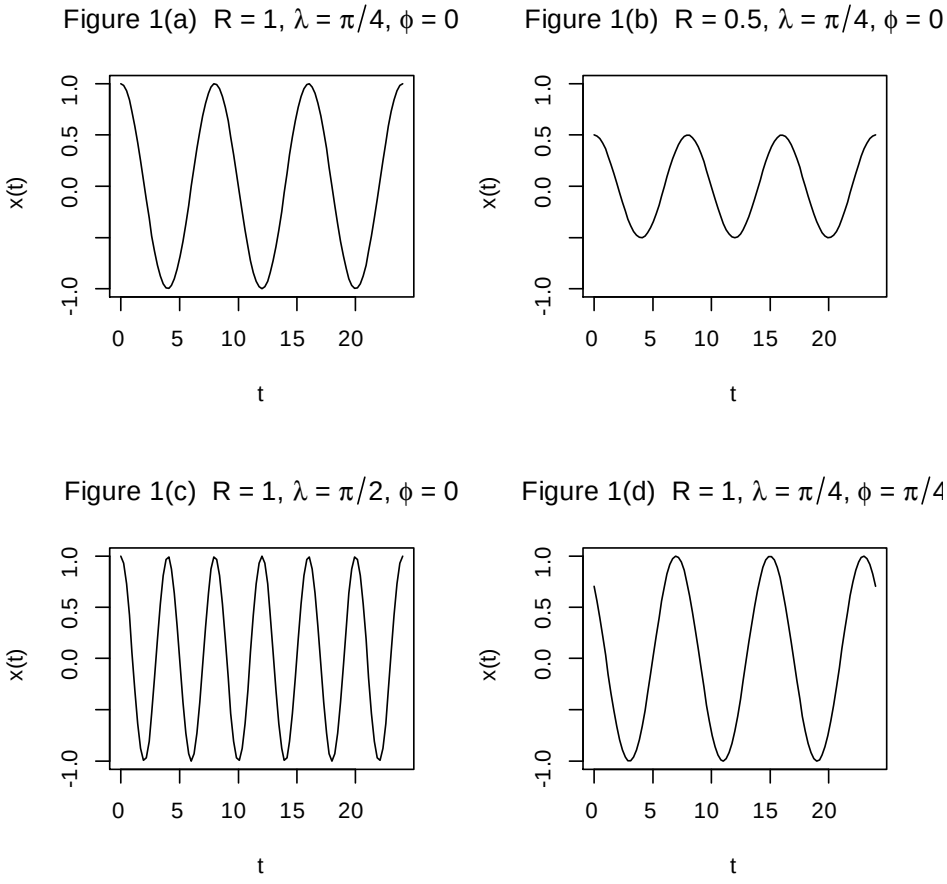
where $\mu, \ R, \ \lambda$ and ϕ are unknown constants and ε_t are the disturbances.

For some readers unfamiliar with sinusoidal functions, we briefly consider a simple sinusoidal function $x : \mathbb{R} \to \mathbb{R}$ defined by

$$x(t) = R\cos(\lambda t + \phi), \quad t \in \mathbb{R}, \tag{2}$$

where $\mathbb{R}$ is the set of real numbers. Graphs of the functions $x(t)$ in (2) for various values

of R, λ and ϕ are shown in Figures 1(a) - 1(d).



The term R in (2) is called the amplitude indicating the hightest and lowest values of the function $x(t)$. The value R is usually assumed to take a positive value so that the highest value of the function $x(t)$ is R and the lowest value is $-R$. Compare Figures 1(a) and 1(b). The term λ is the angular frequency of the function implying how many cycles are completed per unit time. Because the cosine function is periodic with period 2π,

$$\cos(\lambda t + \phi) = \cos(\lambda t + \phi + 2\pi) = \cos\{\lambda(t + 2\pi/\lambda) + \phi\} \text{ for all } t \in \mathbb{R}.$$

Hence, the period of this function is $2\pi/\lambda$, i.e. it takes $2\pi/\lambda$ unit time for the graph of this function to complete its full cycle. A high value of the frequency λ indicates a rapid movement of the function. It takes a relatively short time for the graph of this function to complete its full cycle. On the other hand, a low value of the frequency λ indicates a slow movement of the function. Compare Figures 1(a) and 1(c) The value λ is not the frequency of the cosine wave in the strict sense. Frequency is typically defined as the number of cycles per unit time. For $x(t)$ in (2), the frequency or the number of cycles completed per unit

time should be $\lambda/2\pi$. As the angular frequency is proportional to the frequency in the strtict sense, they both reflect the rate at which a sinusiod completes its full cycle. In this paper, we adopt the convention in time series literature so that the angular frequency λ is simply refered to as frequency. The term ϕ is called the phase indicating the horizontal shift of the graph of the cosine function. Compare Figures 1(a) and 1(d). Since

$$\cos t = \sin\left(t + \pi/2\right) \text{ for all } t \in \mathbb{R},$$

we can rewrite (2) as

$$x\left(t\right) = R\sin\left(\lambda t + \phi_1\right),$$

where $\phi_1 = \phi + \pi/2$. Therefore, we can either employ the cosine or sine function to represent the same sinusoidal movement. The difference arising from the choice of the functions will be a phase shift by $\pi/2$.

Now return to the sinusoid plus noise model in (1). Since the sinusoid in (2) fluctuates around zero, the additional constant μ is introduced so that the function can fluctuate around any constant μ rather than zero. The term ε_t represents an error term or a shock making the path of the process $\{X_t\}$ deviate from its main signal $R\cos\left(\lambda t + \phi\right)$. Developing estimation techniques for the unknown $\mu,\ R,\ \phi$ and particularly the unknown periodicity implied by the unknown λ was an active area of research towards the end of the 19th Century. Despite simplicity of the sinusoid in (1), the main difficulty in obtaining a reasonable estimate of the unknowns is that the sinusoid is nonlinear in both λ and ϕ. One way to mitigate is problem is to work with the following form

$$X_t = \mu + A\cos\lambda t + B\sin\lambda t + \varepsilon_t, \tag{3}$$

where $\mu,\ A,\ B$ and λ are unknown constants, rather than (1). For a given value of λ, the model is now linear in the unknown $\mu,\ A$ and B. Therefore, under (3), the main source of nonlinearity is essentially the unknown frequency λ. The specification in (3) follows directly from (1), since, by a trigonometric identity,

$$R\cos\left(\lambda t + \phi\right) = R\cos\lambda t\sin\phi - R\sin\lambda t\cos\phi.$$

That is $A = R\sin\phi$ and $B = -R\cos\phi$. A simple model in (3) can be generalised to allow for multiple periodic components as

$$X_t = \mu + \sum_{i=1}^{N}\left(A_i\cos\lambda_i t + B_i\cos\lambda_i t\right) + \varepsilon_t, \tag{4}$$

where $\mu,\ A_i,\ B_i,\ \lambda_i$ are unknown constants and ε_t are the noise or error terms.

One implication of the model in (3) and a harmless assumption

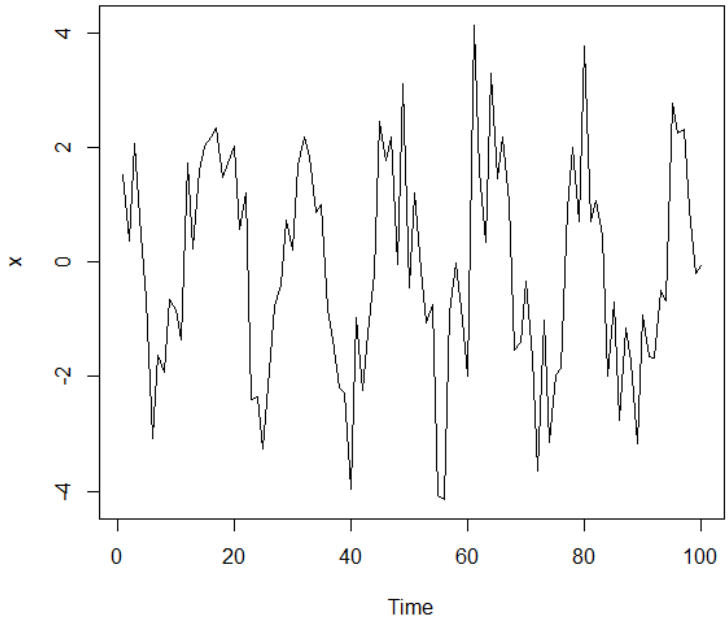
$$\mathbb{E}_t \varepsilon_{t+s} = 0 \quad \text{for all } s > 0 \text{ and all } t \in \mathbb{Z}, \tag{5}$$

is that

$$\mathbb{E}_t X_{t+s} = \mu + A\cos\lambda(t+s) + B\sin\lambda(t+s), \quad \text{for all } s > 0,$$

where $\mathbb{E}_t Y_{t+s}$ denotes conditional expectation of a random variable Y_{t+s}, observed at time $t+s$, given information at time t. If $\mathbb{E}\varepsilon_t^2 < \infty$ for all $t \in \mathbb{Z}$, condition 5 implies that the process $\{\varepsilon_t\}$ is a sequence of uncorrelated random variables. It can easily be shown, by employing basic probability, that $\mathbb{E}_t X_{t+s}$ is actually the best mean squared predictor of X_{t+s} given information available up to time t. If $\mathbb{E}\varepsilon_t^2 = \sigma_\varepsilon^2 < \infty$ for all $t \in \mathbb{Z}$ so that $\{\varepsilon_t\}$ is the, commonly assumed, sequence of white noise errers, then the mean squared error of the best mean squared predictor is simply σ_ε^2 for all $t \in \mathbb{Z}$ and $s > 0$. That is the best mean squared prediction error is constant over time regardless of how far into the future one try to predict the value of X_{t+s}. The readers with experience in empirical economics will find this particular implication of this class of models unrealistic. This implication is unrealistic even for applications in physical science. It is normally the case that short-term prediction is relatively much more accurate than long-term prediction. See Figure 2 for a typical realisation of a sinusoid plus noise process where the disturbances are the standard normal random variables.

Figure 2 Sinusoid plus noise: R = 2, $\lambda = \pi/8$, $\phi = 0$



Moreover, one might argue that prediction with such certainty may not easily arise in

economic applications. Arbitrageurs are likely to take advantage of this high level of predictability so that the highly predictable part of the process will vanish and remain with the one with a relatively high level of uncertainty.

2.2 Stationary and Harmonic Processes

After the breakthroughs from Yule (1927) and Slutzky (1927) and the independent development of functional analysis, a new class of process, certainly much more random than the one in (3), was developed in the 1930s to study the random cyclical movements of various time series. This new class of process is called a class of stationary processes. We say that a time series $\{X_t\}$ is weakly stationary if (i) $\mathbb{E}X_t = \mu$ for all $t \in \mathbb{Z}$, (ii) $\mathbb{E}X_t^2 = \sigma_X^2 < \infty$ for all $t \in \mathbb{Z}$, and (iii)

$$\mathbb{E}\left(X_{t+u} - \mu\right)\left(X_t - \mu\right) = \gamma\left(u\right), \quad \text{for all } t,\ u \in \mathbb{Z}.$$

The physical property of this process is that it fluctuates around a certain constant μ without any tendency to deviate from this constant. Moreover, the correlation structure of the process depends only on the distance between two random values observed at different points in time.

Given the cyclical property of the sinusoids, it is interesting to see whether the sinusoid plus noise process in (1) that can be represented as in (3) can be modified so that it become stationary. First, consider a random sinusoidal wave (without noise) with a random phase. This model has the representation

$$X_t = \mu + R\cos\left(\lambda t + \phi\right), \tag{6}$$

where ϕ is a random variable. By the trigonometric identity given above, this process can be represented as

$$X_t = \mu + A\cos\lambda t + B\sin\lambda t, \tag{7}$$

where $A = R\cos\phi$ and $B = -R\sin\phi$ are new random variables. As the sine and cosine functions are periodic with period 2π, we can restrict ϕ to take values in the interval $[-\pi,\ \pi]$. If ϕ is uniformly distributed on $[-\pi,\ \pi]$, employing properties of sine and cosine functions, it can be shown that $\mathbb{E}A = 0$, $\mathbb{E}B = 0$, $Cov\left(A,\ B\right) = 0$ and $\mathbb{E}A^2 = R^2/2 = \mathbb{E}B^2$. It follows from these results that for a simple harmonic process $\{X_t\}$ defined by (6) $\mathbb{E}X_t = \mu$ for all $t \in \mathbb{Z}$;

$$\begin{aligned}\mathbb{E}X_t^2 &= \mu^2 + \mathbb{E}A^2\cos^2\lambda t + \mathbb{E}B^2\sin^2\lambda t \\ &= \mu^2 + R^2/2 < \infty \ \text{ for all } t \in \mathbb{Z};\end{aligned}$$

and for all $u,\ t \in \mathbb{Z}$,

$$\begin{aligned} Cov\left(X_{t+u},\ X_t\right) &= \mathbb{E}A^2 \cos\lambda\left(t+u\right)\cos\lambda t + \mathbb{E}B^2 \sin\lambda\left(t+u\right)\sin\lambda t \\ &= R^2\cos\left(\lambda t + \lambda u - \lambda t\right)/2 = R^2\cos\left(\lambda u\right)/2. \end{aligned}$$

Since $Cov\left(X_{t+u},\ X_t\right)$ is a function of u only, it follows that $\{X_t\}$ is a stationary process.

The simple random cosine wave in (6) can be generalised to a random sinusoidal wave with mean zero and multiple periodicities as

$$X_t = \sum_{j=1}^{M}\left(A_j\cos\lambda_j t + B_j\sin\lambda_j t\right), \tag{8}$$

where λ_j are distinct and $A_j,\ B_j,\ j = 1,\ 2,\ ...,\ M$ are pairwise uncorrelated random variables such that for $j = 1,\ 2,\ ...,\ M$,

$$\mathbb{E}A_j = \mathbb{E}B_j = 0 \text{ and } \mathbb{E}A_j^2 = \mathbb{E}B_j^2 = f_j/2.$$

With reference to (6) and (7), the term $A_j\cos\lambda_j t + B_j\sin\lambda_j t$ can be regarded as the cosine wave with frequency λ_j and amplitude $\sqrt{f_j}$. Hence, X_t can be regarded as the sum of random sinusoidal waves with M different freqencies.

It follows, from the assumptions on A_j and B_j that $\mathbb{E}X_t = 0$ for all $t \in \mathbb{Z}$; for any $t,\ u \in \mathbb{Z}$,

$$\begin{aligned} \mathbb{E}\left(X_{t+u}X_t\right) &= \sum_{j=1}^{M}\left[\mathbb{E}A_j^2\cos\lambda_j\left(t+u\right)\cos\lambda_j t + \mathbb{E}B_j^2\sin\lambda_j\left(t+u\right)\sin\lambda_j t\right] \\ &= \frac{1}{2}\sum_{j=1}^{M} f_j\cos\lambda_j u. \end{aligned}$$

Hence, $\{X_t\}$ is indeed a stationary process. For $u = 0$, we have

$$\mathbb{E}X_t^2 = Var\left(X_t\right) = \sum_{j=1}^{N} f_j/2. \tag{9}$$

Note that

$$\mathbb{E}\left(A_j\cos\lambda_j t + B_j\sin\lambda_j t\right)^2 = \mathbb{E}A_j^2\cos^2\lambda_j t + \mathbb{E}B_j^2\sin^2\lambda_j t = f_j/2. \tag{10}$$

Equations (9) and (10) give the process of the form (8) an attractive physical meaning. Equation (8) indicates that X_t is a linear combination of random sinusoidal waves with different frequencies. The assumption on the covariance structure of A_j and B_j implies that the random sinusoidal waves are uncorrelated. Equations (9) and (10) indicate that

the variance (variation from the mean) of the time series $\{X_t\}$ is just the sum of the variances of the random sinusoidal waves with different frequencies. The factor $f_j/2$, the variance of A_j and B_j, can be regarded as the contribution from the random sinusoidal wave with frequency λ_j to the variation of the process $\{X_t\}$. If the variances of random sinusoids with low frequencies are relatively high compared with the one with high frequencies, then the process $\{X_t\}$ will tend to be a slowly moving process and vice versa. One may see the connection between this decomposition with the analysis of variance. See Tukey (1961) for more discussion on this connection. Figures 3(a) and 3(b) show a realisation of the following processes

$$\begin{aligned} H_{1t} &= 2A_1 \cos \frac{\pi}{10}t + 2B_1 \sin \frac{\pi}{10}t + A_2 \cos \frac{\pi}{5}t + B_2 \sin \frac{\pi}{5}t, \\ H_{2t} &= A_1 \cos \frac{\pi}{10}t + B_1 \sin \frac{\pi}{10}t + 2A_2 \cos \frac{\pi}{5}t + 2B_2 \sin \frac{\pi}{5}t, \end{aligned}$$

where A_1, B_1, A_2 and B_2 are independent standard normal random variables with mean zero. The process $\{H_{1t}\}$ gives more weight to the sinusoidal wave with the lower frequency compared to the process $\{H_{2t}\}$. It is therefore not surprising to see that the process $\{H_{2t}\}$ has more variation compared with the process $\{H_{1t}\}$.

Figure 3(a) Realisation of H1

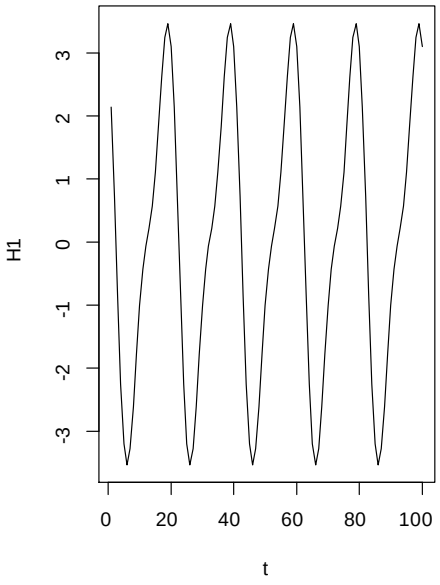
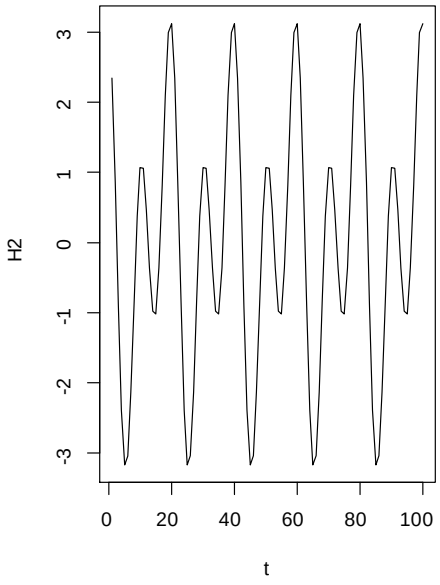


Figure 3(b) Realisation of H2



It is crucial to point out one issue with the specification in (8) and aliasing. For any $\lambda,\ t \in \mathbb{R}$, and any $k \in \mathbb{Z}$,

$$\cos\left(\lambda + 2\pi k\right) t = \cos \lambda t \cos 2\pi k t - \sin \lambda t \sin 2\pi k t.$$

However, for $t \in \mathbb{Z}$, $\cos 2\pi kt = 1$ and $\sin 2\pi kt = 0$ so that

$$\cos(\lambda + 2\pi k)\, t = \cos \lambda t \quad \text{for all } t \in \mathbb{Z}.$$

As our time series of interest $\{X_t\}$ is indexed by the set of all integers, $\mathbb{Z}$, it follows that we cannot differentiate the cyclical movement of the sinusoid with frequency λ and the one with frequency $\lambda + 2\pi k$ for all $k \in \mathbb{Z}$. As a consequence, we need to restrict the values of λ to any particular interval with length 2π. We will adopt the convention in the literature by considering any λ in the Nyquist interval $[-\pi,\ \pi)$.

With this particular Nyquist interval, the readers may be concerned with an interpretation of negative frequencies. Recall that one interpretation of the random wave with frequency λ is the cosine wave with a random phase $\cos(\lambda t + \phi)$. Since cosine is an even function,

$$\cos(-\lambda t + \phi) = \cos(-(\lambda t - \phi)) = \cos(\lambda t - \phi),$$

i.e. the cosine function with frequency $-\lambda$ and phase ϕ is identical to the cosine function with frequency λ and phase $-\phi$. That is a negative frequency indicates a phase shift of a cosine wave with a positive frequency.

In comparison with the multiple sinusoids plus noise in (4), the harmonic process is even less realistic. Consider a simple harmonic process in (6). Suppose μ, λ and R are known, one can determine the realised value of the random phase ϕ given at least two observations from $\{X_t\}$. Once the realised value of ϕ is determined, one can perfectly forecast other values of X_t.

3 Spectral Representation of a Circularly Defined Process

The spectral representation of a stationary process $\{X_t\}$ is essentially the idea of decomposing each X_t as a sum of uncorrelated random sinusoids. The analogous idea in linear algebra is the concept of matrix diagonalisation. We need to formally introduce complex-valued random variables as they will greatly simplify our derivation of a simplified form of the spectral representation theorem. A complex valued random variable Z is a complex-valued function of the form

$$Z = X + iY,$$

where X and Y are real-valued random variables and i is the imaginary number such that $i^2 = -1$. The mean of a complex-valued random variable Z, denoted by $\mathbb{E}Z$, is defined by

$$\mathbb{E}Z = \mathbb{E}X + i\mathbb{E}Y.$$

The variance of a complex-valued random variable Z, denoted by $Var(Z)$, is defined by

$$Var(Z) = \mathbb{E}(Z - \mathbb{E}Z)\left(\overline{Z - \mathbb{E}Z}\right),$$

where $\overline{Z - \mathbb{E}Z}$ is the complex conjugate of $Z - \mathbb{E}Z$. Similarly, the covariance of random variables Z_1 and Z_2, denoted by $Cov(Z_1,\ Z_2)$, is defined by

$$Cov(Z_1,\ Z_2) = \mathbb{E}(Z_1 - \mathbb{E}Z_1)\left(\overline{Z_2 - \mathbb{E}Z_2}\right). \tag{11}$$

A discrete-time time series is typically defined as a family of random variables $\{X_t : t \in \mathbb{Z}\}$ indexed by the set of integers $\mathbb{Z}$. Without any restriction on this family of random variables, we generally have to deal with an infinite-dimensional system. This is the reason why function analysis is key to the development of spectral analysis of stationary processes. To simplify this mathematical difficulty and avoid dealing with an infinite-dimensional system, we follow Hannan (1960) by considering a special class of time series known as a circularly defined process. Given random variables $X_1,\ X_2,\ ...,\ X_N$. For $t > N$ define

$$X_{N+1} = X_1,\ X_{N+2} = X_2,\ ...,\ X_{2N} = X_N,\ X_{2N+1} = X_1,\ ....$$

In general, for $t > 0$ that is not divisible by N, we have that $X_t = X_{t \bmod N}$, where $t \bmod N$ is the remainder of the division of t by N. For t that is divisible by N, $X_t = X_N$. For X_t such that $t \leq 0$, we can also defined them from $X_1,\ ...,\ X_N$ in a similar fashion. However, this is not a necessary step in our discussion.

One interpretation of the spectral representation is that of decomposing a particular random variable X_t from a stationary time series $\{X_t\}$ as a linear combination of uncorrelated random variables associated with sine and cosine waves. For a general stationary process, one has to consider such decomposition for an infinite number of random variables. However, in the case of the circularly defined process, we can reduce and infinite-dimensional problem to a finite-dimensional one. Instead of considering the decomposition of X_t for each $t \in \mathbb{Z}$, we only need to consider the decomposition for $t = 1,\ ...,\ N$. Stationarity of a general stationary process implies that there exists a function γ such that

$$Cov(X_s,\ X_t) = \gamma(s - t), \quad \text{for all } s,\ t \in \mathbb{Z}.$$

Certainly, it is necessary that γ must be even since, for any $u \in \mathbb{Z}$,

$$\gamma(u) = Cov(X_{t+u},\ X_t) = Cov(X_t,\ X_{t+u}) = \gamma(-u).$$

One additional structure imposed by a circularly defined process is that we only have to consider $\gamma(u)$ for $u = 0,\ 1,\ ...,\ N - 1$, since any pair of random variables $(X_s,\ X_t)$ from the circularly defined process can be represented as a pair of random variables $(X_u,\ X_v)$

such that $u,\ v \in \{1,\ ...,\ N\}$. Consider the covariance of the pair $(X_{N+1},\ X_t)$ for any $t \in \{1,\ ...,\ N\}$. Stationarity of a circularly defined process implies that

$$Cov\left(X_{N+1},\ X_t\right) = \gamma\left(N+1-t\right).$$

However, since

$$Cov\left(X_{N+1},\ X_t\right) = Cov\left(X_1,\ X_t\right) = \gamma\left(1-t\right) = \gamma\left(t-1\right),$$

it follows that $\gamma\left(N-(t-1)\right) = \gamma\left(t-1\right)$. This implies that

$$\gamma\left(u\right) = \gamma\left(N-u\right), \quad u = 0,\ 1,\ ...,\ N-1. \tag{12}$$

This additional condition indicates how one can construct a stationary circularly defined process.

Let $X = \left(X_1,\ ...,\ X_N\right)'$, where the prime "$'$" denotes a matrix transposition, and Γ be the covariance matrix of X. As Γ is the covariance matrix, it must be symmetric and positive semidefinite. It follows that Γ is unitarily diagonalisable. That is there exists a unitary matrix P, i.e. $P^* = P^{-1}$, where P^* is the conjugate transpose of P, such that

$$P^*\Gamma P = F,$$

where F is the diagonal matrix whose diagonal elements are eigenvalues of Γ. In addition, eigenvalues of Γ are nonnegative real numbers. With reference to (11), $Z = P^*X$ will become a vector of uncorrelated random variables.

If Γ is any covariance matrix, the choice of a matrix P will generally be arbitary. Simplicity and elegance of a suitable choice of a matrix P arises from stationarity of the process $\{X_t\}$. As discussed earlier, stationarity of a circularly defined process as seen in (12) implies that the covariance matrix Γ can be written as a circulant matrix

$$\Gamma = \begin{pmatrix} \gamma\left(0\right) & \gamma\left(1\right) & \gamma\left(2\right) & \cdots & \gamma\left(N-1\right) \\ \gamma\left(N-1\right) & \gamma\left(0\right) & \gamma\left(1\right) & \cdots & \gamma\left(N-2\right) \\ \gamma\left(N-2\right) & \gamma\left(N-1\right) & \gamma\left(0\right) & \cdots & \gamma\left(N-3\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma\left(1\right) & \gamma\left(2\right) & \gamma\left(3\right) & \cdots & \gamma\left(0\right) \end{pmatrix}. \tag{13}$$

A reader familiar with linear algebra may, at this point, be able to see that X_t can be represented as a linear combination of random sinusoids since a circulant matrix can be diagonalised by a discrete Fourier transform. That is a unitary matrix P in (18) can be

chosen so that the j-th column of P is

$$\frac{1}{\sqrt{N}}\left(e^{i\psi_k},\ ...,\ e^{iN\psi_k}\right)',$$

where $\psi_k = 2\pi k/N,\ k = 1,\ ...,\ N$. That is this choice of matrix P satisfies the property in (18). See the proof in Appendix 1.

There is one rather trivial technical point we need to address at this point. As mentioned in the previous section that thanks to the aliasing issue, we will focus on randam fluctuation with frequency in the Nyquist interval $(-\pi,\ \pi]$. However, the eigenvectors of Γ involve sinusoids with frequencies in $(0,\ 2\pi]$. It turns out that we can easily represent these eigenvectors in terms of sinusoids with frequencies in the standard Nyquist interval. Since for any $\psi \in \mathbb{R}$,

$$e^{i\psi} = e^{i(\psi - 2\pi)}, \tag{14}$$

we can re-arrange $e^{i\psi_k}$, where $\psi_k = 2\pi k/N$ as the following. Let K be the smallest integer less than or equal to $N/2$. If N is even, then $K = N/2$. If N is odd, then $K = (N+1)/2$. We can define

$$\lambda_1 = \psi_{K+1} - 2\pi,\ \lambda_2 = \psi_{K+2} - 2\pi,\ ...,\ \lambda_{N-K} = \psi_N - 2\pi, \tag{15}$$

and

$$\lambda_{N-K+1} = \psi_1,\ \lambda_{N-K+2} = \psi_2,\ ...,\ \lambda_N = \psi_K. \tag{16}$$

By (14),

$$e^{i\lambda_1} = e^{i\psi_{K+1}},\ e^{i\lambda_2} = e^{i\psi_{K+2}},\ ...\ e^{i\lambda_{N-K}} = e^{i(\psi_N - 2\pi)}.$$

Hence

$$p_j = \frac{1}{\sqrt{N}}\left(e^{i\lambda_j},\ ...,\ e^{iN\lambda_j}\right)',\quad j = 1,\ ...,\ N, \tag{17}$$

are orthonormal eigenvectors of Γ. With p_j defined in (17), (15) and (16) as the j-th column of the matrix P, it follows that

$$P^*\Gamma P = F, \tag{18}$$

where F is a diagonal matrix $diag\{f_1,\ ...,\ f_N\}$, where f_j is the eigenvalue of Γ corresponding to the eigenvector p_j as described above. Equation (18) implies that P^*X is a $\mathbb{C}^N$-valued random variable whose elements are uncorrelated. As each element of P^*X can be written as

$$\frac{1}{\sqrt{N}}\sum_{t=1}^{N} e^{-it\lambda_j}X_t,$$

where its variance is f_j, the eigenvalue of Γ corresponding to p_j.

Define $Z = (Z_1,\ ...,\ Z_N)'$ by

$$Z = \frac{1}{\sqrt{N}}P^*X \tag{19}$$

so that each of its element can be written as

$$Z_j = \frac{1}{N}\sum_{t=1}^{N} e^{-it\lambda_j} X_t. \tag{20}$$

It follows that $Z_j,\ j = 1,\ ...,\ N$ are uncorrelated random variables and

$$\mathbb{E}\,|Z_j|^2 = f_j/N, \tag{21}$$

where $|Z_j|$ is the modulus of Z_j and f_j are the corresponding eigenvalues of Γ. It is straightforward from trigonometric identities that the matrix P^* is unitary, i.e. $P^* = P^{-1}$. Hence equation (19) implies that

$$X = \sqrt{N}PZ,$$

i.e.

$$X_t = \sum_{j=1}^{N} e^{it\lambda_j} Z_j, \quad t = 1,\ ...,\ N. \tag{22}$$

The expression in (22) is called the spectral representation of a stationary circularly defined process.

In comparison with the harmonic process in (8), relationship (22) implies that a stationary circularly defined process $\{X_t\}$ can be decomposed into uncorrelated random sinusoids. By Euler's formula, the term

$$e^{it\lambda_j} = \cos t\lambda_j + i \sin t\lambda_j$$

can be interpreted as a complex sinusoid with frequency λ_j. The term Z_j can be interpreted as the random amplitude associated with the complex sinusoid with frequency λ_j. This random amplitude has variance equal to f_j/N. The term f_j/N is called the spectral mass of the process $\{X_t\}$ at frequency λ_j.

Now we summarise this result in the following Proposition.

Proposition 1 *Suppose* $\{X_1,\ ...,\ X_N\}$ *is a stationary circularly defined process with the covariace matrix* Γ. *Then*

$$X_t = \sum_{j=1}^{N} e^{it\lambda_j} Z_j, \quad t = 1,\ ...,\ N,$$

where

$$Z_j = \frac{1}{N}\sum_{t=1}^{N} e^{-it\lambda_j} X_t,$$

with the following properties; (i) $Z_j,\ j = 1,\ ...,\ N$ *are uncorrelated random variables; and*

(ii)

$$\mathbb{E}\left|Z_j\right|^2 = f_j/N,$$

where f_j *are the eigenvalue of* Γ *corresponding the eigenvector* p_j *defined in (17).*

The unitary diagonalisation in (18) implies that

$$\Gamma = PFP^*. \tag{23}$$

That is for $s,\ t = 1,\ ...,\ N$, the $(s,\ t)$-element of Γ is

$$\gamma\left(s-t\right) = \mathbb{E}\left(X_t X_s\right) = \sum_{j=1}^{N} \frac{f_j}{N} e^{i(s-t)\lambda_j}. \tag{24}$$

We can summarise this result as the following.

Proposition 2 *Suppose* $\{X_1,\ ...,\ X_N\}$ *is a stationary circularly defined process with the covariace function* γ*. Then*

$$\gamma\left(u\right) = \frac{1}{N} \sum_{j=1}^{N} e^{iu\lambda_j} f_j, \tag{25}$$

where f_j *are eigenvalues of the covariance matrix* Γ*.*

In particular,

$$\gamma\left(0\right) = \sum_{j=1}^{N} \frac{f_j}{N},$$

i.e. the variation of the process $\{X_t\}$ can be decomposed into contribution from random sinusoids with frequencies $\lambda_j,\ j = 1,\ ...,\ N$ where the contribution from each frequency is f_j/N. The relationship between the autocovariance function γ and its eigenvalues of a stationary circularly defined process $\{X_t\}$ is handy. Equation (25) shows a clear relationship between the time-domain approach, as characterised by the autocovariance function, an the spectral one, as characterised by eigenvalues of the autocovariance function without having to rely on complexity of Fourier inversion.

4 Spectral Representation of Stationary Processes

In this section, we will generalise the spectral representation theorem and the dual relationship between the autocovariance function and its eigenvalues to a general stationary

process. In the previous section, we tried to be as water-tight as possible so that the readers can follow us with a great deal of confidence. As mentioned earlier, the difficulty of the spectral representation theorem for a general stationary process arises from the fact that the process is generally infinite dimensional. The advantage of restricting ourselves to a circularly defined process is to make sure that the system we consider is finite dimensional, i.e. we essentially need to consider the random variables $x_1,\ ...,\ x_N$, rather the whole process $\{x_t : t \in \mathbb{Z}\}$. The reader should be able to see that to deal with a finite dimensional process like a circularly defined process, elementary linear algebra is sufficient. To deal with an infinite dimensional process, we need a working knowledge function analysis. As we do not assume that the readers have a working knowledge of functional analysis, we will not prove any result in this subsection. Instead, we will state some results generalising all the known results for the circularly defined process. Fortunately, the generalised results are just the limiting case of a circularly defined process.

There is a close similarly in what we try to achieve and what the readers may have seen in a first course in probability. It is natural to first introduce a discrete random variable and let the readers become familiar with its properties. Once the readers develop some intuition concerning discrete random variables, continuous random variables are then introduced. Continuous random variables are often introduced as a limiting case of some discrete random variable. A common route is to introduce a Bernoulli random variable X_i,

$$X_i = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } 1-p. \end{cases}$$

Let $X_1,\ ...,\ X_N$ be independent Bernoulli random variables. Then $Y_N = \sum_{i=1}^{N} X_i$ is a Binomial random variable taking values $0,\ 1,\ ...,\ N$. As $N \to \infty$, the random variable

$$Z_N = \frac{1}{\sqrt{N}} (Y_N - Np)$$

will, in the limit, follow the distribution of the normal distribution which is a continuous random variable. Note that Y_N takes a value in $\{1,\ ...,\ N\}$ but $Y_N - Np$ can become negative and is no longer an integer. As $N \to \infty$, $Y_N - Np$ can be arbitrarily large or small. The normalisation by a factor of $\sqrt{N}$ stabilises the variance of Z_N and makes all the possible values of Z_N more dense on the real line $\mathbb{R}$.

Now we will proceed in a similar fashion to the introductory course in probability by trying to convince the readers the more general properties of stationary processes after having established key results concerning circularly defined processes. All the results will be heuristically discussed but the proofs are not given to avoid making use of relatively advanced mathematics. For a rigorous approach to the spectral representation theorem and related results, the readers can consult standard textbooks such as Brockwell and Davis (1991).

The main intuition for generalizing the results established for stationary circularly defined processes to general stationary processes is that when N becomes arbitarily large, periodicity of circularly defined processes starts to disappear and any subset of consecutive observations of stationary processes, say $X_1,\ ...,\ X_K$, should be arbitrarily well approximated by circularly defined processes given large sufficiently large N.

4.1 Wold's Theorem

Recall the dual relationship between an autocovariance function of a stationary circularly defined process and its eigenvalues. Each eigenvalue f_j corresponds to an eigengenvector associated with sinusoids with frequency λ_j. With respect to the definition of λ_j in (15) and (16) as $N \to \infty$, $\lambda_1,\ ...,\ \lambda_N$ will be dense in the interval $(-\pi,\ \pi]$. Hence, as $N \to \infty$, we should be able to approximate the relation in (24) by an integral. If there is no serious concentration of eigenvalues at any particular frequencies, then as $N \to \infty$, the term f_j/N will becomes arbitrarily small so that each frequency in the interval $(-\pi,\ \pi]$ will give a negligible contribution to the variation of the process $\{X_t\}$. But a range of frequencies, say from the interval $(\lambda_1,\ \lambda_2]$ where $\lambda_2 > \lambda_1$ can give a positive contribution to the variance of the process $\{X_t\}$. This generalisation is analogous to the probability density function of a continuous random variable.

Now we need to introduce some concepts concerning different sorts of integration. First, we introduce the intuitive idea of the Riemann-Stieltjes integral. Given a function g and a real valued function F such that F is an increasing function, we can define the integral

$$\int_a^b g(x)\ dF(x)$$

as the limit of the sum

$$\sum_{i=1}^{N} g\left(x_i'\right)\left[F(x_i) - F(x_{i-1})\right],$$

where $x_i' \in [x_i,\ x_{i-1}]$, $a = x_0 < x_1 < \ ... \ < x_N = b$ and $\max_{1\le i\le N} |x_i - x_{i-1}| \to 0$ as $N \to \infty$. Note that as N increases the points $x_1,\ ...,\ x_N$ will become dense in the interval $[a,\ b]$.

With reference to Proposition 2, where we consider a stationary circularly defined process, we can define the function F as

$$F_N(\lambda) = \sum_{j\in A_\lambda} \frac{f_j}{N},$$

where A_λ is the set of all λ_j such that $\lambda_j \le \lambda$. Then

$$\gamma(u) = \frac{1}{N}\sum_{j=1}^{N} e^{iu\lambda_j} f_j = \int_{-\pi}^{\pi} e^{iu\lambda}\ dF_N(\lambda). \tag{26}$$

Note that the function F_N is just a step function with a jump at frequencies $\lambda_1,\ ...,\ \lambda_N$ where the magnitude of the jump is equal to f_j/N where f_j is an eigenvalue of the covariance matrix of $(X_1,\ ...,\ X_N)'$. It shares the properties of a distribution function of a discrete random variable. Now we state a generalisation of this result known as the Wold's Theorem.

Theorem 1 *(Wold's Theorem). The sequence* $\{\gamma(u) : u \in \mathbb{Z}\}$ *is the autocorrelation function for some stationary time series* $\{X_t : t \in \mathbb{Z}\}$ *if and only if there exists a right-continuous, non-decreasing, bounded function* F *on* $[-\pi,\ \pi]$ *and* $F(-\pi) = 0$ *such that*

$$\gamma(u) = \int_{(-\pi,\ \pi]} e^{iu\lambda}\, dF(\lambda) \quad \textit{for all } u \in \mathbb{Z}. \tag{27}$$

The general result in (27) is so similar to the one in (26) that we may intuitively regard F as the limit of F_N as $N \to \infty$. The function F in Theorem 1 is called the spectral distribution function. There two main reasons for this name. First, it meets all the requirements for being a distribution function. Second, it is associated with the eigenvalues of the autocovariance function of the stationary process $\{X_t\}$.

For $u = 0$, Theorem 1 implies that

$$\sigma^2 = \gamma(0) = \int_{(-\pi,\ \pi]} dF(\lambda) = F(\pi).$$

Given that $\sigma^2 > 0$, we can define a new distribution function $F_1 = \sigma^{-2}F$ so that $F_1(\pi) = 1$ and $F_1(-\pi) = 0$. Then F_1 is a probability distribution function of some random variable taking values in the interval $(-\pi,\ \pi]$. Then (27) indicates that the autocorrelation function, $\gamma(u)/\gamma(0)$, of a stationary process $\{x_t\}$ is just the characteristic function of the random variable with the probability distribution function F_1. Recall from probability theory that a probability distribution function can be decomposed into three components where the most important two components are a step function associated with the cummulative mass function of a discrete random variable, and an absolutely continuous function associated wth the cummulative density function of a continuous random variable. The last component is rather pathological and does not make any sense unless one has a working knowledge of measure theory. Moreover, it does not seem to contribute to any real world applications. So we will simply ignore this component in our discussion.

As a result of this, we can essentially decompose a spectral distribution function into a step function and an absolutely continuous function like the cummulative probability function of a continuous random variable. The component associated with the step function is associated with a predictable process like the harmonic process in Section 2. As the

predictable part is unrealistic for economic and many other applications, the core part of classical time series analysis focuses mainly on the type of a spectral distribution function that is absolutely continuous, i.e. there exists a nonnegative function f such that

$$\gamma(u) = \int_{(-\pi,\ \pi]} e^{iu\lambda}\ f(\lambda)\ d\lambda \quad \text{for all } u \in \mathbb{Z}.$$

Again for $u = 0$,

$$Var(X_t) = \gamma(0) = \int_{(-\pi,\ \pi]} f(\lambda)\ d\lambda.$$

This says that the variation of the process $\{X_t\}$ can be composed into contribution from random sinusoids from all frequencies λ in the interval $(-\pi,\ \pi]$ where the contribution from each frequency is $f(\lambda)$. For simplicity of the presentation we assume that persistency of the process $\{X_t\}$ is limited so that $\sum_{u=-\infty}^{\infty} |\gamma(u)| < \infty$. Under this assumption, the following result holds.

Theorem 2 *If $\sum_{u=-\infty}^{\infty} |\gamma(u)| < \infty$, we have that*

$$f(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} e^{-iu\lambda} \gamma(u), \quad \textit{for all } \lambda \in (-\pi,\ \pi]. \tag{28}$$

4.2 Spectral Representation Theorem

To generalise the spectral representation theorem, we can proceed in a similar way as in the Wold's Theorem. Consider the case of circularly defined processes. Recall that for each sinusoid with frequency λ_j, there is a random variable

$$Z_j = \frac{1}{N} \sum_{t=1}^{N} e^{-it\lambda_j} X_t$$

that can be regarded as a random amplitude. For any $\lambda \in (-\pi,\ \pi]$, define

$$Z_N(\lambda) = \sum_{\lambda_j \leq \lambda} Z_j,$$

where the sum is over all $Z_j,\ j = 1,\ ...,\ N$ such that $\lambda_j \leq \lambda$. With the notion of Stieltjes integration, we can re-write (22) as

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda}\ dZ_N(\lambda). \tag{29}$$

We need to introduce some definitions before stating a theorem concerning the spectral

representation theorem of stationary processes.

Definition 1 *An orthogonal-increment process on* $[-\pi,\ \pi]$ *is a complex-valued process* $\{Z(\lambda),\ -\pi \leq \lambda \leq \pi\}$ *such that*

(i) $\mathbb{E}[Z(\lambda)] = 0, \quad -\pi \leq \lambda \leq \pi,$

(ii) $\mathbb{E}|Z(\lambda)|^2 < \infty, \quad -\pi \leq \lambda \leq \pi,$

and (iii) $\mathbb{E}[Z(\lambda_4) - Z(\lambda_3)]\left[\overline{Z(\lambda_2) - Z(\lambda_1)}\right] = 0$ *if* $(\lambda_1,\ \lambda_2] \cap (\lambda_3,\ \lambda_4] = \emptyset$, *where* $\emptyset$ *is the empty set.*

Definition 2 *An orthogonal-increment process* $\{Z(\lambda),\ -\pi \leq \lambda \leq \pi\}$ *is right-continuous if for all* $\lambda \in [-\pi,\ \pi]$,

$$\mathbb{E}|Z(\lambda+\delta) - Z(\lambda)|^2 \to 0 \quad \text{as } \delta \downarrow 0,$$

where $\delta \downarrow 0$ *means positive* δ *get arbitrarily close to zero.*

Now we present a general form of the spectral representation theorem.

Theorem 3 *(The Spectral Representation Theorem) If* $\{x_t\}$ *is a stationary sequence with mean zero and spectral distribution function* F, *then there exists a right-continuous orthgonal-increment process* $\{Z(\lambda),\ -\pi \leq \lambda \leq \pi\}$ *such that (i)* $\mathbb{E}|Z(\lambda) - Z(-\pi)|^2 = F(\lambda), \quad -\pi \leq \lambda \leq \pi$, *and (ii)*

$$X_t = \int_{(-\pi,\ \pi]} e^{it\lambda}\ dZ(\lambda), \tag{30}$$

where the integral is defined in the mean-square sense.

It is important to note the in integral in (30) is

$$\lim_{N\to\infty} \sum_{j=1}^{N} e^{i\lambda_j' t}\ [Z(\lambda_j) - Z(\lambda_{j-1})],$$

where $\lambda_j' \in [\lambda_{j-1},\ \lambda_j]$, $-\pi = \lambda_0 < \lambda_1 < \ ... \ < \lambda_N = \pi$ and as $N \to \infty$, $\max_{1\leq j\leq N} |\lambda_j - \lambda_{j-1}| \to 0$. This is very similar to the way we define the Riemann-Stieltjes integral. The exception is that $Z(\lambda)$ is a random variable from an orthogonal-increment process. As the form of the spectral representation theorem remains the same as in the case for a circularly defined process, the physical interpretation of the general representation theorem remains unchanged.

4.3 White Noise Process

The simplest example of a stationary process is the so called white noise process. It is a sequence of random variables $\{\varepsilon_t\}$ such that

$$\mathbb{E}\varepsilon_t = 0,\ \mathbb{E}\varepsilon_t^2 = \sigma^2 \text{ for all } t \in \mathbb{Z},$$

and for all $s,\ t \in \mathbb{Z}$ such that $t \neq s$

$$Cov\left(\varepsilon_t,\ \varepsilon_s\right) = 0.$$

In this subsection, we will rationalize its name and discuss its properties. First consider its analogue under the class of circularly defined process. Let $E = (\varepsilon_1,\ ...,\ \varepsilon_N)'$. Then it follows that the covariance matrix of E is $\Gamma_\varepsilon = \sigma^2 I_N$ where I_N is the identity matrix of order N. It follows immediately that all eigenvalues of Γ_ε are equal to σ^2. Hence, we have the spectral representation

$$\varepsilon_t = \sum_{j=1}^{N} e^{it\lambda_j} Z_j,$$

where $\mathbb{E}\left|Z_j\right|^2 = \sigma^2/N$ for all $j = 1,\ ...,\ N$. This implies that each random sinusoid, with frequency λ_j, gives the same contribution to the variation of the random variables ε_t. Certainly as $N \to \infty,\ \mathbb{E}\left|Z_j\right|^2 \to 0$ but for all $j,\ k = 1,\ ...,\ N$,

$$\frac{\mathbb{E}\left|Z_j\right|^2}{\mathbb{E}\left|Z_k\right|^2} \to 1 \ \text{ as } N \to \infty.$$

That is although each random sinusoid with a particular frequency λ will give a negligible contribution to the variation of ε_t, it relative contribution compared with those from other frequencies will be the same. From the spectral point of view, a white noise process plays a very similar role to the probability density function of a random variable with uniform distribution.

Now consider the general case of the white noise process. Let γ_ε be the autocovariance of the white noise process $\{\varepsilon_t\}$. It follows from the definiton that $\gamma_\varepsilon(0) = \sigma^2$ and 0 otherwise. Hence $\sum_{u=-\infty}^{\infty}\left|\gamma_\varepsilon(u)\right| = \sigma^2 < \infty$. Then we can apply Theorem 2 to show that for all $\lambda \in (-\pi,\ \pi]$,

$$f_\varepsilon(\lambda) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} e^{-iu\lambda}\gamma(u) = \frac{\sigma^2}{2\pi}. \tag{31}$$

That is each random sinusoid with frequency $\lambda \in (-\pi,\ \pi]$ gives the same relative contribution to the variation of the random variables ε_t. This is analogous to the behaviour of white light which is the mixture of the visible waves with the same intensity.

5 Time Invariant Linear Filters

As mentioned in the introduction, policy makers working with economic time series usually employ various smoothing technique to perform seasonal adjustment or get some sense of a medium-term or long-term trend of the series. From a mathematical point of view, most popular smoothing techniques can be regarded as time-invariant linear filters. Before discussing theoretical implications of time invariant linear filters, we first consider a particular form of filter commonly employed by macroeconomists in Thailand. To shed some light on theoretical results, we will mainly consider the circularly defined process.

5.1 Stationary Circularly Defined Processes

Suppose a monthly economic time series of interest $\{X_t\}$ is a circularly defined stationary process. To eliminate some erratic trends, practitioners in Thailand often apply a simple one-sided moving average linear filter

$$Y_t = \frac{1}{3}\left(X_{t-2} + X_{t-1} + X_t\right).$$

Recall that X_t has the spectral representation

$$X_t = \sum_{j=1}^{N} e^{it\lambda_j} Z_j,$$

where Z_j are uncorrelated complex-valued random variable with zero mean. It follows that

$$\begin{aligned} Y_t &= \frac{1}{3}\sum_{j=1}^{N}\left(e^{i(t-2)\lambda_j} + e^{i(t-1)\lambda_j} + e^{it\lambda_j}\right) Z_j \\ &= \frac{1}{3}\sum_{j=1}^{N} e^{it\lambda_j}\left(e^{-i2\lambda_j} + e^{-i\lambda_j} + 1\right) Z_j. \end{aligned} \tag{32}$$

Equation (32) has an interesting physical interpretation. For the original series $\{X_t\}$, the terms $e^{it\lambda_j}$ represent the sinusoidal wave with freqeuncy λ_j and Z_j are their uncorrelated random amplitudes. The variance of Z_j are the contribution to the variance of $\{X_t\}$ from the random sinusoidal wave with frequency λ_j. Now the transformed series $\{Y_t\}$ has the spectral representation of the form (32) which can be re-written as

$$Y_t = \sum_{j=1}^{N} e^{it\lambda_j} Z_{1j},$$

where $Z_{1j} = \left(e^{-i2\lambda_j} + e^{-i\lambda_j} + 1\right) Z_j/3$. That is the random amplitudes of the sinusiodal waves are modified by a factor of $\left(e^{-i2\lambda_j} + e^{-i\lambda_j} + 1\right)/3$. It follows, from the properties of Z_j, that Z_{1j} remain uncorrelated and have mean zero. However,

$$\begin{aligned} Var\left(Z_{1j}\right) &= \mathbb{E}\left(Z_{1j}\overline{Z_{1j}}\right) \\ &= \frac{f_j}{N}\left|\frac{e^{-i2\lambda_j} + e^{-i\lambda_j} + 1}{3}\right|^2. \end{aligned} \tag{33}$$

This is the contribution to the variation of the transformed process $\{Y_t\}$ from the random sinusoidal wave with frequency λ_j. The function $g\left(\lambda\right) = \left|(e^{-i2\lambda} + i^{-i\lambda} + 1)/3\right|^2$, called the squared gain function, indicates how the contribution to the variance from random sinusoidal wave with each freqeuncy is altered.

Figure 4 Squared Gain Function of the 3-M MA Filter

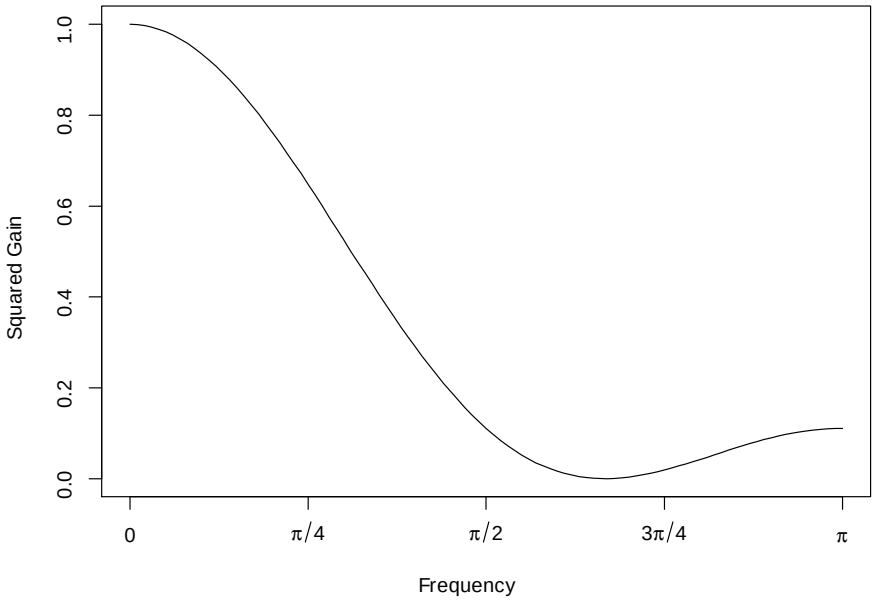


Figure 4 shows the squared gain function of the 3-month moving average filter. Without this filter, the squared gain should be equal to unity for all frequency. Once the filter is applied to the original series, the contribution from other frequency, except at zero frequency, to the variance of the transformed series will be dampened by a factor shown in Figure 4. This implies that the 3-month moving average filter essentially discounts an impact of random sinusoidal waves at high frequency relative to those at low frequency. Approximately 40% of the contribution of the random wave completing its cycle every 8 months will be removed if a 3-month moving average filter is applied to a monthly time series.

5.2 Stationary Processes

As the general theory of invariant linear filters is rather mathematical demanding, we first introduce a simple case in the hope that the reader will develop some intuition.

Assumption 1 *Let $\{X_t\}$ be a stationary time series with mean zero and autocovariance function $\gamma_x(\cdot)$. Assume further that*

$$\sum_{u=-\infty}^{\infty} |\gamma_x(u)| < \infty.$$

Assumption 1 and Theorem 2 imply that the spectral density function of $\{X_t\}$ has the form

$$f_x(\lambda) = \sum_{u=-\infty}^{\infty} e^{-iu\lambda}\gamma_x(u), \quad \lambda \in (-\pi,\ \pi].$$

Consider the one-sided 3-period moving average as in the previous subsection. Define

$$Y_t = \frac{X_{t-2} + X_{t-1} + X_t}{3}$$

It follows that the autocovariance function of $\{Y_t\}$ is

$$\gamma_y(u) = \frac{1}{9}\{\gamma_x(u-2) + 2\gamma_x(u-1) + 3\gamma_x(u) + 2\gamma_x(u+1) + \gamma_x(u+2)\}. \tag{34}$$

Lemma 3 shows that the spectral density function f_y of the process $\{Y_t\}$ is well defined and can be written as

$$\left|\frac{e^{-i2\lambda} + e^{-i\lambda} + 1}{3}\right|^2 f_x(\lambda).$$

This particular form of the spectral density function of $\{Y_t\}$ shares a great deal of similarity with that of a circularly defined case in (33). By taking a one-sided 3-period moving average, we obtain a new stationary process with the spectral density that is a multiple of the spectral density of $\{X_t\}$ by a factor of $\left|\left(e^{-i2\lambda} + e^{-i\lambda} + 1\right)/3\right|^2$. In fact, it can be shown that a similar kind of interpretation concerning the random amplitude of the sinusoids $e^{it\lambda}$ holds for the general case too. However, we omit this discussion concerning the mathematical results related to this interpretation since it requires much more advanced level of mathematics assumed in this article.

Now we jump to a greater level of mathematical generality. First notice that any weighted averages of the time series $\{X_t\}$ can be represented as

$$Y_t = \sum_{\tau=-\infty}^{\infty} g_{t,\tau} X_{t-\tau},$$

where for a given $t \in \mathbb{Z}$, $\{g_{t,\tau}\}$ is a deterministic sequence. The double sequence $\{g_{t,\tau}; t,\ \tau \in \mathbb{Z}\}$ is called a linear filter. In most applications, the deterministic sequence is chosen deliberately to be time independent, i.e. $g_{t,\tau} = g_\tau$ for all $t \in \mathbb{Z}$, so that the linear filter $\{g_\tau; \tau \in \mathbb{Z}\}$ is called time-invariant. The main purpose of this subsection is to explain the effect of applying a time-invariant linear filter to a process $\{X_t\}$. To ensure that the new random variable Y_t is well defined we make the assumption that the weights are absolutely summable, i.e.

$$\sum_{\tau=-\infty}^{\infty} |g_\tau| < \infty. \tag{35}$$

It follows from various results from Chapter 3 in Brockwell and Davis (1991) that the filtered process $\{Y_t\}$ is well defined and also weakly stationary with mean zero and autocovariance function

$$\gamma_y(u) = \sum_{\tau=-\infty}^{\infty} \sum_{\upsilon=-\infty}^{\infty} g_\tau g_\upsilon \gamma_x(h - \tau + \upsilon),$$

where γ_x is the autocovariance function of $\{X_t\}$. As the autocovariance and the spectral distribution functions are two sides of the same coin, once we know that autocovariance function, we can obtain a full knowledge concerning the spectral distribution function.

Theorem 4 *If $\{X_t\}$ is a stationary process with spectral distribution function F_x and a filtered process $\{Y_t\}$ defined by*

$$y_t = \sum_{\tau=-\infty}^{\infty} g_\tau x_{t-\tau} \quad \textit{where} \quad \sum_{\tau=-\infty}^{\infty} |g_\tau| < \infty,$$

then $\{Y_t\}$ is stationary with spectral distribution function

$$F_y(\lambda) = \int_{(-\pi,\ \pi]} \left| \sum_{\tau=-\infty}^{\infty} g_\tau e^{-i\tau\lambda} \right|^2 dF_x(\omega), \quad -\pi < \lambda \leq \pi.$$

With some knowledge of measure theory, one can easily derive the following corollary of Theorem 4.

Corollary 1 *If the processes $\{X_t\}$ and $\{Y_t\}$ satisfy the conditions in Theorem 4 and the spectral density function f_x of $\{X_t\}$ exists, then the spectral density function f_y of $\{Y_t\}$ is well defined and*

$$f_y(\lambda) = \left| \sum_{\tau=-\infty}^{\infty} g_\tau e^{-i\tau\lambda} \right|^2 f_x(\lambda), \quad -\pi < \lambda \leq \pi.$$

Given an absolutely summable time-invariant linear filter $\{g_\tau\}$, the function $T : (-\pi,\ \pi] \to \mathbb{C}$, where $\mathbb{C}$ is the set of complex numbers, defined by

$$T(\lambda) = \sum_{\tau=-\infty}^{\infty} g_\tau e^{-i\tau\lambda}, \quad \lambda \in: (-\pi,\ \pi]$$

is called the transfer function. The function $G : (-\pi,\ \pi] \to \mathbb{R}$ defined by

$$G(\lambda) = |T(\lambda)|, \quad \lambda \in: (-\pi,\ \pi],$$

where $|\cdot|$ is the modulus of a complex number, is called the gain function. Hence, from the spectral representation point of view, effects of time-invariant linear filters on an original time series can be investigated through the transfer and gain functions.

6 Various Applications

The power of spectral analysis and the theory of time invariant linear filters can be appreciated with various examples and applications.

6.1 Autoregressive-Moving Average Models and Yule-Slutsky Effects

Let $\{\varepsilon_t\}$ be a sequence of white noise with variance σ^2. Slutsky (1927), translated into English in Slutsky (1937), considered smoothing $\{\varepsilon_t\}$ to obtain new series and noted that the new processes can capture the movement of the British business cycle quite well. Slutsky (1927) considered various types of smoothed series, for simplicity, we consider

$$X_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2}.$$

Recall from (31) that the spectral density function of the process $\{\varepsilon_t\}$ is just $f_\varepsilon(\lambda) = \sigma^2/2\pi$ for all $-\pi < \lambda \leq \pi$. One can regard $\{X_t\}$ as a filtered process where

$$g_\tau = \begin{cases} 1, & \tau = 0,\ 1,\ 2, \\ 0, & \text{otherwise.} \end{cases}$$

It follows that $\sum_{\tau=-\infty}^{\infty} |g_\tau| < \infty$ and Theorem 4 and Corollary 1 imply that $\{X_t\}$ is stationary with spectral density

$$f_x(\lambda) = \left|1 + e^{-i\lambda} + e^{-i2\lambda}\right|^2 \frac{\sigma^2}{2\pi}.$$

One can generalise Slutsky's smoothing process to introduce a process

$$Y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \ \ldots\ + \theta_q\varepsilon_{t-q},$$

where $\{\varepsilon_t\}$ is a white noise process. By Theorem 4 and Corollary 1, it follows that $\{Y_t\}$ is stationary with spectral density

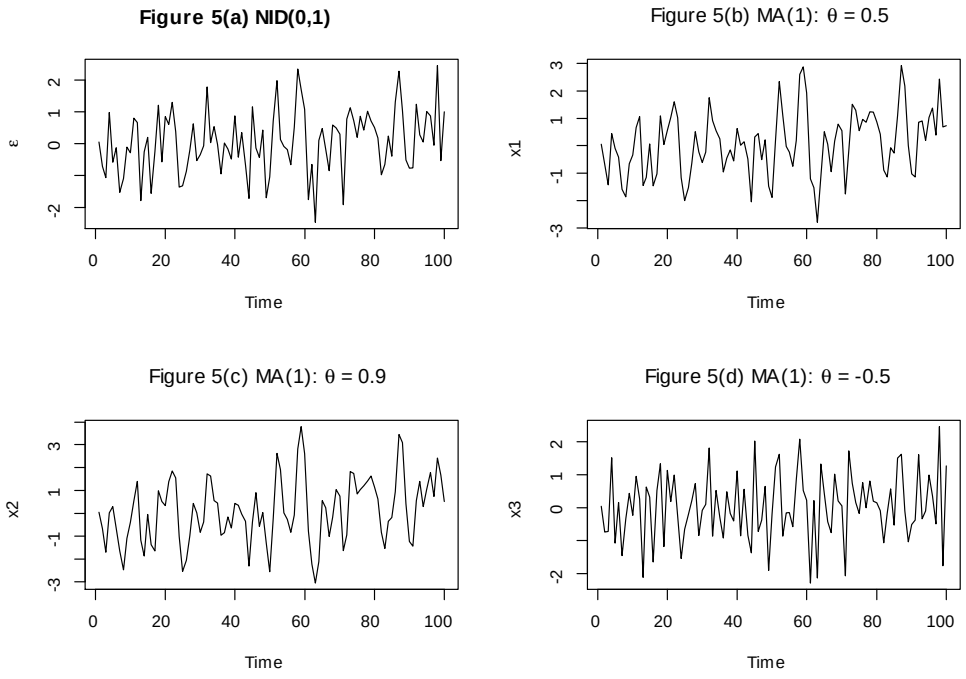
$$f_y(\lambda) = \sigma^2 \left| 1 + \theta_1 e^{-i\lambda} + \theta_2 e^{-i2\lambda} + \ \ldots \ + \theta_q e^{-iq\lambda} \right|^2 .$$

Wold (1953) called $\{Y_t\}$ a Moving Average process of order q, denoted by $MA(q)$.

Now consider $MA(1)$ processes

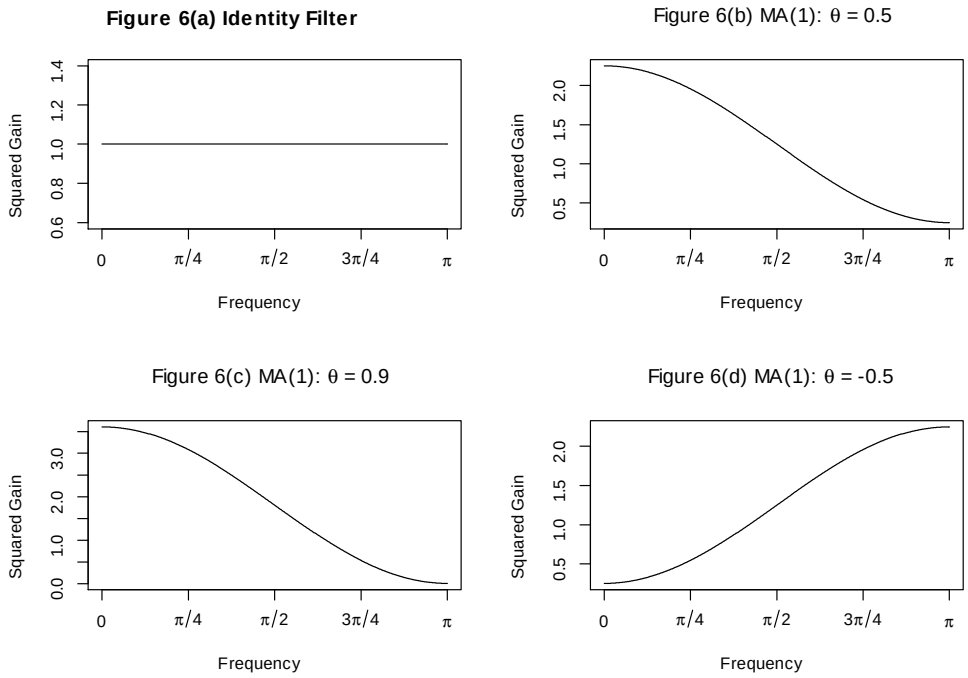
$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

with various values of θ. Figure 5(a) shows a realisation of an independent process $\{\varepsilon_t\}$ with the standard normal distribution. The values of ε_t are employed to create three $MA(1)$ processes with values $\theta = 0.5,\ 0.9$ and -0.5.



It can be seen that the plot in Figure 5(b) is smoother than the one in Figure 5(a). Similarly the plot in Figure 5(c) is even smoother than the one in Figure 5(c). On the other hand, the plot in Figure 5(d) is rougher than the one in Figure 5(a). We can explain this

comparison easily with a simple application of the theory of time invariant linear filters.



Without applying any filter to the white noise process $\{\varepsilon_t\}$, the squared gain function is equal to unity. Once we apply a one-sided moving average filter, we amplify the contribution from the slow moving random sinusoids, the ones with low frequencies, to the variation of the series while dampen the contribution from the fast moving random sinusoids, the one with high frequencies. See Figure 6(b). This explains why the plot in Figure 5(b) is smoother than the one in Figure 5(a). With a higher value of $\theta = 0.9$, we amplify the contribution from the slow moving sinusoids more than at value $\theta = 0.5$. This explains why Figure 5(c) is smoother than Figure 5(b). On the other hand, for a negative value of θ, we amplify the contribution from the fast moving sinusoids while penalise the contribution from the slow moving sinusoids. Hence, the plot in Figure 5(d) is rougher than the one in Figure 5(a).

Yule (1927) also investigated effects of smoothing time series. Rather than the simple moving average, he considered exponential smoothing. In the paper, Yule tried to explain the cyclical behaviour of Wolfer's sunspot data. He consider a process $\{X_{2t}\}$ defined by

$$X_{2t} = \phi_1 X_{2,t-1} + \phi_2 X_{2,t-2} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a white noise process with variance σ^2. For simplicity, we consider a simpler form by assuming that $\phi_2 = 0$ and re-label ϕ_1 simply as ϕ so that we have

$$X_{1t} = \phi X_{1,t-1} + \varepsilon_t. \tag{36}$$

It follows from induction that

$$X_{1t} = \varepsilon_t + \phi\varepsilon_{t-1} + \ ... \ + \phi^p \varepsilon_{t-p} + \phi^{p+1} X_{1,t-p-1}.$$

If $\mathbb{E}\left|X_{1t}\right|^2$ are uniformly bounded and $|\phi| < 1$, then as $p \to \infty$, we have that $\mathbb{E}\left|\phi^{p+1} X_{1,t-p+1}\right|^2 \to 0$. Moreover, it follows that

$$X_{1t} = \varepsilon_t + \phi\varepsilon_{t-1} + \ ... \ + \phi^p \varepsilon_{t-p} + \ ..., \tag{37}$$

where the convergence is in the mean square sense. Let L be a lag operater such that for all $t \in \mathbb{Z}$,

$$LX_{1t} = X_{1,t-1}.$$

It follows that (36) can be re-written as

$$X_{1t} = \phi L X_{1t} + \varepsilon_t.$$

This implies that

$$\left(1 - \phi L\right) X_{1t} = \varepsilon_t.$$

The term $(1 - \phi L)$ can be considered as an operator when applied to $\{X_{1t}\}$, one obtians a white noise sequence $\{\varepsilon_t\}$. Equation (37) suggests that the inverse of an operator $1 - \phi L$ exists since

$$X_{1t} = \left(1 + \phi L + \phi^2 L^2 + \ ... \ + \phi^p L^p + \ ...\right) \varepsilon_t.$$

Hence the operator $\left(1 + \phi_1 L + \phi^2 L^2 + \ ... \ + \phi^p L^p + \ ...\right)$ can be considered as the inverse of $(1 - \phi L)$ under the assumption that $|\phi| < 1$ and $\mathbb{E}\left|X_{1t}\right|^2$ are uniformly bounded. Now we can study the behaviour of $\{X_{1t}\}$ in a context of linear filters. Since

$$X_{1t} = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j},$$

where $\{\varepsilon_t\}$ is a white noise process with variance σ^2. As the series $\sum_{j=0}^{\infty} \phi^j$ is absolutely summable, then Corollary 1 implies that the $\{X_{1t}\}$ is also stationary and its spectral density is

$$f_{x_1}(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} \phi^j e^{-ij\lambda} \right|^2.$$

Recall a well-know result from geometric series that, if $|\phi| < 1$,

$$\sum_{j=0}^{\infty} \phi^j e^{-ij\lambda} = \frac{1}{1 - \phi e^{-ij\lambda}}.$$

Then it follows that for all $\lambda \in (-\pi,\ \pi]$,

$$f_{x_1}(\lambda) = \frac{\sigma^2}{2\pi \left|1 - \phi e^{-i\lambda}\right|^2}.$$

We can generalise the model considered by Yule (1927) to

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} \ + \ ... \ + \phi_p X_{t-p} + \varepsilon_t, \tag{38}$$

where $\{\varepsilon_t\}$ is a white noise process. Wold (1953) called the process defined in (38) an Autoregressive process of order p, denoted by $AR(p)$. We can re-write (38) as

$$\left(1 - \phi_1 L - \phi_2 L^2 - \ ... \ - \phi_p L^p\right) x_t = \varepsilon_t. \tag{39}$$

The heuristic way of solving this equation is to consider the equation

$$1 - \phi_1 z - \phi_2 z^2 - \ ... \ - \phi_p z^p = 0. \tag{40}$$

The left-side of (40) is a polynomial of degree p. Hence, by the fundamental theorem of algebra, there are precise p solutions to this equation. That is there are $\vartheta_1,\ ...,\ \vartheta_p$ such that

$$(1 - \vartheta_1 z)(1 - \vartheta_2 z) \ ... \ (1 - \vartheta_p z) = 0.$$

The composite operater in (39) can be regarded as the product of simple operators

$$(1 - \vartheta_1 L)(1 - \vartheta_2 L) \ ... \ (1 - \vartheta_p L) x_t = \varepsilon_t.$$

Given that $|\vartheta_k| < 1$ for all $k = 1,\ ...,\ p$, then

$$x_t = (1 - \vartheta_1 L)^{-1} (1 - \vartheta_2 L)^{-1} \ ... \ (1 - \vartheta_p L)^{-1} \varepsilon_t.$$

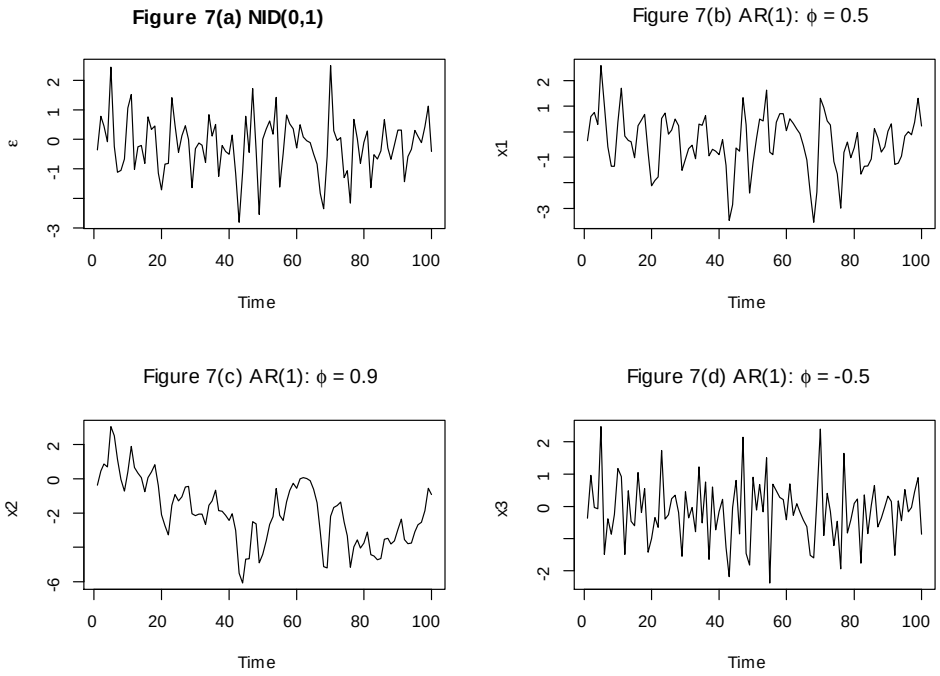
Applying the steps taken for the $AR(1)$ process p times, the readers show be able to show that

$$f_x(\lambda) = \frac{\sigma^2}{2\pi \left|1 - \phi_1 e^{-i\lambda} - \phi_2 e^{-i2\lambda} - \ ... \ - \phi_p e^{-ip\lambda}\right|^2}.$$

Now consider realisations of $AR(1)$ processes with different values of ϕ. Let

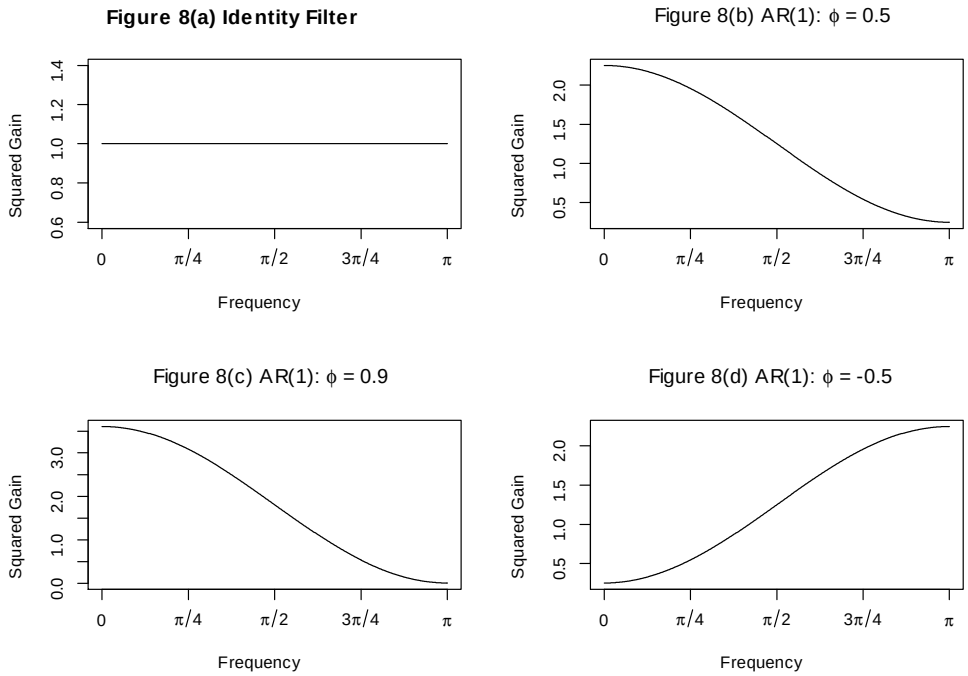
$$X_t = \phi X_{t+1} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is an independent process following the standard normal distribution.



Figures 7(a) - (d) show how smoothness or roughness of the plots can be affected by different values of ϕ. Similar to the explanation given for the moving average smoothing techniqe, the smoothness or roughness of the plots can be explained by the squared gain plot. See

Figures 8(a) - (d).



The idea of moving and exponential smoothing can be combinded as

$$X_t - \phi_1 X_{t-1} - \ \ldots \ - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ \ldots \ + \theta_q \varepsilon_{t-q}, \tag{41}$$

where $\{\varepsilon_t\}$ is a white noise process. The process $\{x_t\}$ satisfying (41) is known as an Autoregressive Moving Average model of order $(p,\ q)$, denoted by $ARMA\,(p,\ q)$. To obtain the spectral density function of $\{X_t\}$, the following proposition can be proven.

Theorem 5 *Let $\{X_t\}$ be an $ARMA\,(p,\ q)$ process satisfying*

$$\phi\,(L)\,X_t = \theta\,(L)\,\varepsilon_t,$$

where $\{\varepsilon_t\}$ is a white noise process with variance σ^2, $\phi\,(z) = 1 - \phi_1 z - \ \ldots \ - \phi_p z^p$ and $\theta\,(z) = 1 + \theta_1 z + \ \ldots \ + \theta_q z^q$ have no common zeros and $\phi\,(z)$ has zeroes outside the unit circle. Then $\{X_t\}$ has spectral density

$$f_x\,(\lambda) = \frac{\sigma^2}{2\pi} \frac{\left|\theta\left(e^{-i\lambda}\right)\right|^2}{\left|\phi\left(e^{-i\lambda}\right)\right|^2}, \quad \lambda \in [-\pi,\ \pi]\,.$$

The statement $\phi(z)$ has zeroes outside the unit circle means

$$\phi(z^*) = 0 \text{ implies } |z^*| > 1.$$

Let $\vartheta_1, \ldots, \vartheta_p$ be the numbers such that $\phi(z) = (1 - \vartheta_1 z) \ldots (1 - \vartheta_p z)$. Then $(1 - \vartheta_k z^*) = 0$ for all $k = 1, \ldots, p$. That is $z^* = \vartheta_k^{-1}$ is a solution to the equation $\phi(z) = 0$. Hence $|z^*| = |\vartheta_k|^{-1}$, and $|z^*| > 1$ implies that $|\vartheta_k| < 1$ for all $k = 1, \ldots, p$. This condition is stated to make sure that the operators $(1 - \vartheta_k L)^{-1}$ is well-defined as discussed earlier.

We have seen that application of either the moving average or the autoregressive filter to an original time series can make some substantial alteration to the original series. We have seen that some sort of trends can be created from these two types of filters. Figures 5(a) - (d) and Figures 7(a) - (d) show how a spurious trend or pattern can be created once these filters are applied to an independent white noise process which is suppose to random and exhibits no trends or patterns.

6.2 Exponential Smoothing and Volatilty of Key Economic Variables

It is sometimes customary to employ an autoregressive filter to monitor stochastic volatility of certain economic time series. Policy maker may need to monitor volatility of the stock indices or foreign exchanges to avoid excessive volatility. One popular approach is to employ the following filter

$$V_t = \phi V_{t-1} + e_t^2,$$

where V_t are daily volatility and e_t are the stock returns or appreciation of a currency. Given that the stock prices or the movement of a currency are believed to to have a unit root, their returns $\{e_t\}$ is therefore stationary. If we assume further that the process $\left\{e_t^2\right\}$ is also stationary, then the volatility process $\{V_t\}$ is just a filtered series of the process $\left\{e_t^2\right\}$. The popularity of this choice of volatility measure may be due to its resemblance with the famous ARCH model.

With reference to the previous subsection, it follows that the spectral density of the process $\{V_t\}$ is

$$f_V(\lambda) = \frac{1}{|1 - \phi e^{-i\lambda}|^2} f_e(\lambda).$$

To see how the original series is affected by this autoregressive filter we consider the squared gain plot. As the value of ϕ is normally chosen to be a positive number close to unity, the squared gain in Figure 8(c) indicates that policy makers may put too much weight on long-run dynamics of the volatility rather than the medium-run or short-run dynamics.

6.3 Seasonal Adjustment

In the interview conducted by Phillips (1988) in 1986, Professor James Durbin, the world renowned statistician and econometrician, mentioned how had gotten involved in research on the topic of seasonal adjustment procedures. It started from Prime Minister Harold Wilson's idea that the strange behaviour of the unemployment time series could be a result of the seasonal adjustment procedure that was being used. After a year of two of pursuing an investigation in this area, James Durbin and Robert Brown found that Harold Wilson was right. There was something wrong with the seasonal adjustment. James Durbin thought it was remarkable that a point like this should be spotted by a prime minister.

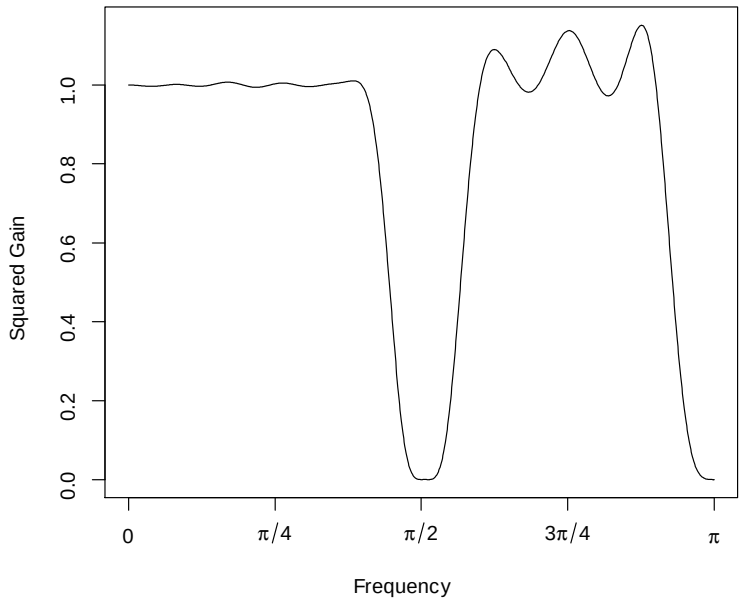
On the other side of the world and forty years later, policy makers in Thailand usually take the issue of seasonal adjustment for granted. Possibly all policy makers notice that seasonal patterns are a common sailient feature in most economic time series. Many of them believe that those seasonal patterns should be removed from the series so that long-term or medium-term fluctuations could become more apparent. However, to our knowledge, the default setups in some famous computer packages are usually employed to deal with seasonal patterns. No special adjustments, other than the default one, seem to have been employed to take into account some extreme events such as the Great Flood in 2011. Most modelers in policy units often treat it as easy and for granted. When one is encountered with a question about what seasonality is, we have not so far got any sensible answer.

There are two popular approaches employed by official statistical units around the world to deal with the issue of seasonality, namely, the X12ARIMA and TRAMO SEATS methodology where the latter methodology has gain more popularity in recent years. Given that X12ARIMA is the only method employed to deal with seasonality in Thailand, we will only briefly discuss the issues related with the X12ARIMA procedure. There are 4 decomposition procedures in the X12 or X13 ARIMA technique, namely, additive, log-additive, multiplicative and pseud-additive decompositions. On the the additive or log-additive can be considered as linear filters. The other are non-linear. Ghysels, Granger and Siklos (1996) showed how non-linearity can have very serious consequences on the deseasonalised series.

Laroque (1977) and Ghysels (1984) showed that the additive and log-additive decompositions can be approximated by the linear filters. The readers can see the original papers to see the numerical values of the weights assigned to the filters. To see efffects of these linear

filters, we focus on the squared gain derived from the quarterly filters.

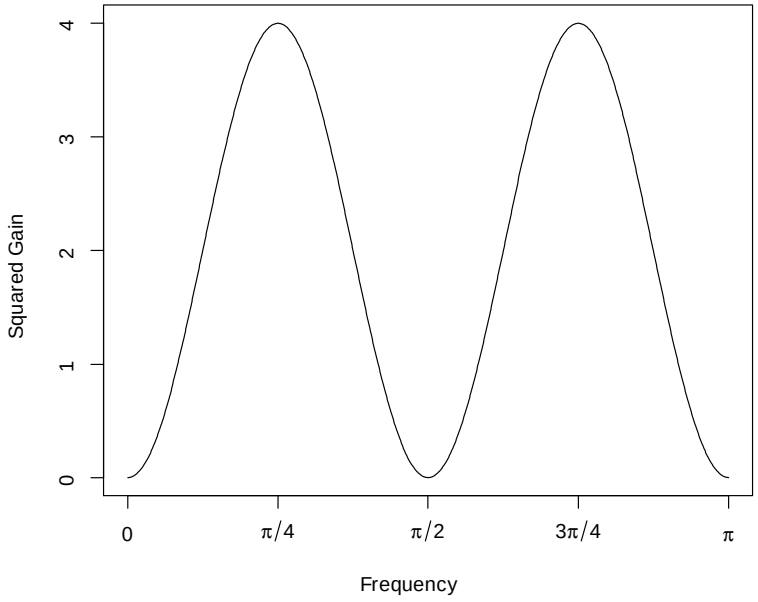
Figure 9 Squared Gain Function of the Q Filter



It can be seen from Figure 9 that the squared gain function remove all fluctuations with frequency around $\pi/2$ and π from the orignal series while leaving fluctuations from other frequencies rather unaffected. Given that the original series is a quarterly data, the sinusoids with frequency $\pi/2$ and π are the ones which complete their cycles in 4 quarters and 2 quarters, respectively. That is the linear quarterly filter remove all fluctuations which repeat themselves on a yearly basis. This conclusion may be regarded as an achievement becuase one can completely remove what most people regard as the seasonal patterns. One can compare the achievement of this technique compared with the crude YoY (year on year)

comparison in Figure 10.

Figure 10 Squared Gain Function of the Y-Y Filter



However, the main drawback of this approach is that the deseasonalised series will have spectral density equal to zero at frequencies $\pi/2$ and π. If policy makers wish to employ this deseasonalised series to construct a model based on the popular Box-Jenkins methodology, one will create an internal inconsistency. Theorem 5 implies that with the standard parameter space associated with the ARMA model, one can never create a spectral density function with $f(\pi/2) = f(\pi) = 0$. Hence, the ARMA model will be inappropriate. This conclusion can be applied to the VAR technique, the multivariate extension of the ARMA model popularly utilised by macroeconomists, too.

6.4 Band-pass and low-pass filters

Macroeconomists are particularly interested in two types of filters, namely, band-pass and low-pass filters. We first discuss the band-pass filter. Some economists believed that the business cycles have periods between xxx to xxx quarters. Hence, to study the behaviour of each macroeconomic time series associated with the business cycle, it may be more appropriate to filter the low frequency and high frequency components out. See Stock and Watson(???). A band-pass filter is the filter $\{g_\tau\}$ such that the squared gain function satisfies the following property:

$$|T(\lambda)|^2 = \begin{cases} 1, & \lambda_1 \leq |\lambda| \leq \lambda_2, \\ 0, & \text{otherwise.} \end{cases} \tag{42}$$

Such filter $\{g_\tau\}$ will remove all random sinusoids with frequency lower than λ_1 and those with frequency higher than λ_2. Recall that sinusoids with negative frequencies can be regarded as those with positve frequencies. Recall that we stick with our convention of interpreting only sinusoids with positive frequency. Recall from Corollary 1 that

$$T(\lambda) = \sum_{\tau=-\infty}^{\infty} g_\tau e^{-i\tau\lambda}. \tag{43}$$

To find the sequence $\{g_\tau\}$ satisfying equations (42) and (43), we need some results from Fourier Analysis. If $\{g_\tau\}$ is absolutely summable, it follows that

$$g_\tau = \frac{1}{2\pi} \int_{-\pi}^{\pi} T(\lambda)\ e^{i\tau\lambda}\ d\lambda. \tag{44}$$

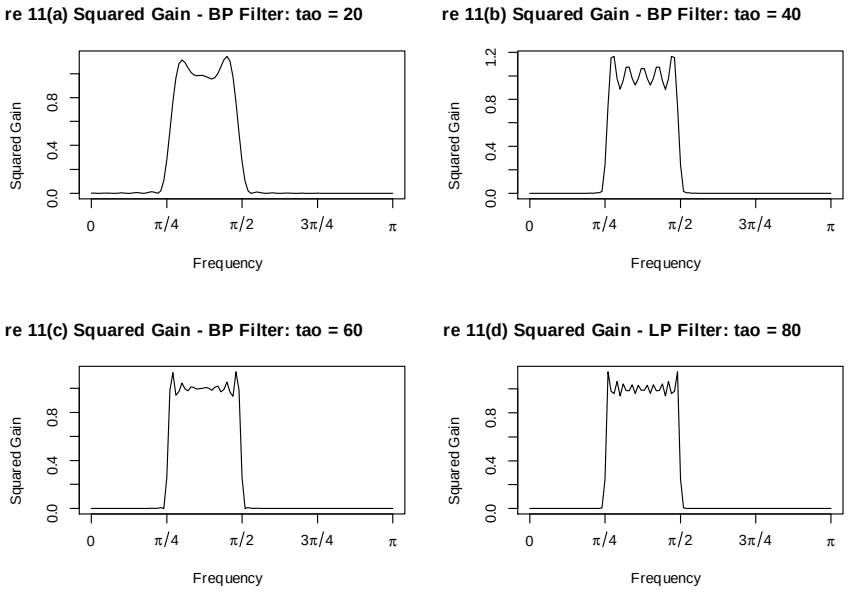
Choosing $T(\lambda) = 1$, for $\lambda_1 \leq |\lambda| \leq \lambda_2$, and 0, otherwise, will make $T(\lambda)$ satisfy (42). Then (44) becomes

$$g_\tau = \frac{1}{2\pi} \int_{\lambda_1}^{\lambda_2} e^{i\tau\lambda}\ d\lambda + \frac{1}{2\pi} \int_{-\lambda_2}^{-\lambda_1} e^{i\tau\lambda}\ d\lambda$$

so that

$$g_\tau = \begin{cases} \dfrac{\lambda_2 - \lambda_1}{\pi} & \tau = 0, \\ \dfrac{\sin \lambda_2\tau - \sin \lambda_1\tau}{\tau\pi} & \tau \neq 0. \end{cases}$$

The problem with this sort of filter is that it requires an infinite number of observations on x_t. The feasible filter such that $g_\tau = 0$ for large values of $|\tau|$ can only give an approximation to the problem. See Baxter and King (???) for a popular approximated band-pass filter. Figure 11(a) - (d) show the squared gain of band-pass filters for various values of τ.



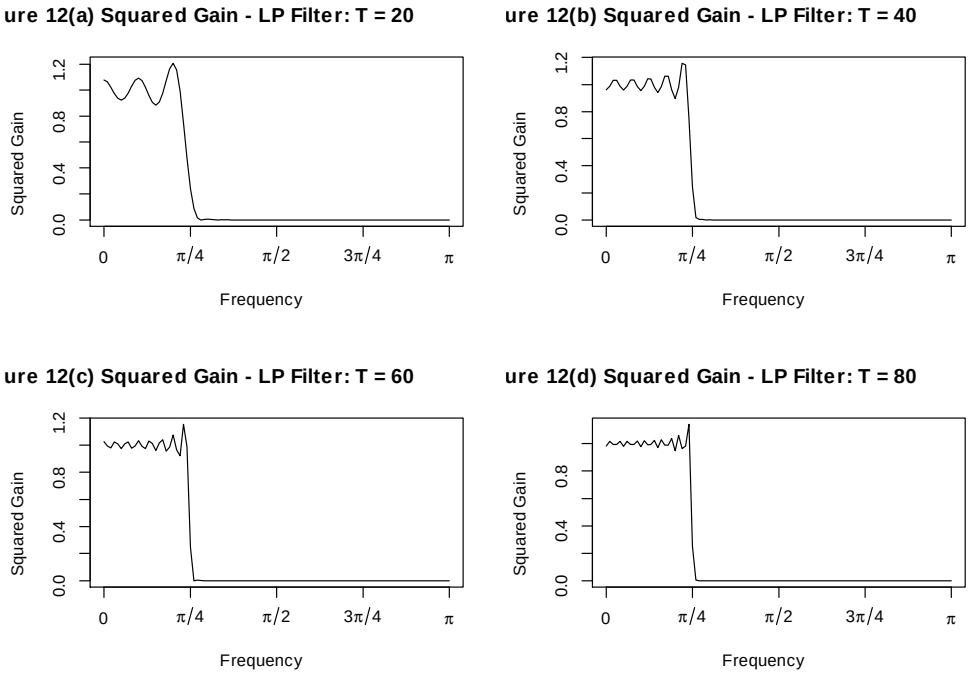
Now we focus on another type of filter useful to macroeconomists. Explain economic intuition. Estimate the trend. Then one can find the so-called output gap once the trend is estimated. It is called a low-pass filter. This filter basically passes all sinusoids with low frequency but suppresses all other sinusoids. We can formalize a low pass filter as a filter $\{g_\tau\}$ such that

$$|T(\lambda)| = \begin{cases} 1, & |\lambda| \leq \lambda_0, \\ 0, & \text{otherwise}, \end{cases}$$

where $T(\lambda)$ is the transfer function. We can apply the result from the band-pass filter by setting $\lambda_2 = \lambda_0$ and $\lambda_1 = 0$ to show that an ideal low-pass filter has the following form

$$g_\tau = \begin{cases} \lambda_0/\pi, & \tau = 0, \\ \dfrac{\sin \lambda_0 \tau}{\tau \pi}, & \tau \neq 0. \end{cases}$$

Alternatively, the readers can start by solving equation (44). As is the case for the band-pass filter, this is just an ideal low-pass filter. In practice, one need to perform the standard truncation. Figures 12(a) - (d) show the squared gain of the low-pass filters for various values of τ.



6.5 HP Filter

One type of a low pass filter that gains a great deal of popularity among macroeconomists is the so-called HP filter popularized by Hodrick and Prescott (1980), later published in Hodrick and Prescott (1997). Hodrick and Prescott (1980) consider a problem of trend

estimation of the form

$$x_t = \tau_t + e_t,$$

where τ_t represent the trend of the time series $\{x_t\}$ and e_t are the irregular components representing deviation from the trend. The time series $\{x_t\}$ is supposed to be seasonally adjusted. Hodrick and Prescott (1980) recommend estimation of the trend components by the penalized least squares smoothing technique where the optimization problem is

$$\min_{\{\tau_t\}_{t=0}^{T+1}} \left\{ \sum_{t=1}^{T} (x_t - \tau_t)^2 + \varrho \sum_{t=1}^{T} \left[(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}) \right]^2 \right\}.$$

The parameter ϱ is a positive number employed to control smoothness of the trend $\{\tau_t\}$. Large values of ρ make the trend $\{\tau_t\}$ smooth. The first order condition of the objective function is

$$\begin{aligned} 0 &= -2(x_t - \tau_t^*) - 4\rho\left(\tau_{t+1}^* - 2\tau_t^* + \tau_{t-1}^*\right) \\ &\quad 2\rho\left(\tau_t^* - 2\tau_{t-1}^* + \tau_{t-2}^*\right) + 2\rho\left(\tau_{t+2} - 2\tau_{t+1} + \tau_t^*\right), \end{aligned}$$

where $\{\tau_t^*\}$ is a sequence of optimal trend. Solving this Euler equation we obtain

$$\begin{aligned} x_t &= \left[\tau_t^* + \rho\left(\tau_{t+2}^* - 4\tau_{t+1}^* + 6\tau_t^* - 4\tau_{t-1}^* + \tau_{t+2}^*\right)\right] \\ &= \left[1 + \rho(1-L)^2\left(1 - L^{-1}\right)^2\right]\tau_t^*, \end{aligned}$$

where L is the lag opeator and L^{-1} is the forward operator, i.e. $L^{-1}\tau_t^*$ denotes τ_{t+1}^*. We can theoretically consider a general process $\{x_t : t \in \mathbb{Z}\}$ so that we have

$$\tau_t^* = \left[1 + \rho(1-L)^2\left(1 - L^{-1}\right)^2\right]^{-1} x_t. \tag{45}$$

That is τ_t^* is a function of x_t for all $t \in \mathbb{Z}$. It is important to note that to show that an inverse operator of $\left[1 + \rho(1-L)^2\left(1 - L^{-1}\right)^2\right]$ exists is non-trivial. For some sketched derivation see King and Rebelo (1993). Equation (45) indicates how we can construct the squared gain function of the HP filter. This is the same as the derivation of the autoregressive filter. It can be shown that the spectral density function of $\{\tau_t^*\}$ is

$$\begin{aligned} f_\tau(\lambda) &= \frac{1}{\left|1 + \rho\left(1 - e^{-i\lambda}\right)^2\left(1 - e^{i\lambda}\right)^2\right|^2} f_x(\lambda) \\ &= \frac{1}{\left|1 + 16\rho\sin(\lambda/2)^4\right|^2} f_x(\lambda), \end{aligned}$$

where f_x is the original spectral density of $\{x_t\}$. Gomez (2001) showed that the HP filter is a special class of the so-called Butterworth filter commonly applied in engineering

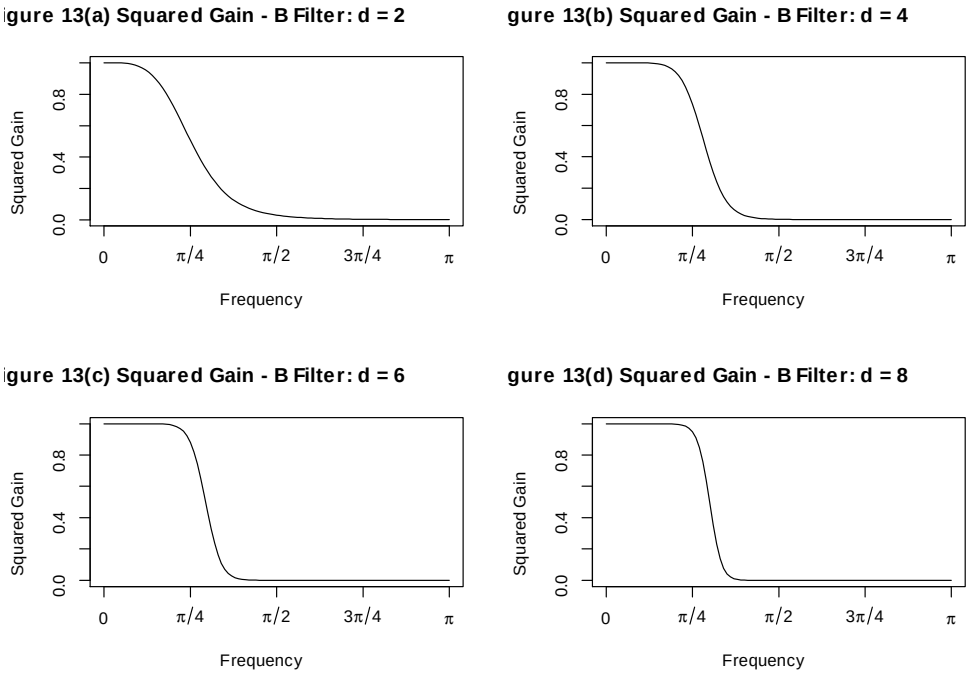
applications. Butterworth filter has the squared gain function of the form

$$|G(\lambda)|^2 = \frac{1}{\left|1 + \left(\frac{\sin(\lambda/2)}{\sin(\lambda_0/2)}\right)^{2d}\right|^2},$$

where λ_0 is the frequency at which $G(\lambda_0) = 1/2$, and $d = 1,\ 2,\ ...$ is a chosen parameter. The higher values d yield sharper results. It can be seen from the squared gain function that the HP filter is a special class of the Butterworth filter where d is set to be equal to 2 and the parameter ρ in the HP filter determines the value of λ_0, i.e.

$$\rho = \left[\frac{1}{2\sin(\lambda_0/2)}\right]^4.$$

Figures 13(a) - (d) show the squared gain of the Butterworth filter for various values of d and λ_0 is set to be equal to unity. The case for $d = 2$ corresponds to the HP filter. It can be seen that high values of d gives sharper results.



As the HP filter is a special class of the Butterworth filter, it would be more optimal to switch from the conventional HP filter to the Butterworth filter. Moreover, the HP filter is a one-sided filter. It is therefore subject to data revision. Furthermore, Harvey and Jaeger (1993) show that application of the HP filter on some nonstationary time series can create spurious cycles and hence is subject to the Yule-Slutsky effect.

6.6 Regression Analysis

The purpose of this subsection is to dispel one myth commonly believed among applied economists working with economic modelling at the central bank in Thailand. It is believed that if one does not remove seasonality, then regression analysis involving economic time series exhibiting seasonal patters will lead to spurious relation or conclusion. The source of spurious relationship is seasonality. To see why this myth is incorrect, we need to introduce an extension of spectral analysis to multivariate time series.

Suppose $\mathbf{X}_t = (X_{1t},\ ...,\ X_{pt})'$ be a vector of time series. If the multivariate time series $\{\mathbf{X}_t\}$ is (covariance) stationary, and

$$\sum_{u=-\infty}^{\infty} \|\Gamma_X(u)\| < \infty,$$

where Γ_X is the autocovariance function of the process $\{\mathbf{X}_t\}$ and $\|\cdot\|$ is a matrix norm, then we can define the spectral density function of $\{\mathbf{X}_t\}$ as

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} e^{-iu\lambda}\, \Gamma_x(u), \quad -\pi < \lambda \leq \pi.$$

In particular, suppose $\mathbf{X}_t = (X_t,\ Y_t)'$. Then we can consider the cross spectral density of $\{X_t\}$ and $\{Y_t\}$ as

$$f_{xy}(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} e^{-iu\lambda} \gamma_{xy}(u), \tag{46}$$

where γ_{xy} is the cross-autocovariance of $\{X_t\}$ and $\{Y_t\}$.

Consider a simple time series regression

$$Y_t = \beta X_t + V_t, \tag{47}$$

where X_t are the regressors and V_t are the disturbances with zero mean. Suppose $\{X_t\}$ and $\{V_t\}$ are stationary with covariance functions γ_x and γ_v respectively. Under the usual assumption of zero correlation between $\{X_t\}$ and $\{V_t\}$, it follows that $\{Y_t\}$ is stationary with covariance function

$$\gamma_y(u) = \beta^2 \gamma_x(u) + \gamma_v(u),$$

where γ_v is the autocovariance function of $\{V_t\}$. Moreover, the cross-autocovariance of $\{Y_t\}$ and $\{X_t\}$ is

$$\gamma_{xy}(u) = \mathbb{E}(X_{t+u} - \mathbb{E}X_{t+u})(Y_t - \mathbb{E}Y_t) = \beta\gamma_x(u).$$

Since the cross-autocovariance between X_{t+u} and Y_t does only on the distance u, we can say that they are jointly stationary. Assuming that $\sum_{u=-\infty}^{\infty} |\gamma_x(u)| < \infty$ and $\sum_{u=-\infty}^{\infty} |\gamma_v(u)| < \infty$, then the linear relationship of $\{Y_t\}$ and $\{X_t\}$ in (47) and the property of the cross spectral

density in (46) imply that

$$f_{xy}(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} e^{-iu\lambda} \beta \gamma_x(u) = \beta f_{xx}(\lambda), \quad \lambda \in (-\pi,\ \pi],$$

where $f_{xx}(\lambda)$ is the spectral density function of $\{X_t\}$. That is under the linear relationship in (47)

$$\beta = \frac{f_{xy}(\lambda)}{f_{xx}(\lambda)}, \quad \lambda \in (-\pi,\ \pi]. \tag{48}$$

Hence the slope parameter of a linear regression in (47) can be interpreted as the ratio between the cross spectral density function f_{xy} and the spectral density function of $\{X_t\}$. Note that this constant ratio holds for all frequency λ in $(-\pi,\ \pi]$. A strong peak of the spectral density functions at some certain frequencies associated with seasonality will therefore be unlikely to cause spurious relationship from regression analysis. It is worth noting that the least squares estimate of β can be written as an average ratio of an approximate analogue of the ratio in (48).

7 Reflections and Conclusions

We have illustrated how the spectral analysis of time series can shed light on many issues and techniques commonly encountered by economists. We have shown that the standard routine commonly employed by policy makers such as the 3-period moving average and the exponential moving average filter can result in undesirable effects. They can potentially create spurious patterns, known as the Yule-Slutsky effect. We have also explained the reason why deseasonalised time series can cause many problems. The issue of trend estimation was discussed where the HP filter is shown to be sub-optimal compared with the Butterworth filter. It is very important to remind the reader that our discussion of the time invariant linear filters only covers stationary series. An extension to nonstationary series such as those with a unit root needs to be done with care.

8 Appendix

8.1 Appendix 1

Let U be the circulant shift matrix defined by

$$u_{ij} = \begin{cases} 1 & \text{for } j = i+1,\ i = 1,\ ...,\ N-1, \\ 1 & j = 1,\ i = N, \\ 0 & \text{otherwise.} \end{cases} \tag{49}$$

That is U is a matrix of the form

$$U = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

It should be noted that U is an orthogonal matrix, i.e. $U' = U^{-1}$. The reason for the name "shift matrix" is that for $X = (X_1,\ X_2,\ ...,\ X_N)'$, it follows that

$$UX = (X_2,\ X_3,\ ...,\ X_N,\ X_1)'.$$

The covariance matrix of UX is $U\Gamma U'$. However, due to conditions (12) and (13), the covariance matrix of UX is also Γ. This implies that $U\Gamma U' = \Gamma$. As U is orthogonal,

$$U\Gamma = \Gamma U. \tag{50}$$

This relation suggests a connection among eigenvectors of Γ and U.

It follows from Lemmas 1 and 2 that eigenvectors of Γ are of the form

$$p_k = \frac{1}{\sqrt{N}} \left(e^{i\omega_k},\ e^{i2\omega_k},\ ...,\ e^{iN\omega_k}\right)',$$

where $\omega_k = 2\pi k/N,\ k = 1,\ ...,\ N$. Let P be a matrix whose j-th column is p_j. Then it can be verify that P is a unitary matrix, i.e. $P^* = P^{-1}$.

8.2 Appendix 2: Technical Lemmas

Lemma 1 *The eigenvalues of the matrix U defined in (49) are*

$$\omega = e^{i2\pi k/N},\ k = 1,\ ...,\ N.$$

Let $\psi_k = 2\pi k/N,\ k = 1,\ ...,\ N$. The vector $\left(e^{i\psi_k},\ e^{i2\psi_k},\ ...,\ e^{iN\psi_k}\right)'$ is an eigenvector associated with the eigenvalue $e^{i\psi_k}$.

Proof. Let I_N be the identity matrix of order N. The characteristic equation of U is $|U - \omega I_N| = 0$, where $|\cdot|$ denotes determinant of a matrix. Let $A = U - \omega I_N$. Then

$$a_{ij} = \begin{cases} 1 & j = i+1,\ i = 1,\ ...,\ N-1, \\ 1 & j = 1,\ i = N, \\ -\omega & j = i,\ i = 1,\ ...,\ N, \\ 0 & \text{otherwise.} \end{cases}$$

Expanding the last row of A, it follows that

$$|A| = (-1)^{N+1} |C_1| - \omega |C_2|,$$

where C_1 is a square matrix of order $N-1$ defined by

$$c_{1ij} = \begin{cases} 1 & j = i \\ -\omega & j = i-1, \\ 0 & \text{otherwise,} \end{cases}$$

and C_2 is a square matrix of order $N-1$ defined by

$$c_{2ij} = \begin{cases} 1 & j = i+1 \\ -\omega & j = i, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that both C_1 and C_2 are triangular matrices. Thus $|C_1| = 1$ and $|C_2| = (-\omega)^{N-1}$. It follows that the characteristic equation is

$$\omega^N = 1.$$

That is

$$\omega = e^{i2\pi k/N},\ k = 1,\ ...,\ N.$$

Let $\psi_k = 2\pi k/N$. Let $y = (y_1,\ ...,\ y_N)'$ be an eigenvector associated with an eigenvalue $\omega = e^{i\psi_k}$. Then the relation $Uy = \omega y$ implies

$$\begin{aligned} y_{k+1} &= \omega y_k, \quad k = 1,\ ...,\ N-1, \\ y_1 &= \omega y_N. \end{aligned}$$

This implies that $\left(e^{i\psi_k},\ e^{i2\psi_k},\ ...,\ e^{iN\psi_k}\right)'$ is an eigenvector associated with the eigenvalue $e^{i\psi_k}$. ■

Lemma 2 *Eigenvectors of U are eigenvectors of Γ defined in (13).*

Proof. Let p_j is an eigenvector of U associated with an eigenvalue ω_j. Then $\Gamma U p_j = \omega_j \Gamma p_j$ for all $j = 1,\ ...,\ N$. Equation (50) implies that

$$U\Gamma p_j = \omega_j \Gamma p_j, \quad j = 1,\ ...,\ N.$$

That is Γp_j is also an eigenvector of U associated with an eigenvalue ω_j. Since ω_j are all

distinct for $j = 1,\ ...,\ N$, it follows that there exist constants c_j such that

$$\Gamma p_j = c_j p_j, \quad j = 1,\ ...,\ N.$$

That is p_j are also eigenvectors of Γ. ■

Lemma 3 *Let $\{y_t\}$ be the process with the autocovariance function in (34) and γ_x is an autocovariance function satisfying Assumption (1). Then the spectral density function f_y of $\{y_t\}$ is well defined and for any $\lambda \in (-\pi,\ \pi]$,*

$$\left| \frac{e^{-i2\lambda} + e^{-i\lambda} + 1}{3} \right|^2 f_x(\lambda).$$

Proof. Assumption (1) implies that the spectral density function f_x of $\{x_t\}$ is well defined. Condition (34) and Assumption (1) imply that

$$\sum_{u=-\infty}^{\infty} \left|\gamma_y(u)\right| < \infty$$

so that the spectral density function f_y of $\{y_t\}$ is well defined. By Theorem 2, for any $\lambda \in (-\pi,\ \pi]$,

$$\begin{aligned} & f_y(\lambda) \\ = \ & \frac{1}{9 \cdot 2\pi} \sum_{u=-\infty}^{\infty} e^{-iu\lambda} \left\{ \gamma_x(u-2) + 2\gamma_x(u-1) + 3\gamma_x(u) + 2\gamma_x(u+1) + \gamma_x(u+2) \right\}. \end{aligned}$$

The first sum involving $\gamma_x(u-2)$ is

$$\begin{aligned} \sum_{u=-\infty}^{\infty} e^{-iu\lambda} \gamma_x(u-2) &= e^{-i2\lambda} \sum_{u=-\infty}^{\infty} e^{-i(u-2)\lambda} \gamma_x(u-2) \\ &= 2\pi e^{-i2\lambda} f_x(u), \end{aligned}$$

where the last equality follows from Theorem 2. We can apply the same manipulation to other terms to show that

$$\begin{aligned} f_y(\lambda) &= \frac{e^{-i2\lambda} + 2e^{-i\lambda} + 3 + 2e^{i\lambda} + e^{i2\lambda}}{9} f_x(\lambda) \\ &= \left| \frac{e^{-i2\lambda} + e^{-i\lambda} + 1}{3} \right|^2 f_x(\lambda). \end{aligned}$$

■

References

Baxter, M., King, R.G., 1999. Measuring business cycles: Approximate band-pass filters for economic time series. The Review of Economics and Statistics 81, 575-593.

Box, G.E.P, Jenkins G.M., 1970. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Brockwell, P.J., Davis R.A., 1991. Time Series: Theory and Methods. Springer-Verlag, New York.

Dagum, E.B., 1980. The X11 ARIMA seasonal adjustment method. Statistics Canada.

Ghysels, E., 1984. The Economics of Seasonality: The Case f the Money Supply, Unpublished Ph.D. dissertation. Department of Managerial Economics and Decision Science, Northwestern University.

Ghysels, E., Granger, C.W.J., Siklos, P., 1996. Is seasonal adjustment a linear or nonlinear data filtering process? Journal of Business and Economic Statistics 14, 374-386.

Gomez, V., 2001. The use of Butterworth filters for trend and cycle estimation in economic time series. Journal of Business and Economic Statistics 19, 365-373.

Gomez, V., Maravall A., 2001. Seasonal adjustment and signal extraction in economic time series. In: Pena, D., Tiao, G.C. Tsay R.S. (Eds.), A Course in Time Series Analysis. J. Wiley and Sons, New York.

Grenander, U., Szego, G., 1958. Toeplitz Forms and Their Applications. University of California Press, Berkeley and Los Angeles.

Hannan, E.J., 1960, Time Series Analysis. Science Paperbacks and Methuen & Co Ltd, London.

Harvey, A.C., Jaeger, A., 1993. Detrending, stylized facts and the business cycle. Journal of Applied Econometrics 8, 231-247.

Hodrick, R.J., Prescott, E.C., 1980. Postwar U.S. business cycles: An empirical investigation. Discussion Paper No. 451, Carnegie-Melow University.

Hodrick, R.J., Prescott, E.C., 1997. Postwar U.S. business cycles: An empirical investigation. Journal of Money, Credit and Banking 29, 1-16.

Laroque, G., 1977. Analyse d'une Methode de Desaisonnalisaton: Le Programme X11 du US Bureau of the Census Version Trimestrielle. Annales de l'INSEE 88, 105-127.

Phillips, P.C.B., 1988. The ET interview: Professor James Durbin. Econometric Theory 4, 125-157.

Slutzky, E., 1927. The summation of random causes as the source of cyclic processes (in Russian). Problems of Economic Conditions, edited by The Conjuncture Institute, Moscow 3, 34-64.

Slutzky, E., 1937. The summation of random causes as the source of cyclic processes. Econometrica 5, 105-146.

Stock, J.H., Watson, M.W., 1999. Business cycle fluctuations in us macroeconomic time series. In: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics. Elsevier,

Amsterdam.

Yule, G.U., 1927. On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. Philosophical Transactions of the Royal Society of London. Series A 226, 267-298.

Tukey, J.W., 1961. Discussion, Emphasizing the connection between analysis of variance and spectrum analysis. Technometrics 3, 191-219.