4 Spectral Analysis: analysis in the frequency domain

4.1 Moment Generating Function

- For a discrete random variable Y, the moment generating function is

$$M_Y(t) = \mathsf{E}\left[e^{tY}\right] = \sum_y e^{ty} P(Y = y)$$

- The mgf is a transform of $P(Y = y)$ which (usually subject to regularity conditions) gives the moments of the random variable Y.

- Thus at least for the discrete case, $P(Y = y)$ and $M_Y(t)$ share a mathematical equivalence, but supply different information about Y.

4.2 Fourier Transform

- For real $-\pi \leq \omega \leq \pi$, the *discrete Fourier transform* of a (possibly complex) function $h(y)$ of a real variable y is

$$H(\omega) = \sum_{y=-\infty}^{\infty} h(y) e^{-i\omega y}$$

- *Inversion*: given a Fourier transform $H(\omega)$ we can recover $h(y)$ as

$$h(y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(\omega) e^{i\omega y} d\omega$$

- *Even Function*: if $h(-y) = h(y)$, then noting

$$e^{-i\omega y} + e^{i\omega y} = 2\cos(\omega y),$$

we see

$$H(\omega) = h(0) + 2\sum_{y=1}^{\infty} h(y)\cos(\omega y).$$

4.3 The Spectrum

The spectral representation of a time series decomposes it into a sum of sinusoidal components with uncorrelated coefficients. The spectral decomposition is thus an analogue for stationary processes of the more familiar Fourier representation of deterministic functions.

The analysis of stationary processes by means of their spectral representation is often referred to as 'frequency domain analysis' or 'spectral analysis'.

The *periodogram* provides a sample-based estimator of the spectral density.

- We have a *stationary random sequence* $\{Y_t\}$ with autocovariance function γ_k and autocorrelation function ρ_k.

- *Autocovariance generating function*: for z a complex variable,

$$G(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k.$$

- Taking $z = e^{-i\omega}$ and noting $\gamma_k = \gamma_{-k}$, and

$$e^{-i\omega y} + e^{i\omega y} = 2\cos(\omega y),$$

the *(power) spectrum* is defined as

$$\begin{aligned} f(\omega) &= \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k} \\ &= \gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k \cos k\omega. \end{aligned}$$

- *Normalized spectrum*:

$$f^*(\omega) = \frac{f(\omega)}{\gamma_0} = 1 + 2\sum_{k=1}^{\infty} \rho_k \cos k\omega$$

- $f(\omega) = f(-\omega)$, $f(\omega + 2\pi m) = f(\omega)$ for all integers m so that the spectrum need only be defined for $0 \leq \omega \leq \pi$.

- *Periodogram and the normalized periodogram*:

$$\begin{aligned} I(\omega) &= g_0 + 2\sum_{k=1}^{n-1} g_k \cos k\omega \\ \frac{I(\omega)}{g_0} &= 1 + 2\sum_{k=1}^{n-1} r_k \cos k\omega \end{aligned}$$

- The spectrum and autocorrelation function *inversion*:

$$\begin{aligned} \gamma_k &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega) e^{i\omega k} d\omega \\ &= \frac{1}{\pi}\int_0^{\pi} f(\omega)\cos k\omega d\omega \quad (**) \end{aligned}$$

Wold's Theorem: (not examinable)

Any non-negative valued function $f(\omega)$ on $(0, \pi)$ defines a legitimate spectrum, and γ_k is therefore a legitimate autocovariance function if and only if it can expressed in the form $(**)$ for some such $f(\omega)$.

Proof: Beyond the scope of this course. [See, for example, Priestly (1981).]#

The important point here is that the permissible form of γ_k is quite severely constrained, whereas that for $f(\omega)$ is not.

Wiener-Khintchine Theorem: (not examinable)

For any real-valued stationary stochastic process with acf, γ_k, there exists a monotonically increasing function $F(\omega)$ such that

$$\gamma_k = \int_0^{\pi} \cos(\omega k) dF(\omega).$$

It can be shown that

$$\gamma(0) = \mathrm{var}(Y_t) = \int_0^{\pi} dF(\omega) = F(\pi).$$

Some Examples

1. White noise (normal case)

- $Y_t = Z_t$, where $Z_t \sim N(0, \sigma^2)$ and the Z_t are independent.

- **autocovariance function**

$$\gamma_k = \begin{cases} \sigma^2 & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

- **autocorrelation function**

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

- **spectrum**

  For all ω,

$$\begin{aligned} f(\omega) &= \sigma^2 \\ f^*(\omega) &= 1 \end{aligned}$$

2. Autoregressive process of order 1

- $Y_t = \alpha Y_{t-1} + Z_t$, where $Z_t \sim N(0, \sigma^2)$ and the Z_t are independent.

- **autocovariance function**: For $|\alpha| < 1$

$$\gamma_k = \begin{cases} \sigma_Y^2 = \sigma^2/(1-\alpha^2) & k = 0 \\ \sigma_Y^2 \alpha^k & k > 0 \end{cases}$$

- **autocorrelation function**

$$\rho_k = \begin{cases} 1 & k = 0 \\ \alpha^k & k > 0 \end{cases}$$

- **spectrum**

  For all ω,

$$\begin{aligned} f(\omega) &= \sigma^2/(1 - 2\alpha\cos\omega + \alpha^2) \\ f^*(\omega) &= (1-\alpha^2)/(1 - 2\alpha\cos\omega + \alpha^2) \end{aligned}$$

4.4 Linear filters

A *filter* is simply a transformation of one random sequence, say U_t into another Y_t.

- A *linear filter* is defined as
$$Y_t = \sum_{j=-\infty}^{\infty} a_j U_{t-j}$$
for some sequence a_j.

- If U_t is stationary, and the number of nonzero a_j is finite, Y_t is also stationary. If the latter is not true, the precise form of the a_j will determine if Y_t is stationary.

- *autocovariance function*: if U_t has acf $\gamma_U(k)$, then
$$\gamma_Y(k) = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} a_j a_i \gamma_U(k+i-j)$$

- the autocovariance generating function of Y_t is
$$G_Y(z) = A(z)A(z^{-1})G_U(z),$$
where $A(z) = \sum_{j=-\infty}^{\infty} a_j z^j$ and $G_U(z)$ is the autocovariance generating function of U_t.

- Putting $z = e^{-i\omega}$ we find
$$f_Y(\omega) = |a(\omega)|^2 f_U(\omega),$$
where
$$a(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j}$$
is called the *transfer function* of the linear filter.

4.5 General Linear Process

- *Recall*: $Y_t = \sum_{j=0}^{\infty} a_j Z_{t-j}$ where Z_t is white noise and a_j are constants.

- *Spectrum*:

$$\begin{aligned} f_Y(\omega) &= \sigma^2 |a(\omega)|^2 = \sigma^2 A(e^{-i\omega}) A(e^{i\omega}) \\ &= \sigma^2 \left(b_0 + \sum_{m=1}^{\infty} b_m \cos(m\omega) \right) \qquad (*) \end{aligned}$$

- *Fourier Analysis*: any real valued continuous function $f(\omega)$ which is even, $f(\omega) = f(-\omega)$, and has period 2π, $f(\omega + 2\pi m) = f(\omega)$, for integer m, can be expressed as $(*)$.

- *Thus* **any** *stationary random sequence with a continuous spectrum has a representation as a general linear process*.

4.6 Moving Average

$$Y_t = \theta(B) Z_t$$

where

$$\theta(B) = \sum_{j=0}^{q} \beta_j B^j$$

with $\beta_0 = 1$ and $\theta(0) = 1$.

- *Spectrum*:

$$\begin{aligned} f(\omega) &= \sigma^2 |\theta(e^{-i\omega})|^2 \\ &= \sigma^2 \left[\left\{ 1 + \sum_{j=1}^{q} \beta_j \cos(j\omega) \right\}^2 + \right. \\ &\qquad \left. \left\{ \sum_{j=1}^{q} \beta_j \sin(j\omega) \right\}^2 \right] \end{aligned}$$

4.7 Autoregressive process

$$\phi(B)Y_t = Z_t$$

where

$$\phi(B) = 1 - \sum_{j=1}^{p} \alpha_j B^j$$

and $\phi(0) = 1$.

- *Spectrum*:

$$|\phi(e^{-i\omega})|^2 f(\omega) = \sigma^2$$

$$f(\omega) = \sigma^2 \left[\left\{ 1 - \sum_{j=1}^{p} \alpha_j \cos(j\omega) \right\}^2 + \left\{ \sum_{j=1}^{p} \alpha_j \sin(j\omega) \right\}^2 \right]^{-1}$$

  There is a potential problem: the term in square brackets may be zero for some ω for specific α_j.

4.8 ARMA processes

$$\phi(B)Y_t = \theta(B)Z_t$$

where

$$\phi(B) = 1 - \sum_{j=1}^{p} \alpha_j B^j, \quad \theta(B) = 1 + \sum_{j=1}^{q} \beta_j B^j$$

and $\phi(0) = 1$ and $\theta(0) = 1$.

- *Spectrum*: must satisfy

$$|\phi(e^{-i\omega})|^2 f(\omega) = \sigma^2 |\theta(e^{-i\omega})|^2$$

  Thus the spectrum is σ^2 times the terms in the MA(q) and AR(p) spectra.

Usefulness of ARMA processes

- Useful *empirical* tool, summarising time series with a small number of parameters (α_i and β_j).
- Of limited value if the underlying mechanisms need to be understood.
- Useful if, for example, forecasting is the aim because of the wide variety of shapes the spectrum can take on for small p and q.

5 ARIMA processes, and fitting ARIMA processes to data

An extension which greatly enhances the value of ARMA processes as empirical descriptors of *non-stationary* time series is the class of *autoregressive integrated moving average processes*:

$$Y_t \sim ARIMA(p, d, q)$$

if the dth difference of Y_t is a *stationary, invertible* ARMA process of order p, q, i.e.,

$$\phi(B)(1 - B)^d Y_t = \theta(B) Z_t.$$

$\phi(B), \theta(B)$ are polynomials of degree p, q, with all roots of the equations $\phi(u) = 0$ and $\theta(u) = 0$ have modulus greater than 1.

Example: the simple random walk

$$Y_t - Y_{t-1} = Z_t$$

is an ARIMA(0,1,0) process.

5.1 Notation for ARIMA processes

ARMA(p, q) is *stationary* if all the roots of $\phi(B) = 0$ have modulus greater than one.

Consider the non-stationary series

$$\varphi(B) Y_t = \theta(B) Z_t$$

where

$$\varphi(B) = \phi(B)(1 - B)^d$$

and $\phi(B)$ has all its roots greater than one in modulus. Thus the full process has d unit roots. Notice that

$$(1 - B)Y_t = Y_t - BY_t = Y_t - Y_{t-1} = DY_t$$

$$(1 - B)^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2} = D^2 Y_t$$

and in general

$$(1 - B)^d Y_t = D^d Y_t$$

so that the unit roots correspond to differencing the original series. The differenced series is therefore a stationary ARMA(p, q) process.

The 'I' stands for integrated or differenced series.

(See separate handout for simulated ARIMA processes.)

This is a specialized form of a non-stationary process, but this and extensions (fractional difference) are useful in practice. These models are used to provide a summary of a time series and to forecast or predict future values. We proceed in three stages:

1. **identification** of plausible model(s),

2. **estimation** or fitting the model(s),

3. **diagnostic checks** of the fit.

5.2 Partial Autocorrelation function

Consider AR(p),

$$Y_t = \sum_{i=1}^{p} \alpha_i Y_{t-i} + Z_t$$

Change notation and write

$$Y_t = \sum_{i=1}^{p} \alpha_{ip} Y_{t-i} + Z_t$$

to indicate a pth order process. Thus if

$$Y_t = \sum_{i=1}^{r} \alpha_{ir} Y_{t-i} + Z_t$$

and $r > p$, we have

$$\alpha_{p+1,p+1} = \alpha_{p+2,p+2} = \cdots = \alpha_{rr} = 0$$

Therefore the extra autoregressive parameters are zero.

In practice, if we estimate $\alpha_{r,r}$ and plot against r, and if the process has order p, there should be a cutoff in the plot at $r = p$, after which these *partial autocorrelations* will be zero.

In R,

```
acf(series, type="partial")
```

produces the sample partial autocorrelations using an efficient algorithm.

As for sample autocorrelations, we have $\widehat{\alpha}_{kk} \sim N(0, 1/n)$ for $k > p$ if the process is AR(p).

The α_{ri} satify the Yule-Walker equations

$$\rho_k = \sum_{i=1}^{r} \alpha_{ri}\rho_{k-i}$$

and these equations, $k = 1, 2, \ldots r$, can be written in matrix form as

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_r \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_{r1} \\ \alpha_{r2} \\ \vdots \\ \alpha_{rr} \end{bmatrix}$$

or

$$\boldsymbol{\rho}_r = P_r\boldsymbol{\alpha}_r$$

Using Cramer's Rule, solutions for α_{rr} for $r = 1, 2$ are

$$\alpha_{11} = \rho_1,\ \alpha_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

This is not the recommended way to solve these equations. There are efficient recursive algorithms.

The partial correlogram is useful in identification of models for time series.

Why is it called a partial autocorrelation? In fact it is a correlation between two variables, given the intermediate values.

1. Identification

- The first step is to look at the data series for evidence of non-stationarity. If data appear stationary, no differencing is required, and provisionally we identify $d = 0$.

- If data appear non-stationary, successively difference the series until the time plot looks stationary (in practice, $d = 1$ or 2 is usually enough).

- If unclear, look at the correlogram. *Why?*

In general, we use the following to identify a model:

- correlogram

- partial correlogram

and the fact that

- for MA processes, the correlogram shows a cut-off

- for MA processes, the partial correlogram decays to zero

- for AR processes, the correlogram decays to zero

- for AR processes, the partial correlogram shows a cut-off

- for ARMA processes, the correlogram decays to zero as for AR, with the first few autocorrelations differing

- for ARMA processes, the partial correlogram decays to zero as for MA, with the first few partial autocorrelations differing.

Usually (p, d, q) are all 0, 1 or 2.

2. Estimation

We consider **maximum likelihood estimation** for an ARMA(p, q) model based on the normal distribution.

Two approaches:

- multivariate normal linear model,
- conditional decomposition of the multivariate normal.

The second approach is based on the fact that the joint density of a random vector $\boldsymbol{Y}$ can be decomposed into a product of marginal and conditional densities, namely

$$f(\boldsymbol{y}) = f(y_1) \prod_{i=2}^{n} f(y_i | y_{i-1}, y_{i-2}, \ldots, y_1)$$

If

$\mathsf{E}\left[Y_t | Y_{t-1}, \ldots, Y_1\right] = \mu_t$, $\mathsf{Var}\left[Y_t | Y_{t-1}, \ldots, Y_1\right] = \sigma^2 f_t$

then as $Y_t | Y_{t-1}, \ldots, Y_1 \sim N(\mu_t, \sigma^2 f_t)$, the log-likelihood is

$$L(\mu, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{y}) = -\frac{1}{2} \sum_{t=1}^{n} \left\{ \log(\sigma^2 f_t) + \frac{e_t^2}{\sigma^2 f_t} \right\}$$

where $e_t = (y_t - \mu_t)$.

In R, a conditional version is used in which the first p terms in the sum are ignored. For ARIMA(p, d, q) the first $p + d$ terms are ignored.

3. Diagnostic checks

- *Overfitting*: if an ARIMA(p, d, q) has a maximized log-likelihood L_0, we fit

  - ARIMA($p + 1, d, q$) with maximized log-likelihood L_1 and

  - ARIMA($p, d, q + 1$) with maximized log-likelihood L_2

  and compare

  $$2(L_1 - L_0), \quad 2(L_2 - L_0)$$

  with $\chi^2(1)$. These test if the additional autoregressive parameter and moving average parameter respectively, are equal to zero (H_0). We reject H_0 if the statistics are bigger than $\chi^2_{1-\alpha}(1)$ for a test of size α.

- *Residuals*: using the conditional log-likelihood, the residuals are $\widehat{e}_t$ and the *standardized residuals* are

  $$\widehat{Z}_t = \frac{\widehat{e}_t}{\sqrt{\widehat{f}_t}}$$

  We use these by

  - plotting $\widehat{Z}_t$ vs t: if all is well this plot should look like white noise,

  - finding the correlogram for the $\widehat{Z}_t$ which should look like white noise,

  - plotting the cumulative periodogram of the residuals which should look like white noise.

- *Portmanteau Statistic*: using the sample autocorrelations of the residuals, if the series is white noise, $r_k \sim N(0, 1/n)$ and so

$$nr_k^2 \sim \chi^2(1), \quad Q_m = n \sum_{k=1}^{m} r_k^2 \sim \chi^2(m)$$

Q_m is called the portmanteau statistic (uncorrected, see earlier for the corrected version).

We reject the hypothesis of 'white noise' if $Q_m > \chi^2_{1-\alpha}(m)$ at level α. Usually this statistic is calculated for several consecutive values of m.

5.3 Forecasting

Forecasting is the prediction of future values of the time series. It is difficult because:

- extrapolation assumes the processes driving the series do not change,
- requires a model which at best can be an approximation to reality (several models may fit the series equally well but produce different forecasts), and
- errors associated with predictions increase as they are made progressively further into the future.

There are some standard approaches which are not always very good. We will focus on the use of ARIMA models, the so-called Box-Jenkins approach.

Consider

$$\varphi(B)Y_t = \theta(B)Z_t, \quad \varphi(B) = \phi(B)(1-B)^d.$$

Then as a general linear process (GLP)

$$\begin{aligned} Y_t &= \varphi(B)^{-1}\theta(B)Z_t \\ &= \sum_{j=0}^{\infty} a_j Z_{t-j} \end{aligned}$$

Let $Y_t(k)$ denote the predicted value of Y_{t+k} based on $Y_1, Y_2, \ldots, Y_t$. We consider linear predictors

$$Y_t(k) = \sum_{i=1}^{t-1} w_i Y_{t-i}$$

for some weights w_i. Thus using the GLP

$$Y_t(k) = \sum_{i=1}^{t-1} w_i \sum_{j=0}^{\infty} a_j Z_{t-i-j} = \sum_{j=0}^{\infty} W_j Z_{t-j}$$

How to select the weights? Minimize the mean square error of prediction,

$$MSEP = \mathsf{E}\left[[Y_{t+k} - Y_t(k)]^2\right]$$

$MSEP$ is minimized if $W_j = a_{j+k}$, so that

$$Y_t(k) = \sum_{j=k}^{\infty} a_j Z_{t+k-j}$$

while

$$Y_{t+k} = \sum_{j=0}^{\infty} a_j Z_{t+k-j}$$

Thus

$$Y_{t+k} = Y_t(k) + R_t(k), \quad R_t(k) = \sum_{j=0}^{k-1} a_j Z_{t+k-j}$$

and $R_t(k)$ are called the forecast errors.

Thus $Y_t(k)$ *is calculated from the definition of* Y_{t+k} *except that future vales* $Z_{t+1}, Z_{t+2}, \ldots, Z_{t+k}$ *are set to zero.* Note

- $Y_{t+1} - Y_t(1) = Z_{t+1}$ so that one step forecast errors are not correlated; this is not true for larger steps.

- $Z_t = Y_t - Y_{t-1}(1)$ which is very useful in actual calculations.

- $\mathsf{E}\left[R_t(k)\right] = 0$ and $\mathsf{Var}\left[R_t(k)\right] = \sigma^2 \sum_{j=0}^{k-1} a_j^2$ so that assuming normality a $1-\alpha$ prediction interval for Y_{t+k} is

$$Y_t(k) \pm z_{1-\alpha/2}\widehat{\sigma}\sqrt{\sum_{j=0}^{k-1} a_j^2}$$

Calculation of forecasts

The Predictor $Y_t(k)$ is actually a conditional expectation,

$$Y_t(k) = \mathsf{E}\left[Y_{t+k}|Y_t, Y_{t-1}, \ldots\right]$$

Now

$$\begin{aligned}\varphi(B) &= \phi(B)(1-B)^d \\ &= (1+\alpha_1 B + \alpha_2 B^2 + \ldots \alpha_p B^p)(1-B)^d \\ &= (1+\varphi_1 B + \varphi_2 B^2 + \ldots \varphi_{p+d} B^{p+d})\end{aligned}$$

for some φ_j. Thus

$$Y_{t+k} = -\sum_{i=1}^{p+d} \varphi_i Y_{t+k-i} + \sum_{j=0}^{q} \beta_j Z_{t+k-j}$$

To find the forecast we take conditional expectations and use

$$\mathsf{E}\left[Y_{t+k-j}|Y_t, Y_{t-1}, \ldots\right] = \begin{cases} Y_{t+k-j} & \text{if } j \geq k \\ Y_t(k-j) & \text{if } j < k \end{cases}$$

and

$$\mathsf{E}\left[Z_{t+k-j}|Y_t, Y_{t-1}, \ldots\right] = \begin{cases} Z_{t+k-j} & \text{if } j \geq k \\ 0 & \text{if } j < k \end{cases}$$

where for $j \geq k$

$$Z_{t+k-j} = Y_{t+k-j} - Y_{t+k-j-1}(1) \qquad (*)$$

Thus our forecast is found from the actual model where

- observed Y_{t+k-j} are left alone
- unobserved Y_{t+k-j} are replaced by earlier forecasts, $Y_t(k-j)$
- observed Z_{t+k-j} are replaced by $(*)$
- unobserved Z_{t+k-j} are replaced by zero (0).

5.4 Seasonal Models

- Earlier in the course we saw differencing at lag 12 removed the seasonal effect for the bronchitis data.

- Logically then, the operator $(1 - B^s)$ where s denotes the period of the seasonal effects (12 for monthly data, 4 for quarterly) would seem to be appropriate.

- $1 - u^s = 0$ has roots $u = e^{i(2\pi k/s)}$, $k = 0, 1, 2, \ldots, s-1$, which are evenly spaced on the unit circle (in the complex plane). Thus this operator is like $(1 - B)^d$ and implies non-stationarity, which seasonal effects naturally are!

- *Multiplicative models*: Consider an ARIMA type model for seasonal effects,

$$\Phi(B^s)(1 - B^s)^D Y_t = \Theta(B^s)U_t$$

where the terms on the left correspond to an AR and a differencing operation and on the left to a MA. The process U_t is NOT white noise because we expect there to be relationships between not only the corresponding seasonal terms but also between observations in previous cycles as well as within the current seasonal cycle. So we have an ARIMA model for the U_t namely,

$$\phi(B)(1 - B)^d U_t = \theta(B)Z_t$$

and putting it all together

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Y_t = \theta(B)\Theta(B^s)Z_t$$

is called an ARIMA $(p, d, q) \times (P, D, Q)$ model; it is a multiplicative model.

- This is only one particular approach. Non multiplicative models are also very useful, and an experienced analyst can propose such models for an application.

- *Identification*: requires study of simple models, but in principle follows as for ARIMA models.

- *Estimation*: as before.

- *Diagnostics*: as before.

- *Forecasting*: same principles apply as for ARIMA models, namely the use of the model and conditional expectations.