ARMA Notes

Eric Desoiza

Chapter 3.1 - Autoregressive Moving Average Models

Autoregressive models are based on the idea of previous values of a series being able to partially predict the current value.

An example of this is the following formula

$$x_t = x_{t-1} - .9x_{t-2} + w_t$$

where w_t is a white gaussian noise with $\sigma_w^2 = 1$.

We can assess the extent of which how much past data can be used to estimate future data but looking at the autocorrelation function. the ACF can display significant correlations at certain lags or points of interest

Autoregressive Model (AR) - of order p, is of the following form

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

where x_t is stationary and w_t white noise$(0, \sigma_w^2)$, and ϕ is a sequence of constants where $\phi_p \neq 0$. The mean can most often be zero but if it is not we can add in an α to identify a constant where

$$\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$$

A useful form of the AR(p) model comes from the backshift operator

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)x_t = w_t \quad \text{or} \quad \phi(B)x_t = w_t$$

Example 1.1 - AR(1) Model

We can begin with a first order model which is given by $x_t = \phi x_{t-1} + w_t$. If we continue to iterate backwardes k times

$$x_t = \phi^k x_{t-k} + \sum_{j=0}^{\infty} \phi^j w_{t-j}$$

This suggests that by continuously iterating backwards and provided that $|\phi| < 1$ and the supremum of variance at any point is less than ∞ we can represent the AR(1) model as a linear process

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}$$

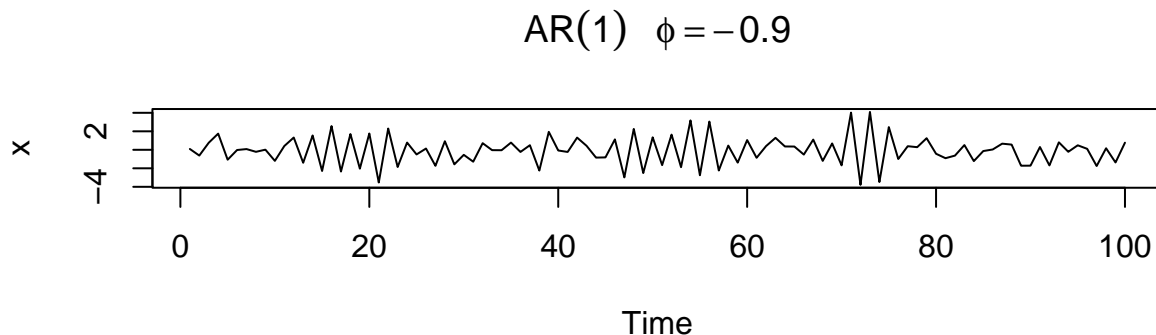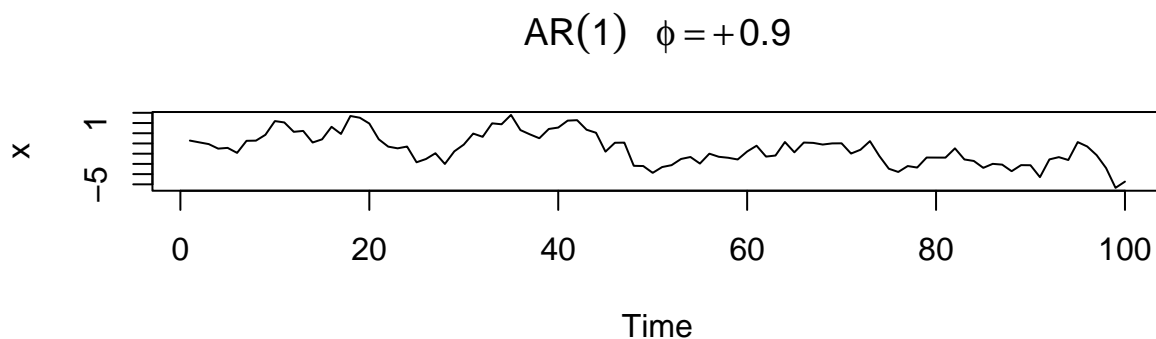AR(1) is defined above is stationary with mean $= 0$.

Which an autocovariance function of

$$\gamma(h) = \frac{\sigma_w^2 \phi^h}{1 - \phi^2}, \quad h \geq 0$$

We can understand that $\gamma(h) = \gamma(-h)$ and the ACF of an AR(1) is

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h, \quad h \geq 0$$

Example 3.2 - Sample Path of AR(1)

AR(1) $\phi = +0.9$

AR(1) $\phi = -0.9$

In an AR(1) model with $\phi = .9$ we can see that values are contiguously close to one another and positively correlated which will lead to a smooth pattern. On the other end we can see $\phi = -.9$ in which there $\rho(h) = (-.9)^k$ which will not be a smooth pass and will actually lead to negative correlation on contiguous pairs but closeness on values separated by one other value.

Example 3.3 - Explosive AR Models and Causality

We understand that the random walk $x_t = x_{t-1} + w_t$ is not stationay. This leads us to wonder if there is a stationary process where $|\phi| > 1$. There are explosive process because of the values in the time series becoming rapidly large in magnitude which is troublesome. These are divergent processes that will not allow to you arrive at the linear process formula of an AR(1). When a process does not depend on the future we can say it is causal but if it is explosibe it can be stationary but it is future dependent and not causal

Example 3.4 - Every Explosion Has a Cause

Excluding explosive models is not bad because the models have causal counterparts. We can actually extract a new autocovariance function

$$\gamma_x(h) = \frac{\sigma_w^2 \phi^{-2} \phi^{-h}}{1 - \phi^{-2}}$$

And the causal process

$$y_t = \phi^{-1} y_{t-1} + v_t, \quad v_t \sim \text{iid } N(0, \sigma_w^2 \phi^{-2})$$

Which is stochastically equal to the x_t process. For example $x_t = 2x_{t-1} + w_t, \sigma_w^2 = 1$ is the "same" as $y_t = .5y_{t-1} + v_t, \sigma_v^2 = .25$. This also generalizes to higher orders.

The backwards iteration works will for p = 1 but not for higher orders. A general technique is about matching coefficients.

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B) w_t$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ and $\psi_j = \phi^j$

Another process is by thinking about AR(1) in operator form, $\phi(B)x_t = w_t$. We then multiply both sides by $\phi^{-1}(B)$ (assuming invertibility)

$$\phi^{-1}(B)\phi(B)x_t = \phi^{-1}(B)w_t \quad \text{or} \quad x_t = \phi^{-1}(B)w_t$$

We now begin to realize that ϕ^{-1} is like $\psi(B)$ and we notice that working with operators is like working with polynomials. We consider the polynomial $\phi(z) = 1 - \phi z$ where z is a complex number and $|\phi| < 1$

$$\phi^{-1}(z) = \frac{1}{1 - \phi z} = 1 + \phi z + \phi^2 z^2 + \cdots + \phi^j z^j + \cdots, \quad |z| \leq 1$$

This makes the connection that the coefficients in B^j in $\phi^{-1}(B)$ are the same as the coefficients of z^j in $\phi^{-1}(z)$. This means we can treat B like a compex number and we can find polynomials that correspond to the general properties

Moving Average Models of order q, MA(q), assumes the white noise on the right-hand side of the equation is linearly combined to form the observed data

Moving Average Model (MA) is defined to be

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$

Where $w_t \sim wn(0, \sigma_w^2)$ and $\theta_q \neq 0$

The system is the same as the infinite moving average defined as the linear process where $\phi_0 = 1, phi_j = \theta_j$. We can also write MA(q) as

$$x_t = \theta(B) w_t$$

Moving Average Operator is

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$

The moving average is in fact stationary for all values of θ

Example 3.5 - The MA(1) Process

Consider the MA(1) model $x_t = w_t + \theta w_{t-1}$.

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma_w^2 & h = 0 \\ \theta\sigma_w^2 & h = 1 \\ 0 & h > 1 \end{cases}$$

and the ACF

$$\rho(h) = \begin{cases} \frac{\theta}{(1+\theta^2)} & h = 1 \\ 0 & h > 1 \end{cases}$$

Note $|\rho(1)| \leq 1/2$ for all values of θ. Also the current value is only correlated with the single one before it. This is in constract to the AR(1) model where all values have some sort of correlation.

Example 3.6 - Non-uniqueness of MA Models and Invertibility

We note that for an MA(1) model, $\rho(h)$ is the same for θ and $\frac{1}{\theta}$. The pair $\sigma_w^2 = 1, \theta = 5$ and $\sigma_w^2 = 25, \theta = 1/5$ yield the same autocovariance function

$$\gamma(h) = \begin{cases} 26 & h = 0 \\ 5 & h = 1 \\ 0 & h > 1 \end{cases}$$

Making the process the same because of normality (i.e. all finite disitrbutions are the same). We can only observe the time series and not the noise so we cannot distrinquish between the models. We will thus choose the invertible process as it is easier to work with!

We can use the previous polynomial thought process from AR(1) to generalize MA processes. For example we can have $x_t = \theta(B)w_t$ where $\theta(B) = 1 + \theta B$. If $|\theta| < 1$ then $\pi(B)x_t = w_t$ where $\pi(B) = \theta^{-1}(B)$. If we let $\theta(z) = 1 + \theta z$ for $|z| \leq 1$ then $\pi(z) = \theta^{-1}(z) = 1/(1+\theta z) = \sum_{j=0}^{\infty}(-\theta)^j z^j$. Finally $\pi(B) = \sum_{j=0}^{\infty}(-\theta)^j B^j$

The Autoregressive Moving Average Models for stationary time series.

ARMA is defined as a stationary time series formed by

$$x_t = \phi_j x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

with $\phi_p \neq 0, \theta_p \neq 0, \sigma_w^2 > 0$. The parameters p and q are called the autoregressive and moving average orders, respectively. If x_t has a nonzero mean then we set $\alpha = \mu(1 - \phi_1 - \cdots + \phi_p)$ and

$$x_t = \alpha + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

where $w_t \sim wn(0, \sigma_w^2)$

We can write the ARMA(p,q) model in a concise form

$$\phi(B)x_t = \theta(B)w_t$$

This form can point to issues of unnecessarily complicating our model

Example 3.7 - Parameter Redundancy

Consider a white noise process $x_t = w_t$. If we multiply both sides by $\eta(B) = 1 - .5B$ then we arrive to

$$x_t = .5x_{t-1} - .5x_{t-1} + w_t$$

which can look like an ARMA(1,1) model. But the x_t is still white noise but we have this issue of parameter redundancy or over-parameterization.

This is an important consideration as we may discover significant parameter estimates where they do not truly exist.

AR and MA polynomials - are defined as

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \quad \phi_p \neq 0$$

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q, \quad \theta_q \neq 0$$

We will require that these equations have no common factors.

Causality with an ARMA model is a time sieres that can be written as a one-sided linear process

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B)w_t$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, and $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and set $\psi_0 = 1$

In AR(1), it is only causal when $|\phi| < 1$. The ARMA process is only causal when the root of $\phi(z) = 1 - \phi z$ is bigger than 1 absolutely.

The coefficient s of the lienar process of an ARMA(p,q) can be determined by

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1$$

Invertbility is only possible when the time series can be written as

$$\pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = w_t$$

where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$ and $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and set $\pi_0 = 1$

The coefficients of pi can be determined by

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1$$

Example 3.8 - Parameter Redundnancy, Causality, Invertibility

Consider the process in operator form

$$(1 - .4B - .45B^2)x_t = (1 + B + .25B^2)w_t$$

At first it appears that we have an ARMA(2,2) process but we can then notice that

$$\phi(B) = (1 + .5B)(1 - .9B) \quad \text{and} \quad \theta(B) = (1 + .5B)^2$$

This allows us to cancel out the common factor and actually arrive at a ARMA(1,1). This process is also causal because the roots of $\phi(z)$ are outside of the unit circle and the model is invertible because the roots of $\theta(z)$ are outside of the unit circle

We can then get the ψ-weights

$$(1 - .9z)(1 + \psi_1 z + \psi_2 z^2 + \cdots + \psi_j z^j + \cdots) = 1 + .5z$$

or

$$1 + (\psi_j - .9)z + (\psi_2 - .9\psi_1)z^2 + \cdots + (\psi_j - .9\psi_{j-1})z^j + \cdots = 1 + .5z$$

This tells us that $\psi_1 - .9 = .5$ and $\psi_j - .9\psi_{j-1} = 0$. This tells us that $\psi_j = 1.4(.9)^{j-1}$

$$x_t = w_t + 1.4 \sum_{j=1}^{\infty} .9^{j-1} w_{t-j}$$

The invertible representation would lead to

$$(1 + .5z)(1 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \cdots) = 1 - .9z$$

and thus

$$x_t = 1.4 \sum_{j=1}^{\infty} (-.5)^{j-1} x_{t-j} + w_t$$

Example 3.9 Causal Conditions for an AR(2) Process

It is pretty simply to identify the root of an AR(2) process as causal from its root. But some processes are not as easy in higher orders. In AR(2) the model is causal if both roots lie outside the unit circle. We can in fact use the quadratic equation on a operators polynomial.

6

Chapter 3.2 - Difference Equations

Below represents the homogeneous difference equation of order 1

$$u_n - \alpha n_{n-1} = 0, \quad \alpha \neq 0, \quad n = 1, 2, \dots$$

We can also use the ACF of an AR(1) process is a sequence

$$\rho(h) - \phi\rho(h-1) = 0, \quad h = 1, 2, \dots$$

To solve the difference equation of order 1

$$u_n = \alpha n_{n-1} = a^n u_0$$

Given an initial condition u_0 = c, we can solve for $u_n = \alpha^n c$

In operator notation we can write the difference equation as $(1 - \alpha B)u_n = 0$. The polynomial associated is $\alpha(z) = 1 - \alpha z$ and the root of this polynomial is $z_a = 1/\alpha$. We can solve the difference equation as

$$u_n = \alpha^n c = (z_0^{-1})^n c$$

Suppose the sequence satisfies

$$u_n - \alpha_1 u_{n-1} - \alpha_2 u_{n-2} = 0, \quad \alpha_2 \neq 0, \quad n = 2, 3, \dots$$

This is the homogeneous difference equation of order 2. The corresponding polynomial is

$$\alpha(z) = 1 - \alpha_1 z - \alpha_2 z^2$$

If the roots are inequivalent then the general solution is

$$u_n = c_1 z_1^{-n} + c_2 z_2^{-n}$$

Given two initial condiitons $u_0 = c_1 + c_2$ and $u_1 = c_1 z_1^{-1} + c_2 z_2^{-1}$. We can use the quadratic formula

When the roots are equal a general solution is

$$u_n = z_0^{-n}(c_1 + c_2 n)$$

Given two initial conditions we can solve for c_1 and c_2

$$u_0 = c_1 \quad \text{and} \quad u_1 = (c_1 + c_2)z_0^{-1}$$

Example 3.10 - The ACF of an AR(2) Process

Suppose that we have a causal AR(2) process. Multiply both sides by x_{t-h} for h > 0 and take expectation to receive

$$\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2), \quad h = 1, 2, \ldots$$

We can divide the process by $\gamma(0)$ to obtain the difference equation for ACF of the process:

$$\rho(h) = -\phi_1\rho(h-1) - \phi_2\rho(h-2) = 0, \quad h = 1, 2, \ldots$$

We now know that $\rho(0) = 1$ and $\rho(-1) = \phi_1/(1-\phi_2)$

we now have the polynomial $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$. We also know this process is causal which is a good note and we can check for 3 cases

1) When z_1 and z_2 are real and distinct

$$\rho(h) = c_1 z_1^{-h} + c_2 z_2^{-h}$$

2) When $z_1 = z_2 (= z_0)$ are real and equal

$$\rho(h) = z_0^{-h}(c_1 + c_2 h)$$

3) When $z_1 = \bar{z}_2$ are a complex conjugate pair then $c_2 = \bar{c}_1$

$$\rho(h) = c_1 z_1^{-h} + \bar{c}_1 \bar{z}_1^{-h}$$

In the above case we write c_1 and z_1 in polar coordinates. Then we can understand that $e^{i\alpha} + e^{-i\alpha} = 2\cos(\alpha)$, the solutions has the form

$$\rho(h) = a|z_1|^{-h}\cos(h\theta + b)$$

Example 3.11 - An AR(2) with Complex Roots

Suppose we have an AR(2) model such as $x_t = 1.5x_{t-1} - .75x_{t-2} + w_t$ with $\sigma_w^2 = 1$ we can see the prcoess exhibits pseudocyclic behavior at the rate of one cycle every 12 points. We can see the roots to be $1 \pm i/\sqrt{3}$ and $\theta = \tan^{-1}(1/\sqrt{3}) = 2\pi/12$ radians per unit time. We can convert from angles to cycles by dividing by 2π and get 1/12 cycles per unit time.

Example 3.12 - The ψ-weights for an ARMA Model

For a causal ARMA $\phi(B)x_t = \theta(B)w_t$ we can write

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$$

The use of the theory of homogeneous difference equaitons can help we can solve for the ψ-weights by matching $\phi(z)\psi(z) = \theta(z)$:

$$(1 - \phi_1 z - \phi_2 z^2 - \cdots)(\psi_0 + \psi_1 z = \psi_2 z^2 + \cdots) = (1 = \theta_1 z + \theta_2 z^2 + \cdots)$$

The psi values follow the following values

$$\psi_0 = 1$$
$$\psi_1 - \phi_1 \psi_0 = \theta_1$$
$$\psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0 = \theta_2$$
$$\psi_3 - \phi_1 \psi_2 - \phi_2 \psi_1 - \phi_3 \psi_0 = \theta_3$$
$$\vdots$$

The ψ-weights satisfy the HDE given by

$$\psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k} = 0, \quad j \geq \max(p, q+1)$$

with initial conditions

$$\psi_j - \sum_{k=1}^{j} \phi_k \psi_{j-k} = \theta_j, \quad 0 \leq j < \max(p, q+1)$$

The general solution depends on the roots of the AR polynomial.

Chapter 3.3 - Autocorrelation and Partial Autocorrelation

We begin with the ACF of an MA(q). Because x_t is a finite linear combination of white noise terms the process is stationary with mean 0

we write $\theta_0 = 1$ and autocovariance of

$$\begin{cases} \sigma_w^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} & 0 \leq h \leq q \\ 0 & h > q \end{cases}$$

We then cut off after q lags and also evaluate for the ACF

$$\rho(h) = \begin{cases} \dfrac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1 + \theta_1^2 + \cdots + \theta_q^2} & 1 \leq h \leq q \\ 0 & h > q \end{cases}$$

with a causal relationship (zeroes outside unit circle)

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$$

It now follows with mean zero that the autocovariance function is

$$\gamma(h) = \sigma_w^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}, \quad h \geq 0$$

We can write the general homogeneous equation for the ACF of a causal ARMA process

$$\gamma(h) - \phi_1 \gamma(h-1) - \cdots - \phi_p \gamma(h-p) = 0, \quad h \geq \max(p, q+1)$$

with initial conditions

$$\gamma(h) - \sum_{j=1}^{p} \phi_j \gamma(h-j) = \sigma_w^2 \sum_{j=h}^{q} \theta_j \psi_{j-h}, \quad 0 \leq h \leq \max(p, q+1)$$

Example 3.13 - The ACF of an AR(p)

We can see the general case of an AR(p) for the ACF

$$\rho(h) - \phi_1 \rho(h-1) - \cdots - \phi_p \rho(h-p) = 0, \quad h \geq p$$

We can also let $z_1, \ldots, z_r$ be the roots of $\phi(z)$ with multiplicites m where $m_1 + \cdots + m_r = p$, the general solution is

$$\rho(h) = z_1^{-h} P_1(h) + z_2^{-h} P_2(H) + \cdots + z_r^{-h} P_r(h), \quad h \geq p$$

where P is a polynomial in h of degree $m_j - 1$

Example 3.14 - The ACF of an ARMA(1,1)

Let us say that the autocovariance function satisfies

$$\gamma(h) - \phi\gamma(h-1) = 0, \quad h = 2, 3, \ldots$$

and the general solution

$$\gamma(h) = c\phi^h, \quad h = 1, 2, \ldots$$

We also solve for $\gamma(0)$ and $\gamma(1)$

$$\gamma(0) = \sigma_w^2 \frac{1 + 2\theta\phi\theta^2}{1 - \phi^2} \quad \text{and} \quad \gamma(1) = \sigma_w^2 \frac{(1 + \theta\phi)(\phi + \theta)}{1 - \phi^2}$$

we can then solve for lag h

$$\gamma(h) = \frac{\gamma(1)}{\phi}\phi^h = \sigma_w^2 \frac{(1+\theta\phi)(\phi+\theta)}{1-\phi^2}\phi^{h-1}$$

We now can solve for the ACF

$$\rho(h) = \frac{(1+\theta\phi)(\phi+\theta)}{1+2\theta\phi+\theta^2}\phi^{h-1}, \quad h \geq 1$$

The Partial Autocorrelation Function (PACF)

The partial correlation between X and Y given Z is obtained by regressing X on Z and Y on Z and then calculating

$$\rho_{XY|Z} = \text{corr}\{X - \hat{X}, Y - \hat{Y}\}$$

If the variables are multivaraite normal, then $\rho_{XY|Z} = \text{corr}\{X,Y|Z\}$

The purpose of the PACF is to figure out the correlation between x_t and x_s while the linear effect of everything in between removed.

To formally define PACF for a mean-zero stationary we start with

$$\hat{x}_{t+h} = \beta_1 x_{t+h-1} + \beta_2 x_{t+2} + \cdots + \beta_{h-1} x_{t+h+1}$$

and we also have

$$\hat{x}_t = \beta_1 x_{t+1} + \beta_2 x_{t+2} + \cdots + \beta_{h-1} x_{t+h-1}$$

Partial Autocorrelation Function (ACF) of a stationary process denoted ϕ_{hh} for $h = 1, 2, \ldots$ is

$$\phi_{11} = \text{corr}(x_{t+1}, x_t) = \rho(1)$$

and

$$\phi_{hh} = \text{corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t), \quad hgeq2$$

It will become clear why double subscript is used. The PACF is the correlation of two points with a lag with the linear dependence in betqeen remove.

Example 3.15 - PACF of an AR(1)

To calculate ϕ_{22} we consider the regression of x_{t+2} on x_{t+1}, say $\hat{x}_{t+2} = \beta x_{t+1}$ we now minimize β

$$E(x_{t+2} - \beta x_{t+1})^2 = \gamma(0) - 2\beta\gamma(1) + \beta^2\gamma(0)$$

If we take derivative with respect to β we identify beta to be equivalent to $\rho(1) = \phi$. And when we consider x_t on x_{t+1}, say $\hat{x}_t = \beta x_{t+1}$ and minimize β, we receive an equivalent equation to above

This means that $\beta = \phi$. Hence,

$$\phi_{22} = \text{corr}(x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t) = \text{corr}(w_{t+2}, x_t - \phi x_{t+1}) = 0$$

Thus we see that $\phi_{22} = 0$ and $\phi_{hh} = 0, \quad h > 1$

Example 3.16 - PACF of an AR(p)

The model implies $x_{t+h} = \sum_{j=1}^{p} \phi_j x_{t+h-j} + w_{t+h}$, where the roots of $\phi(z)$ are outside the unit circle. When h > p, the regression of x_{t+h} on others is

$$\hat{x}_{t+h} = \sum_{j=1}^{p} \phi_j x_{t+h-j}$$

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts of after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Example 3.17 - PACF of an Invertible MA(q)

Example 3.18 - Preliminary Analysis of the Recruitment Series

Chapter 3.4 - Forecasting

Chapter 3.5 - Estimation

Chapter 3.6 - Integrated Models for Nonstationary Data