Massachusetts Institute of Technology
Department of Economics
Time Series 14.384

Guido Kuersteiner

Lecture Note 2 - Stationary Processes

In this lecture we are concerned with models for stationary (in the weak sense) processes. The focus here will be on linear processes. This is clearly restrictive given the stylized facts of financial time series. However linear time series models are often used as building blocks in nonlinear models. Moreover linear models are easier to handle empirically and have certain optimality properties that will be discussed. Linear time series models can be defined as linear difference equations with constant coefficients.

We start by introducing some examples. The simplest case of a stochastic process is one with independent observations. From a second order point of view this transforms into uncorrelatedness.

Example 2.1 (White Noise). *The process* $\{\varepsilon_t\}$ *is called white noise if it is weakly stationary with* $E\varepsilon_t = 0$ *and autocovariance function*

$$\gamma_{\varepsilon\varepsilon}(h) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

and we write $\varepsilon_t \sim WN(0, \sigma^2)$*. A special case is* $\{\varepsilon_t\}$ *with* $\varepsilon_t \sim iid(0, \sigma^2)$*.*

The white noise process is important because it can be used as a building block for more general processes. Consider the following two examples.

Example 2.2 (Moving Average). *The process* $\{x_t\}$ *is called a moving average of order one or MA(1) if* $\{x_t\}$ *is stationary and*

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

and ε_t *is white noise.*

It follows immediately that $\gamma_{xx}(0) = \sigma^2(1 + \theta^2)$, $\gamma_{xx}(1) = \theta\sigma^2$ and $\gamma_{xx}(h) = 0$ for $|h| > 1$. A slightly more complicated situation arises when we consider the following autoregressive process.

Example 2.3 (Autoregression). *The process* $\{x_t\}$ *is called autoregressive of order one or AR(1) if* $\{x_t\}$ *is stationary and satisfies the following stochastic first order difference equation*

$$x_t = \phi x_{t-1} + \varepsilon_t \tag{2.1}$$

and ε_{t-1} *is white noise.*

By iterating on (2.1) we find

$$x_t = \varepsilon_t + \phi\varepsilon_{t-1} + ...\phi^{k-1}\varepsilon_{t-k+1} + \phi^k x_{t-k}.$$

By stationarity Ex_{t-k}^2 is constant and if $|\phi| < 1$ then $E(x_t - \sum_{j=0}^k \phi^j \varepsilon_{t-j})^2 = \phi^{2k} Ex_{t-k}^2$ tends to zero as $k \to \infty$. Therefore

$$x_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \tag{2.2}$$

in mean square and therefore in probability. It can also be shown that (2.2) holds almost surely. Equation (2.2) is the stationary solution to (2.1). It is called causal because it only depends on past innovations.

Under the stationarity assumption we can see that

$$Ex_t = \phi E x_{t-1} + 0 \Rightarrow Ex_t = 0$$

and

$$Ex_t x_{t-h} = \phi E x_{t-1} x_{t-h} + E\varepsilon_t x_{t-h}$$

such that

$$\gamma_{xx}(h) = \phi \gamma_{xx}(h-1). \tag{2.3}$$

By premultiplying equation (2.1) by x_t on both sides and taking expectations we also have $\gamma_{xx}(0) = \phi\gamma_{xx}(1) + \sigma^2$ such that together with $\gamma_{xx}(1) = \phi\gamma_{xx}(0)$ we can solve for $\gamma_{xx}(0) = \sigma^2/(1-\phi^2)$. This now leads to

$$\gamma_{xx}(h) = \frac{\sigma^2 \phi^h}{(1-\phi^2)}.$$

This derivation of the form of $\gamma_{xx}(h)$ is based on solving the Yule Walker equations (2.3). An alternative way to derive this result is to directly calculate the autocovariances based on the solution (2.2).

2.1. Lag Operators

For more general time series models it is less easy to find the solution by repeated substitution of the difference equation. It is therefore necessary to develop a few tools to analyze higher order difference equations. We introduce the lag operator L which maps a sequence $\{x_t\}$ into a sequence $\{y_t\}$ and is defined by its operation on each element in the sequence

$$y_t = Lx_t = x_{t-1} \ \forall t.$$

If we apply L repeatedly then we use the convention that

$$L^p = \underbrace{L \circ L \circ ... \circ L}_{p}.$$

L has an inverse L^{-1} such that $L^{-1}Lx_t = LL^{-1}x_t = x_t$. It is also immediate that L is linear

$$Lax_t = aLx_t = ax_{t-1}.$$

We can use the operator L to define more complex linear operators, the polynomial lag operators. Let $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ then $\phi(L)$ is a polynomial of order p in L. It follows that $\phi(L)$ is again a linear operator. For $p = 1$ we can write the AR(1) model in compact form

$$\phi(L)x_t = \varepsilon_t.$$

In the same way as before $\phi(L)$ has an inverse $\phi(L)^{-1}$ such that $\phi(L)^{-1}\phi(L)x_t = \phi(L)\phi(L)^{-1}x_t = x_t$. For the case of $p = 1$ it is easy to find $\phi(L)^{-1}$ in terms of a polynomial expansion. Since L is a bounded operator and if $|\phi_1| < 1$

$$(1 - \phi_1 L)(1 + \sum_{j=1}^{k} \phi_1^j L^j) = 1 - \phi_1^{k+1} L^{k+1} \to 1 \text{ as } k \to \infty$$

such that

$$\phi(L)^{-1} = (1 + \sum_{j=1}^{\infty} \phi_1^j L^j). \tag{2.4}$$

One way to find the inverse of higher order polynomials is therefore to factorize them into first order polynomials and then use relation (2.4). This is also the idea behind the solution of a p-th order difference equation to which we turn now.

2.2. Linear Difference Equations

We consider solutions $\{x_t\}$ of the p-th order linear difference equation

$$x_t + \alpha_1 x_{t-1} + ... + \alpha_p x_{t-p} = 0 \tag{2.5}$$

where $\alpha_1, ...\alpha_p$ are real constants. In lag polynomial notation we write $\alpha(L)x_t = 0$. A solution then is a sequence $\{x_t\}$ such that (2.5) is satisfied for each t. A set of $m \leq p$ solutions $\{x_t^{(1)}, ..., x_t^{(m)}\}$ are linearly independent if

$$c_1 x_t^{(1)} + ... + c_m x_t^{(m)} = 0 \text{ for all } t = 0, 1, ..., p-1$$

implies $c_1, ..., c_m = 0$. Given p independent solutions and p initial conditions $x_0, ..., x_{p-1}$ we can then solve

$$\begin{bmatrix} x_0^{(1)} & \cdots & x_0^{(p)} \\ \vdots & & \vdots \\ x_{p-1}^{(1)} & \cdots & x_{p-1}^{(p)} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} = \begin{bmatrix} x_0 \\ \vdots \\ x_{p-1} \end{bmatrix}$$

for the vector of coefficients $(c_1, ..., c_p)$. The unique solution to (2.5) is then $c_1 x_t^{(1)} + ... + c_m x_t^{(m)} = x_t$ since this is the only solution x_t that satisfies the initial conditions. All values for x_t, $t > p$ are then uniquely determined by recursively applying (2.5).

From fundamental results in Algebra we know that the equation $\alpha(L) = 0$ has p possibly complex roots such that we can write

$$\alpha(L) = \prod_{i=1}^{j} (1 - \xi_i^{-1} L)^{r_i} \tag{2.6}$$

where ξ_i, $i = 1, ..., j$ are the j distinct roots of $\alpha(L)$ and r_i is the multiplicity of the i-th root. It is therefore enough to find solutions $x_t^{(i)}$ such that $(1 - \xi_i^{-1} L)^{r_i} x_t^{(i)} = 0$. It then follows from (2.6) that $\alpha(L) x_t^{(i)} = 0$. We now show the following result.

Lemma 2.4. *The functions* $h_t^{(k)} = t^k \xi^{-t}$, $k = 0, 1, ..., j-1$ *are linearly independent solutions to the difference equation*

$$(1 - \xi^{-1} L)^j h_t = 0$$

Proof. For $j = 1$ we have

$$\begin{aligned} (1 - \xi^{-1} L)\xi^{-t} &= \xi^{-t} - \xi^{-1} L \xi^{-t} \\ &= \xi^{-t} - \xi^{-1} \xi^{-t+1} = 0. \end{aligned}$$

For $j = 2$ we have $(1 - \xi^{-1} L)^2 \xi^{-t} = 0$ from before and

$$\begin{aligned} (1 - \xi^{-1} L)^2 t\xi^{-t} &= (1 - \xi^{-1} L)(t\xi^{-t} - (t-1)\xi^{-t}) \\ &= t\xi^{-t} - 2(t-1)\xi^{-t} + (t-2)\xi^{-t} = 0 \end{aligned}$$

and similarly for $j > 2$. This follows from repeated application of

$$\begin{aligned} (1 - mL)(a_0 + a_1 t + ... + a_k t^k) m^t &= m^t \left(\sum_{r=0}^{k} a_r (t^r - (t-1)^r) \right) \\ &= (b_0 + b_1 t + ... + b_{k-1} t^{k-1}) m^t \end{aligned}$$

where $b_1, ..., b_{k-1}$ are some constants. Finally we note that $h_t^{(k)}$ are linearly independent since if

$$\left(c_0 + c_1 t + ... + c_{k-1} t^{k-1}\right) \xi^{-t} = 0 \text{ for } t = 0, 1, ..., k-1$$

then the polynomial $\left(c_0 + c_1 t + ... + c_{k-1} t^{k-1}\right)$ of degree $k-1$ has k zeros. This is only possible if $c_1, ..., c_k = 0$. ■

Lemma (2.4) shows that $\alpha(L)x_t = 0$ has p solutions $t^n \xi_i^{-t}, n = 0, 1, ..., r_i - 1, i = 1, ..., j$. The general solution to (2.5) then has the form

$$x_t = \sum_{i=1}^{j} \sum_{n=0}^{r_i - 1} c_{in} t^n \xi_i^{-t}. \tag{2.7}$$

The p coefficients c_{in} are again determined by p initial conditions. The coefficients are unique if the p solutions $t^n \xi_i^{-t}$ are linearly independent. This follows if

$$\sum_{i=1}^{j} \sum_{n=0}^{r_i - 1} c_{in} t^n \xi_i^{-t} = 0 \text{ for } t = 0, 1, 2, ...$$

implies $c_{in} = 0 \; \forall i, n$. For a proof see Brockwell and Davis (1987), p.108.

2.3. The Autocovariance function of the ARMA(p,q) Model

In this section we use the previous results to analyze the properties of the $ARMA(p,q)$ model. The $ARMA(p,q)$ process is defined next.

Definition 2.5. *The process $\{x_t\}$ is called $ARMA(p,q)$ if $\{x_t\}$ is stationary and for every t*

$$x_t - \phi_1 x_{t-1} - ... - \phi_p x_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}. \tag{2.8}$$

A more compact formulation can be given by using lag polynomials. Define

$$\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$$

and

$$\theta(L) = 1 + \theta_1 L + ... + \theta_q L^q.$$

then (2.8) can be written as $\phi(L)x_t = \theta(L)\varepsilon_t$. It is useful to find different representations of this model. For this purpose we introduce the following notions

Definition 2.6. *The ARMA(p,q) process (2.8) is said to be causal if there exists a sequence $\{\psi_i\}_{i=0}^{\infty}$ such that $\sum |\psi_i| < \infty$ and*

$$x_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}.$$

It can be shown, that an ARMA(p,q) process such that $\theta(L)$ and $\phi(L)$ have no common zeros is causal if and only if $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. Then the coefficients ψ_i are determined from

$$\psi(z) = \sum_{i=0}^{\infty} \psi_i z^i = \theta(L)/\phi(L) \tag{2.9}$$

by equating coefficients in the two polynomials. We only discuss the *if* part. Assume $\phi(z) \neq 0$ for $|z| \leq 1$. Then there exists an $\epsilon > 0$ such that $1/\phi(z)$ has a power series expansion

$$1/\phi(z) = \sum_{j=0}^{\infty} \xi_j z^j \text{ for } |z| \leq 1 + \epsilon.$$

This implies $\xi_j (1 + \epsilon/2)^j \to 0$ as $j \to \infty$ so that there exists a positive finite constant K such that $\left|\xi_j\right| < K(1 + \epsilon/2)^{-j}$. Thus $\sum_{j=0}^{\infty} \left|\xi_j\right| < \infty$ and $\xi(z)\phi(z) = 1$. This now justifies writing

$$x_t = \theta(L)/\phi(L)\varepsilon_t.$$

Another related property of the $ARMA(p,q)$ process is invertibility. This notion is defined next.

Definition 2.7. *The ARMA(p,q) process (2.8) is said to be invertible if there exists a sequence* $\{\pi_i\}_{i=0}^{\infty}$ *such that* $\sum |\pi_i| < \infty$ *and*

$$\varepsilon_t = \sum_{i=0}^{\infty} \pi_i x_{t-i}.$$

It can again be shown that (2.8), such that $\theta(L)$ and $\phi(L)$ have no common zeros, is invertible if and only if $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. Invertibility means, that the process can be represented as an infinite order $AR(p)$ process. This property has some importance in applied work since $AR(p)$ models can be estimated by simple projection operators while models with moving average terms need nonlinear optimization.

Causality can be used to compute the covariance function of the $ARMA(p,q)$ process. Under causality we can write

$$x_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

such that

$$\gamma_{xx}(h) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+|h|}.$$

This expression is however not very useful since it does not show the dependence on the underlying parameters of the process. Going back to (2.9) we write $\psi(z)\phi(z) = \theta(z)$ and equate the coefficients on z^i. This leads to

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

and in particular for $j = 0, 1, 2, ..$

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= \theta_1 + \phi_1 \\ \psi_2 &= \theta_2 + \phi_2 + \theta_1\phi_1 + \phi_1^2. \\ &\vdots \end{aligned}$$

We can now obtain the covariance function of x_t by premultiplying both sides of (2.8) by x_{t-h} and using the representation $x_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$. Taking expectations on both sides then gives

$$\begin{aligned} \gamma_{xx}(h) - \phi_1\gamma_{xx}(h-1) - ... - \phi_p\gamma_{xx}(h-p) &= \sigma^2 \sum_{h\leq j\leq q} \theta_j \psi_{j-h}, \\ \text{for } 0 &\leq h < max(p, q+1) \end{aligned} \tag{2.10}$$

and

$$\gamma_{xx}(h) - \phi_1\gamma_{xx}(h-1) - ... - \phi_p\gamma_{xx}(h-p) = 0 \text{ for } h \geq \max(p, q+1). \tag{2.11}$$

Note that if $q + 1 > p$ then there are more initial conditions than linearly independent solutions to the difference equation. In this case the first $q - p + 1$ autocorrelation coefficients are determined from the first $q - p + 1$ initial conditions. The general solution to this system of difference equations is now given by (2.7) as

$$\gamma_{xx}(h) = \sum_{i=1}^{j} \sum_{n=0}^{r_i - 1} c_{in} h^n \xi_i^{-h} \text{ for } h \geq \max(p, q+1) - p$$

where ξ_i are the distinct roots of the AR polynomial $\phi(z)$ and c_{in} are p coefficients determined by the initial conditions (2.10). The covariances $\gamma_{xx}(h)$, $0 \leq h < \max(p, q+1) - p$ are also determined from (2.10).

Example 2.8. *We look at the autocovariance function of the causal $AR(2)$ process*

$$(1-\xi_1^{-1}L)(1-\xi_2^{-1}L)x_t = \varepsilon_t.$$

where $|\xi_1|,|\xi_2| > 1$ *and* $\xi_1 \neq \xi_2$. *Assume* $\sigma^2 = 1$ *w.l.g. The autoregressive parameters are given by* $\phi_1 = \xi_1^{-1} + \xi_2^{-1}$ *and* $\phi_2 = -\xi_1^{-1}\xi_2^{-1}$. *It now follows that* $\psi_0 = 1$ *and* $\psi_1 = \xi_1^{-1} + \xi_2^{-1}$. *Then the boundary conditions are*

$$\begin{aligned}\gamma(0) - \phi_1\gamma(1) - \phi_2\gamma(2) &= 1\\ \gamma(1) - \phi_1\gamma(0) - \phi_2\gamma(1) &= 0.\end{aligned}$$

Now, using the general solution $\gamma(h) = c_1\xi_1^{-h} + c_2\xi_2^{-h}$ *for* $h \geq 0$ *and substituting into the boundary conditions gives*

$$\begin{aligned}c_1 &= \xi_1^3\xi_2^2\left[(\xi_1^2-1)(\xi_2-\xi_1)(\xi_1\xi_2-1)\right]^{-1}\\ c_2 &= \xi_1^2\xi_2^3\left[(\xi_2^2-1)(\xi_2-\xi_1)(\xi_1\xi_2-1)\right]^{-1}\end{aligned}$$

which fully describes the covariance function in terms of underlying parameters. Substituting into the general solution then gives

$$\gamma(h) = \frac{\sigma^2\xi_1^2\xi_2^2}{(\xi_2-\xi_1)(\xi_1\xi_2-1)}[(\xi_1^2-1)^{-1}\xi_1^{1-h} - (\xi_1^2-1)^{-1}\xi_1^{1-h}].$$

Another interesting question is for what values of ϕ_1, ϕ_2 *the roots lie outside the unit circle. Solving for* ξ_1^{-1}, ξ_2^{-1} *in terms of* ϕ_1, ϕ_2 *gives*

$$\xi_{1,2}^{-1} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

such that $\xi_{1,2}^{-1} = 1$ *if* $\phi_2 = 1 - \phi_1$ *and* $\xi_{1,2}^{-1} = -1$ *if* $\phi_2 = 1 + \phi_1$. $\xi_{1,2}^{-1}$ *is complex if* $\phi_1^2 + 4\phi_2 < 0$. *The modulus of the complex roots is larger than one if* $\phi_2 < -1$.

2.4. Linear Projections and Partial Autocorrelations

We begin by reviewing some basic properties of linear vector spaces. A vector space V is a set (here we only consider subsets of $\mathbb{R}^n$ or $\mathbb{C}^n$) with two binary operations, vector addition and scalar multiplication. A vector space is closed under linear transformations, i.e. $\alpha x + \beta y \in V$ for $x, y \in V$ and α, β scalars.

A norm is a function $\|.\| : V \to [0, \infty)$ such that $\|\lambda x\| = |\lambda|\,\|x\|$, $\|x+y\| \leq \|x\| + \|y\|$ and $\|x\| = 0 \Rightarrow x = 0$. A function without the last property is called a seminorm. A normed vector space is a vector space equipped with a norm. We can then define the metric $\rho(x,y) = \|x-y\|$ on V. A vector space is called complete if every Cauchy sequence converges. A complete normed space is called a Banach space. We use the notation $\bar{z}$ for the complex conjugate of z if $z \in \mathbb{C}$.

An inner product on a complex normed vector space is a function $\langle x,y\rangle : V^2 \to \mathbb{C}$ such that $\langle x,y\rangle = \overline{\langle y,x\rangle}$, $\langle \alpha x + \beta y, z\rangle = a\langle x,z\rangle + b\langle y,z\rangle$ for $a, b \in \mathbb{C}$ and $\langle x,x\rangle = \|x\|^2$. A complete inner product space is called a Hilbert space.

Definition 2.9 (Basis of a Vector Space). *A basis of a complex vector space* V *is any set of linearly independent vectors* $v_1, ...v_n$ *that span V, i.e, for any* $v \in V$ *there exist scalars* $\alpha_i \in \mathbb{C}$ *such that* $v = \sum_{i=1}^n \alpha_i v_i$.

A basis is said to be orthonormal if $\langle v_i, v_j\rangle = 1$ for $i = j$ and $\langle v_i, v_j\rangle = 0$ for $i \neq j$.

Definition 2.10. *A subspace of a vector space is a subset $\mathcal{M} \subset V$ such that $\mathcal{M}$ itself is a vector space. If V is an inner product space then we can define the orthogonal complement of $\mathcal{M}$ denoted by $\mathcal{M}^{\perp}$ as*

$$\mathcal{M}^{\perp} = \{x \in V : \langle x, y \rangle = 0, \forall y \in \mathcal{M}\}.$$

Proposition 2.11 (Vector Decomposition). *Any element $x \in V$ can be written as the sum of two vectors y_1, y_2 such that $y_1 \in \mathcal{M}$ and $y_2 \in \mathcal{M}^{\perp}$ for any subspace $\mathcal{M} \subset V$.*

Proof. Let $v_1, ...v_p$ be an orthonormal basis of $\mathcal{M}$. Let $y_1 = \sum_{i=1}^{p} \langle y, v_i \rangle v_i$. Let $y_2 = y - y_1$. Clearly, $y_1 \in \mathcal{M}$. Also, for $j = 1, ..., p$,

$$\langle y_2, v_j \rangle = \langle y, v_j \rangle - \langle y_1, v_j \rangle = \langle y, v_j \rangle - \left\langle \sum_{i=1}^{p} \langle y, v_i \rangle v_i, v_j \right\rangle = \langle y, v_j \rangle - \langle v_j, v_j \rangle \langle y, v_j \rangle = 0.$$

So, y_2 is orthogonal to v_j, $j = 1, ..., p$, thus $y_2 \in \mathcal{M}^{\perp}$. ∎

An alternative way to express this result is to write $V = \mathcal{M} \oplus \mathcal{M}^{\perp}$. Let $x \in V$ and $\mathcal{M}$ a linear subspace of V. A projection $P_{\mathcal{M}}(x)$ of x onto $\mathcal{M}$ is an element of $\mathcal{M}$ such that

$$\|x - P_{\mathcal{M}}(x)\| = \inf_{y \in \mathcal{M}} \|x - y\|.$$

Theorem 2.12 (Projection Theorem). *(a) $P_{\mathcal{M}}(x)$ exists, is unique and is a linear function of x. (b) $P_{\mathcal{M}}(x)$ is the projection of x on $\mathcal{M}$ iff $x - P_{\mathcal{M}}(x) \perp \mathcal{M}$.*

Proof. (a) By the proof that $V = \mathcal{M} \oplus \mathcal{M}^{\perp}$, we can write $x = x_1 + x_2$ where $x_1 \in \mathcal{M}$ and $x_2 \in \mathcal{M}^{\perp}$ and $x_1 = \sum_{i=1}^{p} \langle x, v_i \rangle v_i$. Then for any $y \in \mathcal{M}$ we have

$$\begin{aligned} \|x - y\|^2 &= \langle x - x_1 + x_1 - y, x - x_1 + x_1 - y \rangle \\ &= \|x - x_1\|^2 + \|x_1 - y\|^2 \\ &\geq \|x - x_1\|^2 \end{aligned}$$

where the inequality is strict unless $\|x - y\|^2 = 0$ or $y = x_1$. Hence x_1 is the projection of x onto $\mathcal{M}$ and it is unique.

(b) If $P_{\mathcal{M}}(x)$ is a projection of x onto $\mathcal{M}$ then by part (a) $P_{\mathcal{M}}(x) = x_1$ and $x - P_{\mathcal{M}}(x) = x_2 \in \mathcal{M}^{\perp}$. Conversely if $P_{\mathcal{M}}(x)$ is some element of $\mathcal{M}$ for which $x - P_{\mathcal{M}}(x) \in \mathcal{M}^{\perp}$, then $x_1 + x_2 - P_{\mathcal{M}}(x) \in \mathcal{M}^{\perp}$, $x_1 - P_{\mathcal{M}}(x) \in \mathcal{M}^{\perp}$, since $x_2 \in \mathcal{M}^{\perp}$ and $x_1 - P_{\mathcal{M}}(x) \in \mathcal{M}$ since $x_1, P_{\mathcal{M}}(x) \in \mathcal{M}$. This implies $x_1 - P_{\mathcal{M}}(x) = 0$. Therefore $P_{\mathcal{M}}(x)$ is the projection of x onto $\mathcal{M}$. ∎

In this section we need some properties of complex L^2 spaces defined on the random variables X on $(\Omega, \mathcal{F}, P)$. By definition $E|X|^2 < \infty$ for $X \in L^2(\Omega, \mathcal{F}, P)$. The space L^2 is a complex Hilbert space with inner product

$$\langle X, Y \rangle = E(X\overline{Y}).$$

From the previous definition of a linear subspace $\mathcal{M}$ of a Hilbert space $\mathcal{H}$ we know that $\mathcal{M}$ is a subset $\mathcal{M} \subset \mathcal{H}$ such that $0 \in \mathcal{M}$ and for $x_1, x_2 \in \mathcal{M}$ it follows that $y = a_1 x_1 + a_2 x_2 \in \mathcal{M}$ for all $a_1, a_2 \in \mathbb{C}$. A closed linear subspace is a subspace that contains all its limit points.

Definition 2.13 (Closed Span). *The closed span $\overline{sp}\{x_t, t \in \mathcal{T}\}$ of any subset $\{x_t, t \in \mathcal{T}\}$ of the Hilbert space $\mathcal{H}$ is the smallest closed subspace of $\mathcal{H}$ which contains each element of $\{x_t, t \in \mathcal{T}\}$. The closed span of a finite set $\{x_1, ..., x_n\}$ contains all linear combinations $y = a_1 x_1 + ... + a_n x_n$, $a_1, ..., a_n \in \mathbb{C}$.*

We defined the projection $P_{\mathcal{M}}(x)$ of $x \in \mathcal{H}$ onto the subspace $\mathcal{M}$ as the element $\hat{x} \in \mathcal{M}$ such that $\langle x - \hat{x}, x - \hat{x} \rangle = \inf_{y \in \mathcal{M}} \langle x - y, x - y \rangle$. The projection $\hat{x}$ is unique by the projection theorem. Moreover $x - \hat{x} \in \mathcal{M}^{\perp}$ where $\mathcal{M}^{\perp}$ is orthogonal to $\mathcal{M}$.

It is now obvious from the definition of $P_{\mathcal{M}}(x)$ that the projection onto $\overline{sp}\{x_1, ..., x_n\}$ has the form

$$P_{\overline{sp}\{x_1,...,x_n\}}(x) = a_1x_1 + ... + a_nx_n$$

since $P_{\overline{sp}\{x_1,...,x_n\}}(x) \in \overline{sp}\{x_1, ..., x_n\}$ and the coefficients have to satisfy

$$\sum a_i \langle x_i, x_j \rangle = \langle x, x_j \rangle \text{ for } j = 1, ...n.$$

Using the concept of projection onto linear subspaces we can now introduce the partial autocorrelation function. The partial autocorrelation function measures the correlation between two elements x_{t+k} and x_t of a time series after taking into account the correlation that is explained by $x_{t+1}, ...x_{t+k-1}$. In the following we assume stationarity of x_t and normalize $t = 1$. Formally the partial autocorrelation function of a stationary time series is defined as

$$\alpha(1) = Corr(x_2, x_1) = \rho(1)$$

and

$$\alpha(k) = Corr(x_{k+1} - P_{\overline{sp}\{x_2,...,x_k\}}(x_{k+1}), x_1 - P_{\overline{sp}\{x_2,...,x_k\}}(x_1)).$$

The partial autocorrelation is therefore the correlation between the residuals from a regression of x_{k+1} on $x_2, ..., x_k$ and the residuals from a regression of x_1 onto $x_2, ..., x_k$. An alternative but equivalent definition can be given in terms of the last regression coefficient in a regression of x_t onto the k lagged variables $x_{t-1}, ...x_{t-k}$. If

$$P_{\overline{sp}\{x_1,...,x_k\}}(x_{k+1}) = \sum_{i=1}^{k} \phi_{ik} x_{k+1-i}$$

where

$$\begin{bmatrix} \rho(0) & \rho(1) & \cdots & \rho(k-1) \\ & \ddots & \ddots & \vdots \\ & & \ddots & \rho(1) \\ & & & \rho(0) \end{bmatrix} \begin{bmatrix} \phi_{1k} \\ \vdots \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \vdots \\ \vdots \\ \rho(k) \end{bmatrix} \tag{2.12}$$

then $\alpha(k) \equiv \phi_{kk}$. It can be shown that the two definitions are equivalent. We consider two examples next.

Example 2.14. *Let x_t follow a causal AR(p) process such that*

$$x_t - \phi_1 x_{t-1} - ... - \phi_p x_{t-p} = \varepsilon_t \text{ with } \varepsilon_t \sim WN(0, \sigma^2).$$

Then, for $k \geq p$

$$P_{\overline{sp}\{x_1,...,x_k\}}(x_{k+1}) = \sum_{j=1}^{p} \phi_j x_{k+1-j} \tag{2.13}$$

which can be seen from looking at any $y \in \overline{sp}\{x_1, ..., x_k\}$. By causality $y \in \overline{sp}\{\varepsilon_j, j \leq k\}$ such that $< x_{k+1} - \sum_{j=1}^{p} \phi_j x_{k+1-j}, y >= 0$. By the projection theorem this implies that (2.13) holds. It now follows that for $k > p$

$$\begin{aligned} \alpha(k) &= Corr(x_{k+1} - \sum_{j=1}^{p} \phi_j x_{k+1-j}, x_1 - P_{\overline{sp}\{x_2,...,x_k\}}(x_1)) \\ &= Corr(\varepsilon_{k+1}, x_1 - P_{\overline{sp}\{x_2,...,x_k\}}(x_1)) \\ &= 0. \end{aligned}$$

We see that the partial autocorrelation for the $AR(p)$ process is zero for lags higher than p. The next example considers the $MA(1)$ process. For the invertible case this process can be represented as an $AR(\infty)$. We therefore expect the partial autocorrelations to die out slowly rather than collapsing at a finite lag. This is in fact the case.

Exercise 2.1. *Let x_t be driven by a $MA(1)$ process*

$$x_t = \varepsilon_t - \theta\varepsilon_{t-1} \text{ with } |\theta| < 1,\ \varepsilon_t \sim WN(0, \sigma^2)$$

then we know from before that $\alpha(1) = \rho(1) = -\theta/(1+\theta^2)$. Equations (2.12) now become

$$\begin{bmatrix} 1 & \frac{-\theta}{(1+\theta^2)} & \cdots & 0 \\ \frac{-\theta}{(1+\theta^2)} & \ddots & \ddots & \vdots \\ & \ddots & \ddots & \frac{-\theta}{(1+\theta^2)} \\ 0 & & \frac{-\theta}{(1+\theta^2)} & 1 \end{bmatrix} \begin{bmatrix} \phi_{1k} \\ \vdots \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \frac{-\theta}{(1+\theta^2)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then ϕ_{ik} is the solution to the difference equation

$$-\theta\phi_{i-1k} + (1+\theta^2)\phi_{ik} - \theta\phi_{i+1k} = 0$$

with initial condition

$$(1+\theta^2)\phi_{1k} - \theta\phi_{2k} = -\theta$$

and terminal condition

$$(1+\theta^2)\phi_{kk} - \theta\phi_{k-1k} = 0$$

The difference equation can be written as $(1 - (1+\theta^2)/\theta L + L^2)\phi_i = 0$ with roots θ and $1/\theta$. The general solution is then $\phi_i = c_1\theta^i + c_2\theta^{-i}$. Substitution into the initial and terminal conditions allows to solve for the constants c_1 and c_2, in particular

$$c_1 = \frac{-1}{1-\theta^{2(k+1)}} \text{ and } c_2 = \frac{\theta^{2k+2}}{1-\theta^{2(k+1)}}.$$

The constants depend on k because of the terminal condition. The terminal value ϕ_{kk} is then found from substituting back into the general solution. This leads to

$$\alpha(k) = -\frac{\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}}.$$

We see from the two examples and the results on the autocovariance function that the highest order of AR polynomial can be determined from the point where the partial autocorrelations are zero and the highest order of the MA polynomial can be determined from the autocovariance function in the same way. This has lead to method of identifying the correct specification for an $ARMA(p,q)$ model by looking both at the autocorrelation and partial autocorrelation function of a process. It is clear that for a general model the decay patterns of these functions can be quite complicated. It is therefore usually difficult to reach a clear decision regarding the correct specification by looking at the empirical counterparts of autocorrelations and partial autocorrelations.

Exercise 2.2. *Find the partial autocorrelation function for x_t where*

$$x_t = \varepsilon_t - \varepsilon_{t-1}$$

and ε_t is white noise.