ECONOMETRIC METHODS II: TIME SERIES

LECTURE NOTES 1

KRISTOFFER P. NIMARK

This first set of notes briefly covers some background material that most are likely to have encountered before, perhaps in a first year macro or econometrics course. The purpose of these notes is to establish a bit of terminology and notation as well as to (re-) introduce some basic tools that are useful for modeling stochastic economies. Most of the material can be found in any time series text book, e.g. Brockwell and Davis (2006).

1. Stochastic processes

We will use $\{X_t\}_{t=0}^{\infty} : X_t \in \mathbb{R}^n$ to denote a discrete time vector valued random process. Depending on context, the notation X_t will either denote the random process at time t, or a particular realization of the random process at time t.

1.1. **Uncorrelated, Orthogonal and Independent processes.** Two stochastic processes $\{X_t\}$ and $\{Y_t\}$ are:

Uncorrelated if $E[X_t Y_t'] = E[X_t]E[Y_t'] \; \forall \; t$.

Orthogonal if $E[X_t Y_t'] = 0 \; \forall \; t$.

Independent if $p_{X|Y}(X \mid Y) = p_X(X)$ where $p_{X|Y}$ is the probability density function of X conditional on Y.

1.2. **Markov property.** A stochastic process is said to have the Markov property if $p_{X_{t+1}|X_t}(X_{t+1} \mid X_t) = p_{X_{t+1}|X_t,X_{t-1},...,X_0}(X_{t+1} \mid X_t, X_{t-1}, ..., X_0)$. Loosely speaking, a process is a Markov process if the present is sufficient to optimally predict the future. A related concept is the "state" of the process: The state (or position) of a process completely summarizes all currently available information about future values of the process. If the state of a process can be defined in a time invariant way, the process can usually be redefined as a Markov process, for instance by augmenting the current vector to also include lagged values of itself.

1.3. **Stationarity.** A stochastic process is said to be stationary if the distribution of the process is independent of time. A weaker requirement is that the process is covariance stationary, or wide sense stationary, which only requires the first two moments (i.e. the mean and the covariance) of the process to be (finite and) independent of time. For Gaussian processes stationarity and covariance stationarity are equivalent since Gaussian processes are

Date: May 11, 2010.

completely characterized by their means and covariances. Formally, we call a process $\{X_t\}$ covariance stationarity if the conditions

$$E[X_t] = \mu_X \;\forall\; t \tag{1.1}$$
$$E[X_t X_t'] < \infty \;\forall\; t \tag{1.2}$$
$$E[X_t - \mu_X][X_{t+s} - \mu_X]' = E[X_{t+j} - \mu_X][X_{t+s+j} - \mu_X]' \;\forall\; t, j \tag{1.3}$$

are satisfied.

1.4. **Gaussian Vector Processes.** A Gaussian n dimensional vector process $X \sim N(\mu_X, \Sigma_{XX})$ has the probability density

$$p_X(X) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\Sigma_{XX}|^{1/2}} \exp\left[-\frac{1}{2}(X - \mu_X)\Sigma_{XX}^{-1}(X - \mu_X)'\right] \tag{1.4}$$

The sum of two independent Gaussian processes $X \sim N(\mu_X, \Sigma_{XX})$ and $Y \sim N(\mu_Y, \Sigma_{YY})$ are distributed as $N(\mu_X + \mu_Y, \Sigma_{XX} + \Sigma_{YY})$. If X and Y are jointly Gaussian processes

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_X \\ \mu Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}\right) \tag{1.5}$$

The conditional density $p_{X|Y}$ is Gaussian with mean $\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(Y - \mu_y)$ and covariance $\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$.

2. Linear Stochastic Difference Equations

One of the most useful tools for modeling stochastic economies is the first order linear stochastic difference equation

$$X_t = AX_{t-1} + C\mathbf{u}_t \tag{2.1}$$

where X_t is an $n \times 1$ vector of random variables, u_t is an $m \times 1$ vector of i.i.d. shocks with unit variance, i.e. $E\left[\mathbf{u}_t \mathbf{u}_{t+s}'\right] = I$ if $s = 0$ and $\mathbf{0}$ otherwise. A and C are ($n \times n$ and $n \times m$, respectively) coefficient matrices. We will restrict our attention to processes where the eigenvalues of A all lie inside the unit circle so that (2.1) is a stable process. That the process is a *first order* vector autoregression is not very restrictive as it is straightforward to transform a higher order process to a first order process.

2.0.1. *Examples:* A VAR(p)

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \ldots . A_p y_{t-p} + C_0 \mathbf{u}_t \tag{2.2}$$

can be re-written as

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \cdots & A_p \\ I & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & I & \mathbf{0} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} C_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \mathbf{u}_t \tag{2.3}$$

which is in the form (2.1) if $X_t = \begin{bmatrix} y_t & y_{t-1} & \cdots & y_{t-p+1} \end{bmatrix}'$ and $C = \begin{bmatrix} C_0 & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}'$.

Similarly, an vector ARMA(1,2) process

$$y_t = A_1 y_{t-1} + C_0 \mathbf{u}_t + C_1 \mathbf{u}_{t-1} + C_2 \mathbf{u}_{t-2} \tag{2.4}$$

can be rewritten as

$$\begin{bmatrix} y_t \\ \mathbf{u}_t \\ \mathbf{u}_{t-1} \end{bmatrix} = \begin{bmatrix} A_1 & C_1 & C_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \mathbf{u}_{t-1} \\ \mathbf{u}_{t-2} \end{bmatrix} + \begin{bmatrix} C \\ I \\ \mathbf{0} \end{bmatrix} \mathbf{u}_t \tag{2.5}$$

2.1. **Prediction and discounted geometric sums.** The process (2.1) can be used to make predictions of future values of X_t using that

$$E\left[X_{t+s} \mid X_t\right] = A^s X_t \tag{2.6}$$

Sometimes we might be interested in computing the discounted sum of all future expected values of a vector of variables

$$E\left[\sum_{s=0}^{\infty} \beta^s X_{t+s} \mid X_t\right] = X_t + \beta A X_t + \beta^2 A^2 X_t + ... + \beta^n A^n X_t : n \to \infty \tag{2.7}$$

There is a trick to computing infinite sums like (2.7): Start by subtracting all the terms on the right hand side except X_t from both sides to get

$$E\left[\sum_{s=0}^{\infty} \beta^s X_{t+s} \mid X_t\right] - \left(\beta A X_t + \beta^2 A^2 X_t + ... + \beta^n A^n X_t : n \to \infty\right) = X_t \tag{2.8}$$

which equals

$$(I - \beta A) E\left[\sum_{s=0}^{\infty} \beta^s X_{t+s} \mid X_t\right] = X_t \tag{2.9}$$

since

$$\begin{aligned} \left(\beta A X_t + \beta^2 A^2 X_t + ... + \beta^n A^n X_t\right) &= \beta A \left(X_t + \beta A X_t + \beta^2 A^2 X_t + ... + \beta^n A^n X_t\right) && (2.10) \\ &= \beta A E\left[\sum_{s=0}^{\infty} \beta^s X_{t+s} \mid X_t\right] && (2.11) \end{aligned}$$

Pre-multiplying both sides by $(I - \beta A)^{-1}$

$$E\left[\sum_{s=0}^{\infty} \beta^s X_{t+s} \mid X_t\right] = (I - \beta A)^{-1} X_t \tag{2.12}$$

gives the sum as function of the current state X_t.

2.2. **Autocovariance function.** The autocovariance function $\Gamma_X(s)$ of a stochastic process $\{X_t\}$ is defined as

$$\Gamma_X(s) \equiv E\left[X_t - \mu_X\right]\left[X_{t+s} - \mu_X\right]'$$

To compute the covariance of a process, it is often helpful to rewrite the process as a sum of orthogonal components. The variance of the process can then be computed by summing over the variances of the orthogonal components. The variance of the VAR(1) (2.1) can be decomposed into orthogonal components by noting that a VAR process like (2.1) has an infinite order moving average representation

$$X_t = C\mathbf{u}_t + AC\mathbf{u}_{t-1} + A^2 C\mathbf{u}_{t-2} + ... + A^{\infty} C\mathbf{u}_{t-n} : n \to \infty \tag{2.13}$$

where the terms on the right hand side are orthogonal (by the assumption that $E\left[\mathbf{u}_t\mathbf{u}_{t+s}'\right] = I$ if $s = 0$ and $\mathbf{0}$ otherwise). The covariance of X_t

$$\Gamma_X(0) \equiv \Sigma_{XX} \equiv E\left[X_t - \mu_X\right]\left[X_t - \mu_X\right]' \tag{2.14}$$

is then given by

$$\Sigma_{XX} = CC' + ACC'A' + A^2CC'A^{2\prime} + ... + A^nCC'A^{n\prime} : n \to \infty \tag{2.15}$$

which is again a very long formula. It can be simplified if we pre-multiply both sides of (2.15) with A and post-multiply by A' to get

$$A\Sigma_{XX}A' = ACC'A' + A^2CC'A^{2\prime} + ... + A^nCC'A^{n\prime} : n \to \infty \tag{2.16}$$

Substitute $A\Sigma_{XX}A'$ into (2.15) to get the much neater expression

$$\Sigma_{XX} = A\Sigma_{XX}A' + CC' \tag{2.17}$$

The solution to (2.17) can be computed by iterating on

$$\Sigma_{XX,s+1} = A\Sigma_{XX,s}A' + CC' \tag{2.18}$$

starting from $\Sigma_{XX,0} = 0$. The solution of (2.17) can be used to compute the autocovariance function also for $s \neq 0$ using that

$$\Gamma_X(s) = A^s\Sigma_{XX} \tag{2.19}$$

2.3. **Linear functions of X_t.** If Z_t is a vector of random variables

$$\begin{aligned} Z_t &= \mathbf{a} + DX_t && (2.20)\\ \mu_Z &= \mathbf{a} + D\mu_X && (2.21)\\ E\left[Z_t - \mu_z\right]\left[Z_t - \mu_z\right]' &= D\Sigma_{XX}D' && (2.22)\\ E\left[\sum_{s=0}^{\infty}\beta^s Z_{t+s} \mid X_t\right] &= D\left(I - \beta A\right)^{-1}X_t + D\left(I - \beta A\right)^{-1}\mu_X + \left(1 - \beta\right)^{-1}\mathbf{a} && (2.23) \end{aligned}$$

3. Some useful results and definitions from linear algebra

3.1. **Vector spaces.** A vector space $\mathcal{X}$ is a set of elements and two operations; addition and scalar multiplication. For two elements $x, y \in \mathcal{X}$ the following axioms are assumed to hold:

(1) $x + y = y + x$
(2) $(x + y) + z = x + (y + z)$
(3) $x + \theta = x$ where θ is the null vector
(4) $\alpha(x + y) = \alpha x + \alpha y$
(5) $(\alpha + \beta)\,x = \alpha x + \beta x$
(6) $(\alpha\beta)\,x = \alpha\,(\beta x)$
(7) $0x = 0, 1x = x$

3.2. **Inner product spaces.** A inner product space is a vector space with some additional structure given by an inner product. An inner product of $x, y \in \mathcal{X}$ is denoted $\langle x, y \rangle$ and satisfies the following axioms:

(1) $\langle x, y \rangle = \langle y, x \rangle$
(2) $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$
(3) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$
(4) $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = \theta$.

Example 1. ($\mathbb{R}^n$) The inner product space $\mathbb{R}^n$ is the collection of all n-dimensional vectors with elements in $\mathbb{R}$ *so that if* $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} : x_i \in \mathbb{R} \text{ for } i = 1, 2, ...n \tag{3.1}$$

and the associated inner product

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle &= \mathbf{x}'\mathbf{y} \\ &= \sum_{i=1}^{n} x_i y_i \end{aligned}$$

Example 2. (The inner-product space L^2.) The inner product space L^2 is the collection C of all random variables X with finite variance

$$EX^2 < \infty \tag{3.2}$$

and with inner-product

$$\langle X, Y \rangle \equiv E(XY) : X, Y \in L^2 \tag{3.3}$$

3.3. **Orthogonal vectors.** Two vectors $x, y \in \mathcal{X}$ are said to be orthogonal if their inner product $\langle x, y \rangle$ is zero.

3.4. **Orthogonal subspaces.** Two subspaces $\mathcal{Z}$ and $\mathcal{Y}$ are said to be orthogonal if for any $z \in \mathcal{Z}$ and $y \in \mathcal{Y}$ we have that $\langle z, y \rangle = 0$.

3.5. **The projection theorem.** Let $\mathcal{Y}$ be a subspace of $\mathcal{X}$. An orthogonal projection of $x \in \mathcal{X}$ on Y , denoted $\mathcal{P}_Y x$, is the unique element in $\mathcal{Y}$ satisfying

$$\langle x - \mathcal{P}_Y x, y \rangle = 0 \tag{3.4}$$

for any $y \in \mathcal{Y}$. The orthogonal projection is the element in $\mathcal{Y}$ that minimizes the norm of the projection error

$$\mathcal{P}_Y x = \arg\min_{y \in \mathcal{Y}} \sqrt{\langle (x - \mathcal{P}_Y x), (x - \mathcal{P}_Y x) \rangle}$$

Orthogonal projections have the following useful properties:

(1) The projection $\mathcal{P}_Y x$ coincides with the conditional expectation $E[x \mid \mathcal{Y}]$ in linear models with Gaussian shocks.

(2) Let $\mathcal{Z}$ be a subspace of $\mathcal{X}$ and $\mathcal{Z}^{\perp}$ its orthogonal complement in $\mathcal{X}$. Then each $x \in \mathcal{X}$ has a unique representation as a sum of an element in Z and an element of Y, i.e.

$$x = \mathcal{P}_Z x + \mathcal{P}_{Z^{\perp}} x \tag{3.5}$$

(3) $x \in \mathcal{Z}$ if and only if $\mathcal{P}_{Z^{\perp}} x = 0$, where $\mathcal{Z}^{\perp}$ is the orthogonal complement to $\mathcal{Z}$.
(4) $Z_1 \subseteq Z_2$ if and only if $\mathcal{P}_{Z1} x = \mathcal{P}_{Z1} \mathcal{P}_{Z2} x$ for all $x \in \mathcal{X}$.

Luenberger (1969) is a good reference for those who want to know more about vector spaces and what they can be used for (which is quite a lot, actually).

3.6. **Eigenvalues.** An eigenvalue λ_i and associated eigenvector $x_i \neq 0$ of a square matrix A solves the equation

$$(A - \lambda_i I)\, x_i = 0 \tag{3.6}$$

3.7. **The power of a matrix.** If the matrix A is diagonalizable, that is, if it can be written in the form

$$A = S \Lambda S^{-1} \tag{3.7}$$

where Λ is a diagonal matrix with the eigenvalues of A on the main diagonal the columns of S are the associated eigenvectors, then

$$A^j = S \Lambda^j S^{-1} \tag{3.8}$$

or

$$A^j = S \begin{bmatrix} \lambda_1^j & \mathbf{0} & 0 \\ \mathbf{0} & \ddots & \mathbf{0} \\ 0 & \mathbf{0} & \lambda_n^j \end{bmatrix} S^{-1} \tag{3.9}$$

This generalizes to $j < 0$ and we can also use the eigenvalue-eigenvector decomposition to compute the geometric sum

$$\sum_{j=0}^{\infty} A^j = S \left(I - \Lambda \right)^{-1} S^{-1} \tag{3.10}$$

or

$$\sum_{j=0}^{\infty} A^j = S \begin{bmatrix} \frac{1}{1-\lambda_1} & \mathbf{0} & 0 \\ \mathbf{0} & \ddots & \mathbf{0} \\ 0 & \mathbf{0} & \frac{1}{1-\lambda_n} \end{bmatrix} S^{-1} \tag{3.11}$$

3.8. **Positive semi-definite matrices.** A matrix A is said to be positive definite if for any vector $x \neq 0$ we have that

$$x' A x \geq 0 \tag{3.12}$$

Covariance matrices are positive semi-definite.