

TOPICS IN MACROECONOMICS: MODELING INFORMATION, LEARNING AND EXPECTATIONS

BASICS

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 the Chapter 2 of Ljungqvist and Sargent (2004) and/or the Appendix of Anderson and Moore (1979).

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 realisation 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 $p_{X|Y}(X | Y) = p_X(X)$ 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 $p_{X_{t+1}|X_t}(X_{t+1} | X_t) = p_{X_{t+1}|X_t, X_{t-1}, \dots, X_0}(X_{t+1} | X_t, X_{t-1}, \dots, X_0)$. Loosely speaking, a process is a Markov process

Date: November 2, 2010.

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

$$E[X_t] = \mu_X \quad \forall t \quad (1.1)$$

$$E[X_t X_t'] < \infty \quad \forall t \quad (1.2)$$

$$E[X_t - \mu_X][X_{t+s} - \mu_X]' = E[X_{t+j} - \mu_X][X_{t+s+j} - \mu_X]' \quad \forall t, j \quad (1.3)$$

are satisfied.

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

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

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

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

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

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

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. $E[\mathbf{u}_t\mathbf{u}'_{t+s}] = I$ 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)

$$y_t = A_1y_{t-1} + A_2y_{t-2} + \dots + A_p y_{t-p} + C_0\mathbf{u}_t \quad (2.2)$$

can be re-written as

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

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

Similarly, an vector ARMA(1,2) process

$$y_t = A_1y_{t-1} + C_0\mathbf{u}_t + C_1\mathbf{u}_{t-1} + C_2\mathbf{u}_{t-2} \quad (2.4)$$

can be rewritten as

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

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

$$E[X_{t+s} | X_t] = A^s X_t \quad (2.6)$$

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

$$E \left[\sum_{s=0}^{\infty} \beta^s X_{t+s} | X_t \right] = X_t + \beta A X_t + \beta^2 A^2 X_t + \dots + \beta^n A^n X_t : n \rightarrow \infty \quad (2.7)$$

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

$$E \left[\sum_{s=0}^{\infty} \beta^s X_{t+s} | X_t \right] - (\beta A X_t + \beta^2 A^2 X_t + \dots + \beta^n A^n X_t : n \rightarrow \infty) = X_t \quad (2.8)$$

which equals

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

since

$$(\beta A X_t + \beta^2 A^2 X_t + \dots + \beta^n A^n X_t) = \beta A (X_t + \beta A X_t + \beta^2 A^2 X_t + \dots + \beta^n A^n X_t) \quad (2.10)$$

$$= \beta A E \left[\sum_{s=0}^{\infty} \beta^s X_{t+s} | X_t \right] \quad (2.11)$$

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

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

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

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

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

$$X_t = C\mathbf{u}_t + AC\mathbf{u}_{t-1} + A^2C\mathbf{u}_{t-2} + \dots + A^\infty C\mathbf{u}_{t-n} : n \rightarrow \infty \quad (2.13)$$

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

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

is then given by

$$\Sigma_{XX} = CC' + ACC'A' + A^2CC'A^2' + \dots + A^n CC'A^n' : n \rightarrow \infty \quad (2.15)$$

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

$$A\Sigma_{XX}A' = ACC'A' + A^2CC'A^2' + \dots + A^n CC'A^n' : n \rightarrow \infty \quad (2.16)$$

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

$$\Sigma_{XX} = A\Sigma_{XX}A' + CC' \quad (2.17)$$

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

$$\Sigma_{XX,s+1} = A\Sigma_{XX,s}A' + CC' \quad (2.18)$$

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

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

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

$$Z_t = \mathbf{a} + DX_t \quad (2.20)$$

$$\mu_Z = \mathbf{a} + D\mu_X \quad (2.21)$$

$$E[Z_t - \mu_z][Z_t - \mu_z]' = D\Sigma_{XX}D' \quad (2.22)$$

$$E\left[\sum_{s=0}^{\infty} \beta^s Z_{t+s} \mid X_t\right] = D(I - \beta A)^{-1} X_t + D(I - \beta A)^{-1} \mu_X + (1 - \beta)^{-1} \mathbf{a} \quad (2.23)$$

3. SOME USEFUL RESULTS AND DEFINITIONS FROM LINEAR ALGEBRA

3.1. Linear vector spaces. This Appendix reproduces some results and properties of orthogonal projections on inner-product spaces that are used in Section 3 of the main text above. Proofs and more details can be found in for instance Brockwell and Davis (2006).

Definition 1. *(The inner-product space L^2 .) The inner product space L^2 is the collection C of all random variables X with finite variance*

$$EX^2 < \infty \quad (3.1)$$

and with inner-product

$$\langle X, Y \rangle \equiv E(XY) : X, Y \in L^2 \quad (3.2)$$

3.2. Orthogonal vectors. Two vectors $x, y \in \mathbb{R}^n$ are said to be orthogonal if their inner product $\langle x, y \rangle$ is zero. For linear vector spaces in \mathbb{R}^n the inner product is defined as

$$\langle x, y \rangle \equiv x'y \equiv \sum_{i=0}^n x_i y_i \quad (3.3)$$

$x'y$ is sometimes called the scalar product or the dot product of x and y .

3.3. Orthogonal subspaces. Two subspaces \mathcal{X} and \mathcal{Y} are said to be orthogonal if for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ we have that $\langle x, y \rangle = x'y = 0$.

3.4. The projection theorem. An orthogonal projection of X on Ω , denoted $P_\Omega X$, is the unique element in L^2 satisfying

$$\langle X - P_\Omega X, \omega \rangle = 0 \quad (3.4)$$

for any $\omega \in \Omega$.

Orthogonal projections have the following useful properties:

- (1) The projection $P_\Omega X$ coincides with the conditional expectation $E[X | \Omega]$ in linear models with Gaussian shocks.
- (2) Let Ω' be a subspace of Ω and Ω'^\perp its orthogonal complement in Ω . Then each $\omega \in \Omega$ has a unique representation as a sum of an element in Ω' and an element of Ω'^\perp , i.e.

$$\omega = P_{\Omega'}\omega + P_{\Omega'^\perp}\omega \quad (3.5)$$

- (3) $X \in \Omega^\perp$ if and only if $P_\Omega X = 0$, where Ω^\perp is the orthogonal complement to Ω .
- (4) $\Omega_1 \subseteq \Omega_2$ if and only if $P_{\Omega_1}X = P_{\Omega_1}P_{\Omega_2}X$ for all $X \in L^2$.

3.5. Projections. Let $y_i, x_i \in \mathbb{R}^n$. A linear projection of $Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \end{bmatrix}$ on $X = \begin{bmatrix} x_1 & x_2 & \cdots & x_j \end{bmatrix}$ is given by $X(X'X)^{-1}X'Y$. The matrix $P_x = X(X'X)^{-1}X'$ is a projection matrix if and only if it has the following two properties

$$P^2 = P \quad (3.6)$$

$$P = P' \quad (3.7)$$

That $P^2 = P$ follows from that P projects any conformable matrix onto the closest point in the column space of X . The column space of a matrix X is (as the name suggests...) the space spanned by the columns of X and is denoted $\mathcal{R}(X)$. Since PY is already in $\mathcal{R}(X)$, applying P to PY returns the same point in $\mathcal{R}(X)$ since the point closest to PY in $\mathcal{R}(X)$ is PY itself.

That P is symmetric is easily seen by recognising that $(ABC)' = C'B'A'$ and that $(X'X)^{-1}$ is symmetric.

The projection error $\epsilon = Y - X(X'X)^{-1}X'Y$ is orthogonal to X so that $X'\epsilon = 0$. The matrix $(I - P)$ is also a projection matrix and projects any matrix $Z \in \mathbb{R}^n$ onto the orthogonal complement to $\mathcal{R}(X)$. If X is invertible, then $\mathcal{R}(X)$ spans all of \mathbb{R}^n and $P = I$.

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

$$(A - \lambda_i I)x_i = 0 \quad (3.8)$$

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

$$A = S\Lambda S^{-1} \quad (3.9)$$

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

$$A^j = S\Lambda^j S^{-1} \quad (3.10)$$

or

$$A^j = S \begin{bmatrix} \lambda_1^j & \mathbf{0} & 0 \\ \mathbf{0} & \ddots & \mathbf{0} \\ 0 & \mathbf{0} & \lambda_n^j \end{bmatrix} S^{-1} \quad (3.11)$$

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

$$\sum_{j=0}^{\infty} A^j = S(I - \Lambda)^{-1} S^{-1} \quad (3.12)$$

or

$$\sum_{j=0}^{\infty} A^j = S \begin{bmatrix} \frac{1}{1-\lambda_1} & \mathbf{0} & 0 \\ \mathbf{0} & \ddots & \mathbf{0} \\ 0 & \mathbf{0} & \frac{1}{1-\lambda_n} \end{bmatrix} S^{-1} \quad (3.13)$$

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

$$x'Ax \geq 0 \quad (3.14)$$

Covariance matrices are positive semidefinite.