Chapter 2

Some basic tools

2.1 Time series: Theory

2.1.1 Stochastic processes

A *stochastic process* is a sequence of random variables $\ldots, x_0, x_1, x_2, \ldots$. In this class, the subscript always means time. As always, x can be a vector.

Associated with this stochastic process is its *history*:

$$H_t = \{x_t, x_{t-1}, x_{t-2}, \ldots\} \tag{2.1}$$

The history is just the set of past values of x_t.

The *expected value at time t* of a particular random variable is just:

$$E_t(x_s) = E(x_s|H_t) \tag{2.2}$$

or its expected value conditional on all information available at t. Notice that if $s \leq t$, then $E_t(x_s) = x_s$.

A few results from basic probability theory should be noted here.

1. $E(.)$ is a linear operator, i.e., if X, Y, and Z are random variables then $E(a(Z)X + b(Z)Y|Z = z) = a(z)E(x|Z = z) + b(z)E(y|Z = z)$.
2. The law of iterated expectations. Let $\Omega_0 \subset \Omega_1$ be sets of conditioning variables (or equivalently let Ω_0 and Ω_1 be events and let $\Omega_1 \subset \Omega_0$), and let X be a random variable. Then $E(E(X|\Omega_1)|\Omega_0) = E(X|\Omega_0)$.

3. If $E(Y|X) = 0$, then $E(f(x)Y) = 0$ for any function $f(.)$.

Next we consider two special classes of stochastic process that macroeconomists find particularly useful.

2.1.2 Markov processes

Definition 2.1.1 (Markov process) *A Markov process is a stochastic process with the property that:*

$$\Pr(x_{t+1}|H_t) = \Pr(x_{t+1}|x_t) \tag{2.3}$$

where $\Pr(x)$ *is the probability distribution of* x*.*

In other words, once we know today's state, knowing history previous to today does not change our forecasts of the future.

Markov processes are very convenient to work with, and far less restrictive than one might think. The reason is that any process for which only a finite history matters can be redefined as a Markov process. Let x_t be a random variable such that:

$$\Pr(x_{t+1}|H_t) = \Pr(x_{t+1}|x_t, x_{t-1}, \ldots, x_{t-k}) \tag{2.4}$$

Then we can define y_t as

$$y_t \equiv \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-k} \end{bmatrix} \tag{2.5}$$

Note that the vector y_t is a Markov process.

Definition 2.1.2 (Markov chain) *A Markov chain is a Markov process with a countable state space.*

What does countable mean? Formally, a set A is countable if it is finite, or if there exists a one-to-one mapping between A and the natural numbers. Examples of countably infinite sets include the integers, and the rational

numbers. Examples of uncountable sets include the real numbers, the irrational numbers, and the complex numbers. In practice we will focus on Markov chains with finite state space.

A Markov chain can be defined using three items: an n-dimensional vector $\omega \in R^n$ defining the state space, an n-dimensional vector of initial probabilities π_0:

$$\pi_{0i} = \Pr(x_0 = \omega_i) \tag{2.6}$$

and an n-by-n Markov transition matrix:

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \cdots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix} \tag{2.7}$$

where $p_{ij} = \Pr(x_{t+1} = \omega_j | x_t = \omega_i)$.

Now, in order for everything to be well-defined, we need:

$$\sum_{i=1}^{n} \pi_{0i} = 1 \tag{2.8}$$

$$\sum_{j=1}^{n} p_{i,j} = 1 \tag{2.9}$$

Also let's introduce some additional notation. Let the n-vector π_t be defined as:

$$\pi_{ti} = \Pr(x_t = \omega_i) \tag{2.10}$$

Note that π_t is not random.

Now, why might we find Markov chains convenient for modeling? The first reason is that dynamic programming problems are often easier to solve for a discrete state space than a continuous space. It will often be the case that there is a Markov chain which is "close enough" to another Markov process to serve as more convenient approximation. The second reason is that many of the properties of a given Markov chain can be derived from the transition matrix without much trouble.

An example

Let's construct a simple example, with 2 states. Let x_t be the stochastic process in question, let

$$\omega = \begin{bmatrix} \sigma \\ -\sigma \end{bmatrix} \tag{2.11}$$

Let the transition matrix be given by

$$P = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix} \tag{2.12}$$

Let the initial probability vector be given by:

$$\pi_0 = \begin{bmatrix} p_0 \\ 1-p_0 \end{bmatrix} \tag{2.13}$$

What to do with a Markov chain?

First, how do we find $\Pr(x_{t+k} = \omega_j | x_t = \omega_i)$?

$$\Pr(x_{t+k} = \omega_j | x_t = \omega_i) = \left(P^k\right)_{i,j} \tag{2.14}$$

How do we find the unconditional probability distribution of x_t?

$$\Pr(x_t = \omega_j) = \left(\pi_0' P^t\right)_j \tag{2.15}$$

Another way of writing this is:

$$\pi_t' = \pi_0' P^t \tag{2.16}$$

A quantity that we will often be interested in is the limit of the equation above.

$$\begin{aligned} \pi_\infty &\equiv \lim_{t\to\infty} \pi_t \\ &= \lim_{t\to\infty} \left(\pi_0' P^t\right) \end{aligned} \tag{2.17}$$

This definition, of course, presumes that such a limit exists. There are P such that this limit does not exist, for example

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2.18}$$

Think of π_∞ as providing a measure of the behavior of the system after it has been running for a long time. Notice that the initial condition π_0 appears here. When π_∞ is the same for every π_0, we say that the Markov chain is "asymptotically stationary."

Define a stationary distribution of a given Markov chain as a probability distribution π such that:

$$\pi' = \pi' P \tag{2.19}$$

In other words if the initial distribution is π, the unconditional distribution at time 1 will be π as well. This implies the unconditional distribution at any time t will be π. A stationary distribution is also called an invariant measure or invariant distribution. Finding a stationary distribution is actually fairly easy, it's just a matter of solving a system of linear equations.

We can prove that there will always be at least one stationary distribution for any transition matrix. When is there exactly one?

Theorem 1 *Let P be a Markov transition matrix such that there exists $n \geq 1$ such that $(P^n)_{i,j} > 0$ for all i, j. Then P is asymptotically stationary and has a unique stationary distribution.*

An informal way of putting this condition is that all states "talk to each other." One way this can be satisfied is if every element of P is strictly positive.

Now, if the conditions of the theorem above hold, then for any initial distribution π_0:

$$\lim_{t \to \infty} \pi_t = \pi \tag{2.20}$$

where π solves equation (2.19).

So, if we are looking at a well-behaved macro model whose equilibrium can be expressed as a Markov chain, we can do the following:

1. Check to see that the conditions of the theorem above are satisfied.
2. If so,
    (a) Solve equation (2.19) to get at the long-run or stationary probability distribution over states.

(b) Use the stationary distribution combined with the transition matrix and the state space to calculate means, variances, covariances, autocovariances, etc. for the model.

3. If not, go read more about Markov chains. An introductory textbook on stochastic processes (e.g. Hoel, Port, and Stone) is the place to look.

Forecasting

Note that:

$$E(x_t) = \pi_t' \omega \tag{2.21}$$

Also, we can forecast;

$$E(x_{t+k}|x_t = \omega_i) = \left(P^k \omega\right)_i \tag{2.22}$$

2.1.3 Linear stochastic processes

We will go over linear stochastic processes in scalar form. The textbook generalizes to vectors.

Definition 2.1.3 (White noise) *Define a stochastic process ϵ_t with the following properties:*

$$\begin{aligned} E(\epsilon_t) &= 0 && (2.23)\\ E(\epsilon_t^2) &= \sigma_\epsilon^2 && (2.24)\\ E(\epsilon_t \epsilon_s) &= 0 \qquad \text{if } t \neq s && (2.25) \end{aligned}$$

A process that satisfies these conditions is known as "white noise".

Definition 2.1.4 (Moving average (MA) process) *The stochastic process x_t is a moving average process of order k (usually abbreviated $MA(k)$) if:*

$$x_t = \epsilon_t + c_1 \epsilon_{t-1} + c_2 \epsilon_{t-2} + \ldots + c_k \epsilon_{t-k} \tag{2.26}$$

where $\{c_1, c_2, \ldots, c_k\}$ is a sequence of real numbers and ϵ_t is a white noise process.

An MA process can be rewritten as:

$$x_t = \epsilon_t + \sum_{i=1}^{k} c_i \epsilon_{t-i} \tag{2.27}$$

We can also construct an infinite-order moving average process $MA(\infty)$:

$$x_t = \epsilon_t + \sum_{i=1}^{\infty} c_i \epsilon t - i \tag{2.28}$$

Definition 2.1.5 (Autoregressive (AR) process) *The stochastic process x_t is an autoregressive process of order k, or an $AR(k)$ process, if:*

$$x_t = \epsilon_t + \sum_{i=1}^{k} a_i x_{t-i} \tag{2.29}$$

where a_i are constants, and ϵ_t is a white noise process.

2.1.4 Moments of linear processes

Let x_t be an $MA(k)$ process.

$$\begin{aligned} E(x_t) &= E\left(\sum_{i=0}^{k} c_i \epsilon_{t-i}\right) && (2.30)\\ &= \sum_{i=0}^{k} c_i E(\epsilon_{t-i}) \\ &= 0 \\ E(x_t x_{t+s}) &= E\left[\left(\sum_{i=0}^{k} c_i \epsilon_{t-i}\right)\left(\sum_{i=0}^{k} c_i \epsilon_{t+s-i}\right)\right] && (2.31)\\ &= \sum_{i=0}^{k}\sum_{j=0}^{k} c_i c_j E\left(\epsilon_{t-i}\epsilon_{t+s-j}\right) \\ &= \begin{cases} \sigma_\epsilon^2 \sum_{i=0}^{k-s} c_i c_{i+s} & \text{if } s \le k \\ 0 & \text{otherwise} \end{cases} \\ var(x_t) = E(x_t^2) &= \sigma_\epsilon^2 \sum_{i=0}^{k} c_i^2 && (2.32) \end{aligned}$$

Alternatively, if x_t is an $MA(\infty)$ process, then we have $E(x_t) = 0$ and $var(x_t) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} c_i^2$. In other words, an $MA(\infty)$ process will have finite variance if and only if $\sum_{i=0}^{\infty} c_i^2$ is finite.

An autoregressive process is a little more complicated. For our purposes it will be enough to consider an $AR(1)$ process. Letting $x_t = ax_{t-1} + \epsilon_t$, where ϵ_t is white noise, we have:

$$\begin{aligned} x_t &= a(ax_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= a(a(ax_{t-3} + \epsilon_{t-2}) + \epsilon_{t-1}) + \epsilon_t \\ &= a^s x_{t-s} + \sum_{i=0}^{s-1} a^i \epsilon_{t-i} \\ &= \sum_{i=0}^{\infty} a^i \epsilon_{t-i} \quad \text{if } \lim_{s\to\infty} a^s x_{t+s} = 0 \text{ a.s.} \end{aligned} \tag{2.33}$$

for any $s \geq 0$ (actually, any s). Taking expectations, we have $E(x_t) = 0$. To calculate covariances, we get:

$$\begin{aligned} E(x_t x_{t+s}) &= E\left[\left(\sum_{i=0}^{\infty} a^i \epsilon_{t-i}\right)\left(\sum_{j=0}^{\infty} a^j \epsilon_{t+s-j}\right)\right] \\ &= \sigma_\epsilon^2 a^s \sum_{i=0}^{\infty} a^{2i} \end{aligned} \tag{2.34}$$

$$var(x_t) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} a^{2i} \tag{2.35}$$

Now, in order for $var(x_t)$ to be finite, we need $\sum_{i=0}^{\infty} a^{2i}$ to be finite.

Here is a result we will use a lot: If $|a| < 1$ then $\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}$. Otherwise $\sum_{i=0}^{\infty} a^i$ does not exist (i.e., is not finite).

So, for an AR(1) process with $|a| < 1$:

$$E(x_t x_{t+s}) = \frac{a^s}{1-a^2}\sigma_\epsilon^2 \tag{2.36}$$

Notice that one example of an AR(1) process without a finite variance is the "random walk" $x_t = x_{t-1} + \epsilon_t$.

Stationarity

Definition 2.1.6 (Stationary stochastic process) *A (strictly) stationary stochastic process is one where the joint distribution of x_t and x_s depends only on $|t-s|$.*

This is a fairly strong restriction, we will use a weaker one.

Definition 2.1.7 (Covariance-stationary stochastic process) *A second-order stationary or a covariance-stationary process is one in which:*

$$\begin{aligned} E(x_t) = E(x_s) &= \mu \\ cov(x_t, x_s) = E[(x_t - \mu)(x_s - \mu)] &= \sigma(t-s) \\ var(x_t) = \sigma(0) < \infty & \end{aligned}$$

For the rest of the course we will use "stationary" as a shorthand for "covariance stationary." σ is often called the *autocovariance function.* We can normalize a stationary process so that $\mu = 0$, and drop it from now on.

Why do we care about stationary processes?

- Many standard results in econometrics (e.g., the Central Limit Theorem) require stationarity. Nonstationarity will introduce numerous complications; estimators and test statistics will often have nonstandard asymptotic distributions.
- It will make our models easier to work with.

First, we note the following results which can be proved by inspection:

1. A white noise process is stationary.
2. If x_t and y_t are stationary, then $ax_t + by_t$ is stationary for any (finite) a and b.
3. If x_t is stationary and y_t is nonstationary, then $ax_t + by_t$ is nonstationary for any $b \neq 0$.

4. If x_t and y_t are both nonstationary, then there may or may not be an a and a b such that $ax_t + by_t$ is stationary. For a trivial example, note that $x_t - x_t$ is stationary.
5. Therefore, for finite k, any $MA(k)$ process is stationary.
6. An $MA(\infty)$ process is stationary if $\sum_{i=0}^{\infty} c_i^2$ exists (is finite).
7. An $AR(1)$ process is stationary if and only if $|a| < 1$.

Theorem 2 (Wold representation theorem) *Any covariance-stationary stochastic process can be represented as the sum of a stationary $MA(\infty)$ process and a "deterministic process".*

The part about the deterministic process is something we can ignore here because it will never come up. In other words, any stationary process has a deeply linear structure - its a linear combination of white noise processes.

Prediction

Suppose that x_t is an MA process. How would we predict x_t given information at time $t - s$? Now:

$$x_t = \sum_{i=0}^{k} c_i \epsilon_{t-i} \tag{2.37}$$

Taking E_{t-s} of both sizes, we get:

$$E_{t-s}(x_t) = \sum_{i=0}^{k} c_i E_{t-s}(\epsilon_{t-i}) \tag{2.38}$$

$$= \begin{cases} x_t & s \leq 0 \\ \sum_{i=s}^{k} c_i \epsilon_{t-i} & 0 < s \leq k \\ 0 & s > k \end{cases} \tag{2.39}$$

Notice that a linear stochastic process built with white noise gives us a handy feature: the value of $E_t(\epsilon_s)$ is either zero or ϵ_s.

Now consider x_t be an AR(1) process. We already established that

$$x_t = a^s x_{t-s} + \sum_{i=0}^{s-1} a^i \epsilon_{t-i} \tag{2.40}$$

So $E_{t-s}(x_t) = a^s x_{t-s}$

2.1.5 Nonstationary processes

One of the most common nonstationary stochastic processes is the random walk:

$$x_t = x_{t-1} + \epsilon_t \tag{2.41}$$

where ϵ_t is white noise.

A random walk is a special case of a martingale:

Definition 2.1.8 (Martingale) *A martingale is a stochastic process with the property:*

$$E_t(x_{t+1}) = x_t \tag{2.42}$$

As an aside, the law of iterated expectations imply that any optimal prediction is a martingale. Specifically, let y be any random variable, and let the stochastic process x_t be defined as:

$$x_t \equiv E(y|\Omega_t) \tag{2.43}$$

where Ω_t is a sequence of conditioning variables such that $\Omega_{t-1} \subset \Omega_t$. Then:

$$E_t(x_{t+1}) = x_t \tag{2.44}$$

Definition 2.1.9 (Martingale difference sequence) *A martingale difference sequence is any stochastic process ϵ_t with the property:*

$$E_t(\epsilon_{t+1}) = 0 \tag{2.45}$$

White noise is an example of a martingale difference sequence.

Random walks and martingales are found all over macroeconomics. A simple model of stock prices implies they follow a random walk. A simple version of the permanent income hypothesis implies that the growth rate of consumption follows a martingale.

It's easy to show that a random walk is not stationary, but that its first difference $x_t - x_{t-1}$ is stationary. We call a process with this property *integrated of order 1* or $I(1)$, or we say that x_t has a *unit root*. A stationary process is sometimes called an $I(0)$ process.

2.1.6 Vector stochastic processes

Suppose that x_t is a scalar random variable that follows an $AR(3)$ process:

$$y_t = \epsilon_t + \sum_{i=1}^{3} a_i y_{t-i} \tag{2.46}$$

We can create a vector X_t that follows an $AR(1)$ process.

$$X_t \equiv \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{bmatrix} \tag{2.47}$$

The stochastic process can then be written

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ y_{t-4} \end{bmatrix} + \begin{bmatrix} \sigma_\epsilon \\ 0 \\ 0 \\ 0 \end{bmatrix} w_t \tag{2.48}$$

(where $w_t = \epsilon_t/\sigma_\epsilon$) or

$$X_t = A_0 X_{t-1} + C w_t \tag{2.49}$$

This is the format for a linear stochastic process in vectors used by Sargent, where X_t is a n by 1 vector, A_0 is an $n \times n$ matrix. The $m \times 1$ vector (notice there is no requirement that $m = n$) of shocks w_t is "white noise" which Sargent defines as meeting the conditions:

$$\begin{aligned} E_t(w_{t+1}) &= 0 \\ E_t(w_{t+1} w'_{t+1}) &= I \end{aligned} \tag{2.50}$$

The $n \times m$ matrix takes these orthogonal, normalized (unit variance) shocks, and converts them into shocks which are correlated and have whatever variance you want. By a little linear algebra, the covariance matrix of Cw_{t+1} is

$$E\left(Cw_{t+1}(Cw_{t+1})'\right) = CE(w_{t+1}w'_{t+1})C' = CC' \tag{2.51}$$

As an aside, note that any symmetric, positive definite matrix A has a Cholesky decomposition, an upper triangular matrix U such that $A = U'U$.

So we can set up this model to produce any covariance matrix we want for these shocks by simply taking the transpose of its Cholesky decomposition.

Sargent goes over how to analyze linear stochastic processes for vectors. Most of the ideas associated with scalar processes carry over in vector language; for example the scalar AR(1) process is stationary if and only if $|a| < 1$, while the vector process is stationary if and only if the real parts of the eigenvalues of A_0 are less than one.