LECTURES 6

Markov chains

Macroeconomics 4

A.Y. 2014-15

Markov chains

- A time-invariant, discrete-state Markov chain is characterized by:
  - An n-dimensional state space $S = \{\mathrm{s}_1, \mathrm{s}_2, ..., \mathrm{s}_n\}$.
  - A $n \times n$ non-negative *transition matrix* Π, such that $\sum_{j=1}^{n} \Pi_{ij} = 1$ for $i = 1, 2, ..., n$.
  - A $n \times 1$ non-negative vector π_0, such that $\sum_{i=1}^{n} \pi_{0,i} = 1$, representing the initial (unconditional) probability distribution on s_0:

$$\pi_{0,i} = \mathrm{prob}\left(s_0 = \mathrm{s}_i\right).$$

- The matrix Π is a *right stochastic matrix*, and records the transition prob. from state i into state j:

$$\Pi_{ij} = \mathrm{prob}\left(s_{t+1} = \mathrm{s}_j | s_t = \mathrm{s}_i\right).$$

Markov chains

- Note that:

$$
\begin{aligned}
&\text{prob}\left(s_{t+2} = \mathrm{s}_j | s_t = \mathrm{s}_i\right) \\
&\quad = \sum_{m=1}^{n} \text{prob}\left(s_{t+2} = \mathrm{s}_j | s_{t+1} = \mathrm{s}_m\right) \text{prob}\left(s_{t+1} = \mathrm{s}_m | s_t = \mathrm{s}_i\right) \\
&\quad = \sum_{m=1}^{n} \Pi_{im}\Pi_{mj} = \Pi_{ij}^{(2)},
\end{aligned}
$$

where $\Pi_{ij}^{(2)}$ is the (i, j) element of $\Pi^{(2)}$.

- Hence, in general:

$$
\text{prob}\left(s_{t+k} = \mathrm{s}_j | s_t = \mathrm{s}_i\right) = \Pi_{ij}^{(k)}.
$$

Markov chains

Definition

State i **communicates** with state j if $\Pi_{ij}^{(k)} > 0$ and $\Pi_{ji}^{(k)} > 0$ for some $k \geq 1$. A Markov chain is said to be **irreducible** if every pair (i, j) communicate.

- An irred. Markov chain has the property that it is possible to move from any state to any other in a **countable** number of periods.
- Note that it is not required that this movement is possible in one step, so $\Pi_{ij} = 0$ is permitted.

Markov chains

Definition

The **unconditional distribution** of s_t is given by:

$$\pi_t = \left(\Pi'\right)^t \pi_0,$$

where π_{it} =Prob$(s_t = \mathrm{s}_i)$.

- Note that:

$$\pi_{t+1} = \Pi' \pi_t.$$

- Trivially:

$$\mathbb{E}\left(s_{t+1} \mid s_t = \mathrm{s}_i\right) = \sum_{j=1}^{n} \Pi_{ij} s_j,$$

$$\mathrm{var}\left(s_{t+1} \mid s_t = \mathrm{s}_i\right) = \sum_{j=1}^{n} \Pi_{ij} s_j^2 - \left(\sum_{j=1}^{n} \Pi_{ij} s_j\right)^2.$$

Markov chains

Definition

An unconditional dist. is called **stationary (ergodic)** if it remains constant over time, and satisfies:

$$\pi = \Pi' \pi.$$

- The ergodic dist. can be interpreted in two ways:
  - π_i is the unconditional prob. that the chain is currently in state i,
  - π_i is the prob. that the chain will be in state i in t steps as $t \to \infty$.
- Again, the unconditional moments obtain as:

$$\mathbb{E}(s) = \sum_{j=1}^{n} \pi_j s_j,$$

$$\text{var}(s) = \sum_{j=1}^{n} \pi_j s_j^2 - \left(\sum_{j=1}^{n} \pi_j s_j \right)^2.$$

Markov chains

- The stationarity cond. can be rewritten as:

$$\left(\mathbf{1} - \Pi'\right) \pi = \mathbf{0}.$$

- In other words, π is just an eigenvector associated with a unit eigenvalue of Π', pinned down by the normalization $\sum_{j=1}^{n} \pi_j = 1$.

- The matrix Π is **right stochastic**, i.e. has non-negative elements and rows that sum up to one; this implies that:
  - Π' has at least one (possibly more) unit eigenvalue.
  - There is at least one (again, possibly more) eigenvector satisfying the stat. condition.

Markov chains

Definition

If there is one and only one vector π that satisfies the stat. condition, and:

$$\lim_{t\to\infty} \pi_t = \pi,$$

for all possible π_0, then the Markov chain is **asy. stationary with a unique invariant (ergodic) distribution**.

Theorem

Let Π *be a right stochastic matrix such that* $\Pi_{ij} > 0$ *for all* (i, j)*: the associated Markov chain is irreducible, asy. stationary and has a unique stationary distribution.*

Markov chains

- From a numerical point of view, there are several ways to calculate π given Π.

1. Iterate until convergence on:

$$\pi_{k+1} = \Pi' \pi_k.$$

2. Calculate the eigenvalues and eigenvectors of Π' and take the normalized eigenvector associated to $\lambda = 1$:

$$\pi = \frac{v_1}{\sum_{i=1}^{n} v_{1i}}.$$

Markov chains

3. Define:

$$\hat{\mathbf{A}} \equiv \left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}',$$

where:

$$\underset{(n+1)\times n}{\mathbf{A}} \equiv \begin{bmatrix} \mathbf{I}_n - \Pi' \\ \mathbf{1}_n' \end{bmatrix},$$

and $\mathbf{1}_n$ is a $n \times 1$ vector of ones; it turns out that π is equal to the $n+1$ column of $\hat{\mathbf{A}}$.

4. Note that $\mathbf{1}_{n\times n}\pi = \mathbf{1}_n$, where $\mathbf{1}_{n\times n}$ is a $n \times n$ matrix of ones, since π sums to one. Hence:

$$\mathbf{1}_n = \pi - \Pi'\pi + \mathbf{1}_{n\times n}\pi = \left(\mathbf{I}_n - \Pi' + \mathbf{1}_{n\times n}\right)\pi.$$

This implies that:

$$\pi = \left(\mathbf{I}_n - \Pi' + \mathbf{1}_{n\times n}\right)^{-1}\mathbf{1}_n.$$

Discretizing AR processes

- Consider the following stationary *AR(1)* process:

$$z' = (1 - \rho)\,\mu_z + \rho z + \epsilon,$$

where $|\rho| < 1$ and $\epsilon \sim N\left(0, \sigma_\epsilon^2\right)$.

- Quite often, in Macro we are required to compute expectations of the form:

$$\mathbb{E}\left[V\left(a', z'\right) \mid z\right] = \int_{-\infty}^{+\infty} V\left(a', z'\right) f\left(z' \mid z\right) dz'.$$

- It would be numerically very convenient if we could discretize the continuous distribution of z.

Discretizing AR processes

- In other words, we will approx. the continuous *AR(1)* with a discrete Markov chain, say z with some abuse of notation, that:
  - takes values in a finite set $\mathcal{Z} = \{z_1, z_2, ..., z_n\}$,
  - is characterized by a trans. matrix Π.
- This approx. allows to easily compute the previous expectation as:

$$\mathbb{E}\left[V\left(a', z'\right) \mid \mathrm{z}_i\right] = \sum_{j=1}^{n} \Pi_{ij} V\left(a', \mathrm{z}_j\right).$$

Discretizing AR processes

- To develop an intuition, consider the approach of Tauchen (1986).
- Note that, conditionally on z, $z' \sim N\left(\mu', \sigma_\epsilon^2\right)$, where $\mu' \equiv (1-\rho)\,\mu_z + \rho z$.
- Unconditionally, instead, $\mathbb{E}\left(z\right) = \mu_z$ and $\text{var}\left(z\right) = \sigma_z^2 = \sigma_\epsilon^2 / \left(1-\rho\right)$.
- The first step requires to select the finite set $\mathcal{Z}$; assume that $\mathrm{z}_1 < \mathrm{z}_2 < ... < \mathrm{z}_n$.
- z_1 and z_n are set, respectively, to m uncond. std. dev. on either side of μ_z, and the other z_j are spread uniformly over the interval:

$$\mathrm{z}_1 = \mu_z - m\sigma_z, \quad \mathrm{z}_n = \mu_z + m\sigma_z, \quad \mathrm{z}_j = \mathrm{z}_1 + \frac{\mathrm{z}_n - \mathrm{z}_1}{n-1}\left(j-1\right).$$

Discretizing AR processes

- The trans. matrix Π is chosen to match the prob. of moving from z_i into the interval:

$$\left(\mathrm{z}_j - \mathrm{w}_\mathrm{z}, \mathrm{z}_j + \mathrm{w}_\mathrm{z}\right],$$

where $\mathrm{w}_\mathrm{z} \equiv \frac{1}{2}\frac{\mathrm{z}_n - \mathrm{z}_1}{n-1}$.

- More precisely, we set:

$$\begin{cases} \Pi_{i1} & = \mathrm{prob}\left(\mu_i' + \epsilon \leq \mathrm{z}_1 + \mathrm{w}_\mathrm{z}\right), \\ \Pi_{in} & = 1 - \mathrm{prob}\left(\mathrm{z}_n - \mathrm{w}_\mathrm{z} < \mu_i' + \epsilon\right), \\ \Pi_{ij} & = \mathrm{prob}\left(z_j - \mathrm{w}_\mathrm{z} < \mu_i' + \epsilon \leq z_j + \mathrm{w}_\mathrm{z}\right), \end{cases}$$

where $\mu_i' \equiv (1-\rho)\,\mu_z + \rho z_i$.

Discretizing AR processes

- Operationally, the previous rule boils down to:

$$\Pi_{ij} = \begin{cases} \Phi\left(\frac{\mathrm{z}_1+\mathrm{w}_\mathrm{z}-\mu_i'}{\sigma}\right) & \text{for } j=1, \\ \Phi\left(\frac{\mathrm{z}_j+\mathrm{w}_\mathrm{z}-\mu_i'}{\sigma}\right) - \Phi\left(\frac{\mathrm{z}_j-\mathrm{w}_\mathrm{z}-\mu_i'}{\sigma}\right) & \text{for } 1<j<n, \\ 1-\Phi\left(\frac{\mathrm{z}_n-\mathrm{w}_\mathrm{z}-\mu_i'}{\sigma}\right) & \text{for } j=n. \end{cases}$$

  where Φ is the *CDF* of the standard normal.

- Note that $m = 1$ would cover about 67% of the support of the uncond. dist., $m = 2$ about 96%, and $m = 3$ about 99%; the larger m, the larger n has to be to approx. well the conditional moments.

- Tauchen (1986) suggests that rarely $n > 9$ is needed, but notes that the quality of the approx. decreases sharply when $|\rho| \to 1$.

Discretizing AR processes

- Tauchen and Hussey (1991) improve on Tauchen (1986) by taking advantage of *Gauss-Hermite (G-H) quadrature*.

- Suppose that $z \sim N\left(\mu_z, \sigma_z^2\right)$, i.e. no AR structure.

- Using *G-H* quadrature, we can approx. the expectation as:

$$\mathbb{E}\left[V\left(a', z'\right) \mid z\right] \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^{n} \omega_j V\left(a', \bar{\mathrm{z}}_j\right),$$

  where $\bar{\mathrm{z}}_j \equiv \sqrt{2}\sigma_z h_j + \mu_z$, $\{h_j\}_{j=1}^n$ and $\{\omega_j\}_{j=1}^n$ are, respectively, the G-H nodes (i.e. the roots of the nth order *Hermite polynomial*) and weights.

- If z follows an *AR(1)* process, this would be extremely cumbersome because μ_z, the cond. mean of z', depends on z.

Discretizing AR processes

- Note that:

$$\mathbb{E}\left[V\left(a', z'\right) \mid z\right] = \int_{-\infty}^{+\infty} V\left(a', z'\right) \frac{f\left(z' \mid z\right)}{f\left(z' \mid \mu_z\right)} f\left(z' \mid \mu_z\right) dz',$$

where $f\left(z' \mid \mu_z\right)$ is the density of z' cond. on $z = \mu_z$.

- The approx. becomes the following:

$$\mathbb{E}\left[V\left(a', z'\right) \mid z\right] \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^{n} \omega_j V\left(a', \overline{\mathrm{z}}_j\right) \frac{f\left(\overline{\mathrm{z}}_j \mid z\right)}{f\left(\overline{\mathrm{z}}_j \mid \mu_z\right)}.$$

- For $z = \overline{\mathrm{z}}_i$, we have:

$$\mathbb{E}\left[V\left(a', z'\right) \mid \overline{\mathrm{z}}_i\right] \approx \sum_{j=1}^{n} \bar{\omega}_{i,j} V\left(a', \overline{\mathrm{z}}_j\right),$$

where $\bar{\omega}_{i,j} \equiv \frac{1}{\sqrt{\pi}} \omega_j \frac{f(\overline{\mathrm{z}}_j | \overline{\mathrm{z}}_i)}{f(\overline{\mathrm{z}}_j | \mu_z)}$.

Discretizing AR processes

- The previous discussion suggests that the possible realizations of the discrete Markov chain should be $\mathcal{Z} = \{\bar{\mathrm{z}}_j\}_{j=1}^n$.

- We cannot use the $\bar{\omega}_{i,j}$ directly in the trans. matrix, because $\sum_{j=1}^n \bar{\omega}_{i,j} \neq 1$. Instead, set:

$$\Pi_{ij} = \frac{\bar{\omega}_{i,j}}{\sum_{j=1}^n \bar{\omega}_{i,j}}.$$

- Hence:

$$\mathbb{E}\left[V\left(a', z'\right) \mid \bar{\mathrm{z}}_i\right] \approx \sum_{j=1}^n \Pi_{ij} V\left(a', \bar{\mathrm{z}}_j\right).$$

- Tauchen and Hussey (1991) show that $\lim_{n\to\infty} \sum_{j=1}^n \bar{\omega}_{i,j} = 1$ for all i, so the prev. approx. converges to G-H approx. when $n \to \infty$.

Discretizing AR processes

- For highly persistent processes (say $|\rho| > 0.9$), Kopecky and Suen (2010) advocate Rouwenhorst's method.
- The set $\mathcal{Z}$ consists of n points which are symmetrically and evenly spaced over the interval $[\mu_z - \upsilon, \mu_z + \upsilon]$, for some $\upsilon > 0$.
- Choose some p and q in the $(0, 1)$ interval, with possibly $p = q$.
- For $n = 2$, the trans. matrix becomes:

$$\Pi_2 = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}.$$

Discretizing AR processes

- Π_n for $n \geq 3$ obtains recursively as follows:
  - 1) Compute:

$$\hat{\Pi}_n = p \begin{bmatrix} \Pi_{n-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} + (1-p) \begin{bmatrix} \mathbf{0} & \Pi_{n-1} \\ 0 & \mathbf{0}^T \end{bmatrix} + (1-q) \begin{bmatrix} \mathbf{0}^T & 0 \\ \Pi_{n-1} & \mathbf{0} \end{bmatrix} + q \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \Pi_{n-1} \end{bmatrix},$$

    where $\mathbf{0}$ is a $n-1$ col. vector of zeros.
  - 2) Divide all but the top and bottom rows of $\hat{\Pi}_n$ by 2 to obtain Π_n.

Discretizing AR processes

- Note that setting p different from q would introduce conditional heteroscedasticity in the shocks.

- Regardless of n and $\mathcal{Z}$, the autocor. of this Markov chain is equal to $p+q-1$; hence, assuming conditional homosced., so that $p=q$, we set:
$$p = \frac{1+\rho}{2},$$
in order to replicate the persistence of the continuous process.

- The var. of the Markov chain is simply $\upsilon^2/(n-1)$; hence, for a given n, in order to replicate the var. of the continuous process we set:
$$\upsilon = \sqrt{\frac{n-1}{1-\rho^2}}\sigma_\epsilon.$$

Merging Markov chains

- Suppose that households are subject to two **independent** idiosyncratic shocks, say shocks to labor productivity and preferences.
- The processes are approx. by two discrete Markov chains, $s_t \in \{\mathrm{s}_1, \mathrm{s}_2\}$ and $z_t \in \{\mathrm{z}_1, \mathrm{z}_2\}$, characterized respectively by transition matrices $\mathbf{\Pi}$ and $\mathbf{\Xi}$.
- The joint process can be represented a single Markov chain: construct $q_t = (s_t, z_t) \in \{(\mathrm{s}_1, \mathrm{z}_1), (\mathrm{s}_1, \mathrm{z}_2), (\mathrm{s}_2, \mathrm{z}_1), (\mathrm{s}_2, \mathrm{z}_2)\}$.
- The transition matrix for the joint process becomes:

$$\mathbf{\Lambda} = \begin{bmatrix} \Pi_{11}\Xi_{11} & \Pi_{11}\Xi_{12} & \Pi_{12}\Xi_{11} & \Pi_{12}\Xi_{12} \\ \Pi_{11}\Xi_{21} & \Pi_{11}\Xi_{22} & \Pi_{12}\Xi_{21} & \Pi_{12}\Xi_{22} \\ \Pi_{21}\Xi_{11} & \Pi_{21}\Xi_{12} & \Pi_{22}\Xi_{11} & \Pi_{22}\Xi_{12} \\ \Pi_{21}\Xi_{21} & \Pi_{21}\Xi_{22} & \Pi_{22}\Xi_{21} & \Pi_{22}\Xi_{22} \end{bmatrix}.$$

References I

Kopecky, K. and R. Suen (2010, July). Finite State Markov-chain Approximations to Highly Persistent Processes. *Review of Economic Dynamics 13*(3), 701–14.

Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters 20*(2), 177–181.

Tauchen, G. and R. Hussey (1991, March). Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models. *Econometrica 59*(2), 371–96.