LECTURE NOTES

ECO 613/614

FALL 2007

KAREN A. KOPECKY

1 Discretizing AR(1)'s

When agent's face uncertainty about the future, solving the agent's maximization problem involves computing a conditional expectation. For example consider the following consumption–savings problem:

$$V(a,\lambda) = \max_{c,a'}\{u(c) + \beta \mathsf{E}[V(a',\lambda')|\lambda]\}$$

subject to

$$a(1+r) + w\exp(\lambda) = a' + c,$$

$$c \geq 0, \quad a' \geq 0,$$

$$\lambda' = (1-\rho)\mu_\lambda + \rho\lambda + \varepsilon, \quad \varepsilon \sim N(0,\sigma_\varepsilon^2)$$

Notice that computing the conditional expectation requires computing the integral of the value function weighted by $f(\lambda'|\lambda)$, the conditional density of λ' given λ:

$$\mathsf{E}[V(a',\lambda')|\lambda] = \int_{-\infty}^{\infty} V(a',\lambda') f(\lambda'|\lambda) d\lambda'.$$

This infinite-dimensional object is a function of the current state. We have to decide how to approximate it numerically.

One approach is to discretize the state and approximate the conditional expectation by the finite set of its values at each point in the state space. This technique involves replacing the continuous-valued Markov chain for λ with a finitely-many-discrete-valued Markov chain. Let the new Markov chain be represented by $\tilde{\lambda}$ and the values it can take be members of the finite set $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. Let the elements of the probability transition matrix, P, associated with $\tilde{\lambda}$ be denoted by $p_{i,j}$ for $i,j = 1,\ldots,n$ where

$$p_{i,j} = \textbf{Prob}(\tilde{\lambda}' = \lambda_j | \tilde{\lambda} = \lambda_i).$$

Now computing the conditional expectation no longer involves an integral since

$$\mathsf{E}[V(a', \tilde{\lambda}')|\lambda_i] = \sum_{j=1}^{n} p_{i,j} V(a', \lambda'_j).$$

But how should be choose the set Λ and the elements of P?

1.1 Tauchen (86)

In 1986 Tauchen proposed a method based on the fact that conditional on λ, λ' is normally distributed with mean $(1-\rho)\mu_\lambda + \rho\lambda$ and standard deviation σ_ε.

The idea is as follows. First choose the members of Λ evenly-spaced along the real line. Suppose we set them such that

$$\lambda_1 < \lambda_2 < \cdots < \lambda_n.$$

The upper and lower bounds on the range, λ_1 and λ_n, respectively, are set to m unconditional standard deviations on either side of λ's unconditional mean. I.e., we have

$$\lambda_1 = \mu_\lambda - m\sigma_\lambda$$

and

$$\lambda_n = \mu_\lambda - m\sigma_\lambda$$

where $\sigma_\lambda = \sigma_\varepsilon / \sqrt{1-\rho^2}$. Now the probability transition matrix is determined as follows. First let $w = \lambda_j - \lambda_{j-1}$. If j is between 2 and $n-1$ set

$$p_{i,j} = \textbf{Prob}[\lambda_j - w/2 \leq (1-\rho)\mu_\lambda + \rho\lambda_i + \varepsilon \leq \lambda_j + w/2],$$

if $j = 1$ then set

$$p_{i,1} = \textbf{Prob}[(1-\rho)\mu_\lambda + \rho\lambda_i + \varepsilon \leq \lambda_1 + w/2],$$

and if $j = n$ set

$$p_{i,n} = 1 - \textbf{Prob}[\lambda_n - w/2 \leq (1-\rho)\mu_\lambda + \rho\lambda_i + \varepsilon].$$

This is equivalent to setting the $p_{i,j}$'s as follows

$$p_{ij} = \begin{cases} \Phi\left(\frac{\lambda_1 + w/2 - (1-\rho)\mu_\lambda - \rho\lambda}{\sigma_\varepsilon}\right) & \text{for } j = 1, \\ \Phi\left(\frac{\lambda_j + w/2 - (1-\rho)\mu_\lambda - \rho\lambda}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\lambda_j - w/2 - (1-\rho)\mu_\lambda - \rho\lambda}{\sigma_\varepsilon}\right) & \text{for } 1 < j < n, \\ 1 - \Phi\left(\frac{\lambda_n - w/2 - (1-\rho)\mu_\lambda - \rho\lambda}{\sigma_\varepsilon}\right) & \text{for } j = n, \end{cases}$$

where Φ is the standard normal c.d.f.

Notice that this procedure makes the conditional distribution of $\tilde{\lambda}'$ given $\tilde{\lambda} = \lambda_i$ a discrete approximation to the conditional distribution of λ' given $\lambda = \lambda_i$. As the number of grid points n increases the better the approximation of the distribution becomes. Tauchen comments that the approximation is adequate for most purposes when $n = 9$. Although he notes that the quality of the approximation decreases when λ is close to unity.

1.2 Tauchen and Hussey (91)

In 1991 Tauchen and Hussey developed a method for approximating continuous processes by discrete ones using the ideas of Gauss-Hermite quadrature. Suppose for a moment that $\lambda \sim N(\mu_\lambda, \sigma_\lambda)$. In this case using Gauss-Hermite quadrature to approximate the conditional expectation of V yields

$$\mathsf{E}[V(a', \lambda')|\lambda] \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^{n} w_j V(a', \sqrt{2}\sigma_\lambda \hat{\lambda}'_j + \mu_\lambda),$$

where $\hat{\lambda}'_j, j = 1, \ldots, n$ are the roots of the nth Hermite polynomial and $w_j, j = 1, \ldots, n$ are the corresponding weights.

Now consider the same approach when λ follows the orignial AR(1) process. We want to approximate

$$\mathsf{E}[V(a', \lambda')|\lambda] = \int_{-\infty}^{\infty} V(a', \lambda') f(\lambda'|\lambda) d\lambda',$$

using Gauss-Hermite quadrature. We could approximate this expectation using the same approach as above but notice that the conditional mean of λ' is a function of λ. This means that the we have a different λ' grid for each value of λ and the set of λ's we need to iterate on the expectation can increase without bound. To avoid this problem Tauchen and Hussey suggest making the following transformation

$$\mathsf{E}[V(a', \lambda')|\lambda] = \int_{-\infty}^{\infty} V(a', \lambda') \frac{f(\lambda'|\lambda)}{f(\lambda'|\mu_\lambda)} f(\lambda'|\mu_\lambda) d\lambda',$$

where $f(\lambda'|\mu_\lambda)$ is the density of λ' conditional on λ equal to its unconditional mean, μ_λ. Now the approximation becomes

$$\mathsf{E}[V(a', \lambda')|\lambda] \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^{n} w_j V(a', \lambda_j) \frac{f(\lambda_j|\lambda)}{f(\lambda_j|\mu_\lambda)},$$

where $\lambda_j = \sqrt{2}\sigma_\varepsilon \hat{\lambda}_j + \mu_\lambda$, the variables $\hat{\lambda}_j, j = 1, \ldots, n$ are the roots of the nth Hermite polynomial, and $w_j, j = 1, \ldots, n$ are the corresponding weights. Then for $\lambda = \lambda_i$ we have

$$\mathsf{E}[V(a', \lambda')|\lambda_i] \approx \sum_{j=1}^{n} \tilde{w}_{i,j} V(a', \lambda_j),$$

where

$$\tilde{w}_{i,j} = \frac{1}{\sqrt{\pi}} w_j \frac{f(\lambda_j|\lambda_i)}{f(\lambda_j|\mu_\lambda)}$$

This suggest choosing the set of possible realizations of $\tilde{\lambda}$ to be

$$\Lambda = \{\lambda | \lambda = \sqrt{2}\sigma_\varepsilon \hat{\lambda}_j + \mu_\lambda, j = 1, \ldots, n\}$$

and choosing the elements of P to be the the $\tilde{w}$'s, however,

$$\sum_{j=1}^{n} \tilde{w}_{i,j} \neq 1.$$

So instead we set

$$p_{i,j} = \frac{\tilde{w}_{i,j}}{\sum_{j=1}^{n} \tilde{w}_{i,j}}.$$

Now

$$E[V(a', \tilde{\lambda}')|\lambda_i] = \sum_{j=1}^{n} p_{i,j} V(a', \lambda_j) \tag{1}$$

Note that because of this normalization we are no longer doing Gauss-Hermite quadrature when we compute the conditional expectation. It can be shown, however, that in the limit as n goes to infinity $\sum_{j=1}^{n} \tilde{w}_{i,j}$ goes to one for all i and therefore the approximation given by computing (1) converges to the approximation given by the Gauss-Hermite quadrature formula.

Why use as the weighting function $\omega(\lambda') = f(\lambda'|\mu_\lambda)$? A priori, another reasonable choice for the weighting function is $f(\lambda)$, the unconditional density. The conditional density, $f(\lambda'|\mu_\lambda)$ puts more weight on the central part of the distribution and less weight in the tails then $f(\lambda)$. Hence it balances two conflicting criteria. First good approximations put a lot of weight near the unconditional mean. Second, good approximations are ones in which the ratio $f(\lambda'|\lambda)/\omega(\lambda')$ is well behaved in the tails. In other words, the ratio should not grow too fast in either tail relative to $\omega(\lambda')$.

As is the case for Tauchen (86) the quality of the approximation decreases as the persistence of the process increases. In other words, the higher the persistence the more grid points are required to obtain a discrete process that captures well the behavior of the continuous process.

1.3 Rouwenhorst (95)

For highly persistent processes ($\rho > 0.9$) I recommend using the following discretization method credited to Rouwenhorst.

Let Λ consist of n points which are symmetrically and evenly spaced over the interval $[\mu_\lambda - \nu, \mu_\lambda + \nu]$. Then P is determined as outlined below.

Choose p and q. When $n = 2$ the probability transition matrix is

$$P_2 = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}$$

and when P is 3 it is

$$P_3 = \begin{bmatrix} p^2 & 2p(1-p) & (1-p)^2 \\ p(1-q) & pq + (1-p)(1-q) & q(1-p) \\ (1-q)^2 & 2q(1-q) & q^2 \end{bmatrix}$$

The P_n matrix can be derived recursively from P_{n-1} as follows. First add the $(n \times n)$ matrices

$$p \begin{bmatrix} P_{n-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} + (1-p) \begin{bmatrix} \mathbf{0} & P_{n-1} \\ 0 & \mathbf{0}' \end{bmatrix} \tag{2}$$

$$(1-q) \begin{bmatrix} \mathbf{0}' & 0 \\ P_{n-1} & \mathbf{0} \end{bmatrix} + q \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & P_{n-1} \end{bmatrix} \tag{3}$$

where $\mathbf{0}$ is a $(n-1)$ column vector and $\mathbf{0}'$ is a $(n-1)$ row vector. Then divide all but the top and bottom rows by two so that the conditional probabilities sum to one.

Notice that p^{n-1} is the probability of staying in the lowest state once you are already there while q^{n-1} is the probability of staying in the highest state once you are there. Conversely, $(1-p)^{n-1}$ is the probability of transiting from the lowest to the highest state and $(1-q)^{n-1}$ is the probability of transiting from the highest to the lowest. Note that setting p

different from q will introduce conditional heteroscedasticity in the shocks. It can be shown that regardless of the choice of n and Λ the first-order serial correlation of this process for $\tilde{\lambda}$ will always be $p + q - 1$. Hence for the case of $p = q = \pi$, the value of π can be chosen so that the discrete process has the same first-order persistence as the continuous process.

It can also be shown that the variance of $\tilde{\lambda}$ is $\nu^2/(n-1)$. Hence for a given n a good choice of ν is the value such that the variance of $\tilde{\lambda}$ and λ agree. This would mean setting

$$\nu = \sqrt{\frac{n-1}{\rho^2 - 1}}\sigma_\varepsilon.$$

1.4 Computing the Stationary Distribution

Often times we will need the stationary distribution of $\tilde{\lambda}$, π^*. Three methods for computing it are outlined below. Note that we assume that the invariant distribution exists and is unique.

The easiest way to compute π^* is to iterate on

$$\pi'_{t+1} = \pi'_t P.$$

This is done by starting at some initial distribution $\pi^{(0)}$ and iterating until the π's don't change much.

Another procedure is to use a Monte Carlo simulation. To do this one needs a uniform random number generator. Suppose we have one that generates uniformly distributed numbers in the interval $[0, 1]$. Pick a value in Λ for $\lambda^{(0)}$, say $\lambda^{(0)} = \lambda_l$. Then obtain the random number ν. Set $\lambda^{(1)} = \lambda_{\hat{l}}$ where $\hat{l}$ is the smallest integer such that

$$\nu \leq \sum_{j=1}^{\hat{l}} p_{l,j}.$$

Continue this process for T periods and set π_i^* to the relative frequency of λ_i for $i = 1, \ldots, n$. The relative frequencies will converge slowly to the stationary distribution by the Law of Large Numbers.

Finally the most direct approach is to notice that π^* is the normalized eigenvector associated with the eigenvalue 1 of the matrix $-P'$.

References

- **Burnside, Craig.** 1999. "Discrete State-Space Methods for the Study of Dynamic Economies," in *Computational Methods for the Study of Dynamic Economies.* 1999. Ramon Marimon and Andrew Scott, eds. New York: Oxford University Press.

- **Tauchen, George.** 1986. "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions" *Economic Letters.* 20: 177–181.

- **Tauchen George and Robert Hussey.** 1991. "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models." *Econometrica.* 59(2): 317–396.

- **Rouwenhorst, K. Geert.** "Asset Pricing Implications of Equilibrium Business Cycle Models," in *Frontiers of Business Cycle Research.* 1995. Thomas F. Cooley, ed. Princeton, New Jersey: Princeton University Press.