Economics Letters 20 (1986) 177–181
North-Holland

FINITE STATE MARKOV-CHAIN APPROXIMATIONS TO UNIVARIATE AND VECTOR AUTOREGRESSIONS

George TAUCHEN
Duke University, Durham, NC 27706, USA

Received 9 August 1985

The paper develops a procedure for finding a discrete-valued Markov chain whose sample paths approximate well those of a vector autoregression. The procedure has applications in those areas of economics, finance, and econometrics where approximate solutions to integral equations are required.

1. Introduction

There is interest in economics, finance, and econometrics in the solutions to functional equations where the arguments of the solution functions are the values of an autoregressive process. A typical problem is to characterize the price of an asset, where the law of motion for the dividend is a logarithmic AR(1) process. For example, with an additively separable intertemporal utility function the functional equation for the asset price p is

$$u'(h)\,p(h) = \beta \int u'(h')\left[\,p(h') + h'\right] F(\mathrm{d}h' \mid h), \tag{1}$$

where $u'(h)$ is the marginal utility of consumption, h is the dividend, β is the subjective rate of discount, and $F(h' \mid h)$ is the conditional distribution of the dividend. The law of motion for the state variable $y = \log(h)$ is $y' = \lambda y + \epsilon$, where ϵ is white noise. Under regularity conditions (1) admits a unique solution for the asset price as a function $p(y)$ of the log of the dividend. Such pricing functions are studied by Lucas (1978), Brock (1982), and others.

Generally the solutions to (1) are not available in an analytically simple closed form. Instead, the solutions are given as the limit of a sequence of computationally difficult iterations motivated by contraction mapping theorems. However, there are instances where calculation of the exact solution – or a good approximation to it – is important. For example, Prescott and Mehra (1985) examine the quantitative aspects of asset pricing in their study of the puzzle of 'high' equity returns, while Donaldson and Mehra (1984) study the qualitative features of multivariate asset pricing functions. The strategy adopted in these papers and in other work is to use a finite-state discrete Markov chain for the state variables and to restrict the number of possible values of the state variables to be very small, usually only two or three. In the discrete case the problem of solving the functional equation (1) becomes the simpler problem of just inverting a matrix.

If the range space of state variables is small, then one can find ad hoc, though presumably 'realistic' numerical values for the transition probabilities and the state variables. However, the

difficulty with using a small state space is that one is never sure that unusual and interesting characteristics of the solution are not simply artifacts of the coarseness of the state space. This suggests using finer state spaces. Of course, the size of the matrix to be inverted will become much larger, but with the large-scale computational resources that are, or will soon be available [National Science Foundation (1984)], the inversion of very large matrices is practicable.

This paper develops a method for choosing values for the state variables and the transition probabilities so that the resulting finite-state Markov chain mimics closely an underlying continuous-valued autoregression. The motivation for the method is the well-known fact that the statistical properties of many economic time series are captured adequately by vector autoregressions, after an adjustment for trend. Thus it is possible to use the method to calculate explicitly the solutions to functional equations like (1) using dynamics for the state variables that are close to those actually observed in the economy. In fact, as the state space becomes finer and finer the solution to the functional equation will, under regularity conditions, become arbitrarily close to the solution for the continuous case. This is the Kantorovich approach [Wouk (1979, pp. 120–142)] to solving functional equations. Below, evidence is presented indicating that the approximation error from the method should be small for moderate sized state spaces. Also, the method works well in an ongoing project [Tauchen (1985)] that investigates the small sample properties of generalized method of moments estimator [Hansen (1982)].

2. The scalar case

Let y_t be generated by the autoregressive scheme,

$$y_t = \lambda y_{t-1} + \epsilon_t, \tag{2}$$

where ϵ_t is a white noise process with variance σ_ϵ^2. Let the distribution function of ϵ_t be $\Pr[\epsilon_t \leqslant u] = F(u/\sigma_\epsilon)$, where F is a cumulative distribution with unit variance. Let $\tilde{y}_t$ denote the discrete-valued process that approximates the continuous-valued process (2), and let $\bar{y}^1 < \bar{y}^2, \ldots, < \bar{y}^N$ denote the values that $\tilde{y}_t$ may take on. A method for selecting the values $\bar{y}^j$ is to let $\bar{y}^N$ be a multiple m of the unconditional standard deviation $\sigma_y = (\sigma_\epsilon^2/(1-\lambda^2))^{1/2}$. Then let $\bar{y}^1 = -\bar{y}^N$, and let the remaining be equispaced over the interval $[\bar{y}^1, \bar{y}^N]$.

The method for calculating the transition matrix $p_{jk} = \Pr[\tilde{y}_t = \bar{y}^k \mid \tilde{y}_{t-1} = \bar{y}^j]$ follows. Put $w = \bar{y}^k - \bar{y}^{k-1}$. For each j, if k is between 2 and $N-1$, set

$$\begin{aligned} p_{jk} &= \Pr\left[\bar{y}^k - w/2 \leqslant \lambda \bar{y}^j + \epsilon_t \leqslant \bar{y}^k + w/2\right] \\ &= F\left(\frac{\bar{y}^k - \lambda \bar{y}^j + w/2}{\sigma_\epsilon}\right) - F\left(\frac{\bar{y}^k - \lambda \bar{y}^j - w/2}{\sigma_\epsilon}\right), \end{aligned} \tag{3a}$$

otherwise,

$$p_{j1} = F\left(\frac{\bar{y}^1 - \lambda \bar{y}^j + w/2}{\sigma_\epsilon}\right) \quad \text{and} \quad p_{jN} = 1 - F\left(\frac{\bar{y}^N - \lambda \bar{y}^j - w/2}{\sigma_\epsilon}\right). \tag{3b}$$

The rationale for this assignment of the transition probabilities can be understood by considering a random variable of the form $v = \lambda \bar{y}^j + \epsilon$ where $\lambda \bar{y}^j$ is fixed and ϵ is distributed as ϵ_t. Then the assignments (3) for p_{jk} make the distribution of $\tilde{y}_t$ conditional on $\tilde{y}_{t-1} = \bar{y}^j$ be a discrete approxima-

Table 1

Case number	Number of grid points N	Continuous process		Discrete process	
		λ	σ_y	$\bar{\lambda}$	$\sigma_{\tilde{y}}$
(1)	9	0.10	0.101	0.100	0.103
(2)	9	0.80	0.167	0.798	0.176
(3)	9	0.90	0.229	0.898	0.253
(4)	5	0.90	0.229	0.932	0.291

tion to the distribution of the random variable v. If the partitioning of the real line formed by the $\bar{y}^k$ is reasonably fine, then the conditional distribution of $\tilde{y}_t$ given $\tilde{y}_{t-1} = \bar{y}^j$ will approximate closely in the sense of weak convergence that of y_t given $y_{t-1} = \lambda \bar{y}^j$.

It is recognized that other integration rules, e.g., Gaussian quadrature, may lead to a placement of the grid points that in principle is more efficient than the equispaced scheme outlined above. The advantages of the above scheme, however, are computational speed and numerical stability, especially in the multivariate case given below. A rule based on Gaussian quadrature would essentially use the method of moments to determine the grid points and the transition probability matrix. This would necessarily entail the inversion of very large Vandermonde matrices, a problem which is notoriously time consuming and numerically unstable. The above scheme, on the other hand, can be coded very easily and the approximating Markov chains have been found to be quickly computable for a large number of sets of parameter values for the underlying autoregression.

To assess the adequacy of the approximation (3) note that the discrete process $\tilde{y}_t$ admits a representation of the form, $\tilde{y}_t - \bar{\lambda}\tilde{y}_{t-1} = \tilde{\epsilon}_t$, with $\operatorname{cov}(\tilde{\epsilon}_t, \tilde{y}_{t-1}) = 0$, where $\bar{\lambda} = \operatorname{cov}(\tilde{y}_t, \tilde{y}_{t-1})/\operatorname{var}(\tilde{y}_t)$, and $\sigma_{\tilde{\epsilon}}^2 = (1 - \bar{\lambda}^2)\operatorname{var}(\tilde{y}_t)$. The parameters $\bar{\lambda}$ and $\sigma_{\tilde{\epsilon}}^2$ are functions of the second moments of the $\tilde{y}_t$ process, and these moments can be computed from the transition matrix and the $\{\bar{y}^j\}$. Table 1 shows the induced population statistics $\bar{\lambda}$ and $\sigma_{\tilde{y}}$ for a range of values for λ and N with $\sigma_\epsilon = 0.1$.

The transition probabilities were computed under a normality assumption for ϵ_t and using the value $m = 3$ for determining the grid width. The approximation of $\bar{\lambda}$ and $\sigma_{\tilde{y}}$ to λ and σ_y is clearly adequate for most purposes when $N = 9$. Experimentation showed that the quality of the approximation remains good except when λ is very close to unity.

Monte Carlo studies showed that fitting standard linear time series models to the discrete-valued approximating process $\tilde{y}_t$ gives results very similar to what one would expect if the models had been fitted to realizations of the continuous-valued process y_t itself. Generating 51 pseudo observations on $\tilde{y}_t$ for the parameters in case no. (2) in table 1 and then fitting linear autoregressive models to these data gives

$$\underset{}{\tilde{y}_t =} \underset{(0.02)}{0.03} + \underset{(0.09)}{0.78}\,\tilde{y}_{t-1}, \qquad s^2 = 0.0095, \quad NOBS = 50, \tag{4a}$$

$$\underset{}{\tilde{y}_t =} \underset{(0.02)}{0.03} + \underset{(0.15)}{0.84}\,\tilde{y}_{t-1} - \underset{(0.14)}{0.08}\,\tilde{y}_{t-2}, \qquad s^2 = 0.0096, \quad NOBS = 49, \tag{4b}$$

where standard errors are in parentheses. These regressions are about what one would expect to get if the continuous-valued process y_t had been simulated. Interestingly, Kolmorgov–Smirnov tests accept normality of the residuals, indicating that the distance in the K–S metric between the error distribution and the normal distribution is not large.

3. The vector case

Suppose the vector process is

$$y_t = Ay_{t-1} + \epsilon_t, \qquad \text{var}(\epsilon_t) = \Sigma_\epsilon, \quad \text{a diagonal matrix}, \tag{5}$$

where y_t is an $M \times 1$ vector, A is an $M \times M$ matrix, and ϵ_t is an $M \times 1$ vector white noise process. It is assumed that the elements ϵ_{it} of ϵ_t are mutually independent with distribution $\Pr[\epsilon_{it} \leqslant u] = F_i(u/\sigma(\epsilon_t))$, where F_i is a standardized distribution function. After taking appropriate linear combinations any VAR model can be put in the form (5). Assume that after taking such linear combinations the elements of ϵ_t are also independent.

Let $\tilde{y}_t$ denote the approximating discrete-valued vector Markov chain for y_t in (5). Each component $\tilde{y}_{it}$ takes on one of N_i values: $\bar{y}_i^1 < \bar{y}_i^2 < \cdots < \bar{y}_i^{N_i}$. As in the univariate case, $\bar{y}_i^1$ and $\bar{y}_i^{N_i}$ are set to minus and plus a small integer m times the unconditional standard deviation of y_{it}. The remaining $\bar{y}_i^l$ satisfy $\bar{y}_i^{l+1} = \bar{y}_i^l + w_i$, $l = 1, 2, \ldots, N_i - 1$, where $w_i = 2m\sigma(y_i)/(N_i - 1)$. The $\sigma(y_i)$ are the square roots of the diagonal elements of the matrix Σ_y that satisfies $\Sigma_y = A\Sigma_y A' + \Sigma_\epsilon$, which can be found by iterating $\Sigma_y(r) = A\Sigma_y(r-1)A' + \Sigma_\epsilon$, with convergence as $r \to \infty$ guaranteed so long as (5) is stationary.

There are $N^* = N_1 \cdot N_2 \cdot \ldots \cdot N_M$ possible states for the system. Enumerate these states using the index $j = 1, 2, \ldots, N^*$ as a label for the states. Let $\tilde{l}(j)$ be an $M \times 1$ vector of integers associated with state j such that when the system is in state j at time $t-1$ the components $\tilde{y}_{i,t-1}$ assume the values $\tilde{y}_{i,t-1} = \bar{y}_i^p$, where $p = \tilde{l}_i(j)$, for $i = 1, 2, \ldots, M$.

To calculate transition probabilities p_{jk} let $\tilde{y}(j)$ be the $M \times 1$ vector of values for the $\tilde{y}$'s when the system is in state j, and put $\mu = A\tilde{y}(j)$. For each i let $h_i(j, l) = \Pr[\tilde{y}_{it} = \bar{y}_i^l \mid \text{state } j \text{ at } t-1]$ for $1 \leqslant l \leqslant N_i$, which, analogously to (3), is taken to be

$$\begin{aligned} h_i(j, l) &= F_i\left(\bar{y}_i^l - \mu_i + w_i/2\right) - F_i\left(\bar{y}_i^l - \mu_i - w_i/2\right) && \text{if} \quad 2 \leqslant l \leqslant N_i - 1, \\ &= F_i\left(\bar{y}_i^1 - \mu_i + w_i/2\right) && \text{if} \quad l = 1, \\ &= 1 - F_i\left(\bar{y}_i^{N_i} - \mu_i - w_i/2\right) && \text{if} \quad l = N_i. \end{aligned} \tag{6}$$

Given the h's the transition probabilities $p(j, k) = \Pr[\text{in state } k \mid \text{in state } j]$ are, by independence of the ϵ's, the products of the appropriate h's,

$$p(j, k) = \prod_{i=1}^{M} h_i(j, \tilde{l}_i(k)) \quad \text{for } j, k = 1, 2, \ldots, N^*.$$

Using this method a discrete approximation was taken to the process

$$y_{1t} = 0.70y_{1,t-1} + 0.30y_{2,t-1} + \epsilon_{1t}, \qquad y_{2t} = 0.20y_{1,t-1} + 0.50y_{2,t-1} + \epsilon_{2t}, \tag{7}$$

where ϵ_{1t} and ϵ_{2t} are iid normal (0, 0.1) random variables. The unconditional covariance matrix of this y_t process is

$$\Sigma_y = \begin{bmatrix} 0.332 & 0.126 \\ 0.126 & 0.185 \end{bmatrix}. \tag{8}$$

The N_i were each set to nine, yielding 81 states for $\tilde{y}_t$. The values $\bar{y}_i^l$ were computed with $m = 3$

standard deviations. As in the univariate case it is possible to check the accuracy of the approximation by finding the induced representation $\tilde{y}_t = \bar{A}\tilde{y}_{t-1} + \tilde{\epsilon}_t$, where $\bar{A} = (\mathrm{E}[\tilde{y}_t \tilde{y}'_{t-1}])(\mathrm{E}[\tilde{y}_{t-1}\tilde{y}'_{t-1}])^{-1}$, with the expectations computed using the transition matrix. In this case,

$$\bar{A} = \begin{bmatrix} 0.699 & 0.299 \\ 0.200 & 0.499 \end{bmatrix}, \qquad \Sigma_{\tilde{\epsilon}} = \begin{bmatrix} 0.373 & 0.139 \\ 0.139 & 0.200 \end{bmatrix}. \tag{9}$$

The elements of $\bar{A}$ are very close to the corresponding autoregressive parameters in (7), though those of $\Sigma_{\tilde{\epsilon}}$ are not quite as close to those of Σ_{ϵ} in (8). Experimentation showed that increasing moderately the number of grid points N_l often improved the accuracy of $\Sigma_{\tilde{\epsilon}}$.

Generating 51 observations on this $\tilde{y}_t$ and fitting a VAR to these data gives

$$\tilde{y}_{1t} = \underset{(0.06)}{0.11} + \underset{(0.11)}{0.61}\,\tilde{y}_{1,t-1} + \underset{(0.17)}{0.34}\,\tilde{y}_{2,t-1}, \qquad s^2 = 0.15, \quad NOBS = 50,$$

$$\tilde{y}_{2t} = \underset{(0.05)}{0.07} + \underset{(0.08)}{0.16}\,\tilde{y}_{1,t-1} + \underset{(0.13)}{0.44}\,\tilde{y}_{2,t-1}, \qquad s^2 = 0.09, \quad NOBS = 50.$$

Comparing the estimates here with (7)–(9) shows that the discrete Markov chain imitates quite well the statistical properties underlying a first-order vector process.

4. Conclusion

This paper has developed a method for finding a discrete Markov chain that approximates in the sense of weak convergence a continuous-valued univariate or multivariate autoregression. The method should be useful in both economics and finance where discrete state spaces are used for finding numerical solutions to integral equations.

References

Brock, W., 1982, Asset prices in a production economy, in: J. McCall, ed., The economics of information and uncertainty (The University of Chicago Press, Chicago, IL).

Donaldson, J.B. and R. Mehra, 1984, Comparative dynamics of an intertemporal asset pricing model, Review of Economic Studies 51, 491–508.

Hansen, Lars P., 1982, Generalized method of moments estimators, Econometrica 50, 1029–1055.

Lucas, R.E., Jr., 1978, Asset prices in an exchange economy, Econometrica 46, 1429–1445.

Michener, R., 1984, Permanent income in general equilibrium, Journal of Monetary Economics 13, 279–305.

National Science Foundation, 1984, Access to supercomputers (NSF Office of Scientific Computing, Washington, DC).

Prescott, E. and R. Mehra, 1985, The equity premium: A puzzle, Journal of Monetary Economics 15, 145–162.

Tauchen, George, 1985, Statistical properties of generalized method of moments estimators of structured parameters using financial market data, Working paper (Duke University, Durham, NC).

Wouk, A., 1979, A first course in functional analysis (Wiley, New York).