

## A Moment-Matching Method for Approximating Vector Autoregressive Processes by Finite-State Markov Chains

Nikolay Gospodinov and Damba Lkhagvasuren

Working Paper 2013-5

September 2013

**Abstract:** This paper proposes a moment-matching method for approximating vector autoregressions by finite-state Markov chains. The Markov chain is constructed by targeting the conditional moments of the underlying continuous process. The proposed method is more robust to the number of discrete values and tends to outperform the existing methods for approximating multivariate processes over a wide range of the parameter space, especially for highly persistent vector autoregressions with roots near the unit circle.

JEL classification: C15, C32, C60, E13, E32, E62

Key words: Markov chain, vector autoregressive processes, numerical methods, moment matching, non-linear stochastic dynamic models state space discretization, stochastic growth model, fiscal policy

---

The authors thank the coeditor Fabio Canova, three anonymous referees, Yongsung Chang, Paul Gomme, Andrei Jirnyi, Paul Klein, and Owen Wu for helpful comments and suggestions. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Nikolay Gospodinov, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, N.E., Atlanta, GA 30309-4470, 404-498-7892, [nikolay.gospodinov@atl.frb.org](mailto:nikolay.gospodinov@atl.frb.org), or Damba Lkhagvasuren (contact author), Department of Economics, Concordia University, 1455 de Maisonneuve Boulevard West, Montreal, Quebec, H3G 1M8 Canada, 514-848-2424 (ext. 5726), 514-848-4536 (fax), [damba.lkhagvasuren@concordia.ca](mailto:damba.lkhagvasuren@concordia.ca).

Federal Reserve Bank of Atlanta working papers, including revised versions, are available on the Atlanta Fed's website at [frbatlanta.org/pubs/WP/](http://frbatlanta.org/pubs/WP/). Use the WebScriber Service at [frbatlanta.org](http://frbatlanta.org) to receive e-mail notifications about new papers.

# 1 Introduction

Nonlinear dynamic macroeconomic and asset pricing models often imply a set of integral equations that do not admit explicit solutions. The finite-state Markov chain approximation methods developed by [Tauchen \(1986a\)](#) and [Tauchen and Hussey \(1991\)](#) prove to be an effective tool for reducing the complexity of solving these equations where the state variables follow autoregressive processes ([Burnside, 1999](#)). However, it is well known that these methods do not perform well for highly persistent autoregressive (AR) processes or processes with characteristic roots close to unity (see, e.g., [Tauchen, 1986a](#), [Tauchen and Hussey, 1991](#), and [Floden, 2008](#)). Although, the methods can generate a better approximation at the cost of a finer state space, this type of approach is not always feasible, especially in the multivariate case. The latter is important, as persistent multivariate structural shocks have become an increasingly popular device in accounting for business cycle fluctuations (e.g., [Cúrdia and Reis, 2010](#), and [Caldara, Fernandez-Villaverde, Rubio-Ramirez and Yao, 2012](#)).

The poor approximation of the methods by [Tauchen \(1986a\)](#) and [Tauchen and Hussey \(1991\)](#) for strongly autocorrelated processes has spurred a renewed research interest given the prevalence of highly persistent shocks in dynamic macroeconomic models. [Rouwenhorst \(1995\)](#) proposes a Markov-chain approximation of an AR(1) process constructed by targeting its first two conditional moments. Some recent advances in the literature on Markov-chain approximation methods include [Adda and Cooper \(2003\)](#), [Floden \(2008\)](#) and [Kopecky and Suen \(2010\)](#). While these methods provide substantial improvements in approximating the first-order univariate autoregressions, their extension to vector autoregressions (and higher-order autoregressive processes), which is of great practical interest to macroeconomists, is not readily available and possibly highly non-trivial. As a result, the method by [Tauchen \(1986a\)](#) continues

to be employed almost exclusively by researchers for approximating multivariate processes by finite-state Markov chains. The only alternative method that is available for approximating multivariate processes is the method proposed by [Galindev and Lkhagvasuren \(2010\)](#) for models with correlated AR(1) shocks. Although this method can be applied to vector autoregressions (VAR) by decomposing the latter into a set of interdependent AR(1) shocks, the state space generated by the method is not finite, except for the special case of equally-persistent underlying shocks. Therefore, to the best of our knowledge, a general method for approximating VAR processes by a finite-state Markov chain with appealing approximation properties over the whole parameter region of interest (including highly persistent parameterizations) is not yet available in the literature.

This paper fills this gap and proposes a moment-matching method for approximating vector autoregressions by a finite-state Markov chain. The main idea behind this method is to construct the Markov chain by targeting conditional moments of the underlying continuous process as in [Rouwenhorst \(1995\)](#), rather than directly calculating the transition probabilities using the distribution of the continuous process as in the existing methods. More specifically, we obtain the Markov-chain transitional probabilities by mixing a set of probability mass functions associated with the conditional distributions of finite-state univariate processes. To target the conditional moments in constructing the Markov chain, we use key elements of the Markov chains generated by the methods of [Tauchen \(1986a\)](#) and [Rouwenhorst \(1995\)](#). Therefore, the proposed method extends the multivariate methods of [Tauchen \(1986a\)](#) and [Tauchen and Hussey \(1991\)](#) to highly persistent cases and [Rouwenhorst's \(1995\)](#) scalar method (see [Kopecky and Suen, 2010](#)) to vector cases, while still maintaining a finite number of states.

Our method yields accurate approximations without relying on a large number of grid points for the state variables. In particular, the method expands the finite-state

Markov chain approximation to a much wider range of the parameter space. While the largest gains of the proposed approach arise when the characteristic roots of the underlying process are close to unity, it tends to outperform (in terms of bias and variance) the existing methods even when the persistence is moderate or low. Finally, the method can be readily adapted to accommodate other important features of the conditional distribution of the continuous-valued process.

The rest of the paper is organized as follows. Section 2 presents the continuous- and discrete-valued versions of the multivariate model and reviews the existing approximation methods. Our approximation method is introduced in Section 3. We show that the approximation is achieved by matching the first two conditional moments of the underlying process and describe the construction of the transition probability matrix and the Markov chain. Section 4 investigates the numerical properties of the method in the context of a stochastic growth model with technology and government expenditure shocks. Section 5 discusses possible extensions of the proposed method. Section 6 offers practical recommendations for the implementation of the method.

## 2 Model and Notation

### 2.1 Continuous VAR process

Let  $\mathbf{y}_t$  be an  $M \times 1$  vector containing the values that variables,  $y_1, y_2, \dots, y_M$ , assume at date  $t$ . We consider the following vector autoregressive (VAR) process:

$$\mathbf{y}_t = A\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{1}$$

where  $A$  (with a generic element  $a_{i,j}$ ) is an  $M \times M$  matrix with roots that lie strictly outside (although arbitrarily close to) the unit circle and the  $M \times 1$  vector  $\boldsymbol{\varepsilon}_t$  is i.i.d.  $\mathcal{N}(0, \Omega)$  with  $\Omega = \text{diag}(\omega_1^2, \omega_2^2, \dots, \omega_M^2)$  being a diagonal matrix. Extending the analysis to a non-diagonal  $\Omega$  and non-Gaussian errors is discussed later in the paper. Our focus on the

zero-mean, first-order VAR is primarily driven by expositional and notational simplicity and deterministic terms as well as higher-order dynamics can be easily incorporated at the expense of additional notation. Let  $\Sigma$  be the unconditional covariance matrix of the process  $\mathbf{y}_t$  and  $\sigma_i$  denote the unconditional standard deviation of  $y_i$  for each  $i$ . Then, the  $i$ -th diagonal element of  $\Sigma$  is given by  $\sigma_i^2$ .

## 2.2 Finite-state Markov chain

Let  $\tilde{\mathbf{y}}_t$  denote the finite-state Markov chain that approximates  $\mathbf{y}_t$  in (1). Each component  $\tilde{y}_{i,t}$  takes on one of the  $N_i$  discrete values denoted by  $\bar{y}_i^{(1)}, \bar{y}_i^{(2)}, \dots, \bar{y}_i^{(N_i)}$ . Therefore, at each point in time, the entire system will be in one of the  $N^* = N_1 \times N_2 \times \dots \times N_M$  states. Let  $\bar{y}^{(1)}, \bar{y}^{(2)}, \dots, \bar{y}^{(N^*)}$  label these  $N^*$  states and  $\Pi$  denote the  $N^* \times N^*$  transition matrix whose [row  $j$ , column  $k$ ] element  $\pi_{j,k}$  measures the probability that in the next period the system will be in state  $k$  conditional on the current state  $j$ . Our goal is to construct a finite number of grid points for each element of  $\tilde{\mathbf{y}}_t$  and calculate the associated transition probability matrix  $\Pi$  so that the characteristics of the generated process closely mimic those of the underlying process  $\mathbf{y}$ .

Define

$$h_i(j, l) = \Pr(\tilde{y}_{i,t} = \bar{y}_i^{(l)} | \tilde{\mathbf{y}}_{t-1} = \bar{\mathbf{y}}^{(j)}) \quad (2)$$

for  $i = 1, 2, \dots, M$ ,  $l = 1, 2, \dots, N_i$  and  $j = 1, 2, \dots, N^*$ . For any  $i$ , let  $L_i$  be an integer-valued function such that  $\tilde{y}_{i,t} = \bar{y}_i^{(L_i(j))}$  when the system is in state  $j$  at time  $t$ . Since the components of  $\boldsymbol{\varepsilon}_t$  are independent, the transition probability  $\pi_{j,k}$  can be written as the product of the individual transition probabilities:

$$\pi_{j,k} = \prod_{i=1}^M h_i(j, L_i(k)). \quad (3)$$

This means that, for each pair  $(i, j)$ , we need to construct  $N_i$  transition probabilities

$$H_i(j) = \{h_i(j, 1), h_i(j, 2), \dots, h_i(j, N_i)\} \quad (4)$$

over the grid points  $\bar{y}_i^{(1)}, \bar{y}_i^{(2)}, \dots, \bar{y}_i^{(N_i)}$ . Since  $\sum_{l=1}^{N_i} h_i(j, l) = 1$  for each  $(i, j)$ ,  $H_i(j)$  can be regarded as a probability mass distribution defined over the discrete values  $\bar{y}_i^{(1)}, \bar{y}_i^{(2)}, \dots, \bar{y}_i^{(N_i)}$ .

For any  $i$  and  $j$ , let  $\mu_i(j)$  denote the expected value of process  $y_{i,t+1}$ , conditional on  $\mathbf{y}_t = \bar{y}^{(j)}$ , i.e.,

$$\mu_i(j) = a_{i,1}\bar{y}_1^{(L_1(j))} + a_{i,2}\bar{y}_2^{(L_2(j))} + \dots + a_{i,M}\bar{y}_M^{(L_M(j))}. \quad (5)$$

The method that we propose below targets the first and second conditional moments of the process  $\mathbf{y}_t$  by minimizing the distance of the following moment conditions (for  $i = 1, \dots, M$  and  $j = 1, \dots, N^*$ )

$$\sum_{l=1}^{N_i} h_i(j, l)\bar{y}_i^{(l)} - \mu_i(j) \quad (6)$$

$$\sum_{l=1}^{N_i} h_i(j, l)(\bar{y}_i^{(l)} - \mu_i(j))^2 - \omega_i^2 \quad (7)$$

from zero, subject to some restrictions on the transition probabilities. Thus, the approach that we adopt in this paper requires that the Markov chain adequately approximates the conditional mean and variance of the continuous-valued process  $\mathbf{y}_t$ . In this respect, we give our method a moment-matching interpretation and refer to it as moment-matching (MM) method.

More specifically, our proposed method generates a set of discrete distributions using the [Rouwenhorst \(1995\)](#) method and mixes these distributions to construct conditional distributions while targeting the conditional mean and conditional variance of each  $\tilde{y}_i$  at each  $j$ . Thus, the MM method deals with  $M \times N^*$  conditional distributions and  $2M \times N^*$  conditional moments given by equations (6) and (7). With  $2M \times N^*$  free parameters, this gives rise to an over-identification problem and, in general, the conditional mean and variance cannot be both perfectly matched.

## 2.3 Existing methods

The existing finite-state methods for approximating vector autoregressions by [Tauchen \(1986a\)](#) and [Tauchen and Hussey \(1991\)](#) share the common feature that they use continuous probability distribution functions for calculating the transition probabilities defined over discrete grids. As mentioned in the introduction, the finite-state extension to multivariate processes of the recently proposed methods for improving the Markov chain approximation in near-nonstationary region of univariate AR processes is not readily available. In what follows, we consider explicitly the method proposed by [Tauchen \(1986a\)](#) as a representative of the existing methods since, according to [Floden \(2008\)](#), it tends to be more robust to the parameters of the underlying process than its version in [Tauchen and Hussey \(1991\)](#).

There are two free parameters that underlie the approximation accuracy of [Tauchen's \(1986a\)](#) method: the number of grid points  $N_i$  and a parameter  $m_i$  which is positively related to the distance between the grid points. Specifically, the distance between two consecutive nodes of  $\tilde{y}_i$  in [Tauchen's \(1986a\)](#) method is  $2m_i\sigma_i/(N_i - 1)$  for some  $m_i > 0$ , where  $\sigma_i$  denotes the unconditional standard deviation of  $y_i$ . First, if the the number grid points  $N_i$  is relatively large, then the conditional distribution of  $\tilde{y}_{i,t}$  given state  $j$  at time  $t - 1$  is expected to approximate closely (in the sense of weak convergence) the conditional distribution of  $y_{i,t}$  given  $y_{i,t-1} = \mu_i(j)$ . However, calculating the transition probabilities using the continuous distribution functions does not always deliver a desirable approximation. Specifically, [Tauchen's \(1986a\)](#) method fails to approximate the variability in  $\mathbf{y}_t$  as one or more of the roots of the underlying continuous-valued VAR process  $\mathbf{y}_t$  approach the unit circle. This problem arises because [Tauchen's \(1986a\)](#) method targets only the first conditional moment of the continuous-valued process  $\mathbf{y}_t$ .

Figure 1 illustrates this point numerically. It shows that as the persistence of the

underlying process increases, [Tauchen's \(1986a\)](#) method fails to generate sufficient variability in  $y_2$  (an online appendix provides a formal discussion of this issue). It should be noted that despite some numerical and methodological differences across the existing finite-state Markov-chain approximations for the vector case, all these methods suffer from the same problem since they calculate the transition matrices using distribution functions around the first conditional moment. In other words, regardless of the way the grid points are constructed, there is a non-zero distance between any two grid points and the above issue always arises in these methods.

Second, the choice of parameter  $m_i$  involves a sharp trade-off (especially in the presence of high persistence) between targeting unconditional variance and conditional variance and the quality of the approximation appears to be highly sensitive to the value of  $m_i$  ([Kopecky and Suen \(2010\)](#)). If the value of  $m_i$  is too small (say  $m_i = 2$ ), the resulting truncation of the grid space can be quite severe and [Tauchen's \(1986a\)](#) method performs poorly for approximating the unconditional variance, as well as other higher-order moments, of the underlying process. On the other hand, if the value of  $m_i$  is too large, the distance between the grid points increases which reduces the quality of approximating the conditional variance of the underlying process. In contrast, our proposed method breaks the tight link between the conditional and unconditional variance which is inherent in the existing finite-state VAR methods.

### 3 A Moment-Matching Markov Chain Method

#### 3.1 Probability mass functions

Before constructing the Markov chain, we introduce the following notation. Let  $\mathbf{y}_t = (y_{t,1}, \dots, y_{t,i}, \dots, y_{t,M})'$  and consider the scalar zero-mean AR(1) process  $y_{t,i}$  with persis-

tence parameter  $\rho_i$  and unconditional standard deviation  $\sigma_i$ , i.e.,

$$y_{t,i} = \rho_i y_{t-1,i} + \varepsilon_{t,i}, \quad (8)$$

where  $|\rho_i| < 1$ ,  $\varepsilon_{t,i}$  is i.i.d.  $\mathcal{N}(0, (1 - \rho_i^2)\sigma_i^2)$  and  $\sigma_i^2 = \text{Var}(y_{t,i})$ . Let  $\tilde{y}_i(N_i, \rho_i, \sigma_i)$  be the  $N_i$ -state symmetric Markov chain process, constructed by the method of [Rouwenhorst \(1995\)](#), that approximates the AR(1) process equation (8). Furthermore, let  $\bar{y}_i(N_i, \sigma_i) = \{\bar{y}_i^{(1)}, \bar{y}_i^{(2)}, \dots, \bar{y}_i^{(N_i)}\}$  denote the grid points and  $\Pi(N_i, \rho_i)$  be the probability transition matrix of  $\tilde{y}_i(N_i, \rho_i, \sigma_i)$  which is the  $i$ -th  $N_i \times N_i$  block on the diagonal of matrix  $\Pi$ . Suppose that the [row  $k$ , column  $l$ ] element of  $\Pi(N_i, \rho_i)$ ,  $\pi_{k,l}(N_i, \rho_i)$ , denotes the probability that the  $N_i$ -state process switches from  $\bar{y}_i^{(k)}$  to  $\bar{y}_i^{(l)}$ .

Now consider the  $k$ -th row of  $\Pi(N_i, \rho_i)$ :

$$\pi_k(N_i, \rho_i) = \{\pi_{k,1}(N_i, \rho_i), \pi_{k,2}(N_i, \rho_i), \dots, \pi_{k,N_i}(N_i, \rho_i)\}. \quad (9)$$

The key observation is that this row can be interpreted as a probability mass function associated with the nodes  $\{\bar{y}_i^{(1)}, \bar{y}_i^{(2)}, \dots, \bar{y}_i^{(N_i)}\}$ . It is easy to show that  $E(\tilde{y}_{t,i} | \tilde{y}_{t-1,i} = \bar{y}_i^{(k)}) = \rho_i \bar{y}_i^{(k)}$  and  $\text{Var}(\tilde{y}_{t,i} | \tilde{y}_{t-1,i} = \bar{y}_i^{(k)}) = (1 - \rho_i^2)\sigma_i^2$  so that the mean and the variance of the probability mass distribution are  $\rho_i \bar{y}_i^{(k)}$  and  $(1 - \rho_i^2)\sigma_i^2$ , respectively. Note that the conditional mean and variance of the Markov chain is independent of the number of grid points. This stands in sharp contrast with the existing methods (including [Tauchen's \(1986a\)](#) method) which are very sensitive to the number of grid points, especially in approximating near unit root processes.

On the other hand, the grid points  $\bar{y}_i(N_i, \sigma_i)$  and the transition matrix  $\Pi(N_i, \rho_i)$  are directly determined by the input parameters:  $N_i$ ,  $\rho_i$ , and  $\sigma_i$ . Specifically, let  $l \in \{0, 1, \dots, N_i - 1\}$ . For each  $l$ , consider the following  $N_i - 1$  probabilities:  $\{q_{l,j}(\rho_i)\}_{j=1}^{N_i-1}$ , where  $q_{l,j}(\rho_i) = (1 + \rho_i)/2$  if  $j \leq l$  or  $q_{l,j}(\rho_i) = (1 - \rho_i)/2$  otherwise. Then, the elements of  $\Pi(N_i, \rho_i)$  are given recursively by  $\pi_{l+1,1}(N_i, \rho_i) = \prod_{j=1}^{N_i-1} (1 - q_{l,j}(\rho_i))$ , and

$$\pi_{l+1,k+1}(N_i, \rho_i) = \frac{1}{k} \sum_{n=1}^k (-1)^{n-1} d_{l,n}(N_i, \rho_i) \pi_{l+1,k-n+1}(N_i, \rho_i), \quad (10)$$

$k = 1, 2, \dots, N_i - 1$ , where  $d_{l,n}(N_i, \rho_i) = \sum_{j=1}^{N_i-1} \left( \frac{q_{l,j}(\rho_i)}{1 - q_{l,j}(\rho_i)} \right)^n$ .

The grid points  $\bar{y}_i(N_i, \sigma_i)$  are given by the following  $N_i$  equally-spaced points:

$$\bar{y}_i^{(k)} = -\sigma_i \sqrt{N_i - 1} + 2\sigma_i \frac{k - 1}{\sqrt{N_i - 1}} \quad (11)$$

for  $k = 1, 2, \dots, N_i$ . Therefore, using different combinations of grid points and transition matrices constructed by the method of Rouwenhorst (1995), one can generate a class of probability mass functions with a wide range of means and variances. Below, we discuss how to construct the Markov chain of the VAR process in equation (1) by mixing these univariate probability mass functions.

### 3.2 Markov chain construction

For a given  $i$ ,  $A$ ,  $\Omega$  and  $\Sigma$ , equation (5) holds for each  $j \in \{1, 2, \dots, N^*\}$ . Assume that  $0 < \rho_i < 1$  and let  $\rho_i = \sqrt{1 - \omega_i^2 / \sigma_i^2}$ . Below, we consider two cases:  $\mu_i(j) \in [\rho_i \bar{y}_i^{(1)}, \rho_i \bar{y}_i^{(N_i)}]$  and  $\mu_i(j) \notin [\rho_i \bar{y}_i^{(1)}, \rho_i \bar{y}_i^{(N_i)}]$ .

1. In the typical case,  $\mu_i(j) \in [\rho_i \bar{y}_i^{(1)}, \rho_i \bar{y}_i^{(N_i)}]$ . Let  $r_{i,j}$  denote a strictly positive number below one:  $0 < r_{i,j} < 1$ . For now, set  $r_{i,j} = \rho_i$ . Then, consider the following mixture distribution

$$\tilde{\pi}(N_i, r_{i,j}) = \lambda(r_{i,j}) \pi_k(N_i, r_{i,j}) + (1 - \lambda(r_{i,j})) \pi_{k+1}(N_i, r_{i,j}), \quad (12)$$

where  $k$  is such that  $r_{i,j} \bar{y}_i^{(k)} \leq \mu_i(j) \leq r_{i,j} \bar{y}_i^{(k+1)}$  and  $\lambda(r_{i,j}) = \frac{r_{i,j} \bar{y}_i^{(k+1)} - \mu_i(j)}{r_{i,j} \bar{y}_i^{(k+1)} - r_{i,j} \bar{y}_i^{(k)}}$ . The mean and variance of this mixture distribution are, respectively,  $\mu_i(j)$  and

$$\tilde{\omega}^2(r_{i,j}) = \sigma_i^2(1 - r_{i,j}^2) + \sigma_i^2 r_{i,j}^2 \frac{4\lambda(r_{i,j})(1 - \lambda(r_{i,j}))}{N_i - 1}. \quad (13)$$

Note that the restrictions above ensure that  $\lambda(r_{i,j}) \in [0, 1]$  and we consider the boundary case ( $\lambda(r_{i,j}) = 0$  or  $1$ ) and the case of  $0 < \lambda(r_{i,j}) < 1$  separately. In the boundary case when  $\mu_i(j) = r_{i,j} \bar{y}_i^{(k)}$  or  $\mu_i(j) = r_{i,j} \bar{y}_i^{(k+1)}$ , we set  $H_i(j) \equiv \pi_k(N_i, r_{i,j})$  or  $H_i(j) \equiv \pi_{k+1}(N_i, r_{i,j})$  and both the conditional mean  $\mu_i(j)$  and conditional variance  $\omega_i^2$  are matched.

In the more common situation when  $0 < \lambda(r_{i,j}) < 1$ , the second term on the right hand side of equation (13) is positive. Therefore, although the mean of the mixture distribution  $\tilde{\pi}(N_i, r_{i,j})$  hits the target  $\mu_i(j)$ , the variance of the distribution is greater than the targeted conditional variance  $\omega_i^2$ . It should be noted that since  $r_{i,j} < 1$  and  $\lambda(r_{i,j})(1 - \lambda(r_{i,j})) \leq 0.25$ , the second term on the right-hand side of equation (13) converges to zero as the number of grid points increases. Hence, if the number of grid points is large,  $\tilde{\omega}^2(r_{i,j})$  is close to  $\omega_i^2$  and the procedure can be terminated at this step by setting  $H_i(j) \equiv \tilde{\pi}(N_i, r_{i,j})$ . In fact, experimentation shows that for a moderate number of state space (e.g.,  $N_i = 9$ ), setting  $H_i(j) \equiv \tilde{\pi}(N_i, r_{i,j})$  already provides a reasonable quality of approximation. Alternatively, one could further improve the quality of approximation by minimizing the distance  $|\tilde{\omega}^2(r_{i,j}) - \omega_i^2|$  over  $r_{i,j} \in (\rho_i, 1)$  as in step 1(a) below.

- 1(a). Let  $r_{i,j}^* = \arg \min_{r_{i,j} \in (\rho_i, 1)} |\tilde{\omega}^2(r_{i,j}) - \omega_i^2|$ . Let  $\tilde{\pi}(N_i, r_{i,j}^*)$  be the mixture distribution obtained by substituting the value of  $r_{i,j}^*$  for  $r_{i,j}$  in equation (12). Then, setting  $H_i(j) \equiv \tilde{\pi}(N_i, r_{i,j}^*)$  matches the conditional mean while achieving the best possible value for the conditional variance.
2. In the case when  $\mu_i(j) \notin [\rho_i \bar{y}_i^{(1)}, \rho_i \bar{y}_i^{(N_i)}]$ , we set  $H_i(j) \equiv \pi_1(N_i, \rho_i)$  if  $\mu_i(j) < \rho_i \bar{y}_i^{(1)}$  or  $H_i(j) \equiv \pi_{N_i}(N_i, \rho_i)$  if  $\mu_i(j) > \rho_i \bar{y}_i^{(N_i)}$ . In both situations, the conditional variance  $\omega_i^2$  is matched exactly while the conditional mean attains the value closest to  $\mu_i(j)$  given the grid points.

The probability mass functions  $\{H_i(1), H_i(2), \dots, H_i(N^*)\}_{i=1}^M$  are then obtained by repeating this procedure for  $i = 1, 2, \dots, M$ . The asymptotic validity of the method for approximating conditional expectations of general nonlinear functions that often arise in economic models is provided in an online appendix.

## 4 Application: Stochastic Growth Model

### 4.1 Main setup

To evaluate the method, we consider a version of the simple stochastic growth model of [Christiano and Eichenbaum \(1992\)](#) that allows the technology and demand shocks to be correlated.<sup>1</sup> Suppose that the social planner chooses streams of consumption services and capital stock,  $\{C_t, K_{t+1}\}_{t=0}^{\infty}$ , to maximize the expected utility function  $E_0 \sum_{t=0}^{\infty} \beta^t \ln(C_t)$  subject to  $C_t + G_t + K_{t+1} = \exp(z_t)K_t^\alpha + (1 - \delta)K_t$ , where  $z_t$  is the aggregate technology shock and  $G_t$  is the government expenditure at time  $t$  and the parameters  $\beta$ ,  $\alpha$ ,  $\delta$  are the time discount factor, the share of capital income in total output and the depreciation rate, respectively. Let  $Y_t$  denote the output at time  $t$ :  $Y_t = \exp(z_t)K_t^\alpha$ . Moreover, let  $g_t = \ln(G_t/\bar{G})$ , where  $\bar{G} > 0$  is the government expenditure at the steady state.

We assume that  $\mathbf{y}_t = (z_t, g_t)'$  evolves according to the VAR process given by equation (1) for  $M = 2$ . Let  $F(\cdot, \cdot | z, g)$  denote the implied bivariate distribution of  $(z_t, g_t)$  given  $z_{t-1} = z$  and  $g_{t-1} = g$ , i.e.,  $F(z', g' | z, g) = \Pr(z_t < z', g_t < g' | z_{t-1} = z, g_{t-1} = g)$ . Then, the social planner's problem is given by the following functional equation:

$$V(K, z, g) = \max_{K'} \left\{ \ln(\exp(z)K^\alpha + (1 - \delta)K - K' - \exp(g)\bar{G}) + \beta \int V(K', g') dF(z', g' | z, g) \right\}. \quad (14)$$

### 4.2 Parameters and shock dynamics

In the numerical analysis, we use the following values of the structural parameters  $\alpha$ ,  $\beta$ ,  $\delta$  and  $\bar{G}$ . The share of capital income and the discount factor are set to  $\alpha = 0.283$  and  $\beta = 0.986$ , respectively, as in [Gomme and Rupert \(2007\)](#). The value for the depreciation rate,

---

<sup>1</sup>For studies that use a finite-state Markov chains for solving stochastic dynamic models with multivariate persistent processes, see, for example, [Burnside \(1999\)](#), [Bayer and Juessen \(2012\)](#), [Caldara et al. \(2012\)](#), and [Lkhagvasuren \(2012\)](#).

$\delta = 0.0183$ , is obtained as a weighted average of the depreciation rates for (i) market structures and (ii) equipment and software, constructed by [Gomme and Lkhagvasuren \(2013\)](#) for the period 1954:Q1-2010:Q4. To calibrate the government expenditure at the steady state  $\bar{G}$  (obtained when  $z_t = g_t = 0$  for all  $t$  in the model), we use quarterly U.S. federal government current expenditures and U.S. gross domestic product over the period of 1948:Q1-2010:Q4 from the U.S. Bureau of Economic Analysis. We then target the average expenditure-to-output ratio (0.253), computed from the data, for the steady state expenditure-to-output ratio in the model which gives  $\bar{G} = 0.594$ .

The parameters of the driving process for aggregate technology and government expenditure are estimated from fitting a VAR(1) with a constant and a linear trend to the U.S. data of the period 1948:Q1-2010:Q4. We scale government expenditures by the U.S. civilian non-institutionalized population aged 16 and over and convert it into real terms using the consumption deflator ([Gomme and Rupert \(2007\)](#)). The logarithm of real per-capita expenditure is used for  $g_t$ . The Solow residual series  $z_t$  is obtained as in [Gomme and Lkhagvasuren \(2013\)](#).

The estimated VAR(1) model for  $z_t$  and  $g_t$ , using a bootstrap bias-corrected procedure, is given by

$$\begin{pmatrix} z_t \\ g_t \end{pmatrix} = \begin{pmatrix} 0.9809 & 0.0028 \\ 0.0410 & 0.9648 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ g_{t-1} \end{pmatrix} + \begin{pmatrix} e_{z,t} \\ e_{g,t} \end{pmatrix}, \quad (15)$$

where  $e_{z,t}$  and  $e_{g,t}$  are mean-zero, *iid* random variables with standard deviations of 0.0087 and 0.0262, respectively. Given equation (15), the unconditional covariance matrix of the process  $\mathbf{y}_t = (z_t, g_t)'$  is  $\begin{pmatrix} \sigma_z^2 & \sigma_{zg} \\ \sigma_{zg} & \sigma_g^2 \end{pmatrix} = \begin{pmatrix} 0.00235 & 0.00241 \\ 0.00241 & 0.01274 \end{pmatrix}$  and the implied correlation coefficient of the two shocks is  $\rho_{z,g} = 0.4347$ .

### 4.3 Approximation of the VAR process

In this section, we assess numerically how well our proposed MM method works in terms of approximating the autoregressive process in equation (15) for various degrees of fineness of the discrete space. Let  $\tilde{\mathbf{y}}_t = (\tilde{z}_t, \tilde{g}_t)$ , for  $t = 1, \dots, \tau$ , denote the simulated time series either from the Markov chain approximation by Tauchen (1986a) or the MM method proposed in this paper. In order to shed additional light on the performance of the moment-matching method and illustrate the value added from each of its implementation steps, we report two versions of the method: (i) the baseline version of the method that does not include the minimization step 1(a) in section 3.2 (denoted by MM0) and (ii) the full version of the method that includes the minimization step 1(a) in section 3.2 (denoted by MM).

#### 4.3.1 Unconditional moments

First, the evaluation of the accuracy of the two approximations is based on some unconditional moments and parameters of the underlying process in equation (1) and the simulated processes. The parameters of interest are the unconditional variances of  $z_t$  and  $g_t$  (denoted by  $\sigma_z^2$  and  $\sigma_g^2$ ), the correlation coefficient between  $z_t$  and  $g_t$  (denoted by  $\rho_{z,g}$ ), and the persistence measures  $1 - \varsigma_1$  and  $1 - \varsigma_2$ , where  $\varsigma_1$  and  $\varsigma_2$  are the two roots (eigenvalues) of matrix  $A$  associated with equation (15). As in Tauchen (1986a) and Tauchen and Hussey (1991), the simulated counterpart of  $A$ ,  $\hat{A}$ , is obtained by fitting the VAR model in equation (1) to  $\{\tilde{\mathbf{y}}_t\}_{t=1}^\tau$ . The evaluation of the approximation accuracy is based on 1,000 simulated series of length  $\tau = 2,000,000$ , with the first 200,000 observations of each sample being discarded to remove the effect of the initial condition. As in Tauchen (1986a), we choose nine grid points for each component:  $\bar{N} = N_1 = N_2 = 9$ . We also consider other cases in which the state space is much finer:  $\bar{N} = 15$  and  $\bar{N} = 21$ . When using Tauchen's (1986a) method, we use, for each

$i$ , equispaced grid points on the interval  $[-m_i\sigma_i, m_i\sigma_i]$ , where  $m_i = 1.2 \ln N_i$  (Floden, 2008).<sup>2</sup>

Table 1 reports the root mean squared error (RMSE), bias and standard deviation of these parameters relative to their true values. The results suggest that our MM method dominates the method by Tauchen (1986a) in terms of bias and RMSE for all parameters of interest across all three levels of fineness. For example, the relative bias for  $\bar{N} = 9$  of the estimated  $1 - \varsigma_1$ ,  $\sigma_1^2$  and  $\sigma_2^2$ , using data generated by Tauchen’s (1986a) method, is more than 30%, whereas the corresponding biases for the MM method are approximately 1%. Increasing the number of grid points from 9 to 15 and 21 improves the performance of Tauchen’s (1986a) method but its numerical properties remain rather poor. The version of our proposed method that does not include the minimization step 1(a), MM0, significantly outperforms Tauchen’s (1986a) method and delivers accurate approximations. However, the advantages of performing this additional minimization step become obvious in removing the bias in the estimates of  $\sigma_z^2$  and  $\sigma_g^2$ .

### 4.3.2 Conditional moments

Potentially important information about the quality of the approximation is also contained in the conditional moments. Hence, it would be interesting also to report the approximation accuracy of the first two moments, conditional on the state of the process.

Given the constructed grid points and transition probabilities, the implied conditional mean and variance are  $\hat{\mu}_i(j) = \sum_{l=1}^{N_i} h_i(j, l) \bar{y}_i^{(l)}$  and  $\hat{\omega}_i^2(j) = \sum_{l=1}^{N_i} h_i(j, l) (\bar{y}_i^{(l)} - \hat{\mu}_i(j))^2$ , where  $i \in \{1, 2, \dots, M\}$  and  $j \in \{1, 2, \dots, N^*\}$ . Then, for each  $i$  and  $j$ , the distances between the true and the generated conditional moments can be measured

---

<sup>2</sup>In an online appendix, we adjust these grid points by targeting unconditional variances as in Kopecky and Suen (2010). Our numerical results suggest that when the persistence is low, adjusting the unconditional variance improves the approximation of the conditional moments. However, in the case of high persistence, this adjustment tends to compromise the approximation accuracy of the conditional variance.

by  $|\hat{\mu}_i(j) - \mu_i(j)|$  and  $|\hat{\omega}_i^2(j)/\omega_i^2 - 1|$ . To assess the overall accuracy of the conditional moments, we consider weighted averages of these distances across the  $N^*$  states using the frequencies of each state as weights. The weights are constructed from a simulated process of length  $\tau = 1,800,000$ . The results are presented in the lower panel of Table 1 and show that the MM method performs extremely well across all parameterizations. Again, this is not surprising since, by construction, this method targets the first two conditional moments of the underlying process. More importantly, the results show that calculating the transition probabilities using the conditional distribution as in Tauchen (1986a) generates a substantial bias in the second conditional moment. These numerical results lend support to our discussion in section 2.3. The MM0 method partially correct the bias in the second conditional moments of Tauchen’s (1986a) approximation while the MM method (that includes the minimization step 1(a)) completely eliminates the bias.

## 4.4 Planner’s problem

Next, we consider the performance of the MM and Tauchen’s (1986a) methods for solving the planner’s problem in equation (14).

### 4.4.1 Solution with finite-state shocks

The Bellman equation of the social planner is solved over the discrete space for  $K$ ,  $z$  and  $g$ . Given equation (15), let  $\mathcal{D}^2$  be the  $N_1 \times N_2$  bivariate discrete grid for  $(z, g)$  used in the Markov chain approximation by Tauchen (1986a) or the method proposed in this paper. Following Coleman (1990), for the capital stock, we consider a discrete grid, denoted by  $\mathcal{K}$ , of fifty (equispaced in logarithmic scale) points that span the interval  $(0.7K_{ss}, 1.3K_{ss})$ , where  $K_{ss}$  is the steady state capital stock.

Solving the discrete analog of the Bellman equation requires finding the optimal choice of the capital stock at each point of the trivariate grid  $\mathcal{D}^2 \times \mathcal{K}$ , while replacing

the conditional distribution  $F(\cdot, \cdot | z, g)$  by the transition matrix  $\Pi$ . The decision rule is computed iteratively by using the solution from the previous iteration to evaluate the conditional expectation. For capital stock values that do not belong to  $\mathcal{K}$ , we use univariate interpolation along  $K$ . The iterations are repeated until the algorithm reaches the convergence criterion such that the gap between the decision rules obtained from two consecutive iterations become negligibly small at each point of the discrete space  $\mathcal{D}^2 \times \mathcal{K}$ . Once the policy function is obtained, the simulation of the economy amounts to generating a time series of capital (using univariate interpolation along  $K$ ) for the sequence of shocks simulated by section 4.3.

#### 4.4.2 Solution with continuous shocks

[Kopecky and Suen \(2010\)](#) show that the appropriate way to evaluate finite-state approximation methods is to compare the numerical results for these different methods with those obtained from a highly accurate, albeit computationally demanding, method. Following their approach, we evaluate the above finite-state methods by considering an alternative solution method that takes explicitly into account the continuous nature of the shocks. Unfortunately, the presence of two shocks in our case does not allow us to solve the Bellman equation over a grid as fine as the one used in [Kopecky and Suen \(2010\)](#). Given this important limitation, we first find the policy rule on a coarse trivariate discrete grid (similar to  $\mathcal{D}^2 \times \mathcal{K}$ ), while performing numerical integration using the continuous conditional distributions, and then interpolate the rule to a continuous process directly generated by equation (15). Since the numerical integration and simulated processes are the two key elements of any finite-state approximation method, this alternative method serves as an efficient benchmark for accuracy of the above methods. Moreover, as shown below, the alternative method is remarkably robust to the coarseness of the grid points.

For the shocks of  $z$  and  $g$ , we use  $N_1$  and  $N_2$  equispaced grid points on the intervals

$(-3\sigma_z, 3\sigma_z)$  and  $(-3\sigma_g, 3\sigma_g)$  (Coleman, 1990; Gomme and Rupert, 2007), respectively, and for the capital stock, we consider the same grid as in the previous case. We calculate the numerical integration using a bivariate Gauss-Hermite quadrature rule with  $9 \times 9$  quadrature nodes for  $(z, g)$  (see, for example, Judd, 1998, for the quadrature nodes and associated weights of the rule). Note that both calculating the conditional expectation and generating the capital stock are now computationally more demanding. Specifically, they require trivariate interpolation along  $(K, z, g)$ , as opposed to univariate interpolation along  $K$  in the finite-state method. For brevity, this method will be referred to as the continuous method.

#### 4.4.3 Results

After solving equation (14) using the transition matrices constructed by the two methods, we simulate the time series of output, capital, consumption and government expenditure for  $\tau = 2,000,000$ . We also generate a sequence of length 2,000,000 using the actual VAR(1) process. As before, the first 10 percent of the observations are discarded.

The basic results are presented in the upper panel of Figure 2. It shows that the MM method is less sensitive to the number of grid points compared with Tauchen's (1986a) method. More importantly, as the number of grid points increases, the moments generated by Tauchen's (1986a) method approach those obtained by the continuous and MM methods. This suggests that the simulated data obtained by the MM method describes more accurately the underlying dynamics of the variables.<sup>3</sup>

It is important to note that differences in moments obtained by the two methods arise not only because of differences in the simulated shocks, but also due to differences

---

<sup>3</sup>The computing times of solving the full model (using Matlab on a 3.4 GHz, 64-bit Intel PC) for Tauchen's (1986a) and MM methods with  $\bar{N} = 9$  ( $\bar{N} = 21$ ) are 890.8 (4757.5) seconds and 867.5 (4820.5) seconds, respectively. Note, however, that the degree of accuracy for the MM approximation for  $\bar{N} = 9$  is much higher than the degree of accuracy of Tauchen's (1986a) method for  $\bar{N} = 21$  (see panel A of Figure 2). This suggests that the computational advantages of the MM method could be substantial for the same level of approximation error.

in the decision rules. For example, panel B of Figure 2 plots the optimal policy rules for capital (as a function of technology and government expenditure shocks) for the different solution methods. Figure 2 shows that the decision rule obtained by the MM method tracks very closely the decision rule obtained by the continuous method. On the other hand, the decision rule from Tauchen’s (1986a) method appears to be much less susceptible to the two shocks as indicated by the relatively flatter curves generated by the method.

Table 2 summarizes other moments of the simulated data such as volatility, persistence and correlation with the current and lead value of output. The main results that emerge from this exercise are consistent with our previous findings. Tauchen’s (1986a) method provides a poor approximation to key moments of the underlying process and the approximation errors for some moments remain relatively large even for  $\bar{N} = 9$ . In contrast, the approximation by the MM method tends to be extremely reliable across moments and different grids.

## 4.5 GMM estimation

In this section, we use 10,000 simulated series for  $K_t$ ,  $Y_t$  and  $C_t$  ( $t = 1, \dots, T$ ) from the model to estimate the true structural parameters  $\beta = 0.986$ ,  $\alpha = 0.283$  and  $\delta = 0.0183$  by generalized method of moments (GMM). We consider empirically relevant sample sizes,  $T = 120$  and  $T = 240$ , which correspond to 30 and 60 years of quarterly observations. The goal of this exercise is to see whether differences in the simulated data from the MM and Tauchen’s (1986a) approximation methods will translate in differences in the estimates (and their associated variability) of the structural parameters.

We first turn our attention to the properties of the simulated data obtained from the MM and Tauchen’s (1986a) methods with  $\bar{N} = 9$ , along with the continuous method as a benchmark. Table 3 reports the standard deviations of the variables  $C_{t+1}/C_t$  and

$Y_{t+1}/K_{t+1}$  that are used in the subsequent analysis as well as the correlation between  $C_{t+1}/C_t$  and  $Y_{t+1}/K_{t+1}$ . While both (MM and [Tauchen's \(1986a\)](#)) methods appear to provide a good approximation to the variability of the output-capital ratio, the MM method's approximation of  $\text{std}(C_{t+1}/C_t)$  and  $\text{Corr}(C_{t+1}/C_t, Y_{t+1}/K_{t+1})$  is much closer to the corresponding statistics for the continuous method.

The Euler equation for the stochastic growth model is given by

$$E_t \left[ \beta \frac{C_t}{C_{t+1}} \left( \alpha \frac{Y_{t+1}}{K_{t+1}} + (1 - \delta) \right) - 1 \right] = 0, \quad (16)$$

where  $E_t[\cdot]$  denotes the expectation conditional on the information set at time  $t$ . In what follows, we fix the depreciation parameter at its true value and estimate only  $(\alpha, \beta)'$ . Furthermore, we use  $z_t = (1, Y_t/K_t)'$  as instruments and transform (by the law of iterated expectations) the above conditional restriction into two unconditional restrictions which gives rise to the following just-identified system of equations:

$$\frac{1}{T-1} \sum_{t=1}^{T-1} \left[ \beta \frac{C_t}{C_{t+1}} \left( \alpha \frac{Y_{t+1}}{K_{t+1}} + (1 - \delta) \right) - 1 \right] z_t = \mathbf{0}_2. \quad (17)$$

The values  $\hat{\alpha}$  and  $\hat{\beta}$  that solve these two equations are the GMM estimates. Since this is a just-identified model, the choice of a weighting matrix is irrelevant. We also note that the choice of this particular set of instruments and estimated parameters is made to sidestep some identification issues that arise in this model. [Table 3](#) shows that although the differences in the average estimates for the three methods are fairly small, the MM method produces up to 8–9% less volatile estimates than [Tauchen's \(1986a\)](#) method.<sup>4</sup> This finding reinforces the practical relevance of the proposed MM method for macroeconomic and econometric analysis.

---

<sup>4</sup>Unreported results reveal that as the sample size increases, the upward bias of  $\hat{\alpha}$  shrinks towards zero and the differences in the variability of the GMM estimates between the MM and [Tauchen's \(1986a\)](#) methods become even larger.

## 5 Extensions

### 5.1 Non-diagonal covariance matrix

While the procedure in Section 3 is developed under the assumption of a diagonal covariance matrix  $\Omega$ , the proposed method can be easily extended to the case of a non-diagonal covariance matrix. Suppose now that the underlying continuous-valued process follows

$$\mathbf{x}_t = b + B\mathbf{x}_{t-1} + \eta_t, \quad (18)$$

where  $\eta_t$  is i.i.d.  $(0, \Psi)$  and  $\Psi$  is a non-diagonal matrix (see also [Terry and Knotek II, 2011](#)). Let  $D$  be a lower triangular matrix such that  $\Omega = D\Psi D^{-1}$  is a diagonal matrix. Define the transformations ([Tauchen, 1986b](#)),  $\mathbf{x}_t \rightarrow D[\mathbf{y}_t - (I_M - B)^{-1}b]$ ,  $B \rightarrow A = DBD^{-1}$  and  $\eta_t \rightarrow D\varepsilon_t$ . Then, we have the same model as in equation (1). After computing the discrete Markov-chain approximation for this modified model, we reverse the transformations above in order to obtain the discrete process corresponding to equation (18). We should note that since any stationary  $\text{AR}(p)$  process can be expressed in a companion form as a  $\text{VAR}(1)$  process, our method effectively extends the method by [Rouwenhorst \(1995\)](#) to higher-order scalar autoregressive processes.

### 5.2 Targeting higher-order moments

The normality (and log-normality, in the case of modeling shocks with stochastic volatility as in [Fernandez-Villaverde, Guerron-Quintana, Rubio-Ramirez and Uribe, 2011](#) and [Caldara et al., 2012](#)) assumption is routinely used in describing the properties of the shocks in macroeconomic models. Nevertheless, in some applications, the normality assumption of the error term in equation (1) may seem restrictive.

In general, targeting higher-order conditional moments requires a much finer state space. The reason is that due to the finite-state approximation itself, the innovation

of the finite-state process takes on a finite number of values. For example, when the conditional mean is close to the upper and lower bounds of the grid, the conditional distribution function is highly asymmetric and the overall skewness of the error term is distorted. Moreover, when the persistence is high, the probability that the current state repeats itself increases. As a result, the innovation will be highly concentrated at zero and will jump to another state within a finite distance with low probability, which gives rise to a leptokurtic distribution. Hence, non-zero skewness and excess kurtosis inherently arise in any finite-state approximation. Therefore, to obtain a reasonable approximation of higher-order conditional moments, one has to employ a much larger number of grid points. Keeping this in mind, we make the following modifications to our procedure in Section 3.2 that allow us to target skewness and excess kurtosis.

To generate non-zero conditional skewness, we use the first row of the transition matrix  $\Pi(N_i, \rho_i)$ :

$$\pi_1(N_i, \rho_i) = \{\pi_{1,1}(N_i, \rho_i), \pi_{1,2}(N_i, \rho_i), \dots, \pi_{1,N_i}(N_i, \rho_i)\}, \quad (19)$$

where  $N_i \geq 3$ . Since this probability mass distribution is associated with the lowest discrete value of the scalar AR(1) process and  $\rho_i > 0$ , it is positively skewed. Moreover, the skewness increases with  $\rho_i$ .

Now let us consider  $\tilde{N}_i > N_i$  grid points constructed by [Rouwenhorst's \(1995\)](#) method for an autoregressive process with unconditional variance  $\sigma_i^2$ . Then, the transition matrix associated with these grid points is constructed using  $\pi_1(N_i, \rho_i)$ . Specifically, we set the  $k$ -th row of the  $\tilde{N}_i \times \tilde{N}_i$  matrix  $\tilde{\Pi}(\tilde{N}_i, \rho_i)$  to

$$\tilde{\pi}_k(\tilde{N}_i, \rho_i) = \begin{cases} (\mathbf{0}'_{k-1} \quad \pi_1(N_i, \rho_i) \quad \mathbf{0}'_{\tilde{N}_i - N_i - k + 1}) & \text{if } k \leq \tilde{N}_i - N_i + 1, \\ (\mathbf{0}'_{\tilde{N}_i - N_i} \quad \pi_1(N_i, \rho_i)) & \text{otherwise,} \end{cases} \quad (20)$$

where  $k = \{1, 2, \dots, \tilde{N}_i\}$  and  $\mathbf{0}_k$  denotes a  $k \times 1$  zero vector. Note that by reversing the order of the elements of  $\tilde{\pi}_k(\tilde{N}_i, \rho_i)$ , one can target negative skewness. It can be seen that the transition matrix  $\tilde{\Pi}(\tilde{N}_i, \rho_i)$ , along with the grid points, yields a scalar Markov

chain whose conditional distribution has the same skewness as the mass distribution function (19). Therefore, to construct the probability transition functions  $H_i(j)$  as in Section 3.2, one can use  $\tilde{\Pi}(\tilde{N}_i, \rho_i)$  instead of  $\Pi(N_i, \rho_i)$ .

To generate a conditional distribution with excess kurtosis, one can use a mixture distribution approach. More specifically, let  $\pi_{(1)}(N_i, \rho_i)$  and  $\pi_{(2)}(N_i, \rho_i)$  be two discrete probability distributions defined over  $N_i$  equally-distanced grid points that have a common mean but different variances  $\sigma_1^2$  and  $\sigma_2^2$ . Consider now the mixture  $\tilde{\pi}_1(N_i, \rho_i) = \tilde{\lambda}\pi_{(1)}(N_i, \rho_i) + (1 - \tilde{\lambda})\pi_{(2)}(N_i, \rho_i)$ , where  $0 \leq \tilde{\lambda} \leq 1$ . Setting both  $\sigma_1/\sigma_2$  and  $1 - \tilde{\lambda}$  to low values would result in excess kurtosis for the conditional distribution  $\tilde{\pi}_1(N_i, \rho_i)$ . Then, substituting this conditional distribution for  $\tilde{\pi}_1(N_i, \rho_i)$  in (20) gives the  $k$ -th ( $k = 1, 2, \dots, \tilde{N}_i$ ) row of the desired transition matrix  $\tilde{\Pi}(\tilde{N}_i, \rho_i)$ .

## 6 Conclusions and Practical Recommendations

We conclude with some remarks and recommendations regarding the implementation of our proposed method. First, several interesting observations regarding the computational costs associated with the different approximation methods emerge from our numerical analysis. When the largest roots of the process are much closer to one as in the case at higher (monthly) frequency, Tauchen's (1986a) method requires an extremely large number of states in order to achieve the level of accuracy comparable to that of the MM method with far fewer states. This means that as the roots of the process approach the nonstationary boundary, the number of grid points for Tauchen's (1986a) method must increase sharply. Consequently, the computation involved becomes prohibitively time consuming or even infeasible (see Figure 1) and this curse of dimensionality (Burnside, 1999) becomes even more severe for non-linear dynamic models. In contrast, the accuracy of approximation of the MM method is less sensitive to the number of grid points. For instance, Figure 2 shows that the quality of the ap-

proximation obtained by the MM method using  $N_1 = N_2 = 9$  is much higher than the one obtained by Tauchen’s (1986a) method using  $N_1 = N_2 = 21$ . More importantly, unlike Tauchen’s (1986a) method, the MM method can always generate a time-varying process (see Figure 1).

Second, the computational gains of the MM method do not come at the cost of increased complexity in the implementation of our approximation. The practical recommendations for implementing the proposed MM method can be summarized as follows. The number of grid points  $N^*$  can be determined by selecting  $N_i$  ( $i = 1, \dots, M$ ) for each individual series depending on the properties of the data. For highly persistent data,  $N_i = 9$  tends to provide a very precise approximation of the underlying process while for less persistent data,  $N_i = 5$  appears to be a reasonable choice. However, it should be emphasized that in approximating multivariate processes, adjusting the number of grid points of the individual components of  $\tilde{\mathbf{y}}$  to accommodate their persistence is not always the best approach. The reason is that, unlike in the scalar case, one must also target the cross correlations between the different components which are reflected in the conditional mean of the process (see equation (5)). Therefore, the number of grid points for one component of  $\tilde{\mathbf{y}}$  will affect the conditional moments of the other variables.

The choice which version of our proposed method should be employed is mainly dictated by the trade-off between speed and accuracy. If the main goal is the quality of approximation for a given number of grid points, the full version of the MM method that includes the minimization step 1(a) is the preferred option. On the other hand, if the computational costs are of primary concern, especially for large dimensional VAR processes, the baseline version of the proposed method without the minimization step 1(a) will deliver a simple, fast and reliable approximation for the same number of grid points. In order to improve the degree of approximation, the baseline version of the MM method requires a much finer grid of nodes which will increase the computing time. To

put it differently, while the minimization step 1(a) involves additional computations, it effectively reduces the number of grid points that are needed to achieve the same quality of approximation.

## References

- Adda J, Cooper R. 2003. *Dynamic Economics*. MIT Press, Cambridge, MA.
- Bayer C, Juessen F. 2012. On the dynamics of interstate migration: Migration costs and self-selection. *Review of Economic Dynamics* **15**: 377–401.
- Burnside C. 1999. Discrete state-space methods for the study of dynamic economies. In Marimon R, Scott A (eds.) *Computational Methods for the Study of Dynamic Economies*. New York: Oxford University Press, 95–113.
- Caldara D, Fernandez-Villaverde J, Rubio-Ramirez JF, Yao W. 2012. Computing DSGE models with recursive preferences and stochastic volatility. *Review of Economic Dynamics* **15**: 188–206.
- Christiano LJ, Eichenbaum M. 1992. Current real business cycle theories and aggregate labor market fluctuations. *American Economic Review* **82**: 430–450.
- Coleman W II. 1990. Solving the stochastic growth model by policy-function iteration. *Journal of Business and Economic Statistics* **8**: 27.
- Cúrdia V, Reis R. 2010. Correlated disturbances and U.S. business cycles. Working Paper 15774, NBER.
- Fernandez-Villaverde J, Guerron-Quintana P, Rubio-Ramirez JF, Uribe M. 2011. Risk matters: The real effects of volatility shocks. *American Economic Review* **101**: 2530–2561.
- Floden M. 2008. A note on the accuracy of Markov-chain approximations to highly persistent AR(1) processes. *Economics Letters* **99**: 516–520.
- Galindev R, Lkhagvasuren D. 2010. Discretization of highly persistent correlated AR(1) shocks. *Journal of Economic Dynamics and Control* **34**: 1260–1276.

- Gomme P, Lkhagvasuren D. 2013. Calibration and simulation of DSGE models. In Hashimzade N, Thornton M (eds.) *Handbook of Research Methods and Applications in Empirical Macroeconomics*. Edward Elgar Publishing.
- Gomme P, Rupert P. 2007. Theory, measurement, and calibration of macroeconomic models. *Journal of Monetary Economics* **54**: 460–497.
- Judd KL. 1998. *Numerical Methods in Economics*. Cambridge: The MIT Press.
- Kopecky KA, Suen RM. 2010. Finite state Markov-chain approximations to highly persistent processes. *Review of Economic Dynamics* **13**: 701–714.
- Lkhagvasuren D. 2012. Big locational unemployment differences despite high labor mobility. *Journal of Monetary Economics* **59**: 798–814.
- Rouwenhorst GK. 1995. Asset pricing implications of equilibrium business cycle models. In Cooley T (ed.) *Structural Models of Wage and Employment Dynamics*. Princeton: Princeton University Press.
- Tauchen G. 1986a. Finite state Markov-chain approximations to univariate and vector autoregressions. *Economics Letters* **20**: 177–181.
- Tauchen G. 1986b. Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data. *Journal of Business and Economic Statistics* **4**: 397–416.
- Tauchen G, Hussey R. 1991. Quadrature-based methods for obtaining approximate solutions to linear asset pricing models. *Econometrica* **59**: 371–396.
- Terry SJ, Knotek II ES. 2011. Markov-chain approximations of vector autoregressions: Application of general multivariate-normal integration techniques. *Economics Letters* **110**: 4–6.

**Table 1:** Approximation Accuracy: The Shock Process

	$\bar{N} = 9$			$\bar{N} = 15$			$\bar{N} = 21$		
	Tau.	MM0	MM	Tau.	MM0	MM	Tau.	MM0	MM
<i>Root mean squared error</i>									
$\hat{\sigma}_z^2$	0.433	0.099	0.009	0.410	0.081	0.008	0.314	0.068	0.009
$\hat{\sigma}_g^2$	0.363	0.139	0.010	0.306	0.114	0.007	0.218	0.095	0.007
$\rho_{z,g}$	0.040	0.023	0.011	0.011	0.012	0.009	0.017	0.011	0.009
$1 - \hat{\zeta}_1$	0.323	0.016	0.011	0.048	0.009	0.009	0.009	0.009	0.009
$1 - \hat{\zeta}_2$	0.160	0.006	0.006	0.016	0.005	0.005	0.005	0.006	0.005
<i>Bias</i>									
$\hat{\sigma}_z^2$	0.433	0.099	-0.005	0.410	0.080	0.000	0.313	0.068	0.001
$\hat{\sigma}_g^2$	0.362	0.138	-0.007	0.306	0.114	0.000	0.217	0.094	0.000
$\rho_{z,g}$	-0.038	-0.022	-0.006	0.006	-0.008	0.000	0.015	-0.005	0.000
$1 - \hat{\zeta}_1$	-0.323	0.013	0.007	-0.048	0.001	0.000	-0.003	-0.001	-0.001
$1 - \hat{\zeta}_2$	-0.160	0.003	0.002	-0.016	0.000	0.000	0.000	0.000	0.000
<i>Standard deviation</i>									
$\hat{\sigma}_z^2$	0.012	0.008	0.008	0.011	0.009	0.008	0.011	0.009	0.008
$\hat{\sigma}_g^2$	0.008	0.007	0.006	0.008	0.007	0.007	0.008	0.008	0.007
$\rho_{z,g}$	0.010	0.009	0.009	0.009	0.009	0.009	0.009	0.010	0.009
$1 - \hat{\zeta}_1$	0.006	0.009	0.009	0.008	0.009	0.009	0.009	0.009	0.009
$1 - \hat{\zeta}_2$	0.005	0.006	0.005	0.005	0.005	0.005	0.005	0.006	0.005
<i>Distance between simulated and true conditional moments</i>									
$10 \times \hat{\mu}_z$	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
$10 \times \hat{\mu}_g$	0.006	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
$(\hat{\omega}_z/\omega_z)^2$	0.100	0.106	0.000	0.319	0.080	0.000	0.310	0.066	0.000
$(\hat{\omega}_g/\omega_g)^2$	0.269	0.163	0.000	0.304	0.122	0.000	0.205	0.101	0.000

*Notes.* This table reports key moments of the finite state process associated with the bivariate VAR(1) model in equation (15). The root mean squared error, the bias and the standard deviation are relative to their true values. ‘‘Tau.’’ denotes the approximation obtained by the method of Tauchen (1986a), whereas ‘‘MM’’ denotes the Markov chain approximation method developed in this paper. ‘‘MM0’’ denotes the version of the method that does not include the minimization step.  $\bar{N}$  stands for the number of grid points used for each component, i.e. for  $z$  and  $g$ .  $\hat{\sigma}_z^2$  and  $\hat{\sigma}_g^2$  denote the simulated unconditional variance of  $z$  and  $g$ .  $\rho_{z,g}$  is the correlation coefficient between  $z$  and  $g$ , while  $\hat{\zeta}_1$  and  $\hat{\zeta}_2$  are the eigenvalues of matrix  $\hat{A}$ . For each  $x \in \{z, g\}$ , the numbers in row  $10 \times \hat{\mu}_x$  are the weighted average of  $10 \times |\hat{\mu}_x(j) - \mu_x(j)|$  which uses the realized frequencies of states  $j = 1, 2, \dots, N^*$  as weights. The frequencies are constructed using a simulated process of length  $\tau = 1,800,000$ . Similarly, for each  $x \in \{z, g\}$ , the numbers in row  $\hat{\omega}_x^2/\omega_x^2$  are the weighted average of  $|\hat{\omega}_x^2(j)/\omega_x^2 - 1|$  which uses the same frequencies. The numbers that are smaller than 0.0005 are denoted by 0.000.

**Table 2:** Approximation Accuracy: Key Economic Variables

	$\bar{N} = 9$		$\bar{N} = 15$		$\bar{N} = 21$	
	Tau.	MM	Tau.	MM	Tau.	MM
<i>Volatility</i>						
Std( $\ln(Y_t)$ )	0.210	0.018	0.193	0.005	0.149	0.001
Std( $\ln(K_t)$ )	0.209	0.028	0.187	0.009	0.152	0.003
Std( $\ln(C_t)$ )	0.283	0.015	0.189	0.003	0.151	0.003
Std( $\ln(G_t)$ )	0.167	0.009	0.135	-0.003	0.092	-0.004
<i>Persistence</i>						
1 – Corr( $\ln(Y_t), \ln(Y_{t-1})$ )	-0.391	-0.034	-0.070	-0.012	-0.012	0.000
1 – Corr( $\ln(K_t), \ln(K_{t-1})$ )	-0.250	0.000	0.000	0.000	0.000	0.000
1 – Corr( $\ln(C_t), \ln(C_{t-1})$ )	-0.194	-0.048	-0.049	-0.016	-0.066	-0.016
1 – Corr( $\ln(G_t), \ln(G_{t-1})$ )	-0.083	-0.018	0.015	0.004	0.007	0.007
<i>Correlation with current output</i>						
1 – Corr( $\ln(Y_t), \ln(K_t)$ )	-0.245	-0.035	-0.080	-0.011	-0.043	-0.008
1 – Corr( $\ln(Y_t), \ln(C_t)$ )	-0.143	-0.023	-0.114	-0.014	-0.094	-0.010
1 – Corr( $\ln(Y_t), \ln(G_t)$ )	0.039	-0.014	-0.025	-0.005	0.001	0.005
<i>Correlation with lead output</i>						
1 – Corr( $\ln(Y_{t+1}), \ln(K_t)$ )	-0.237	-0.035	-0.082	-0.011	-0.044	-0.007
1 – Corr( $\ln(Y_{t+1}), \ln(C_t)$ )	-0.139	-0.024	-0.112	-0.015	-0.092	-0.009
1 – Corr( $\ln(Y_{t+1}), \ln(G_t)$ )	0.040	-0.014	-0.026	-0.005	0.000	0.005

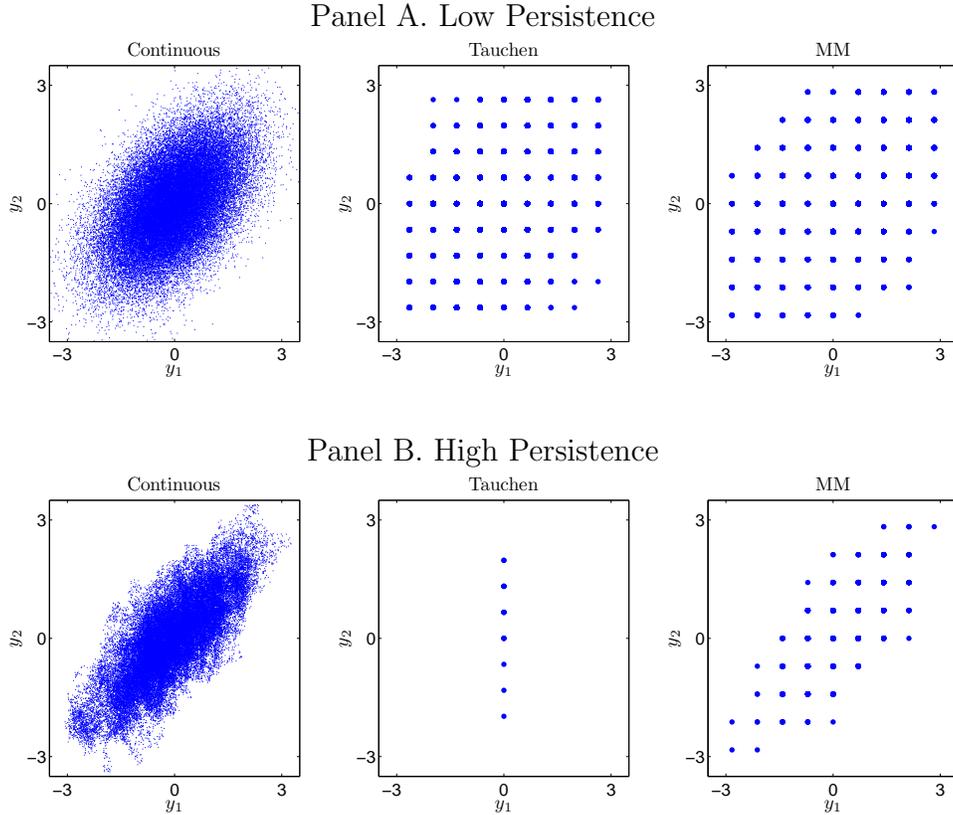
*Notes.* The approximation measure is calculated as the percentage difference between the moments generated by the discrete and continuous methods. For example, the numbers close to zero mean a better approximation. When evaluating the approximation quality of a correlation coefficient, say  $\text{Corr}(\ln(Y_t), \ln(Y_{t-1}))$ , we consider  $1 - \text{Corr}(\ln(Y_t), \ln(Y_{t-1}))$ . This is for the purpose of capturing the differences between correlation coefficients that are inherently very close to one. The numbers that are smaller than 0.0005 (i.e. approximation errors less than 0.05%) in absolute terms are denoted by 0.000 with their appropriate signs.

**Table 3:** Moments of Economic Variables and GMM Estimates

	$T = 120$			$T = 240$		
	Cont.	Tau.	MM	Cont.	Tau.	MM
<i>Moments</i>						
$\text{std}(C_{t+1}/C_t)$	0.0081	0.0093	0.0080	0.0082	0.0094	0.0081
$\text{std}(Y_{t+1}/K_{t+1})$	0.0031	0.0031	0.0031	0.0036	0.0036	0.0036
$\text{Corr}(C_{t+1}/C_t, Y_{t+1}/K_{t+1})$	0.3347	0.2994	0.3342	0.3203	0.2849	0.3206
<i>Estimates of structural parameters</i>						
$\beta = 0.986$	0.9827 (0.0130)	0.9819 (0.0140)	0.9821 (0.0129)	0.9825 (0.0112)	0.9820 (0.0121)	0.9823 (0.0111)
$\alpha = 0.283$	0.3162 (0.1178)	0.3219 (0.1263)	0.3195 (0.1162)	0.3176 (0.1017)	0.3207 (0.1094)	0.3181 (0.1002)

*Notes.* Standard deviations are in parentheses.

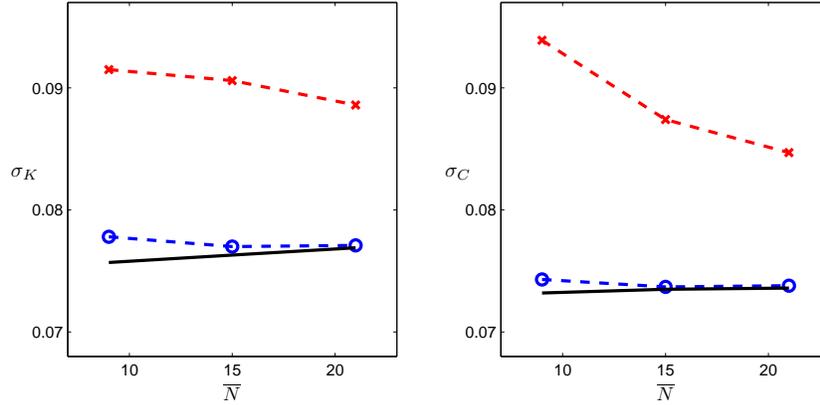
**Figure 1:** Continuous Process and Discrete Approximation Methods



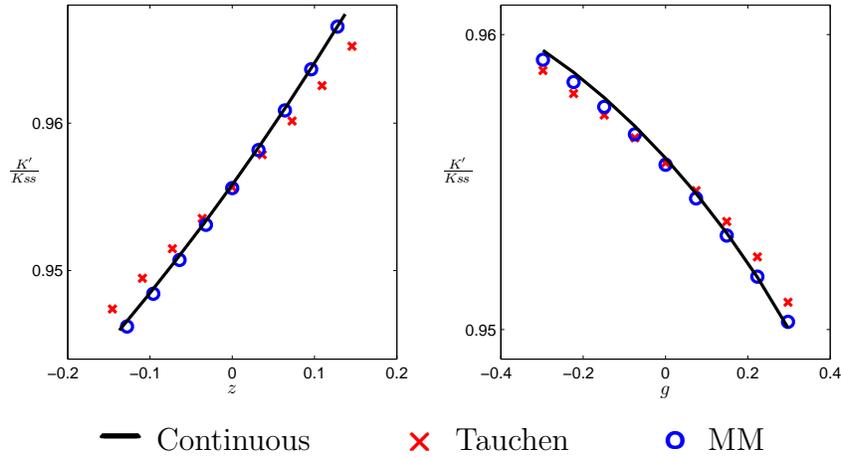
*Notes.* The left column of the figure plots the scatter diagram of  $y_{1,t}$  and  $y_{2,t}$  in equation (1) ( $M = 2$ ) for cases of low and high persistence. The middle and right columns plot the scatter diagrams of  $\tilde{y}_{1,t}$  and  $\tilde{y}_{2,t}$  using Tauchen's and MM methods, respectively. The number of grid points used for each component is nine, i.e.,  $N_i = 9$ ,  $i \in \{1, 2\}$ . The less persistent case is the VAR process considered by Tauchen (1986a). The more persistent case is obtained by replacing the matrix  $A$  of the less persistent case (see equation (1)) by  $A_h$  where  $A_h^{100} = A$ . In both cases, the true variances of  $y_{1,t}$  and  $y_{2,t}$  are normalized to unity. The diagrams are based on a series of 50,000 periods. For legibility purposes, heavier dots are used for the discrete methods.

**Figure 2:** Stochastic Growth Model

Panel A. Volatility



Panel B. Policy Rule



*Notes.* Panel A plots the volatility of the logarithm of the capital stock and consumption. The horizontal axis is the number of grid points used for each of the two underlying shocks. The moments are calculated using simulated series of  $\tau = 2,000,000$  periods. Panel B shows the policy rules obtained under different solution methods for a set of states. For all cases considered here, the current capital stock,  $K$ , is set to the steady state capital stock  $K_{ss}$ . The lower left diagram shows the optimal capital stock,  $K'$ , (relative to  $K_{ss}$ ) as a function of the technology shock  $z$  when  $g = 0$ . Similarly, the lower right diagram shows the same variable as a function of the current expenditure shock  $g$  while setting  $z = 0$ .