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

Nikolay Gospodinov
Concordia University and CIREQ

Damba Lkhagvasuren†
Concordia University and CIREQ

September 7, 2012

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.

Keywords: Markov Chain, Vector Autoregressive Processes, Numerical Methods, Moment Matching, Non-Linear Stochastic Dynamic Models, Solving Functional Equation, Discretization of State Space

JEL Codes: C15, C53, C60

*We thank Paul Gomme for helpful comments.

†Corresponding author: Damba Lkhagvasuren, Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8 Canada; Tel: (514) 848 2424 (ext. 5726); Fax: (514) 848-4536; E-mail: damba.lkhagvasuren@concordia.ca.

1 Introduction

The finite-state Markov chain approximation methods developed by Tauchen (1986a) and Tauchen and Hussey (1991) are widely used in solving functional equations where the state variables follow autoregressive processes. Nonlinear dynamic macroeconomic and asset pricing models often imply a set of integral equations (moment conditions) that do not admit explicit solutions. Discrete-valued approximations prove to be an effective tool for reducing the complexity of the problem (Burnside, 1999). Also, there is a substantial interest in these methods for generating simulation data from nonlinear dynamic models in evaluating the sampling properties of generalized method of moments estimators (Tauchen, 1986b; Hansen, Heaton and Yaron, 1996; Stock and Wright, 2000; among others). The Markov-chain approximation methods choose discrete values for the state variables and construct transition probabilities such that the characteristics of the generated process mimic those of the underlying process. However, both Tauchen (1986a) and Tauchen and Hussey (1991) point out that these methods do not perform well for highly persistent autoregressive (AR) processes or processes with characteristic roots close to unity. Although these methods can generate a better approximation at the cost of a finer state space, this type of approach is not always feasible.[1]

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

[1]See Burnside (1999) for how rapidly the computational cost increases with the number of states, and how severe this curse of dimensionality is in the vector autoregressive case. More importantly, as we show in Proposition 1 below, these existing methods cannot always generate a meaningful approximation even when the number of states is very large.

(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). However, this method is developed for a particular class of multivariate autoregressive processes: correlated AR(1) shocks, i.e., a set of AR(1) shocks whose innovation terms are correlated with each other. 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 express the Markov-chain transitional probabilities as the solution of a nonparametric (empirical likelihood) problem subject to moment restrictions. 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 procedures of Tauchen (1986a) and Tauchen and Hussey (1991) to highly persistent cases and those of Rouwenhorst (1995) and 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 introduces the continuous- and discrete-valued versions of the multivariate model and the main notation. Section 3 reviews the existing approximation methods and demonstrates that they fail to deliver a reasonable approximation as the roots of the continuous-valued process approach the unit circle. The reason for this is that the existing methods calculate transition probabilities defined over discrete grids using continuous probability density functions. Therefore, the quality of the approximation deteriorates sharply when the standard deviation of the error terms becomes comparable to or smaller than the distance between the grid points. Our approximation method is introduced in Section 4. 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 5 investigates the numerical properties of the method in a bivariate VAR(1) process with varying degrees of persistence. Section 6 concludes. The proofs and some additional theoretical results are presented in Appendices A to C.

2 Model

In this section we present the underlying continuous-valued vector autoregressive process and introduce the main structure and notation for the finite-state Markov chain used for approximating the continuous process.

2.1 Continuous VAR process

Let $\mathbf{y}_t$ be an $M \times 1$ vector containing the values that variables, $y_1, y_2, \cdots, y_M$, 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 $\Omega = \text{diag}(\omega_1^2, \omega_2^2, ..., \omega_M^2)$ being a diagonal matrix. Extending the analysis to a non-diagonal Ω is relatively straightforward and is discussed later in the paper.[2] 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 Σ be the unconditional covariance matrix of the process $\mathbf{y}_t$. Let σ_i denote the unconditional standard deviation of y_i for each i. Then, the i-th diagonal element of Σ is given by σ_i^2.

[2]Results for non-Gaussian errors, that target also the conditional skewness and kurtosis of the underlying process, are presented in Appendix C. Since 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) assumption is routinely used in describing the properties of the shocks in macroeconomic models, the current version of the paper presents the construction of the finite-state Markov chain only for this benchmark case. We should also note that regardless of the true values of skewness and kurtosis of the error terms, the unconditional skewness and kurtosis of the process converge to those of the normal distribution when the persistence approaches the nonstationary boundary.

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 $\bar{y}_i^{(1)}, \bar{y}_i^{(2)}, \cdots, \bar{y}_i^{(N_i)}$. Therefore, at each point in time, the entire system will be in one of the $N^* = N_1 \times N_2 \times \cdots \times N_M$ states. Let $\bar{y}^{(1)}, \bar{y}^{(2)}, \cdots, \bar{y}^{(N^*)}$ label these N^* states and Π 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 to calculate the associated transition probability matrix Π so that the characteristics of the generated process closely mimic those of the underlying process $\mathbf{y}$.

Define

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

for $i = 1, 2, \cdots, M$, $l = 1, 2, \cdots, N_i$ and $j = 1, 2, \cdots, N^*$. 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:

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

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

$$H_i(j) = \{h_i(j,1), h_i(j,2), \cdots, h_i(j,N_i)\} \tag{4}$$

over the grid points $\bar{y}_i^{(1)}, \bar{y}_i^{(2)}, \cdots, \bar{y}_i^{(N_i)}$. 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 $\bar{y}_i^{(1)}$, $\bar{y}_i^{(2)}$, $\cdots$, $\bar{y}_i^{(N_i)}$.

The problem of determining the probability weights associated with this probability mass distribution can be expressed as a nonparametric likelihood problem. In particular, given the grids $(\bar{y}^{(1)}, \bar{y}^{(2)}, \cdots, \bar{y}^{(N^*)})$, the nonparametric (or empirical) likelihood estimate of the transition probability matrix can be obtained as the solution to the constrained maximization problem:

$$\max_{h_i(j,l)\in[0,1]} \sum_{i=1}^{M} \sum_{j=1}^{N^*} \sum_{l=1}^{N_i} \ln(h_i(j,l)) \tag{5}$$

subject to the constraint

$$\sum_{l=1}^{N_i} h_i(j,l) = 1 \text{ for } i = 1, \cdots, M. \tag{6}$$

To avoid trivial solutions, this optimization problem needs to be augmented with additional restrictions that best describe the statistical properties of model (1) for $\mathbf{y}_t$. 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.,

$$\mu_i(j) = a_{i,1}\bar{y}_1^{(L_1(j))} + a_{i,2}\bar{y}_2^{(L_2(j))} + \cdots + a_{i,M}\bar{y}_M^{(L_M(j))}. \tag{7}$$

The new method that we propose below targets the first and second conditional moments of the process $\mathbf{y}_t$ by imposing the following restrictions:

$$\sum_{l=1}^{N_i} h_i(j,l)\bar{y}_i^{(l)} = \mu_i(j) \tag{8}$$

and

$$\sum_{l=1}^{N_i} h_i(j,l)(\bar{y}_i^{(l)} - \mu_i(j))^2 = \omega_i^2 \tag{9}$$

for $i = 1, \cdots, M$ and $j = 1, \cdots, N^*$. Equations (8) and (9) require that the Markov

chain adequately approximate the conditional mean and variance of the continuous-valued process $\mathbf{y}_t$.

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).

The construction of the transition probabilities and the Markov chain for Tauchen's (1986a) method can be described as follows. For each i, Tauchen (1986a) chooses equispaced grid points over the interval $[-m\sigma_i, m\sigma_i]$ for some $m > 0$,[3] where σ_i denotes the unconditional standard deviation of y_i. Specifically, for each i, the grid points are chosen according to the following rule:

$$\bar{y}_i^{(l)} = -m\sigma_i + (l-1)\triangle_i, \tag{10}$$

where

$$\triangle_i = 2m\sigma_i/(N_i - 1) \tag{11}$$

[3]According to Tauchen (1986b), $m = 3$ works well in practice. Footnote 4 below discusses how the value of m affects conditional and unconditional variances differently.

and $l = 1, 2, \cdots, N_i$. Note that $\triangle_i$ measures the distance between two consecutive nodes of $\tilde{y}_i$.[4]

Given the above grid points, consider the following partition of the real line for each i: $C_i^{(1)} =]-\infty, \bar{y}_i^{(1)} + \triangle_i/2]$, $C_i^{(N_i)} =]\bar{y}_i^{(N_i)} - \triangle_i/2, \infty[$, and $C_i^{(l)} =]\bar{y}_i^{(l)} - \triangle_i/2, \bar{y}_i^{(l)} + \triangle_i/2]$, where $l = 2, 3, \cdots, N_i - 1$. Tauchen (1986a) calculates the transition probabilities as

$$h_i(j,l) = \Pr\left(\mu_i(j) + \varepsilon_i \in C_i^{(l)}\right). \tag{12}$$

Denoting the cumulative distribution function of the standard normal variable ε_i/ω_i by Φ_i, equation (12) can be rewritten as

$$h_i(j,l) = \begin{cases} \Phi_i\left(\frac{\bar{y}_i^{(1)} - \mu_i(j) + \triangle_i/2}{\omega_i}\right) & \text{if } l = 1, \\ 1 - \Phi_i\left(\frac{\bar{y}_i^{(N_i)} - \mu_i(j) - \triangle_i/2}{\omega_i}\right) & \text{if } l = N_i, \\ \Phi_i\left(\frac{\bar{y}_i^{(l)} - \mu_i(j) + \triangle_i/2}{\omega_i}\right) - \Phi_i\left(\frac{\bar{y}_i^{(l)} - \mu_i(j) - \triangle_i/2}{\omega_i}\right) & \text{otherwise.} \end{cases} \tag{13}$$

According to Tauchen (1986a), the rationale for equations (7) and (12) is that if the partitioning $\left(C_i^{(1)}, C_i^{(2)}, \cdots, C_i^{(M)}\right)$ is reasonably fine, then the conditional distribution of $\tilde{y}_{i,t}$ given state j at time $t-1$ will approximate closely (in the sense of weak convergence) the conditional distribution of $y_{i,t}$ given $y_{i,t-1} = \mu_i(j)$.

[4]Note that there are two free parameters that underlie the approximation accuracy of Tauchen's method: the number of grid points N_i and the parameter m which is positively related to the distance between the grid points. First, while the quality of Tauchen's approximation improves as the number grid points N_i increases, this type of approach is not always feasible as stated in Proposition 1 below. Second, the choice of parameter m 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 (Kopecky and Suen (2010)). If the value of m is too small (say $m = 2$), the resulting truncation imposed by $[-m\sigma_i, m\sigma_i]$ can be quite severe and Tauchen (1986a)'s 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 is too large, the distance between the grid points increases (see equation (11)) which reduces the quality of approximating the conditional variance of the underlying process. It should be noted that our proposed method breaks the tight link between the conditional and unconditional variance inherent in the existing finite-state VAR methods.

Given the finite-state Markov chain $\tilde{\mathbf{y}}_t$, let

$$\tilde{\boldsymbol{\varepsilon}}_t = \tilde{\mathbf{y}}_t - A\tilde{\mathbf{y}}_{t-1}, \tag{14}$$

$\tilde{\Omega}$ be the covariance matrix of $\tilde{\boldsymbol{\varepsilon}}$ and $\tilde{\omega}_i$ denote the square root of the i-th diagonal element of this matrix. Since the conditional probabilities for this Markov chain are obtained by centering the density of $\boldsymbol{\varepsilon}$ around $A\tilde{\mathbf{y}}_{t-1}$, we have

$$\mathrm{E}(\tilde{\mathbf{y}}_t|\tilde{\mathbf{y}}_{t-1}, \tilde{\mathbf{y}}_{t-2}, \cdots) = A\tilde{\mathbf{y}}_{t-1}, \tag{15}$$

and

$$\mathrm{E}(\tilde{\boldsymbol{\varepsilon}}_t) = E\{E[(\tilde{\mathbf{y}}_t - A\tilde{\mathbf{y}}_{t-1})|\tilde{\mathbf{y}}_{t-1}, \tilde{\mathbf{y}}_{t-2}, \cdots]\} = 0_M. \tag{16}$$

This implies, by the law of iterated expectations, that $\tilde{\boldsymbol{\varepsilon}}_t$ is uncorrelated with $\tilde{\mathbf{y}}_{t-s}$ for $s = 1, 2, \cdots$. However, the conditional covariance matrix of $\tilde{\boldsymbol{\varepsilon}}_t$, $\mathrm{Var}(\tilde{\boldsymbol{\varepsilon}}_t|\tilde{\mathbf{y}}_{t-1}, \tilde{\mathbf{y}}_{t-2}, \cdots)$, depends on $\tilde{\mathbf{y}}_{t-1}$ and thus, $\tilde{\boldsymbol{\varepsilon}}_t$ and $\tilde{\mathbf{y}}_{t-1}$ are dependent (Anderson, 1989). This clearly suggests that targeting the first and second conditional moments will improve the quality of the Markov-chain approximation and it serves as a main motivation for the new method proposed in this paper.

We now show that calculating the transition probabilities using the continuous distribution functions does not always deliver meaningful approximations.

Proposition 1. *Let $\tilde{\omega}_i^2$ denote the conditional variance of the i-th element of $\tilde{\boldsymbol{\varepsilon}}$ in equation (14), where $\tilde{\mathbf{y}}_t$ is the finite-state Markov chain constructed using the Tauchen's (1986a) method with the standard normal CDF Φ_i. Then, for any set of integers $(N_1, N_2, \cdots, N_M)$ and any arbitrarily small positive number ϵ, there always exists a highly persistent vector autoregressive process for which $\tilde{\omega}_i/\omega_i < \epsilon$ for all i.*

Proof. See Appendix A. □

Proposition 1 is an extension of the results in Galindev and Lkhagvasuren (2010). The main implication of the result in Proposition 1 is that Tauchen's (1986a) method will fail 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 the method targets only the first conditional moment of the continuous-valued process $\mathbf{y}_t$.

Moreover, despite some numerical and methodological differences across the existing Markov-chain approximations, all these methods suffer from the same problem as in Tauchen (1986a) 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 thus one can directly extend Proposition 1 to these methods.

4 A Moment-Matching Markov Chain Method

4.1 Main idea

Unlike the existing finite-state methods for multivariate processes that calculate the transition probabilities using the conditional distribution function of $\mathbf{y}_t$, our proposed method chooses the transition probabilities by targeting the key conditional moments of $\mathbf{y}_t$. In this respect, we give our method a moment-matching interpretation and refer to it as moment-matching (MM) method. More specifically, it approximates the underlying process by targeting the $2M \times N^*$ moment conditions given by equations (8) and (9). This means that the method chooses the grid points $\{\bar{y}_i^{(1)}, \bar{y}_i^{(2)}, \cdots, \bar{y}_i^{(N_i)}\}_{i=1}^{M}$ and the associated probability mass functions $\{H_i(1), H_i(2), \cdots, H_i(N^*)\}_{i=1}^{M}$ such that, for all i and j, the mean and the variance of distribution $H_i(j)$ target $\mu_i(j)$ and ω_i^2, respectively.

The grid points and the probability mass functions are constructed by mixing a set of probability mass functions associated with the conditional distributions of the finite-state processes generated by the method of Rouwenhorst (1995). The reason for this particular choice of probability mass functions is that their mean and variance (or the conditional mean and variance of the Rouwenhorst process) are perfectly matched.[5] While explicitly incorporating information about the conditional moments is expected to deliver efficiency gains compared to Tauchen's (1986a) method over the whole permissible parameter range, this method has some appealing properties when the underlying process is highly persistent or near unit root.

It should be stressed that our MM method does not treat the underlying VAR process as a collection of M univariate AR processes and then approximates these scalar processes by the method of Rouwenhorst (1995). While this approach might seem to be a natural way to extend the method of Rouwenhorst (1995) to a multivariate setting, it introduces $M-1$ additional continuous variables as shown in Galindev and Lkhagvasuren (2010). In contrast, our proposed method generates a set of discrete distributions (not a set of discrete processes) using the Rouwenhorst (1995) method and mix these distributions to target 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.

Furthermore, unlike the method by Galindev and Lkhagvasuren (2010) which cannot (unless the shocks have the same persistence) generate a finite-state Markov chain

[5] In particular, for an AR(1) process $y_t = \rho y_{t-1} + \varepsilon_t$, where $|\rho| < 1$, ε_t is i.i.d. $\mathcal{N}(0, (1-\rho^2)\sigma^2)$ with $\sigma^2 = \text{Var}(y_t)$, it can be shown that $\text{E}(\tilde{y}_t) = 0$ and $\text{Var}(\tilde{y}_t) = \sigma^2$. More importantly, the conditional mean and variance of the Markov chain is also independent of the number of grid points: $\text{E}(\tilde{y}_t | \tilde{y}_{t-1} = \bar{y}^{(k)}) = \rho \bar{y}^{(k)}$ and $\text{Var}(\tilde{y}_t | \tilde{y}_{t-1} = \bar{y}^{(k)}) = (1-\rho^2)\sigma^2$. This stands in sharp contrast with the existing methods (including Tauchen's (1986a) method) which are very sensitive to the number of grid points in approximating near unit root processes.

when adapted for approximating vector autoregressions,[6] the moment-matching method developed in this paper yields a finite-state approximation regardless of the degree of persistence of the different components of $\mathbf{y}$ while using only M state variables (i.e., $\tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_M$). This could potentially offer substantial computational gains when solving functional equations. Finally, the flexibility of the MM method allows us to generalize it easily to more complex setups such as nonlinear and non-Gaussian VAR processes.

4.2 Probability mass functions

Before constructing the Markov chain, let us introduce the following notation. Consider a zero-mean AR(1) process with a persistence parameter r and unconditional standard deviation s. Let $\tilde{x}(n, r, s)$ be the n-state symmetric Markov chain process constructed by the method of Rouwenhorst (1995) to approximate the AR(1) process. Let $\bar{x}(n, s) = \{\bar{x}^{(1)}(n, s), \bar{x}^{(2)}(n, s), \cdots, \bar{x}^{(n)}(n, s)\}$ denote the grid points and $P(n, r)$ be the probability transition matrix of $\tilde{x}(n, r, s)$. Suppose that the [row k, column l] element of $P(n, r)$, $p_{k,l}(n, r)$, denotes the probability that the n-state process switches from $\bar{x}_i^{(k)}(n, s)$ to $\bar{x}_i^{(l)}(n, s)$.

Now consider the k-th row of $P(n, r)$,

$$p_k(n, r) = \{p_{k,1}(n, r), p_{k,2}(n, r), \cdots, p_{k,n}(n, r)\}, \tag{17}$$

where $1 \leq k \leq n$. The key observation is that this row can be interpreted as a probability mass function associated with the nodes $\bar{x}(n, s) = \{\bar{x}^{(1)}(n, s), \bar{x}^{(2)}(n, s), \cdots, \bar{x}^{(n)}(n, s)\}$.

[6] In general, in order to approximate an M-variate process $\mathbf{y}_t$ given by equation (1), the method by Galindev and Lkhagvasuren (2010) uses $2M - 1$ state variables, of which $M - 1$ are continuous. Therefore, their method requires an additional approximation step, $M - 1$ dimensional numerical interpolation, when applied to functional equations. As a result, there is no explicit transition matrix and even the implicit transition probabilities are model-dependent. From a practical point of view, the method by Galindev and Lkhagvasuren (2010) is difficult to implement due to the large dimensionality of the problem and the continuous nature of the additional variables which hampers the use of matrix operations in solving functional equations.

The mean and the variance of the probability mass distribution are $r\bar{x}^{(k)}$ and $(1-r^2)s^2$, respectively. On the other hand, the grid points $\bar{x}(n,s)$ and the transition matrix $P(n,r)$ are analytically related to the input variables, n, r, and s.[7] 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. We can now construct the Markov chain of the VAR process in equation (1) by mixing these univariate probability mass functions.

4.3 Markov chain construction

4.3.1 Grid points

For each $i \in \{1, 2, \cdots, M\}$, the grid points of $\{\tilde{y}_i^1, \tilde{y}_i^2, \cdots, \tilde{y}_i^{N_i}\}$ are given by $\bar{x}(N_i, \sigma_i)$. Specifically,

$$\bar{y}_i^{(k)} = -\sigma_i\sqrt{N_i - 1} + 2\sigma_i\frac{k-1}{\sqrt{N_i - 1}} \tag{18}$$

for $k = 1, 2, \cdots, N_i$.

4.3.2 Transition probabilities

Let $\rho_i = \sqrt{1 - \omega_i^2/\sigma_i^2}$. Given i, the following four *distinct* cases are considered for each $j \in \{1, 2, \cdots, N^*\}$:

1. If $\mu_i(j) < \rho_i\bar{y}_i^{(1)}$, set $H_i(j) \equiv p_1(N_i, \rho_i)$.

2. If $\mu_i(j) > \rho_i\bar{y}_i^{(N_i)}$, set $H_i(j) \equiv p_{N_i}(N_i, \rho_i)$. In these two cases, the conditional variance ω_i^2 is matched while the conditional mean attains the value closest to $\mu_i(j)$ given the grid points.

[7] When $n = 2$, the probability transition matrix is given by $P(2,r) = \begin{pmatrix} (1+r)/2 & (1-r)/2 \\ (1-r)/2 & (1+r)/2 \end{pmatrix}$. For higher values of n, the transition probability matrix is constructed recursively using the elements of $P(2,r)$. The grid points $\{\bar{x}^{(1)}(n,s), \bar{x}^{(2)}(n,s), \cdots, \bar{x}^{(n)}(n,s)\}$ are given by n equally-spaced points on the interval $[-s\sqrt{n-1}, s\sqrt{n-1}]$. See Rouwenhorst (1995) for details.

3. If $\mu_i(j) = \rho_i \bar{y}_i^{(k)}$ for some k, set $H_i(j) \equiv p_k(N_i, \rho_i)$. In this case, both the conditional mean $\mu_i(j)$ and conditional variance ω_i^2 are matched.

4. If any of the above three conditions is not met, there must be an integer k such that

$$\rho_i \bar{y}_i^{(k)} < \mu_i(j) < \rho_i \bar{y}_i^{(k+1)}. \tag{19}$$

Then, consider the following mixture distribution on $\{\bar{y}_i^{(1)}, \bar{y}_i^{(2)}, \cdots, \bar{y}_i^{(N_i)}\}$:

$$\tilde{p}_k(N_i, \rho_i) = \lambda p_k(N_i, \rho_i) + (1 - \lambda) p_{k+1}(N_i, \rho_i), \tag{20}$$

where $\lambda = \frac{\rho_i \bar{y}_i^{(k+1)} - \mu_i(j)}{\rho_i \bar{y}_i^{(k+1)} - \rho_i \bar{y}_i^{(k)}}$. The mean and variance of this mixture distribution are, respectively, $\tilde{\mu}_k(\rho_i) = \mu_i(j)$ and

$$\tilde{\omega}_k^2(\rho_i) = \omega_i^2 + \rho_i^2 \lambda (1 - \lambda) \left(\bar{y}_i^{(k+1)} - \bar{y}_i^{(k)} \right)^2. \tag{21}$$

Since $0 < \lambda < 1$, the second term on the right hand side is positive. Therefore, although the mean of the mixture distribution $\tilde{p}_k(N_i, \rho_i)$ hits the target $\mu_i(j)$, the variance of the distribution is greater than the targeted conditional variance ω_i^2. Given the grid points, $|\tilde{\omega}_k^2(\rho_i) - \omega_i^2|$ can be minimized by choosing a value higher than $\sqrt{1 - \omega_i^2/\sigma_i^2}$ for ρ_i in equation (20) while, if necessary, adjusting k to satisfy inequality (19).[8] Let $\tilde{p}_{k^*}(N_i, \rho_i^*)$ be the mixture distribution obtained in such a manner. Then, setting $H_i(j) \equiv \tilde{p}_{k^*}(N_i, \rho_i^*)$ matches the conditional mean while achieving the best possible value for the conditional variance.

[8] As equation (21) shows, the gap $|\tilde{\omega}_k^2(\rho_i) - \omega_i^2|$ decreases with the number of grids, N_i. Thus, given a sufficient number of grids, the above minimization may be redundant in certain cases. In fact, experimentation shows that for a moderate number of state space (e.g., $N_i = 9$), setting $H_i(j) \equiv \tilde{p}_k(N_i, \rho_i)$ already provides a reasonable quality of approximation.

Repeating this procedure for $i = 1, 2, \cdots, M$, we obtain the probability mass functions $\{H_i(1),\ H_i(2),\ \cdots,\ H_i(N^*)\}_{i=1}^{M}$. The asymptotic validity of the method for approximating conditional expectations of general nonlinear functions that often arise in economic models is discussed in Appendix B.

4.4 Extensions

While the procedure above is developed under the assumption of a diagonal covariance matrix Ω, 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

$$\mathbf{x}_t = b + B\mathbf{x}_{t-1} + \eta_t, \tag{22}$$

where η_t is i.i.d. $(0, \Psi)$ and Ψ is a non-diagonal matrix.[9] Let G be a lower triangular matrix such that $\Omega = G\Psi G^{-1}$ is a diagonal matrix. Define the transformations (Tauchen, 1986b),

$$\mathbf{x}_t \to G[\mathbf{y}_t - (I_M - B)^{-1}b], \tag{23}$$

$$B \to A = GBG^{-1}, \tag{24}$$

and

$$\eta_t \to G\varepsilon_t. \tag{25}$$

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 (22).

Furthermore, since any stationary AR(p) process can be expressed in a companion

[9] See also Terry and Knotek II (2011), who extend Tauchen (1986a) to processes with arbitrary positive-semidefinite covariance structures for the error term.

form as a VAR(1) process, our method effectively extends the method by Rouwenhorst (1995) to higher-order scalar autoregressive processes. Another appealing feature of our method is its flexibility which allows us to deal with possibly nonlinear VAR processes and targeting additional moments of the conditional distribution such as skewness and kurtosis.

Below we examine numerically how well the method works in terms of approximating autoregressive processes for various degrees of persistence of the discrete space. We show that the MM method outperforms the method by Tauchen (1986a), especially for processes whose characteristic roots are close to one.

5 Numerical Evaluation

For our main numerical evaluation, we consider the bivariate VAR(1) case ($M = 2$) with

$$\boldsymbol{\varepsilon}_t \sim i.i.d.\ \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}\right) \tag{26}$$

and $A = A_0^K$, where

$$A_0 = \begin{pmatrix} 0.995619 & 0.005335 \\ 0.003557 & 0.992063 \end{pmatrix} \tag{27}$$

and K is a positive integer set to 1, 10 and 100.[10] It is straightforward to see that higher values of K imply lower persistence. As in Tauchen (1986a), we choose nine grid points for each component: $\overline{N} = N_1 = N_2 = 9$. We also consider another case in which the state space is much finer: $\overline{N} = 19$. Given the trade-off between the approximation of

[10]The matrix A_0 is chosen for comparison purposes. Specifically, when $K = 100$,

$$A = A_0^{100} = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.5 \end{pmatrix}.$$

Therefore, the vector autoregressive process coincides with the one considered in Tauchen (1986a).

unconditional and conditional variance (see footnote 3), we set $m = 3$ which, according to Tauchen (1986b), works well in practice.

5.1 Approximation accuracy

Let $\{\widetilde{\mathbf{y}}_t\}_{t=1}^{\tau}$ denote the simulated time series either from the Markov chain approximation by Tauchen (1986a) or the method proposed in this paper. The accuracy of the two approximations can then be examined by estimating the key parameters of the initial process in equation (1). The parameters of interest are the unconditional variances of y_1 and y_2 (denoted by σ_1^2 and σ_2^2), the correlation coefficient between y_1 and y_2, and the persistence measures $1-\xi_1$ and $1-\xi_2$, where ξ_1 and ξ_2 are the two roots (eigenvalues) of matrix A. As in Tauchen (1986a) and Tauchen and Hussey (1991), the simulated counterpart of A, $\hat{A}$, is obtained by fitting the linear autoregressive model in equation (1) to $\{\widetilde{\mathbf{y}}_t\}_{t=1}^{\tau}$. The evaluation of the approximation accuracy is based on 1,000 Monte Carlo replications of length $\tau = 2,000,000$.[11] Tables 1 and 2 report the root mean squared errors (RMSE) as well as the biases and standard deviations 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 degrees of persistence. For example, for the least persistent case ($K = 100$), the relative bias for $\overline{N} = 9$ of the estimated $1-\xi_1$, σ_1^2 and σ_2^2, using data generated by Tauchen's (1986a) method, is 3.5%, 6.6% and 4.4%, respectively, whereas the corresponding biases for the MM method are 0.9%, -0.8% and -0.5%. For the moderate degree of persistence ($K = 10$), the biases for the method of Tauchen (1986a) become -19.3%, 35.6% and 28.7%, while those of the

[11] Note that the length of the time series is much larger than that considered by Tauchen (1986a). The main reason is that, for smaller number of observations, Tauchen's method fails to generate time-varying data for the examples considered here and, thus, renders the numerical evaluation of the methods impossible. Put differently, for shorter time series, the numerical results will be much more favorable for the method developed in this paper.

MM method remain almost constant at 1.7%, -0.7% and -0.9%, respectively. However, the advantages of our method become particularly striking for the high persistence case ($K = 1$). For this degree of persistence, Tauchen's (1986a) method fails to produce any time variation in the approximate Markov chain process, which is consistent with our theoretical results in Proposition 1. For example, the average probability of switching from the current state to any other state (with $\overline{N} = 9$) is only 0.03% for the method by Tauchen (1986a). This results in substantially large biases and inflated RMSEs for the parameters of interest. At the same time, our method continues to perform extremely well with very low biases and RMSEs. Increasing the number of grid points from 9 to 19 improves the performance of Tauchen's (1986a) method in the less persistent cases but its numerical properties in the highly persistent case remain rather poor.

Kopecky and Suen (2010) prove that the invariant distribution of the Markov chain constructed by Rouwenhorst's (1995) method is a binomial distribution. A direct consequence of this result is that the invariant distribution of the Markov chain constructed by Rouwenhorst's (1995) method converges asymptotically (as the number of states goes to infinity) to a normal distribution. This is not surprising because the method by Rouwenhorst (1995) targets only the first two conditional moments of the underlying process. Therefore, it might be instructive to see how our method and Tauchen's (1986a) method approximate the higher-order moments (skewness and excess kurtosis) of the continuously-valued process. The results (not reported here to conserve space) show that the higher-order moments of the two methods do not differ much when the persistence is low. When persistence is high, the MM method outperforms (often substantially) Tauchen's method by generating skewness and excess kurtosis much closer to their true values. In highly persistent cases, the method by Tauchen (1986a) often fails to generate any variation in some of the components of $\tilde{y}$ (see Proposition 1) and thus their higher-order moments are not defined.

5.2 Conditional moments

The evaluation of the approximation accuracy in Tables 1 and 2 is based on unconditional moments of the underlying and simulated processes. 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 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 $i \in \{1, 2, \cdots, M\}$ and $j \in \{1, 2, \cdots, N^*\}$. Then, for each i and j, the distances between the targeted and the generated conditional moments can be measured 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 the 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 $\tau = 2,000,000$. The results are presented in Table 3 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 conditional moments. This numerical finding lends support to our theoretical result in Proposition 1.

5.3 Solving functional equations

Next, we consider the performance of the MM and Tauchen's methods for solving functional equations. For ease of presentation, we consider the following simple model.[12]

[12]For studies that use a finite-state Markov chains for solving a dynamic model with a multivariate autoregressive processes, see, for example, Tauchen (1986b), Tauchen and Hussey (1991), Burnside (1999) and Bayer and Juessen (2012).

Consider a firm whose per-period profit is denoted by z. Suppose that each period the firm is dissolved at an exogenous rate $d < 1$. Then, the firm's (effective) time discount factor is given by $\delta = \frac{1-d}{1+r}$, where $r > 0$ is the interest rate. We assume that z and δ follow a bivariate VAR(1) process denoted by F, i.e.

$$\Pr(z_{t+1} < z', \delta_{t+1} < \delta' | z_t = z, \delta_t = \delta) = F(z', \delta' | z, \delta). \tag{28}$$

Then, given the current state of the firm (z, δ), the expected present value of its profits is defined by the following functional equation:

$$V(z, \delta) = z + \delta \int V(z', \delta') dF(z', \delta' | z, \delta). \tag{29}$$

In the numerical implementation, we assume that $z = y_1$ and $\delta = \frac{1}{1+\exp(y_2)}$, where y_1 and y_2 evolve according to equation (1). The following values are considered for the parameters of equation (1): $a_{1,1} = 0.990$, $a_{1,2} = 0$, $a_{2,1} = 0.083$, $a_{2,2} = 0.95$ and $\omega_{1,1} = \omega_{2,2} = 1$. After solving equation (29) using the transition matrices constructed by the two methods, we simulate the time series of V for $\tau = 2,000,000$ periods. To evaluate the quality of the approximation, we compute the mean, standard deviation and autocorrelation of $\{V_t\}_{t=1}^{\tau}$. The results are presented in Figure 1. Figure 1 shows that the MM method is less sensitive to the number of grid points compared with Tauchen's method. More importantly, as the number of grid points increases, the mean, standard deviation and autocorrelation of $\{V_t\}_{t=1}^{\tau}$ generated by Tauchen's method approach those obtained by the MM method. This suggests that the numerical solution to the functional equation obtained by the MM method describes more accurately the underlying dynamics of the value of the firm.

5.4 Computational implications

These numerical results have important implications for the computational costs associated with the different approximation methods. First, it should be noted that for the purpose of demonstrating how the quality of the approximation changes as the number of states increases, we focus on the process that can be approximated by Tauchen's method for the grid points considered here. Therefore, for the processes that have roots much closer to (but less than) one, Tauchen's method requires an extremely large number of states in order to achieve the level of accuracy comparable to that obtained by the new method with far fewer observations. This means that, under Tauchen's method, as the persistence goes up, the number of grid points must increase sharply. Consequently, the computation involved becomes more time consuming or infeasible (see Proposition 1). For example, given equation (1), the computation time for solving linear equations is approximately proportional to $(N^1 \times N^2 \times \cdots \times N^M)^3$ (Burnside, 1999). Therefore, depending on the exact nature of the problem, this curse of dimensionality will be 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 1 shows that the quality of the approximation obtained by the MM method using $N_1 = N_2 = 9$ is much higher than the one obtained by Tauchen's method using $N_1 = N_2 = 49$. More importantly, unlike Tauchen's method, the MM method can always generate a time-varying process.

6 Conclusion

This paper proposes a new 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 a set of conditional moments of the underlying process rather than

calculating the transition probabilities directly from an assumed distribution, centered around the first conditional moment, as in the existing methods. Our moment-matching method yields accurate approximations for a wide range of the parameter space, without relying on a large number of grid points for the state variables. The improved approximation accuracy of the proposed method is expected to have important quantitative implications for solving dynamic stochastic models as well as multivariate functional equations.

APPENDICES

A Proof of Proposition 1

Let $(\tilde{\mathbf{y}}_1^{(n)}, \tilde{\mathbf{y}}_2^{(n)}, \cdots, \tilde{\mathbf{y}}_t^{(n)}, \cdots, \tilde{\mathbf{y}}_T^{(n)})$ be a realization of the n-state Markov chain of length T approximated over n grid points. In what follows, we keep n fixed and perform the analysis as $T \to \infty$. The VAR model that describes the dynamics of the underlying continuous-valued process is given by

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

where $\boldsymbol{\varepsilon}_t$ is i.i.d. $\mathcal{N}(0, \Omega)$, Ω is a diagonal matrix with an i-th diagonal element ω_i^2 and Σ is the unconditional covariance matrix of $\mathbf{y}_t$ with an i-th diagonal element σ_i^2. Since we are interested in the behavior of highly persistent processes, it is convenient to reparameterize the matrix A as local-to-unity (see Phillips, 1987, for example). In particular, the matrix A is reparameterized as a function of T as (Elliott, 1998)

$$A = I_M - \frac{C}{T}, \tag{A.2}$$

where $C = \text{diag}(c_1, c_2, ..., c_M)$ with $c_i > 0$ being fixed constants for all $i = 1, ..., M$.[13] This is an artificial statistical device in which the parameter space for each individual process is a shrinking neighborhood of one as T increases. This parameterization proves to be very useful for studying the properties of strongly dependent processes as $T \to \infty$.

First note that using this reparameterization, the innovation variance matrix for the

[13] We can also allow for non-zero off-diagonal elements of C (see Gospodinov, Maynard and Pesavento, 2011) provided that this does not induce nonstationarity and preserves the stability of the process. The proof that we present below goes through for this more general specification but at the cost of more complicated notation.

continuous-valued process can be expressed as

$$\Omega = \frac{C\Sigma + \Sigma C'}{T} - \frac{C\Sigma C'}{T^2}. \tag{A.3}$$

and the variance for the i-th innovation is

$$\omega_i^2 = \frac{2c_i\sigma_i^2}{T} - \frac{c_i^2\sigma_i^2}{T^2}. \tag{A.4}$$

For Tauchen's (1986a) method, the probability that the process y_i switches from state j (corresponding to grid point $\bar{y}_i^{(j)}$) to any other state is given by

$$1 - \pi_{j,j}^{(i)} = 1 - \Pr\left(\left|\varepsilon_i - \frac{c_i\bar{y}_i^{(j)}}{T}\right| \leq 2\triangle_i\right), \tag{A.5}$$

where $\pi_{j,j}^{(i)}$ is the j-th diagonal element of the i-th $N_i \times N_i$ block of matrix Π and $\triangle_i$ denotes the distance between the grid points. As $T \to \infty$, the persistence of the process increases and $0 < \bar{y}_i^{(j)}/T < 2\triangle_i$ (for all j) with probability approaching one.[14] Therefore,

$$1 - \pi_{j,j}^{(i)} \leq 1 - \Pr\left(|\varepsilon_i| \leq 2\triangle_i\right) = 2\left(1 - \Phi\left(\frac{2\triangle_i}{\sqrt{2c_i\sigma_i^2/T - c_i^2\sigma_i^2/T^2}}\right)\right)$$

and thus,

$$\frac{1 - \pi_{j,j}^{(i)}}{\omega_i^2} < \frac{2\left(1 - \Phi\left(\frac{\triangle_i\sqrt{2T}}{\sigma_i\sqrt{c_i}}\right)\right)}{2c_i\sigma_i^2/T - c_i^2\sigma_i^2/T^2} \tag{A.6}$$

for all j. Since

$$\Phi\left(\frac{\triangle_i\sqrt{2T}}{\sigma_i\sqrt{c_i}}\right) \to 1 \text{ as } T \to \infty \tag{A.7}$$

[14]Note that Δ_i is fixed. While one can reduce the speed of the convergence by making m a decreasing function of the persistence, such an adjustment will severely distort the unconditional variances. See footnote 3.

by l'Hopital's rule,

$$\lim_{T\to\infty} \frac{1-\pi_{j,j}^{(i)}}{\omega_i^2} = \frac{\triangle_i}{2\sigma_i^3 c_i^{3/2}\pi^{1/2}} \frac{1}{(1/T^{3/2} - c_i/T^{5/2})\exp(2c_i\,\triangle_i^2\,T/\sigma_i^2)} = 0. \tag{A.8}$$

Hence, since the limiting behavior of the conditional variance of the Markov-chain approximation is determined by the limiting behavior of $1-\pi_{j,j}^{(i)}$,

$$\frac{\tilde{\omega}_i^2}{\omega_i^2} \to 0 \text{ as } T \to \infty. \tag{A.9}$$

B Asymptotic Validity of the Method

In this Appendix, we establish the asymptotic validity of the proposed moment-matching method for approximating conditional expectations of nonlinear functions and solving functional equations. For notational simplicity, we present the results for a scalar continuous-valued process with conditional density $f(y'|y)$ although the results can be extended to the vector case $f(\mathbf{y}|\mathbf{x})$, where $\mathbf{y} \in \mathcal{R}^M$ and $\mathbf{x} = (\mathbf{y}_{-1}, ..., \mathbf{y}_{-L}) \in \mathcal{R}^{M\cdot L}$. Consider the function

$$e_g(y) = \int g(y')f(y'|y)dy, \tag{B.1}$$

where $g(y) \in C_0[a,b]$ and $C_0[a,b]$ denotes the space of continuous functions on $[a,b]$ with $a < b$ and both a and b are finite. Assume that the support of $f(y'|y)$ is a subset of $[a,b]\times[a,b]$ and $f(y'|y)$ is jointly continuous in y' and y. Let $\tilde{y}$ denote the n-state Markov-chain approximation proposed that takes on the discrete values $\{\bar{y}^{(1)}, \bar{y}^{(2)}, \cdots, \bar{y}^{(n)}\}$ and transition probabilities $\pi_{j,k}^{(n)} = \Pr(\tilde{y}' = \bar{y}^{(k)}|\tilde{y} = \bar{y}^{(j)})$. Let

$$e_{gn}(y) = \sum_{k=1}^{n} g(\bar{y}^{(k)})\pi_{j,k}^{(n)}. \tag{B.2}$$

Following Tauchen and Hussey (1991), we need to show the uniform convergence result

$$\sup_{y\in[a,b]} |e_{gn}(y) - e_g(y)| \xrightarrow{p} 0 \tag{B.3}$$

as $n \to \infty$.

The pointwise convergence of the conditional distribution of the Markov chain $\tilde{y}'$ given $\tilde{y} = \bar{y}^{(j)}$ to the conditional distribution of y' given $y = \mu(j)$ can be inferred from noting that the transition probability matrix for our method can be expressed in a polynomial form (see Kopecky and Suen, 2010) and by appealing to the Stone-Weierstrass approximation theorem. Finally, the condition that $e_{gn}(y)$ is uniformly bounded converts the pointwise convergence into uniform convergence. As a result, $e_{gn}(y)$ is equicontinuous which is a sufficient condition for the uniform convergence result

$$\sup_{y\in[a,b]} |e_{gn}(y) - e_g(y)| \xrightarrow{p} 0 \text{ as } n \to \infty. \tag{B.4}$$

C Targeting Higher-Order Moments

In some applications, the normality assumption of the error term in equation (1) may seem restrictive. Unfortunately, accounting for non-Gaussian features (for example, non-zero skewness and excess kurtosis) in the conditional distribution of the underlying process appears to be highly non-trivial. To see the source of the problem, note that the innovation of the n-state scalar Markov process generated by the Rouwenhorst (1995) method is the sum of innovations of $n-1$ independent, two-state Markov processes and as the number of grid points for each component of the vector autoregressive process goes to infinity, the conditional skewness and excess kurtosis of the components approach zero. Thus, in order to target non-zero conditional skewness and excess kurtosis for any y_i, one has to consider an alternative way of constructing the probability

transition functions $H_j(i)$, $j \in \{1, 2, \cdots, N^*\}$.

It turns out that 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 jumps 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 such as skewness and kurtosis, 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 4.3.2 that allow us to target skewness and excess kurtosis.

C.1 Conditional skewness

To generate non-zero conditional skewness, or equivalently, asymmetric conditional distribution, we use the first row of the transition matrix $P^{(\tilde{n})}(\rho)$:

$$p_1^{(\tilde{n})}(\rho) = (p_{1,1}^{(\tilde{n})}(\rho)\ p_{1,2}^{(\tilde{n})}(\rho)\ \cdots\ p_{1,\tilde{n}}^{(\tilde{n})}(\rho)), \tag{C.1}$$

where $\tilde{n} \geq 3$. Since this probability mass distribution is associated with the lowest discrete value of the scalar AR(1) process, it is positively skewed unless $\rho \leq 0$. Moreover, when $\rho > 0$, the skewness increases with ρ.

Now let us consider n grid points constructed by Rouwenhorst's (1995) method

for an autoregressive process with unconditional variance σ_i^2: $[z^{(1)}, z^{(2)}, \cdots, z^{(n)}]$ with $n > \tilde{n}$. We construct the transition matrix associated with these grid points using $p_1^{(n)}(\rho)$. Specifically, we set the i-th row of the $n \times n$ matrix $Q^{(n)}$ to

$$q_i = \begin{cases} (\mathbf{0}'_{i-1} \quad p_1^{(\tilde{n})}(\rho) \quad \mathbf{0}'_{n-\tilde{n}-i+1}) & \text{if } i < n - \tilde{n}, \\ (\mathbf{0}'_{n-\tilde{n}} \quad p_1^{(\tilde{n})}(\rho)) & \text{otherwise,} \end{cases} \tag{C.2}$$

where $\mathbf{0}_i$ denotes an $i \times 1$ zero vector.[15] It can be seen that the transition matrix $Q^{(n)}$, along with the grid points, yields a scalar Markov chain whose conditional distribution has the same skewness as the mass distribution function (C.1). Therefore, to construct the probability transition functions $H_i(j)$ as in Section 4.3.2, one can use $Q^{(n)}$ instead of $P^{(n)}$.

C.2 Conditional kurtosis

To generate a conditional distribution with excess kurtosis, one can use a mixture distribution approach. More specifically, let $\tilde{p}_1^{(\tilde{n})}$ and $\tilde{p}_2^{(\tilde{n})}$ be two discrete probability distributions defined over $\tilde{n}$ equally-distanced grid points that have a common mean but different variances σ_1^2 and σ_2^2. Consider now the following mixture:

$$p^{(\tilde{n})} = \tilde{\lambda}\tilde{p}_1^{(\tilde{n})} + (1 - \tilde{\lambda})\tilde{p}_2^{(\tilde{n})}, \tag{C.3}$$

where $0 \leq \tilde{\lambda} \leq 1$. Setting both σ_1/σ_2 and $1 - \tilde{\lambda}$ to low values would result in excess kurtosis for the conditional distribution $p^{(\tilde{n})}$. Then, substituting this conditional distribution for $p_1^{(\tilde{n})}(\rho)$ in (C.2) gives the i-th $(i = 1, 2, ..., n)$ row of the desired transition matrix $Q^{(n)}$.

[15] Using the row q_i, one can target the negative skewness with the same absolute value by setting $q_{i,k} = q_{i,n-k+1}$ where $k \in \{1, 2, ..., n\}$.

References

Adda J, Cooper R. 2003. *Dynamic Economics*. MIT Press, Cambridge, MA.

Anderson TW. 1989. Second-order moments of a stationary Markov chain and some applications. Technical Report No. 22, Department of Statistics, Stanford University.

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.

Elliott G. 1998. On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica* **66**: 149–158.

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.

Gospodinov N, Maynard A, Pesavento E. 2011. Sensitivity of impulse responses to small low-frequency comovements: Reconciling the evidence on the effects of technology shocks. *Journal of Business and Economic Statistics* **29**: 455–467.

Hansen LP, Heaton J, Yaron A. 1996. Finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics* **14**: 262–280.

Kopecky KA, Suen RM. 2010. Finite state Markov-chain approximations to highly persistent processes. *Review of Economic Dynamics* **13**: 701–714.

Phillips PCB. 1987. Towards a unified asymptotic theory for autoregression. *Biometrika* **74**: 535–547.

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.

Stock JH, Wright J. 2000. GMM with weak identification. *Econometrica* **68**: 1055–1096.

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: RMSE

		$\hat{\sigma}_1^2$		$\hat{\sigma}_2^2$		$\text{Corr}(\tilde{y}_1, \tilde{y}_2)$		$1 - \hat{\xi}_1$		$1 - \hat{\xi}_2$	
K	$\overline{N}$	Tau.	MM	Tau.	MM	Tau.	MM	Tau.	MM	Tau.	MM
100	9	0.066	0.008	0.044	0.005	0.017	0.006	0.035	0.010	0.003	0.001
	19	0.036	0.002	0.025	0.002	0.002	0.002	0.003	0.003	0.001	0.001
10	9	0.356	0.010	0.287	0.011	0.047	0.006	0.193	0.019	0.121	0.003
	19	0.278	0.008	0.195	0.006	0.004	0.003	0.008	0.008	0.004	0.003
1	9	0.993	0.025	0.215	0.021	NA	0.010	216.099	0.032	0.993	0.010
	19	0.634	0.025	0.585	0.020	0.079	0.009	0.963	0.026	0.748	0.010

Notes. This table reports the root mean squared error (RMSE) of the key parameters of the bivariate VAR(1) model relative to their true values. (See Section 5 for details). "Tau." denotes the approximation obtained by the method of Tauchen (1986a) whereas "MM" denotes the Markov chain approximation method developed in this paper. Higher values of K imply less persistence. $\overline{N}$ stands for the number of grid points used for each component of $\mathbf{y}$. $\hat{\sigma}_i^2$ denote the simulated unconditional variance of $\tilde{y}_i$ where $i \in \{1, 2\}$. $\text{Corr}(\tilde{y}_1, \tilde{y}_2)$ is the correlation coefficient between $\tilde{y}_1$ and $\tilde{y}_2$. $\hat{\xi}_1$ and $\hat{\xi}_2$ are the eigenvalues of matrix $\hat{A}$. NA indicates that, in some cases, there is no variation in $\tilde{y}_1$ and, therefore, $\hat{\sigma}_1^2 = 0$ and the correlation coefficient $\text{Corr}(\tilde{y}_1, \tilde{y}_2)$ is not defined. The fact that the RMSE of $\hat{\sigma}_1^2$ relative to its true value is very close to 1 indicates that, for most of the Monte Carlo experiments, there is no variation in $\tilde{y}_1$.

Table 2. Approximation Accuracy: Bias and Standard Deviation

		$\hat{\sigma}_1^2$		$\hat{\sigma}_2^2$		$\text{Corr}(\tilde{y}_1, \tilde{y}_2)$		$1 - \hat{\xi}_1$		$1 - \hat{\xi}_2$	
K	$\overline{N}$	Tau.	MM	Tau.	MM	Tau.	MM	Tau.	MM	Tau.	MM
Bias											
100	9	0.066	-0.008	0.044	-0.005	-0.016	-0.005	0.035	0.009	0.003	0.000
	19	0.036	-0.000	0.025	0.000	0.000	0.000	0.002	0.000	0.000	-0.000
10	9	0.356	-0.007	0.287	-0.009	-0.046	-0.005	-0.193	0.017	-0.121	0.001
	19	0.277	-0.000	0.195	-0.000	0.003	-0.000	0.002	0.000	-0.003	0.000
1	9	-0.993	-0.001	-0.167	-0.006	NA	-0.005	67.092	0.018	-0.993	0.001
	19	0.604	-0.001	0.578	-0.001	-0.071	-0.001	-0.963	0.002	-0.748	-0.000
Standard Deviation											
100	9	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.001	0.001
	19	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.001	0.001
10	9	0.010	0.007	0.007	0.006	0.003	0.003	0.006	0.008	0.003	0.003
	19	0.010	0.008	0.007	0.006	0.003	0.003	0.008	0.008	0.003	0.003
1	9	0.025	0.025	0.135	0.020	NA	0.009	205.42	0.026	0.001	0.010
	19	0.192	0.025	0.091	0.020	0.035	0.009	0.005	0.026	0.005	0.009

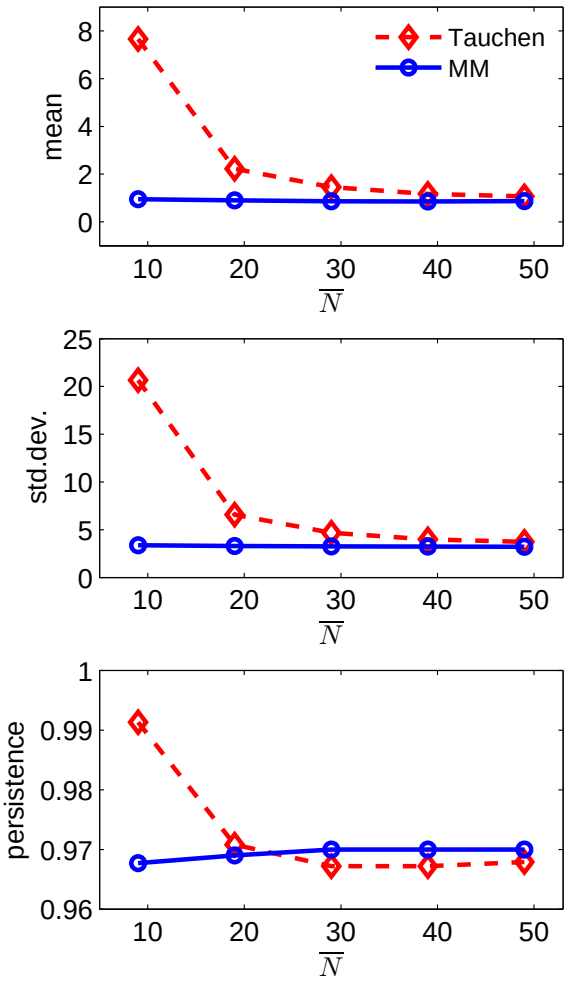
Notes. This table reports the bias and the standard deviation of the parameters relative to their true values. For the bias, the numbers that are much smaller than 0.0005 (0.05%) in absolute terms are denoted by 0.000 with their appropriate signs. See also notes to Table 1.

Table 3. The Distance between Generated and True Conditional Moments

		$\hat{\mu}_1 - \mu_1$		$\hat{\mu}_2 - \mu_2$		$\hat{\omega}_1^2/\omega_1^2 - 1$		$\hat{\omega}_2^2/\omega_2^2 - 1$	
K	$\overline{N}$	Tau.	MM	Tau.	MM	Tau.	MM	Tau.	MM
100	9	0.0010	0.0002	0.0009	0.0001	0.1164	0.0000	0.0599	0.0000
	19	0.0000	0.0000	0.0000	0.0000	0.0425	0.0000	0.0233	0.0000
10	9	0.0179	0.0001	0.0041	0.0001	0.0524	0.0117	0.3428	0.0005
	19	0.0010	0.0000	0.0001	0.0000	0.3201	0.0001	0.1634	0.0000
1	9	0.0163	0.0000	0.0240	0.0000	1.0000	0.0217	0.9852	0.0032
	19	0.0171	0.0000	0.0139	0.0000	0.9515	0.0010	0.4298	0.0000

Notes. This table reports the distance between generated and true conditional moments. Specifically, for $i \in \{1, 2\}$, the numbers in column $\hat{\mu}_i - \mu_i$, are the weighted average of $|\hat{\mu}_i(j) - \mu_i(j)|$ which uses the frequencies of states $j = 1, 2, ..., N^*$ as weights. The frequencies are constructed using a simulated process of length $\tau = 2,000,000$. Similarly, the numbers in column $\hat{\omega}_i^2/\omega_i^2 - 1$ for $i \in \{1, 2\}$ are the weighted average of $|\hat{\omega}_i^2(j)/\omega_i^2 - 1|$ which uses the same frequencies as in columns $\hat{\mu}_i - \mu_i$. The numbers that are smaller than 0.00005 are denoted by 0.0000.

Figure 1: Simulated Moments of the Value of a Firm



Notes: The figure plots the key simulated moments of the value of a firm, V, defined in section 5.3. The horizontal axis is the number of grid points used for each of the two underlying shocks. The moments are calculated using a simulated process of $\tau = 2,000,000$ periods. The upper and middle panels plot the mean and the standard deviation of $\{V_t\}_{t=1}^{\tau}$, while the lower one plots the autocorrelation of the simulated series, $\text{corr}(V_t, V_{t+1})$. The solid line connects the values generated by the moment-matching method developed in this paper, while the dashed line links those generated by Tauchen's method.