Discretization of the Markov Regime Switching AR(1) Process

Yan Liu*
Wuhan University

First version: January 17, 2013
This version: March 22, 2017

Abstract

We propose a simple and theory based method to discretize the Markov regime switching AR(1) process into a first order Markov chain. The method is based on closed form expressions of the regime conditional first and second moments of the process, in conjunction with the Rouwenhorst method for constructing a proper state space and transition matrix. The resulting discrete Markov chain exactly replicates the regime conditional and unconditional means and variances, and the regime conditional autocorrelations, of the original process. The benchmark method is subject to a bias in the unconditional autocorrelation approximation; however, simulation results show that the magnitude of the bias is small. At a cost of compromising regime conditional autocorrelations accuracy, two modifications of the benchmark method with respect to construction of the transition matrix may improve the unconditional autocorrelation approximation with other moments unaffected, especially when the original process is persistent unconditionally.
Key words: Discretization, Markov regime switching process.
JEL codes: C63.

I Introduction

A most widely used approach in solving recursive models in economics is the discrete approximation of the original dynamic programming problems. This approach requires discretization of the state space of the original model, which can be continuous in both endogenous state variables and exogenous shock variables. It is typical to model a shock process as a Markov process, especially an AR(1) process. Since AR(1) processes are easy to estimate and provide a simple way of capturing persistent effect of shocks, it is no surprise that such a choice

*Department of Finance, Economics and Management School. Email: yanliu.ems@whu.edu.cn. I am grateful to comments from participants of 3rd HenU/INFER Workshop on Applied Macroeconomics, and to the financial support from NSFC grant (No. 71503191). Usual disclaimer applies.

becomes predominant in applications of recursive methods. Correspondingly, a host of discretization methods for an AR(1) process is available in the literature, with famous ones such as Tauchen (1986), Tauchen and Hussey (1991), Deaton (1991), and Rouwenhorst (1995).[1] In particular, Rouwenhorst's method provides an almost complete solution to the discretization problem of an AR(1) process, in the sense that it delivers *exact* replications of the most concerned first and second moments, both conditional and unconditional, of the AR(1) process under consideration.

However, most of the seminal discretization methods deals only with the AR(1) process, with a few extensions to the discretization of VAR processes at most. Such a situation poses an obstacle for researchers attempting to model the shock process with an alternative Markov process other than the AR(1) process. One such case occurs in the field of macroeconomics. Ever since the seminal work of Hamilton (1989), the use of the autoregressive models with Markov regime switching coefficients is pervasive, in both empirical and theoretical works; see Hamilton (2015) for an up-to-date survey. Despite its widely use in many areas in macroeconomics, there is by far no theory-based method for the discretization of such a process, which we believe have became an restriction of using the otherwise important process in formulating and solving recursive dynamic models.[2]

In this paper, we propose a theory-based method to discretize a Markov regime switching AR(1) process, henceforth MRS AR(1). The method is based on an analytical characterization of the conditional moments of the MRS AR(1) process. In particular, we give closed form expressions for the regime-conditional mean, variance, and autocorrelation of the MRS AR(1) process. Based on thse regime-conditional moments, we modify the Rouwenhorst method to construct an appropriate discrete Markov chain which *exactly* replicate these moments. The method ensure the local dynamic properties, as captured by the conditional moments, of the MRS AR(1) process to be approximated accurately. In addition, we show that the discretization also replicates the unconditional mean and variance of the MRS AR(1) process, while the unconditional autocorrelation differs in general. The latter property is a natural outcome of preserving the local dynamics of the MRS AR(1) process, and further numerical examples illustrate the bias in the unconditional autocorrelation to be small. Taking together, the paper provides the first and an easy-to-implement discretization method for the popular Markov regime switching AR(1) process.

[1] Deaton's method is made widely known by the textbook of Adda and Cooper (2003). More recent works on this topic include Flodén (2008), Kopecky and Suen (2010), Galindev and Lkhagvasuren (2010), Gospodinov and Lkhagvasuren (2014), Tanaka and Toda (2013, 2015), and Farmer and Toda (2017).

[2] Bai and Zhang (2010) is one of a few quantitative works modeling a shock process with a fully specified AR(1) Markov regime switching process. However, they discretize the process in an ad hoc way. Storesletten, Telmer, and Yaron (2004) is another example of incorporating Markov regime switching process into quantitative works, yet they only consider regime switching in the variance of the innovation, and thus the discretization difficulty is non-essential.

The paper proceeds as follows. Section 2 introduce the basic setup. Section 3.1 provides the closed form expressions for various moments of the MRS AR(1) process; section 3.2 shows how to modify the Rouwenhorst for the purpose of our discretization; section 3.3 describes the discretization method; section 3.4 discusses the properties of the discretized Markov chain; section 3.5 devotes to further discussion on the bias in the unconditional autocorrelation. Section 4 contains several numerical examples.

II Basic Setup

Consider the following simple model of an AR(1) process with regime switching

$$X_t = (1 - \rho(S_t))\mu(S_t) + \rho(S_t)X_{t-1} + \sigma(S_t)\varepsilon_t, \tag{1}$$

where regime S_t follows a Markov chain with state space $\mathcal{S} \equiv \{1, \ldots, K\}$, and $\rho(\cdot)$, $\mu(\cdot)$ and $\sigma(\cdot)$ are functions of S_t. We assume that $\varepsilon_t \overset{\text{iid}}{\sim} N(0,1)$ and the Markov chain $\{S_t\}$ is homogeneous and ergodic.[3] It follows that the range $\mathcal{X}$ of X_t, i.e., the state space, is the entire real line $\mathbb{R}$. One crucial assumption for tractability, as stressed in Hamilton (1990, p. 43), is that S_t be independent of X_{t-1} when conditioning on S_{t-1}. For the investigation of the discretization method in the next section, the assumption of ε_t having normal distribution can be relaxed to: (i) ε_t is iid with zero mean and unit variance; (ii) has strictly positive density over its support; and (iii) induces conditional independence between S_t and X_{t-1}. No result is affected by the specific distributional assumption on ε_t.

Note that for the degenerate case in which $S_\tau \equiv k \in \mathcal{S}$, $\forall \tau \in \mathbb{Z}$, X_t is a standard AR(1) process, hence a Markov chain with state space $\mathcal{X}$. However, given S_t follows a Markov chain, there exists no simple expression of the distribution of X_t conditional on X_{t-1} alone. Nonetheless, it is possible to derive the joint transition kernel for (X_t, S_t), upon which probabilistic properties of X_t can be investigated. To begin with, we use a succinct notation $p(\cdot)$ to denote both the density and probability of a continuous and discrete distribution respectively. In this way, first note that the autoregressive equation (1) implies the following conditional distribution of X_t on X_{t-1} and S_t:

$$X_t | (X_{t-1}, S_t) \sim N((1 - \rho(S_t))\mu(S_t) + \rho(S_t)X_{t-1}, \sigma^2(S_t)),$$

which leads to an analytic expression

$$\begin{aligned} p(X_t = y | X_{t-1} = x, S_t = \ell) &= f(y, x, \ell) \\ &\equiv \frac{1}{\sqrt{2\pi}\sigma(\ell)} \exp\left(- \frac{\big(y - \big(1 - \rho(\ell)\big)\mu(\ell) - \rho(\ell)x\big)^2}{2\sigma^2(\ell)} \right), \end{aligned}$$

[3] In general, we call a Markov chain to be ergodic if it is irreducible, aperiodic and positive recurrent; cf. Meyn and Tweedie (2009, ch. 13) for terminology. We note in passing that when a chain is finite, then irreducibility implies recurrence, hence positive recurrence.

Second, let $p_{k\ell} \equiv p(S_t = \ell | S_{t-1} = k)$, $1 \leq \ell, k \leq K$, denote the regime transition probabilities, and correspondingly $\boldsymbol{P} \equiv [p_{k\ell}]$ the transition matrix. It is then straightforward to show that

$$\begin{aligned} & p(X_t = y, S_t = \ell | X_{t-1} = x, S_{t-1} = k) \\ = & \, p(X_t = y | X_{t-1} = x, S_t = \ell, S_{t-1} = k) p(S_t = \ell | X_{t-1} = x, S_{t-1} = k) \\ = & \, p(X_t = y | X_{t-1} = x, S_t = \ell) p(S_t = \ell | S_{t-1} = k) \\ = & \, f(y, x, \ell) p_{k\ell}, \end{aligned} \tag{2}$$

which can be easily verified as a Markov transition kernel defined over the product state space $\mathcal{X} \times \mathcal{S}$. It is worth to remind that the derivation rests on two facts: (i) the distribution of X_t does not dependent on S_{t-1} when conditioning on X_{t-1} and S_t and (ii) S_t is independent of X_{t-1} conditional on S_{t-1}.

By ergodicity, S_t has a unique invariant distribution $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ and $\pi_k > 0$ for all k.[4] Given that $\varepsilon_t \sim N(0,1)$ and S_t is ergodic, the transition kernel defined in (2) guarantees that the joint process (X_t, S_t) is ergodic as well. This fact follows directly from the results of Yao and Attali (2000, thm. 1), which deals with a general nonlinear MRS autoregressive processes. To be more specific, Yao and Attali (2000) identify three conditions for the ergodicity of (X_t, S_t): (i) a moment condition on ε_t, i.e., $\mathbb{E}|\varepsilon_t|^m < \infty$ for some $m > 0$; (ii) a stability condition on $\rho(\cdot)$, i.e., $\sum_k \pi_k \log |\rho(k)| < 0$;[5] and (iii) an irreducibility condition. The moment condition is clearly satisfied in our setup, and we shall assume the stability condition holds throughout the paper. Note that the stability condition allows $|\rho(\ell)| > 1$ for some ℓ, as long as the mean of $\log |\rho(k)|$ is negative. For the third condition, irreducibility of S_t and strict positivity of the density of ε_t guarantee φ-irreducibility of (X_t, S_t) where φ is the product measure over $\mathcal{X} \times \mathcal{S}$ (i.e., counting $\otimes$ Lebesgue). The ergodicity of the joint Markov chain $\{(X_t, S_t)\}$ implies a *unique* invariant distribution $\nu(\cdot, \cdot)$ over $\mathcal{X} \times \mathcal{S}$.

In what follows, we shall assume (X_t, S_t) starts from ν, and all moments related to X_t are taken under ν, denoted by $\mathbb{E}^\nu$. Meanwhile, we also use $\mathbb{E}^\pi$ whenever it is desirable to indicate the expectation being taken for S_t under $\boldsymbol{\pi}$. It is worth to stress that $\{X_t\}$ alone is

[4] Under the stated assumptions, S_t admits a unique invariant distribution, which is also the marginal distribution of ν on S_t. To be precise, the main results derived in the paper only require the existence of *one* invariant distribution of S_t that is strictly positive for all regimes. This condition in turn is equivalent to (positive) recurrence of S_t; see Timmermann (2000) and Yang (2000) for related works under similar conditions. Instead, working under the ergodic assumption allows us to be more precise on the probabilistic properties of X_t and its approximation. In addition, ergodic chains of S_t prevail in econometric analyses of MRS processes, either as assumptions or estimation results.

[5] It is understood that if $\rho(k) = 0$ for some k, then $\log \rho(k) = -\infty$ and $\mathbb{E} \log \rho(S_t) = -\infty$ as $\pi(k) > 0$. The sufficiency of the specific stability condition dates back to Brandt (1986), and Bougerol and Picard (1992) establish the converse for case where S_t is iid.

not Markovian, despite that $\{(X_t, S_t)\}$ constitutes a Markov chain jointly.[6]

III Discretization

We now describe how to approximate the joint process $\{(X_t, S_t)\}$ by a finite Markov chain. Since the regimes are already discrete, the basic idea is to find a set of discrete state space $\mathcal{Z}(k) = \{z_i(k)\}$ for each regime k, and then compute the associated transition probability both within and cross regimes given $\mathcal{Z} = \cup_k \mathcal{Z}(k)$. The result out of these two steps is a finite Markov chain $\{Z_t\}$ with state space $\mathcal{Z}$ and transition matrix $\boldsymbol{Q}$. The target is not only to have $(\mathcal{Z}, \boldsymbol{Q})$ approximate well the stationary properties of X_t, but perhaps even more so to ensure the conditional dynamics of X_t given S_t is captured by $(\mathcal{Z}, \boldsymbol{Q})$. Thus, it is important to account for the difference among the regime conditional distributions in both choosing $\mathcal{Z}(k)$ and computing $\boldsymbol{Q}$.

From both the practical and theoretical perspective, it is convenient, and very often sufficient, to capture the main properties of conditional distributions through the first and second moments (including the autocorrelations). In what follows, we first present closed form formulas for the first two moments of X_t, conditional on each regime S_t. Then we utilize the conditional moments to determine state space $\mathcal{Z}$ and transition matrix $\boldsymbol{Q}$, through a slightly generalized Rouwenhorst method.

III.A Moment Formula

To begin with, we first present closed form expressions for the regime conditional mean $\mathbb{E}^\nu[X_t|S_t]$ and variance $\mathrm{var}(X_t|S_t)$.[7] Let $\mathbb{E}^\nu[X_t|\boldsymbol{S}]$ and $\mathbb{E}^\nu[X_t^2|\boldsymbol{S}]$ denote the column vectors of $\mathbb{E}^\nu[X_t|S_t = k]$ and $\mathbb{E}^\nu[X_t^2|S_t = k]$ for $k = 1, \ldots, S$ respectively, then we have:

$$
\begin{aligned}
\mathbb{E}^\nu[X_t|\boldsymbol{S}] &= \mathrm{diag}^{-1}(\boldsymbol{\pi})\Big(\boldsymbol{\pi}\mathrm{diag}(\boldsymbol{\mu})\mathrm{diag}(\boldsymbol{\iota}-\boldsymbol{\rho})\big(\boldsymbol{I}-\boldsymbol{P}\mathrm{diag}(\boldsymbol{\rho})\big)^{-1}\Big)', \\
\mathbb{E}^\nu\big(X_t^2\big|S_t=\boldsymbol{S}\big) &= \mathrm{diag}^{-1}(\boldsymbol{\pi})\Big(\boldsymbol{\pi}\mathrm{diag}\big(\boldsymbol{\sigma}^2\big)\big(\boldsymbol{I}-\boldsymbol{P}\mathrm{diag}\big(\boldsymbol{\rho}^2\big)\big)^{-1} \\
&\quad + \boldsymbol{\pi}\mathrm{diag}\big(\boldsymbol{\mu}^2\big)\mathrm{diag}\big((\boldsymbol{\iota}-\boldsymbol{\rho})^2\big)\big(\boldsymbol{I}-\boldsymbol{P}\mathrm{diag}\big(\boldsymbol{\rho}^2\big)\big)^{-1} \\
&\quad + 2\boldsymbol{\pi}\mathrm{diag}(\boldsymbol{\mu})\mathrm{diag}(\boldsymbol{\iota}-\boldsymbol{\rho})\big(\boldsymbol{I}-\boldsymbol{P}\mathrm{diag}(\boldsymbol{\rho})\big)^{-1}\boldsymbol{P}\mathrm{diag}(\boldsymbol{\rho}) \\
&\quad \cdot \mathrm{diag}(\boldsymbol{\mu})\mathrm{diag}(\boldsymbol{\iota}-\boldsymbol{\rho})\big(\boldsymbol{I}-\boldsymbol{P}\mathrm{diag}\big(\boldsymbol{\rho}^2\big)\big)^{-1}\Big)'.
\end{aligned}
$$

In the above expressions, we adopt the following notation definition: for any vector $\boldsymbol{a} = [a_1, \ldots, a_K]$, $\mathrm{diag}(\boldsymbol{a})$ denotes the diagonal matrix where the i'th diagonal element is a_i, $\boldsymbol{a}^2$ denotes $[a_1^2, \ldots, a_K^2]$, and $\boldsymbol{\iota} = [1, \ldots, 1]$. Accordingly, we have closed form expression for the conditional variance $\mathrm{var}(X_t|S_t = k) = \mathbb{E}^\nu[X_t^2|S_t = k] - \mathbb{E}^\nu[X_t|S_t = k]^2$.

[6] Indeed, Francq and Zakoïan (2001) and Zhang and Stine (2001) prove that, up to the covariance structure alone and ignoring higher order moments, X_t has an ARMA(p, q) representation, where $p, q \geq K - 1$ depends on the number of regimes K. In particular, whenever $K \geq 3$, $p, q \geq 2$, so that X_t is necessarily not first-order Markovian, as X_{t-2} affects the firs two moments of X_t.

[7] See Liu (2015) for the relevant derivations.

In addition, in the benchmark method, we use the regime conditional autocorrelation to capture the *local* dynamic property of X_t. In specific, conditioning on two consecutive regimes (S_{t+1}, S_t), the conditional first order autocorrelation

$$\phi(S_{t+1}, S_t) = \frac{\text{cov}(X_{t+1}X_t|S_{t+1}, S_t)}{\sqrt{\text{var}(X_{t+1}|S_{t+1}, S_t)\text{var}(X_t|S_{t+1}, S_t)}}$$

characterizes the persistence of X_t across the given regimes. To calculate the related conditional moment, we note that all coefficients of the AR(1) equation (1) become constant once conditioning on S_{t+1}, hence the randomness of X_{t+1} only comes from X_t and ε_{t+1}. As a result, it is straightforward to derive

$$\text{cov}(X_{t+1}, X_t|S_{t+1}, S_t) = \rho(S_{t+1})\text{var}(X_t|S_t), \tag{3}$$

$$\text{var}(X_{t+1}|S_{t+1}, S_t) = \rho^2(S_{t+1})\text{var}(X_t|S_t) + \sigma^2(S_{t+1}), \tag{4}$$

where we have used the fact that S_{t+1} and X_t are independent conditional on S_t, which also implies $\text{var}(X_t|S_{t+1}, S_t) = \text{var}(X_t|S_t)$. As a result, the conditional autocorrelation can be written as

$$\phi(S_{t+1}, S_t) = \frac{\rho(S_{t+1})}{\sqrt{\rho^2(S_{t+1}) + \sigma^2(S_{t+1})/\text{var}(X_t|S_t)}}. \tag{5}$$

III.B The Rouwenhorst Method

We rely heavily on the discretization method developed first by Rouwenhorst (1995) (henceforth Rouwenhorst method) for constant coefficient AR(1) process. Rouwenhorst method has regained considerable attention in the recent literature on Markov process discretization, and Kopecky and Suen (2010) provide a comprehensive examination of analytical properties of the method. The main advantage of Rouwenhorst method is the exact replication of the variance and auto-covariance, thus autocorrelation, of the candidate AR(1) process.[8]

The original Rouwenhorst method is designed for a setup of Markov chain. To serve our purpose, we show below that the key properties of the constructs can be extended to a setting of two continuously valued random variables W and Y, with any finite means and variances, and an arbitrary correlation coefficient ϕ. Now, suppose we want to find two discrete random variables W^* and Y^*, so as to replicate respectively the mean and variance of W and Y, while jointly match the correlation coefficient ϕ. Equivalently, this requires us to figure out a proper state space of each variable and a proper joint distribution of the two. We now demonstrate how to adapt Rouwenhorst method to this setting.

[8]This makes it possible for an accurate discretization of highly persistent AR(1) shock process, e.g., with an autocorrelation of 0.99, a parameter region which is repeatedly encountered in calibration exercises while standard methods such as Tauchen (1986) perform poorly. Kopecky and Suen (2010) show that Rouwenhorst method also replicate the conditional mean and variance exactly for AR(1) process, which turns out to be the source of inaccuracy for standard methods.

Let us first fix the number of discrete values W^* and Y^* to be the same $N \geq 2$. Denote μ_W and μ_Y the means, and σ_W and σ_Y the variances, of W and Y. Following Rouwenhorst method, we choose the N values $\{w_1, \ldots, w_N\}$ of W^* to be equally spaced between $\mu_W - \sigma_W\sqrt{N-1}$ and $\mu_W + \sigma_W\sqrt{N-1}$, and choose $\{y_1, \ldots, y_N\}$ in exactly the same way. Next, we choose the marginal distribution of W^* to be binomial, i.e., $\Pr(W^* = w_i) = 2^{-(N-1)}\binom{N-1}{i-1} = 2^{-(N-1)}\frac{(N-1)!}{(i-1)!(N-i)!}$, and the *conditional distribution* of Y^* to be $\Pr(Y^* = y_j | W^* = w_i) = \lambda_{ij}$, where λ_{ij} is the (i,j)'th element of a matrix $\mathbf{\Lambda}$. The key ingredient of Rouwenhorst method is the recursive construction of $\mathbf{\Lambda}$:

1. Denote $\psi = (\phi + 1)/2$. For $n = 2$, let
$$\mathbf{\Lambda}_2 = \begin{bmatrix} \psi & 1-\psi \\ 1-\psi & \psi \end{bmatrix}.$$

2. Given $\mathbf{\Lambda}_{n-1}$, first construct
$$\begin{aligned} \widehat{\mathbf{\Lambda}}_n = \psi \begin{bmatrix} \mathbf{\Lambda}_{n-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} + (1-\psi) \begin{bmatrix} \mathbf{0} & \mathbf{\Lambda}_{n-1} \\ 0 & \mathbf{0}' \end{bmatrix} \\ + (1-\psi) \begin{bmatrix} \mathbf{0}' & 0 \\ \mathbf{\Lambda}_{n-1} & \mathbf{0} \end{bmatrix} + \psi \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \mathbf{\Lambda}_{n-1} \end{bmatrix}, \end{aligned}$$
where $\mathbf{0}$ is a column vector of zeros, then divide all except the first and last row of $\widehat{\mathbf{\Lambda}}_n$ by 2, and the resulting matrix is $\mathbf{\Lambda}_n$.

3. Repeat the last step until $n = N$ and $\mathbf{\Lambda} = \mathbf{\Lambda}_N$.

Rouwenhorst (1995) points out that $\mathbf{\Lambda}$ is a Markov transition matrix, thus each row indeed gives rise to a conditional distribution. Moreover, the binomial distribution $\left\{2^{-(N-1)}\binom{N-1}{i-1}\right\}$ constitutes the unique ergodic distribution of $\mathbf{\Lambda}$, therefore the marginal distribution of Y^* is the same binomial as $\Pr(Y^* = y_i) = 2^{-(N-1)}\binom{N-1}{i-1}$. It is then not difficult to demonstrate that the above constructions satisfy $\mathbb{E}W^* = \mu_W$, $\mathbb{E}Y^* = \mu_Y$, $\mathrm{var}W^* = \sigma_W^2$, $\mathrm{var}Y^* = \sigma_Y^2$, and $\mathrm{cov}(W^*, Y^*) = \phi\sigma_W\sigma_Y$. It is worth to stress that these properties are independent on the number of states N chosen at the beginning.

Rouwenhorst's original paper dose not contain formal proofs on these facts; see Kopecky and Suen (2010) for formal proofs in the setting of a Markov chain,[9] which can be identified with the condition of $\mu_W = \mu_Y$ and $\sigma_W = \sigma_Y$ in the current more general setting. All the proofs directly apply to our setting because (i) the state spaces of W^* and Y^* differ only by a scaling factor after removing the means and (ii) the marginal distributions are the same.

[9] Especially the appendix of their working paper version.

III.C Discretizing the MRS Process

We are now ready to construct the discrete state space $\mathcal{Z}$ and the transition matrix $\boldsymbol{Q}$. Let $N \geq 2$ be the fixed number of states chosen for all regimes, so that there are NK states in total for the approximating Markov chain. For each regime k, let $\mathcal{Z}(k) = \{z_1(k), \ldots, z_N(k)\}$ be N points equally spaced from $\mathbb{E}^\nu(X_t|S_t = k) - \sqrt{\text{var}(X_t|S_t = k)(N-1)}$ to $\mathbb{E}^\nu(X_t|S_t = k) + \sqrt{\text{var}(X_t|S_t = k)(N-1)}$.

Let $\boldsymbol{Q}_{k\ell}$ denote an $N \times N$ matrix and be the (k,ℓ)'th square block of $\boldsymbol{Q}$, with $1 \leq k, \ell \leq K$. Each $\boldsymbol{Q}_{k\ell}$ equals to $p_{k\ell}\boldsymbol{\Lambda}_N(\phi(k,\ell))$, where $p_{k\ell}$ is the transition probability from regime k to ℓ, $\boldsymbol{\Lambda}_N(\phi(k,\ell))$ is the Rouwenhorst transition matrix with a correlation of $\phi(k,\ell)$, and lastly $\phi(k,\ell)$ denotes the conditional autocorrelation $\phi(S_{t+1}, S_t)$ with $S_t = k$ and $S_{t+1} = \ell$. It is readily verifiable that the so constructed $\boldsymbol{Q}$ is indeed a Markov transition matrix, and its unique invariant distribution $\boldsymbol{\eta}$ has a simple structure. Let

$$\boldsymbol{\xi} = \begin{bmatrix} 2^{-(N-1)}\binom{N-1}{1-1} & \cdots & 2^{-(N-1)}\binom{N-1}{N-1} \end{bmatrix}$$

denote the vector of N binomial probabilities, and recall $\boldsymbol{\pi}$ is the invariant distribution of $\boldsymbol{P}$. Straightforward calculation then shows that the following row vector

$$\boldsymbol{\eta} = \boldsymbol{\pi} \otimes \boldsymbol{\xi} = (\pi_1, \ldots, \pi_K) \otimes (\xi_1, \ldots, \xi_N)$$

satisfies $\boldsymbol{\eta} = \boldsymbol{\eta}\boldsymbol{Q}$, thus $\boldsymbol{\eta}$ is the unique invariant distribution of $\boldsymbol{Q}$. We shall use $\mathbb{E}^\eta$ to denote the expectation of Z_t under $\boldsymbol{\eta}$.

III.D Properties of the Discretization

By construction, the discrete Markov chain $(\mathcal{Z}, \boldsymbol{Q})$ exactly replicates the conditional mean and variance of X_t on each regime, i.e., $\mathbb{E}^\nu(X_t|S_t) = \mathbb{E}^\eta(Z_t|S_t)$ and $\text{var}(X_t|S_t) = \text{var}(Z_t|S_t)$, where conditional moments such as $\mathbb{E}^\eta(Z_t|S_t = k)$ refer to Z_t taking values in $\mathcal{Z}(k)$ under the marginal distribution derived from $\boldsymbol{\xi}$. It follows directly that the chain also replicates the unconditional mean, as

$$\mathbb{E}^\nu X_t = \mathbb{E}^\pi \mathbb{E}^\nu(X_t|S_t) = \mathbb{E}^\pi \mathbb{E}^\eta(Z_t|S_t) = \mathbb{E}^\eta Z_t.$$

A less straightforward fact is that the chain also replicates the unconditional variance of X_t. To show this, note that $\text{var}Y = \mathbb{E}Y^2 - (\mathbb{E}Y)^2$ for any random variable Y and $\mathbb{E}^\nu X_t = \mathbb{E}^\eta Z_t$, therefore we only need to verify $\mathbb{E}^\nu X_t^2 = \mathbb{E}^\eta Z_t^2$. The last equality can be established as follows:

$$\begin{aligned} \mathbb{E}^\nu X_t^2 &= \mathbb{E}^\pi \mathbb{E}^\nu(X_t^2|S_t) = \mathbb{E}^\pi\big(\text{var}(X_t|S_t) + [\mathbb{E}^\nu(X_t|S_t)]^2\big) \\ &= \mathbb{E}^\pi\big(\text{var}(Z_t|S_t) + [\mathbb{E}^\eta(Z_t|S_t)]^2\big) = \mathbb{E}^\pi \mathbb{E}^\eta(Z_t^2|S_t) = \mathbb{E}^\eta Z_t^2. \end{aligned}$$

It is worth to stress that the unconditional variance $\text{var}X_t$ does not equal to the average of the conditional variance $\text{var}(X_t|S_t)$ weighted by $\boldsymbol{\pi}$, and consequently one can not conclude $\text{var}Z_t = \text{var}X_t$ simply from the fact that $\text{var}(Z_t|S_t) = \text{var}(X_t|S_t)$.[10]

The case for unconditional auto-correlations is more involved. As unconditional variance, identical autocorrelations of X_t and Z_t conditional on consecutive regimes (S_{t+1}, S_t) do not imply identical unconditional autocorrelations. Actually, we shall demonstrate in a moment that the unconditional autocorrelation ρ_X and ρ_Z do not equal with each other in general.

Since $\text{var}(X_t) = \text{var}(Z_t)$ and the autocorrelation equals to the autocovariance divided by the variance, it suffices to consider the autocovariance of X_t and Z_t. Moreover, since $\mathbb{E}^\nu X_t = \mathbb{E}^\eta Z_t$, $\text{cov}(X_{t+1}, X_t) = \mathbb{E}^\nu X_{t+1}X_t - [\mathbb{E}^\nu X_t]^2$ under the invariance distribution, and $\text{cov}(Z_{t+1}, Z_t) = \mathbb{E}^\eta Z_{t+1}Z_t - [\mathbb{E}^\eta Z_t]^2$ analogously, we only need to compare $\mathbb{E}^\nu X_{t+1}X_t$ and $\mathbb{E}^\eta Z_{t+1}Z_t$. In turn, these two unconditional moments are related via the conditional moments.

To spell out the detail, first note that

$$\begin{aligned}
\mathbb{E}^\nu X_{t+1}X_t &= \mathbb{E}^\pi \mathbb{E}^\nu (X_{t+1}X_t|S_{t+1}, S_t) \\
&= \mathbb{E}^\pi [\text{cov}(X_{t+1}X_t|S_{t+1}, S_t) + \mathbb{E}^\nu (X_{t+1}|S_{t+1}, S_t)\mathbb{E}^\nu (X_t|S_{t+1}, S_t)] \\
&= \mathbb{E}^\pi [\text{cov}(X_{t+1}X_t|S_{t+1}, S_t) + \mathbb{E}^\nu (X_{t+1}|S_{t+1}, S_t)\mathbb{E}^\nu (X_t|S_t)].
\end{aligned}$$

Next, recall that $\{Z_t\}$ is constructed so that

$$\text{cov}(Z_{t+1}, Z_t|S_{t+1}, S_t) = \phi(S_{t+1}, S_t)\sqrt{\text{var}(Z_t|S_{t+1}, S_t)\text{var}(Z_{t+1}|S_{t+1}, S_t)},$$

and since $\text{var}(Z_t|S_{t+1}, S_t) = \text{var}(Z_t|S_t)$, $\text{var}(Z_{t+1}|S_{t+1}, S_t) = \text{var}(Z_{t+1}|S_{t+1})$, it follows that

$$\text{cov}(Z_{t+1}, Z_t|S_{t+1}, S_t) = \phi(S_{t+1}, S_t)\sqrt{\text{var}(Z_t|S_t)\text{var}(Z_{t+1}|S_{t+1})}.$$

Employing the definition of $\phi(S_{t+1}, S_t)$ in (3)–(5) and the fact that $\text{var}(Z_t|S_t) = \text{var}(X_t|S_t)$, we can write the last expression as

$$\text{cov}(Z_{t+1}, Z_t|S_{t+1}, S_t) = \frac{\text{cov}(X_{t+1}, X_t|S_{t+1}, S_t)}{\sqrt{\text{var}(X_{t+1}|S_{t+1}, S_t)/\text{var}(X_{t+1}|S_{t+1})}}.$$

To simplify notation, let

$$\chi(S_{t+1}, S_t) \equiv \sqrt{\text{var}(X_{t+1}|S_{t+1}, S_t)/\text{var}(X_{t+1}|S_{t+1})},$$

[10] Jensen's inequality implies that $\mathbb{E}^\nu[\mathbb{E}^\nu(X_t|S_t)]^2 > [\mathbb{E}^\nu \mathbb{E}^\nu(X_t|S_t)]^2 = [\mathbb{E}^\nu X_t]^2$, therefore $\text{var}(X_t) = \mathbb{E}^\nu X_t^2 - [\mathbb{E}^\nu X_t]^2 > \mathbb{E}^\nu \mathbb{E}^\nu(X_t^2|S_t) - \mathbb{E}^\nu[\mathbb{E}^\nu(X_t|S_t)]^2 = \mathbb{E}^\nu[\text{var}(X_t|S_t)]$, despite $\mathbb{E}^\nu X_t^2 = \mathbb{E}^\nu \mathbb{E}^\nu(X_t^2|S_t)$.

so that $\mathrm{cov}(X_{t+1}, X_t|S_{t+1}, S_t) = \chi(S_{t+1}, S_t)\mathrm{cov}(Z_{t+1}, Z_t|S_{t+1}, S_t)$. As a result,

$$\begin{aligned}
\mathbb{E}^\nu X_{t+1}X_t &= \mathbb{E}^\pi[\mathrm{cov}(X_{t+1}X_t|S_{t+1}, S_t) + \mathbb{E}^\nu(X_{t+1}|S_{t+1}, S_t)\mathbb{E}^\nu(X_t|S_t)] \\
&= \mathbb{E}^\pi[\chi(S_{t+1}, S_t)\mathrm{cov}(Z_{t+1}, Z_t|S_{t+1}, S_t) + \mathbb{E}^\nu(X_{t+1}|S_{t+1}, S_t)\mathbb{E}^\nu(X_t|S_t)] \\
&= \mathbb{E}^\pi\chi(S_{t+1}, S_t)[\mathrm{cov}(Z_{t+1}, Z_t|S_{t+1}, S_t) + \mathbb{E}^\eta(Z_t|S_t)\mathbb{E}^\eta(Z_{t+1}|S_{t+1})] \\
&\quad + \mathbb{E}^\pi[\mathbb{E}^\nu(X_{t+1}|S_{t+1}, S_t)\mathbb{E}^\nu(X_t|S_t) - \chi(S_{t+1}, S_t)\mathbb{E}^\eta(Z_{t+1}|S_{t+1})\mathbb{E}^\eta(Z_t|S_t)] \\
&= \mathbb{E}^\pi\chi(S_{t+1}, S_t)\mathbb{E}^\eta(Z_{t+1}Z_t|S_{t+1}, S_t) \\
&\quad + \mathbb{E}^\pi[\mathbb{E}^\nu(X_{t+1}|S_{t+1}, S_t) - \chi(S_{t+1}, S_t)\mathbb{E}^\nu(X_{t+1}|S_{t+1})]\mathbb{E}^\nu(X_t|S_t). \qquad (6)
\end{aligned}$$

Since $\mathbb{E}^\nu(X_{t+1}|S_{t+1}, S_t) = (1 - \rho(S_{t+1}))\mu(S_{t+1}) + \rho(S_{t+1})\mathbb{E}^\nu(X_t|S_t)$,

$$\mathbb{E}^\nu(X_{t+1}|S_{t+1}, S_t) - \chi(S_{t+1}, S_t)\mathbb{E}^\nu(X_{t+1}|S_{t+1}) \neq 0$$

in general. Recall that $\mathbb{E}^\eta Z_{t+1}Z_t = \mathbb{E}^\pi\mathbb{E}^\eta(Z_{t+1}Z_t|S_{t+1}, S_t)$, it is thus evident from (6) that $\mathbb{E}^\nu X_{t+1}X_t \neq \mathbb{E}^\eta Z_{t+1}Z_t$ in general as well. To illustrate, consider a particular case with $\mu(\cdot) \equiv 0$, so that $\mathbb{E}^\nu X_t = \mathbb{E}^\eta Z_t = 0$, hence $\mathbb{E}^\nu X_{t+1}X_t = \mathbb{E}^\pi\chi(S_{t+1}, S_t)\mathbb{E}^\eta(Z_{t+1}Z_t|S_{t+1}, S_t)$. Observe that $\chi(S_{t+1}, S_t)$ is generically correlated with $\mathbb{E}^\eta(Z_{t+1}Z_t|S_{t+1}, S_t)$, thus only under rather special circumstances will $\mathbb{E}^\nu X_{t+1}X_t = \mathbb{E}^\eta Z_{t+1}Z_t$ hold.

III.E Further Discussion

A careful check of the derivation of (6) reveals the reasons underlying the disparity between $\mathbb{E}^\nu X_{t+1}X_t$ and $\mathbb{E}^\eta Z_{t+1}Z_t$. The first reason lies in the fact that $\chi(S_{t+1}, S_t) \neq 1$ in general, or more precisely, the conditional variance of X_t

$$\mathrm{var}(X_{t+1}|S_{t+1}, S_t) = \rho^2(S_{t+1})\mathrm{var}(X_t|S_t) + \sigma^2(S_{t+1})$$

differs from $\mathrm{var}(Z_{t+1}|S_{t+1}, S_t) = \mathrm{var}(Z_{t+1}|S_{t+1})$, which then equals to $\mathrm{var}(X_{t+1}|S_{t+1})$. The second reason is similar to the first one in that

$$\mathbb{E}^\nu(X_{t+1}|S_{t+1}, S_t) = (1 - \rho(S_{t+1}))\mu(S_{t+1}) + \rho(S_{t+1})\mathbb{E}^\pi(X_t|S_t)$$

differs from $\mathbb{E}^\eta(Z_{t+1}|S_{t+1}, S_t) = \mathbb{E}^\eta(Z_{t+1}|S_{t+1}) = \mathbb{E}^\nu(X_{t+1}|S_{t+1})$. Both facts points to the nature of the problem: the distribution of X_{t+1} conditional S_{t+1} is affected by S_t (indeed, on S_{t-j} for all $j \geq 0$), or alternatively, past regimes always contain useful information regarding current X_t; in contrast, distribution of Z_{t+1} conditional on S_{t+1} is fixed by construction irrespective of any past regime S_{t-j}, $j \geq 0$.

The latter property is present whenever the state space chosen for the discretization targets only the conditional distribution of X_t on *current* regime S_t only, so that the impact of past regimes disappears, while an essential feature of the MRS autoregressive process is the influence of the entire history of regimes on current observation. To what extent such a deficiency will jeopardize the performance of the discretized Markov chain ultimately depends

on the specific setting where discretization is required. Nonetheless, it seems justifiable to us to first target *local* properties, i.e., regime specific, of the MRS process in question, as the very advantage of an MRS model is to provide a simple yet flexible setup for capturing heterogeneous local dynamics observed in time series data.

Notwithstanding this essential feature of the discretization procedure, we can still modified the construction of the transition matrix so that $\{Z_t\}$ perfectly replicates the unconditional mean, variance and autocorrelation in together with conditional mean and variance of X_t on each regime. To this end, we only need to modify the transition matrix so that $\boldsymbol{Q}_{k\ell} = p_{k\ell}\boldsymbol{\Lambda}_N\big(\bar{\phi}(k,\ell)\big)$ for $1 \leq k, \ell \leq N$, where

$$\bar{\phi}(S_{t+1}, S_t) = \frac{\mathrm{cov}(X_t, X_{t+1}|S_{t+1}, S_t) + [\mathbb{E}^\nu(X_{t+1}|S_{t+1}, S_t) - \mathbb{E}^\nu(X_{t+1}|S_{t+1})]\mathbb{E}^\nu(X_t|S_t)}{\sqrt{\mathrm{var}(X_t|S_t)\mathrm{var}(X_{t+1}|S_{t+1})}}, \quad (7)$$

with $S_t = k, S_{t+1} = \ell$. With $\mathcal{Z}$ unchanged and the binomial distribution $\boldsymbol{\xi}$ remaining the conditional distribution of Z_t, it follows that both the conditional and unconditional mean and variance of Z_t stay the same, hence equal to those of X_t. An analogous derivation as in (6) shows that $\mathbb{E}^\nu X_{t+1}X_t = \mathbb{E}^\eta Z_{t+1}Z_t$, thus the unconditional autocorrelations of X_t and Z_t are the same. However, the drawback of (7) is that $\bar{\phi}(S_{t+1}, S_t)$ needs not to be less than unity in absolute value, which renders the possibility of negative entries in the transition matrix based on $\bar{\phi}(S_{t+1}, S_t)$. As a result, matching the unconditional autocorrelation by such a method may not be feasible. This is an important caveat to keep in mind in any practical exercise.

IV Numerical Examples

References

Adda, J., and R. Cooper (2003): *Dynamic Economics: Quantitative Methods and Applications*. The MIT Press, Cambridge. [2]

Bai, Y., and J. Zhang (2010): "Solving the Feldstein-Horioka Puzzle with Financial Frictions," *Econometrica*, 78(2), 603–632. [2]

Bougerol, P., and N. Picard (1992): "Strict Stationarity of Generalized Autoregressive Processes," *Annals of Probability*, 20(4), 1714–1730. [4]

Brandt, A. (1986): "The Stochastic Equation $Y_{n+1} = A_nY_n + B_n$ with Stationary Coefficients," *Advances in Applied Probability*, 18(1), 211–220. [4]

Deaton, A. (1991): "Saving and Liquidity Constraints," *Econometrica*, 59(5), 1221–1248. [2]

Farmer, L. E., and A. A. Toda (2017): "Discretizing Nonlinear, Non-Gaussian Markov Processes with Exact Conditional Moments," *Quantitative Economics*, forthcoming. [2]

Flodén, M. (2008): "A Note on the Accuracy of Markov-Chain Approximations to Highly Persistent AR(1) Processes," *Economics Letters*, 99(3), 516–520. [2]

Francq, C., and J.-M. Zakoïan (2001): "Stationarity of Multivariate Markov-Switching ARMA Models," *Journal of Econometrics*, 102(2), 339–364. [5]

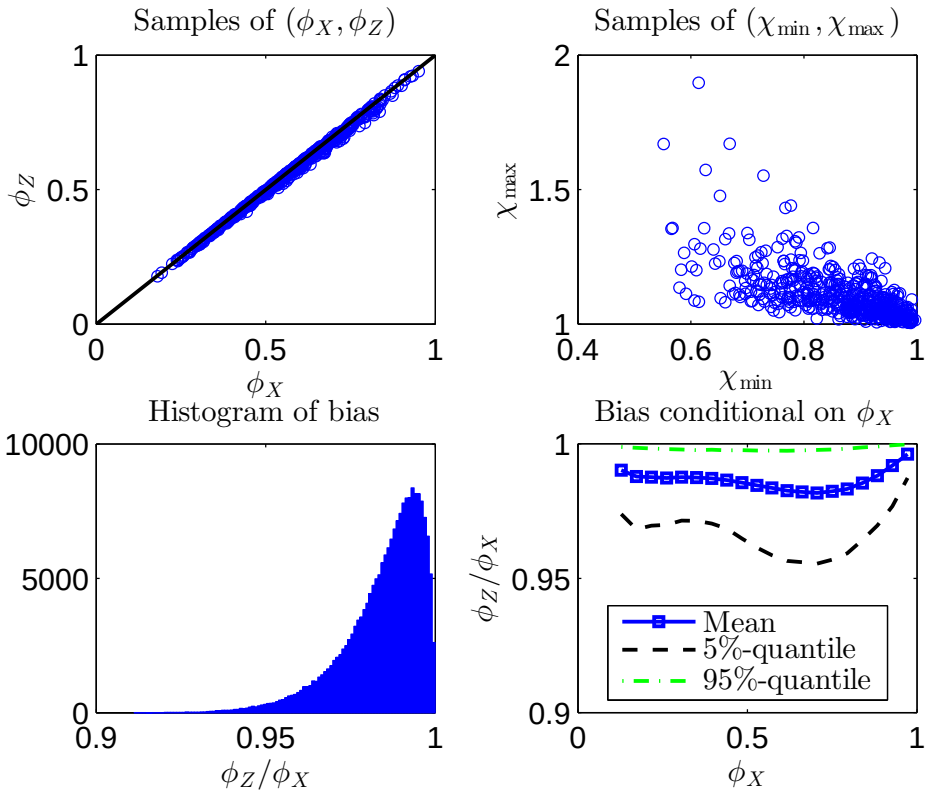

Figure 1: Bias of unconditional autocorrelation approximation

Galindev, R., and D. Lkhagvasuren (2010): "Discretization of Highly Persistent Correlated AR(1) Shocks," *Journal of Economic Dynamics and Control*, 34(7), 1260–1276. [2]

Gospodinov, N., and D. Lkhagvasuren (2014): "A Moment-Matching Method For Approximating Vector Autoregressive Processes By Finite-State Markov Chains," *Journal of Applied Econometrics*, 29(5), 843–859. [2]

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384. [2]

——— (1990): "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45(1-2), 39–70. [3]

——— (2015): "Macroeconomic Regimes and Regime Shifts," in *Handbook of Macroeconomics*, vol. 2. Elsevier. [2]

Kopecky, K. A., and R. M. H. Suen (2010): "Finite State Markov-Chain Approximations to Highly Persistent Processes," *Review of Economic Dynamics*, 13(3), 701–714. [2, 6, 7]

Liu, Y. (2015): "Estimation of Time Series Model with Markov Regime Switching: A GMM Approach," Working paper, Wuhan University. [5]

Meyn, S., and R. L. Tweedie (2009): *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, 2 edn. [3]

Rouwenhorst, K. G. (1995): "Asset Pricing Implications of Equilibrium Business Cycle Models," in *Frontiers of Business Cycle Research*, ed. by T. F. Cooley, chap. 10, pp. 294–330. Princeton University Press, Princeton. [2, 6, 7]

Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Consumption and Risk Sharing over the Life Cycle," *Journal of Monetary Economics*, 51(3), 609–633. [2]

Tanaka, K., and A. A. Toda (2013): "Discrete Approximations of Continuous Distributions by Maximum Entropy," *Economics Letters*, 118(3), 445–450. [2]

——— (2015): "Discretizing Distributions with Exact Moments: Error Estimate and Convergence Analysis," *SIAM Journal on Numerical Analysis*, 53(5), 2158–2177. [2]

Tauchen, G. (1986): "Finite State Markov-chain Approximations to Univariate and Vector Autoregressions," *Economics Letters*, 20(2), 177–181. [2, 6]

Tauchen, G., and R. Hussey (1991): "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica*, 59(2), 371–396. [2]

Timmermann, A. (2000): "Moments of Markov Switching Models," *Journal of Econometrics*, 96(1), 75–111. [4]

Yang, M. (2000): "Some Properties of Vector Autoregressive Processes with Markov-Switching Coefficients," *Econometric Theory*, 16(1), 23–43. [4]

Yao, J., and J.-G. Attali (2000): "On Stability of Nonlinear AR Processes with Markov Switching," *Advances in Applied Probability*, 32(2), 394–407. [4]

Zhang, J., and R. A. Stine (2001): "Autocovariance Structure of Markov Regime Switching Models and Model Selection," *Journal of Time Series Analysis*, 22(1), 107–124. [5]