Supplement to "Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments": Appendix

(*Quantitative Economics*, Vol. 8, No. 2, July 2017, 651–683)

Leland E. Farmer
Department of Economics, University of California, San Diego

Alexis Akira Toda
Department of Economics, University of California, San Diego

Appendix A: Proofs

Proof of Theorem 2.1. (i) The constraint set in (P) is nonempty if and only if $\bar{T} \in \operatorname{co} T(D_N)$. Since $\operatorname{co} T(D_N)$ is nonempty, compact, convex, and the objective function in (P) is strictly convex (a well known property of the Kullback–Leibler information), the claim is trivial.

(ii) The "if" part is Theorem 2 of Tanaka and Toda (2013). To show the "only if" part, suppose that λ_N is a solution to (D). Since the objective function is differentiable, by taking the derivative we get

$$\bar{T} - \sum_{n=1}^{N} \frac{q_n \mathrm{e}^{\lambda_N' T(x_n)}}{\sum_{n=1}^{N} q_n \mathrm{e}^{\lambda_N' T(x_n)}} T(x_n) = 0.$$

Letting p_n be as in (2.3), this equation shows $\bar{T} = \sum_{n=1}^{N} p_n T(x_n)$, $\sum_{n=1}^{N} p_n = 1$, and $p_n > 0$ for all n. Therefore, $\bar{T} \in \operatorname{int} \operatorname{co} T(D_N)$.

(iii) This is Theorem 1 of Tanaka and Toda (2013). □

Proof of Theorem 3.2. The proof is a special case of the following theorem by setting $\Sigma_t = D$ (constant). □

Theorem A.1. *Let* $\{y_t\}$ *be a* VAR *with stochastic volatility*

$$y_t = A y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim (0, \Sigma_{t-1}),$$

Leland E. Farmer: lefarmer@ucsd.edu
Alexis Akira Toda: atoda@ucsd.edu

DOI: 10.3982/QE737

where all eigenvalues of A are less than 1 in absolute value and $\{\Sigma_t\}$ is an exogenous, stationary, ergodic finite-state Markov chain. Let $z_t = (y_t, \Sigma_t)$. Suppose that $z_t^d = (y_t^d, \Sigma_t)$ is a stationary and ergodic Markov chain approximation of z_t such that the conditional mean and variance of y_t are exact, so

$$\mathrm{E}\big[y_t^d | z_{t-1}^d\big] = \mathrm{E}\big[y_t | z_{t-1}^d\big] = A y_{t-1}^d,$$
$$\mathrm{Var}\big[y_t^d | z_{t-1}^d\big] = \mathrm{Var}\big[y_t | z_{t-1}^d\big] = \Sigma_{t-1}.$$

Then the unconditional *mean, variance, and all autocovariance* (*hence the spectrum*) *of $\{y_t\}$ and $\{y_t^d\}$ are identical, and so are all k-step ahead conditional mean and variance.*

PROOF. By assumption, $\Sigma := \mathrm{E}[\Sigma_t]$ exists and $\mathrm{E}[y_t] = 0$.

Define the discretized error term $\varepsilon_t^d := y_t^d - A y_{t-1}^d$. First we prove that the first two unconditional moments are exact. Since by assumption the conditional mean is exact, we have

$$\mathrm{E}\big[\varepsilon_t^d | z_{t-1}^d\big] = \mathrm{E}\big[y_t^d | z_{t-1}^d\big] - A y_{t-1}^d = A y_{t-1}^d - A y_{t-1}^d = 0,$$

and hence $\mathrm{E}[\varepsilon_t^d] = 0$. Since by assumption $\{y_t^d\}$ is stationary and the eigenvalues of A are less than 1 in absolute value, taking the unconditional expectation of both sides of $y_t^d = A y_{t-1}^d + \varepsilon_t^d$, we get $\mathrm{E}[y_t^d] = 0$. Therefore, the unconditional mean is exact. To compute the variance, note that

$$\begin{aligned} y_t^d \big(y_t^d\big)' &= \big(A y_{t-1}^d + \varepsilon_t^d\big)\big(A y_{t-1}^d + \varepsilon_t^d\big)' \\ &= A y_{t-1}^d \big(y_{t-1}^d\big)' A' + A y_{t-1}^d \big(\varepsilon_t^d\big)' + \varepsilon_t^d \big(y_{t-1}^d\big)' A' + \varepsilon_t^d \big(\varepsilon_t^d\big)'. \end{aligned}$$

Since $\mathrm{E}[\varepsilon_t^d | z_{t-1}^d] = 0$ and the conditional variance is exact, taking the conditional expectation we obtain

$$\mathrm{E}\big[y_t^d \big(y_t^d\big)' | z_{t-1}^d\big] = A y_{t-1}^d \big(y_{t-1}^d\big)' A' + \Sigma_{t-1}.$$

Taking the unconditional expectation, using the law of iterated expectations, and noting that $\{y_t^d\}$ is stationary, we get

$$\begin{aligned} \mathrm{Var}\big[y_t^d\big] = \mathrm{E}\big[\mathrm{E}\big[y_t^d \big(y_t^d\big)' | z_{t-1}^d\big]\big] &= A\,\mathrm{E}\big[y_{t-1}^d \big(y_{t-1}^d\big)'\big] A' + \mathrm{E}[\Sigma_{t-1}] \\ &= A\,\mathrm{Var}\big[y_{t-1}^d\big] A' + \Sigma = A\,\mathrm{Var}\big[y_t^d\big] A' + \Sigma. \end{aligned}$$

But the variance matrix of the true process $\{y_t\}$ satisfies the same equation. Since the eigenvalues of A are less than 1 in absolute value, the solution is unique. Therefore, $\mathrm{Var}[y_t^d] = \mathrm{Var}[y_t]$.

Let $\Gamma(k) = \mathrm{E}[y_{t+k} y_t']$ be the true kth order autocovariance matrix and let $\Gamma^d(k) = \mathrm{E}[y_{t+k}^d (y_t^d)']$ be that of the discretized process. Multiplying $(y_t^d)'$ from the right to both sides of $y_{t+k+1}^d = A y_{t+k}^d + \varepsilon_{t+k+1}^d$ and taking expectations, we obtain $\Gamma^d(k+1) = A\Gamma^d(k)$. By iteration, we get $\Gamma^d(k) = A^k \Gamma^d(0)$. Similarly, $\Gamma(k) = A^k \Gamma(0)$. Since $\Gamma(0) = \mathrm{Var}[y_t] =$

$\mathrm{Var}[y_t^d] = \Gamma^d(0)$, it follows that $\Gamma^d(k) = \Gamma(k)$ for all k. Therefore, all autocovariances of $\{y_t\}$ are exact, and so is the spectrum.

To evaluate the k-step ahead conditional moments, note that

$$y_{t+k}^d = \varepsilon_{t+k}^d + \cdots + A^{k-1}\varepsilon_{t+1}^d + A^k y_t^d.$$

Since $\{y_t^d\}$ is a Markov process, we have

$$\mathrm{E}\big[\varepsilon_{t+j}^d | z_t^d\big] = \mathrm{E}\big[\mathrm{E}\big[\varepsilon_{t+j}^d | y_{t+j-1}^d\big] | z_t^d\big] = 0$$

for any $j \geq 1$. Therefore, $\mathrm{E}[y_{t+k}^d | z_t^d] = A^k y_t^d$, so the k-step ahead conditional mean is exact. The proof for the conditional variance is analogous. □

Remark. If the conditional variance of ε_t is unknown at $t-1$, say $\varepsilon_t \sim (0, \Sigma_t)$, then the same result holds by replacing Σ_{t-1} in the proof by $\mathrm{E}[\Sigma_t | \Sigma_{t-1}]$.

Proof of Proposition 3.3. Let $\rho(M)$ denote the spectral radius of the matrix M. Since $\rho(|A|) < 1$, there exists $\delta > 0$ such that $\alpha := \rho(\delta I + |A|) < 1$. By the Perron–Frobenius theorem, $\delta I + |A|$ has a strictly positive eigenvector $v = (v_1, \ldots, v_K) \gg 0$. Take a tensor grid D_N with convex hull $\operatorname{co} D_N = [-v_1, v_1] \times \cdots \times [-v_K, v_K]$. Let $\bar{y}_n$ be any grid point of D_N and let $T(x) = x$ be the moment defining function for the conditional mean (therefore it is the identity map). Then $T(D_N) = D_N$ and

$$\bar{T}_n := \mathrm{E}\big[T(y_t) | y_{t-1} = \bar{y}_n\big] = \mathrm{E}[y_t | y_{t-1} = \bar{y}_n] = A\bar{y}_n.$$

Taking absolute values element by element, since $0 < \alpha < 1$ we get

$$|\bar{T}_n| \leq |A||\bar{y}_n| \leq |A|v \leq \big(\delta I + |A|\big)v = \alpha v \ll v,$$

so $\bar{T}_n \in \operatorname{int} \operatorname{co} T(D_N)$. □

Proof of Proposition 3.4. Let $D = \{\bar{x}_n\}_{n=1}^N$ be the set of grid points and let $M = \max_n |\bar{x}_n|$. Suppose $x_{t-1} = x$, where $x \in D$. By symmetry, without loss of generality we may assume $x \geq 0$. Then the conditional first and second (uncentered) moments of x_t are ρx and $(\rho x)^2 + 1$, respectively. The moment defining function is $T(x) = (x, x^2)$. By Theorem 2.1, it suffices to show that $(\rho x, (\rho x)^2 + 1) \in \operatorname{int} \operatorname{co} T(D)$.

Define the points $P = (M, M^2)$, $Q = (-M, M^2)$, $X = (x, x^2)$, and $X' = (\rho x, (\rho x)^2 + 1)$. If $x = M$, for $X' \in \operatorname{int} \operatorname{co} T(D)$ it is necessary that X' lies below the segment PQ, so we need

$$(\rho M)^2 + 1 < M^2 \quad \Longleftrightarrow \quad M > \frac{1}{\sqrt{1-\rho^2}},$$

which is condition (i) in Proposition 3.4. Therefore X' lies below PQ. Now take any $x \in D$ and set $\mu = \rho x$. Take two grid points $a_1 < a_2 \in D$ such that $\mu \in [a_1, a_2]$. Let $A_1 = (a_1, a_1^2)$ and $A_2 = (a_2, a_2^2)$. If X' lies above the segment $A_1 A_2$, then X' is in the interior of the

quadrilateral A_1A_2PQ, which is a subset of $\operatorname{co} T(D)$. Therefore, it suffices to show that X' lies above A_1A_2. The equation of the straight line A_1A_2 is

$$y = \frac{a_2^2 - a_1^2}{a_2 - a_1}(x - a_1) + a_1^2 = (a_1 + a_2)(x - a_1) + a_1^2.$$

Therefore, X' lies above A_1A_2 if and only if

$$\mu^2 + 1 > (a_1 + a_2)(\mu - a_1) + a_1^2 \quad \Longleftrightarrow \quad (\mu - a_1)(a_2 - \mu) < 1. \tag{A.1}$$

First, consider the case in which the maximum distance between neighboring points is $d < 2$. Take a_1 and a_2 as neighboring points. By the arithmetic mean-geometric mean inequality, we have

$$(\mu - a_1)(a_2 - \mu) \le \left(\frac{(\mu - a_1) + (a_2 - \mu)}{2}\right)^2 = \left(\frac{a_2 - a_1}{2}\right)^2 \le (d/2)^2 < 1,$$

so (A.1) holds. Next we show (3.3). Setting $a_2 = x$ and $\mu = \rho x$ in (A.1) and solving the inequality, a sufficient condition for existence is

$$\rho x = \mu \ge a_1 > \rho x - \frac{1}{(1-\rho)x},$$

which is (3.3) by setting $x = \bar{x}_n$ and $a_1 = \bar{x}_{n'}$. □

Proof of Corollary 3.5. Since the grid $\{\bar{x}_n\}_{n=1}^N$ spans from $-M$ to M and is evenly spaced, the grid size is $d = \frac{2M}{N-1}$. Suppose that $M > \sigma = 1/\sqrt{1-\rho^2}$, so condition (i) of Proposition 3.4 holds. Note that the grid has at least three points 0 and $\pm M$, so $N \ge 3$. *Case* (i). $[\rho \le 1 - \frac{2}{N-1}]$. By Proposition 3.4, it suffices to show $d < 2 \iff M < N - 1$. Since $M \le \sqrt{2}\sigma\sqrt{N-1}$ by assumption, it suffices to show

$$\frac{\sqrt{2}\sqrt{N-1}}{\sqrt{1-\rho^2}} < N - 1 \quad \Longleftrightarrow \quad \rho^2 < 1 - \frac{2}{N-1}.$$

But this inequality is trivial because $\rho^2 < \rho \le 1 - \frac{2}{N-1}$. *Case* (ii). $[\rho > 1 - \frac{2}{N-1}]$. Let $-M = \bar{x}_1 < \cdots < \bar{x}_N = M$ be the grid points. By Proposition 3.4, it suffices to show that (3.3) holds for all n such that $\bar{x}_n > 0$, which means that the interval $(\rho\bar{x}_n - \frac{1}{(1-\rho)\bar{x}_n}, \rho\bar{x}_n)$ contains a grid point. Since the length of this interval is $d_n := \frac{1}{(1-\rho)\bar{x}_n}$, if $d < d_n$, then the interval contains a grid point. Furthermore, since $d_n = \frac{1}{(1-\rho)\bar{x}_n}$ is decreasing in $\bar{x}_n$, it follows that if $d < d_n$ for some n, then $d < d_{n'}$ for all $n' < n$ such that $\bar{x}_{n'} > 0$.

Consider the point $n = N - 1$. Since $d = \frac{2M}{N-1}$, we have $\bar{x}_{N-1} = M - d = M\frac{N-3}{N-1}$. Hence $d_{N-1} = \frac{1}{M(1-\rho)}\frac{N-1}{N-3}$. Therefore,

$$d < d_{N-1} \quad \Longleftrightarrow \quad \frac{2M}{N-1} < \frac{1}{M(1-\rho)}\frac{N-1}{N-3} \quad \Longleftrightarrow \quad M < \frac{N-1}{\sqrt{2(1-\rho)(N-3)}}.$$

Since $M \leq \sigma\sqrt{N-1}$ by assumption, to show $d < d_{N-1}$, it suffices to show

$$\frac{\sqrt{N-1}}{\sqrt{1-\rho^2}} < \frac{N-1}{\sqrt{2(1-\rho)(N-3)}} \quad \Longleftrightarrow \quad 1+\rho > \frac{2(N-3)}{N-1}$$

$$\Longleftrightarrow \quad \rho > 1 - \frac{4}{N-1},$$

which trivially holds because $\rho > 1 - \frac{2}{N-1}$.

Therefore it remains to show that the two inequalities in (3.3) also hold for $n = N$, the boundary point. Take $n' = N-1$. Since $\bar{x}_{N-1} = M\frac{N-3}{N-1}$, the right inequality holds because

$$\bar{x}_{N-1} \leq \rho\bar{x}_N \quad \Longleftrightarrow \quad M\frac{N-3}{N-1} \leq \rho M \quad \Longleftrightarrow \quad \rho \geq 1 - \frac{2}{N-1},$$

which is trivial. The left inequality is equivalent to

$$\rho\bar{x}_N - \frac{1}{(1-\rho)\bar{x}_N} < \bar{x}_{N-1} \quad \Longleftrightarrow \quad \rho M - \frac{1}{(1-\rho)M} < M\frac{N-3}{N-1}$$

$$\Longleftrightarrow \quad M^2\left(\rho - \frac{N-3}{N-1}\right) < \frac{1}{1-\rho}.$$

Since $M \leq \sigma\sqrt{N-1}$, it suffices to show

$$\frac{N-1}{1-\rho^2}\left(\rho - \frac{N-3}{N-1}\right) < \frac{1}{1-\rho} \quad \Longleftrightarrow \quad (N-1)\rho - (N-3) < 1+\rho$$

$$\Longleftrightarrow \quad \rho < 1,$$

which is trivial. □

Appendix B: Accuracy of discretization

The accuracy of discretization has traditionally been evaluated by simulating the resulting Markov chain (Tauchen, 1986, Gospodinov and Lkhagvasuren, 2014). However, we think that such simulations have limited value, for the following reason. According to Theorem 3.2, for VARs the first two population moments—both k-step ahead conditional and unconditional—are exact whenever the one-step ahead conditional moments are exact. Since the population moments will be identical for such discretizations, any difference in the simulation performance must be due to sampling error.

A better approach is to directly compare the population moments of interest of the true process with those of the discretized Markov chains. For example, suppose that $(x_t, y_t)_{t=0}^{\infty} \subset \mathbb{R}^K \times \mathbb{R}$ is generated by some covariance stationary process such that

$$y_t = \beta' x_t + \varepsilon_t,$$

where $\mathrm{E}[x_t\varepsilon_t] = 0$. Then the population OLS coefficient is

$$\beta = \mathrm{E}\big[x_t x_t'\big]^{-1}\mathrm{E}[x_t y_t].$$

If $(x_t^d, y_t^d)_{t=0}^{\infty}$ is a discretized Markov chain, then we can define its OLS coefficient by

$$\beta^d = \mathrm{E}\big[x_t^d \big(x_t^d\big)'\big]^{-1} \mathrm{E}\big[x_t^d y_t^d\big],$$

where the expectation is taken under the ergodic distribution of the Markov chain. Then the bias of the discretization is $\beta^d - \beta$. Here we used the OLS coefficient as an example, but it can be any quantity that is defined through the population moments.

B.1 VAR(1)

As a concrete example, following Gospodinov and Lkhagvasuren (2014), consider the two-dimensional VAR(1) process

$$x_t = Bx_{t-1} + \eta_t,$$

where

$$x_t = \begin{bmatrix} z_t \\ g_t \end{bmatrix}, \qquad \eta_t = \begin{bmatrix} e_{z,t} \\ e_{g,t} \end{bmatrix}, \qquad B = \begin{bmatrix} 0.9809 & 0.0028 \\ 0.0410 & 0.9648 \end{bmatrix},$$

and the shocks $e_{z,t}, e_{g,t}$ are uncorrelated, i.i.d. over time, and have standard deviations 0.0087 and 0.0262, respectively. The implied unconditional variance–covariance matrix is

$$\begin{bmatrix} \sigma_z^2 & \sigma_{zg} \\ \sigma_{zg} & \sigma_g^2 \end{bmatrix} = \begin{bmatrix} 0.00235 & 0.00241 \\ 0.00241 & 0.01274 \end{bmatrix}$$

and the eigenvalues of the coefficient matrix B are $\zeta_1 = 0.9863$ and $\zeta_2 = 0.9594$.

To evaluate the accuracy of discretization, we compute the Markov chain counterpart θ^d of the parameter $\theta = \sigma_z^2, \sigma_g^2, \sigma_{zg}, 1-\zeta_1, 1-\zeta_2$ and calculate the $\log_{10}$ relative bias $\log_{10}|\theta^d/\theta - 1|$ for various number of nodes in each dimension, $N = 5, 9, 15, 21$. For our method, we consider the evenly spaced, quantile, and Gauss–Hermite quadrature grids, which we label as ME-Even, ME-Quant, and ME-Quad, respectively. As a comparison, we consider the existing methods of Tauchen (1986), Tauchen and Hussey (1991) (TH), and Gospodinov and Lkhagvasuren (2014) (GL).[1] The GL method has two versions: one that is the VAR generalization of the Rouwenhorst method (referred to as GL0) and another that fine-tunes this method by targeting the first and second conditional moments (referred to as GL). Table B.1 shows the results.

We can make a few observations from Table B.1. First, as is well known, the accuracy of discretization for the Tauchen and Tauchen–Hussey methods are poor, with relative bias of order about 10^0. Consistent with Gospodinov and Lkhagvasuren (2014), the GL methods improve upon earlier methods by several orders of magnitude.

[1] For the Tauchen method, we need to specify the grid spacing. To give it the best chance, following Kopecky and Suen (2010) we set the grid size proportional to the unconditional standard deviation of the VAR, and choose the constant of proportionality so as to make the unconditional variance as close to the true VAR as possible.

TABLE B.1. The $\log_{10}$ relative bias of VAR discretization.

		Existing Methods				ME Methods		
N	Param.	Tauchen	TH	GL0	GL	Even	Quant	Quad
5	σ_z^2	−0.106	−0.052	−1.061	−1.500	−3.062	−1.465	−0.138
	σ_g^2	−0.106	−0.087	−0.918	−1.331	−2.369	−0.772	−0.138
	σ_{zg}	−0.001	−0.006	−4.394	−1.015	−2.408	−0.811	−0.138
	$1-\zeta_1$	1.641	1.178	−1.100	−1.235	−7.932	−8.178	−7.604
	$1-\zeta_2$	1.158	0.657	−1.865	−1.949	−9.303	−8.554	−8.538
9	σ_z^2	−0.106	−0.098	−1.004	−2.342	−9.321	−8.126	−0.379
	σ_g^2	−0.106	−0.166	−0.859	−2.156	−8.918	−9.372	−0.372
	σ_{zg}	−0.001	−0.021	−1.024	−1.915	−9.337	−7.787	−0.373
	$1-\zeta_1$	1.639	0.950	−1.904	−2.171	−8.690	−7.694	−8.410
	$1-\zeta_2$	1.157	0.396	−2.487	−2.713	−9.271	−9.077	−8.292
15	σ_z^2	−0.106	−0.170	−1.093	−3.730	−8.712	−9.085	−1.454
	σ_g^2	−0.106	−0.285	−0.944	−3.545	−8.783	−9.086	−0.760
	σ_{zg}	−0.001	−0.059	−1.052	−3.357	−10.015	−9.082	−0.800
	$1-\zeta_1$	1.639	0.696	−3.188	−3.664	−8.424	−8.774	−8.846
	$1-\zeta_2$	1.156	0.093	−3.650	−4.106	−8.729	−9.627	−9.790
21	σ_z^2	−0.106	−0.244	−1.174	−4.369	−9.539	−9.171	−8.966
	σ_g^2	−0.106	−0.403	−1.025	−4.140	−9.694	−8.538	−11.359
	σ_{zg}	−0.001	−0.114	−1.129	−4.240	−10.124	−8.524	−8.672
	$1-\zeta_1$	1.638	0.494	−4.517	−5.195	−9.373	−9.202	−8.589
	$1-\zeta_2$	1.156	−0.157	−4.894	−5.563	−9.665	−9.226	−9.301

Note: N denotes the number of discrete points in each dimension; TH denotes the Tauchen and Hussey (1991) method; GL and GL0 denote the Gospodinov and Lkhagvasuren (2014) methods with or without moment targeting; ME denotes maximum entropy methods. The ME methods target the first two conditional moments. For ME-Even, the grid for the $\{y_t\}$ process (3.1) spans $[-\sigma\sqrt{N-1}, \sigma\sqrt{N-1}]$ in each dimension, where σ^2 is the smallest eigenvalue of the unconditional variance of $\{y_t\}$.

Second, the relative bias of ME-Even and ME-Quant is substantially smaller (of order about 10^{-9}, except when $N=5$), which makes our method about 4–6 orders of magnitude more accurate than the GL methods. The reason why the bias is not exactly zero—although it should theoretically be zero if the regularity condition (2.7) holds—is because our method involves the numerical minimization of the dual function in (D'_n), in which we set the error tolerance to 10^{-10}.[2] Therefore, this result suggests that for this particular example, ME-Even and ME-Quant match all first and second conditional moments of the VAR.

Third, our method with Gauss–Hermite quadrature grid (ME-Quad) is poor for $N = 5, 9, 15$, especially for the unconditional variance. This is because, by construction, the quadrature method uses the Gauss–Hermite quadrature nodes of the *conditional* variance. When the process is highly persistent (as in this case since the spectral radius is $\zeta_1 = 0.9863$, which is close to 1), the *unconditional* variance is much larger than the conditional variance. Since the grid is much smaller than typical values of the true process,

[2]This point also explains why the accuracy does not monotonically improve as N gets larger for ME-Even and ME-Quant: since the relative bias is essentially the error tolerance (which is constant), it need not be monotonic in N. In contrast, since the relative bias is not zero for existing methods and ME-Quad, the accuracy of these methods monotonically improves with larger N.

TABLE B.2. Computation time for discretizing the VAR(1) process in seconds.

	Existing Methods				ME Methods		
N	Tauchen	TH	GL0	GL	Even	Quant	Quad
5	0.490	0.008	0.013	0.559	0.684	0.616	1.017
9	1.198	0.016	0.047	2.107	1.397	1.268	1.851
15	3.487	0.049	0.265	5.910	3.212	3.031	3.525
21	8.324	0.078	0.730	12.074	5.561	5.616	6.301

Note: The table shows the computing time in seconds for discretizing the VAR(1) process in this section using a Windows 10 laptop computer with 2.2-GHz Intel Core i5 processor. The Tauchen method matches the unconditional variance. The codes for the ME methods are available as discussed in Appendix E. The GL methods use the codes supplied in the online appendix to Gospodinov and Lkhagvasuren (2014).

the regularity condition (2.7) may be violated and a solution to the dual problem may not exist. Note that ME-Quad is still quite accurate for the parameters $\theta = 1-\zeta_1, 1-\zeta_2$. The reason is that since $1-\zeta_1$and $1-\zeta_2$ depend only on the coefficient matrix B and not on the variance, if the discretization method is able to match all first conditional moments, then the coefficient matrix will be exact. But B in this example satisfies the assumption of Proposition 3.3, so we can match $1-\zeta_1$ and $1-\zeta_2$ exactly.

While Table B.1 shows the high accuracy of discretization by ME methods, is it computationally efficient? Table B.2 shows the computing time for discretizing the VAR(1) process using various methods and numbers of grid points in each dimension. The TH and GL0 methods, which require no optimization, are clearly very fast. All other methods involve solving optimization problems. According to the table, the ME methods are faster than the GL method, probably because we solve the unconstrained dual problem using the Newton algorithm by supplying the analytical gradient and Hessian.

B.2 AR(1) *with stochastic volatility*

Next, we consider the accuracy of the stochastic volatility discretization in Section 3.2. As a comparison, we construct an alternative approximation that uses the Rouwenhorst method to discretize the x_t process and the Tauchen method to discretize the conditional distributions $y_t|x_{t-1}, y_{t-1}$. This is the most logical choice since x is just an AR(1) process (for which the Rouwenhorst method is accurate) and there is no obvious way to discretize the y process except by the Tauchen method. We choose the spacing of the y process to target the unconditional variance σ_y^2. As in the simple autoregressive case, when discretizing the log variance process (x_t), we use $\sqrt{N-1}$ standard deviations for the Rouwenhorst method and either the evenly spaced grid, Gauss–Hermite quadrature grid or the quantile grid for our method. A similar type of discretization is considered in Caldara et al. (2012), although they use Tauchen's method to discretize both the log variance and the level of the process.

Following Caldara et al. (2012), we set the parameter values to $\lambda = 0.95$, $\rho = 0.9$, and $\sigma = 0.06$, and choose $\mu = -9.9426$ to make the conditional standard deviation of the y process equal to 0.007. As a robustness check, we also vary λ, the persistence of technology shocks, between 0 and 0.99. We focus on characteristics of the time series of y_t (the

TABLE B.3. The $\log_{10}$ relative bias of stochastic volatility discretization.

N	λ	TR		ME-Even		ME-Quant		ME-Quad	
Parameter		$1-\lambda$	σ_y^2	$1-\lambda$	σ_y^2	$1-\lambda$	σ_y^2	$1-\lambda$	σ_y^2
9	0	$-\infty$	−9.781	$-\infty$	−6.101	$-\infty$	−5.034	$-\infty$	−5.282
	0.5	−1.819	−9.352	−9.556	−6.102	−9.997	−5.034	−8.755	−5.281
	0.9	−0.982	−8.265	−9.458	−6.102	−9.790	−5.034	−8.857	−5.281
	0.95	−0.718	−9.666	−9.117	−6.102	−9.153	−5.034	−9.409	−5.281
	0.99	−1.381	−8.034	−8.390	−6.102	−8.091	−5.034	−8.455	−5.281
15	0	$-\infty$	−11.15	$-\infty$	−7.371	−14.33	−5.203	−14.70	−6.060
	0.5	−2.189	−8.943	−9.079	−7.367	−9.647	−5.203	−9.630	−6.060
	0.9	−1.337	−8.502	−9.376	−7.364	−9.845	−5.203	−9.269	−6.060
	0.95	−1.061	−8.334	−9.902	−7.363	−9.245	−5.203	−9.158	−6.060
	0.99	−0.540	−8.112	−8.652	−7.399	−7.777	−5.204	−8.059	−6.067
21	0	$-\infty$	−9.336	−14.78	−8.625	−15.96	−5.317	−15.66	−6.898
	0.5	−2.436	−9.821	−10.09	−8.668	−9.813	−5.317	−10.46	−6.900
	0.9	−1.575	−8.693	−9.663	−8.700	−9.556	−5.317	−9.725	−6.900
	0.95	−1.296	−9.755	−10.44	−8.645	−9.993	−5.317	−10.24	−6.899
	0.99	−0.705	−8.193	−9.537	−8.750	−7.823	−5.319	−8.974	−6.909

OLS coefficient λ and the unconditional variance σ_y^2), because the component approximations of x_t are just the standard autoregressive processes we studied before. For each discretization procedure, we vary N (the number of log variance and technology points) between 9, 15, and 21. Table B.3 shows the results.

Since the state space of the volatility process is continuous, Theorem A.1 does not apply, so the unconditional moments need not be exact. However, Table B.3 shows that our method is highly accurate, with a relative bias on the order of 10^{-8} or less for $1-\lambda$ and 10^{-5} or less for σ_y^2. This is likely because the finite-state Markov chain approximation of the volatility process is so accurate that Theorem A.1 "almost" applies. As expected, the Tauchen–Rouwenhorst (TR) method does extremely well for the unconditional variance because it is designed to match by construction. However, it does very poorly compared to the ME methods for the persistence, and this gap widens as λ gets closer to 1.

APPENDIX C: SOLVING ASSET PRICING MODELS

C.1 *Analytical solution with* AR(1)/VAR(1) *shocks*

Burnside (1998) iterates (4.2) forward and obtains a closed-form solution as follows. So as to be consistent with the notation in Section 3, let

$$x_t = (I - B)\mu + Bx_{t-1} + \eta_t,$$

where μ is the unconditional mean of $\{x_t\}$ and $\eta_t \sim N(0, \Psi)$. Let

$$\tilde{\Psi} = (I - B)^{-1}\Psi\left(I - B'\right)^{-1},$$

$$\Psi_n = \sum_{k=1}^{n} B^k \tilde{\Psi} \big(B'\big)^k,$$
$$C_n = B\big(I - B^n\big)(I - B)^{-1},$$
$$\Omega_n = n\tilde{\Psi} - C_n\tilde{\Psi} - \tilde{\Psi}C_n' + \Psi_n.$$

Then we have

$$V(x) = \sum_{n=1}^{\infty} \beta^n \exp\left(n\alpha'\mu + \alpha' C_n(x-\mu) + \frac{1}{2}\alpha'\Omega_n\alpha \right). \tag{C.1}$$

A similar formula can be derived even if the shock distribution is non-Gaussian. For example, for the AR(1) case (so $C_t = D_t$), Tsionas (2003) shows that the price–dividend ratio is

$$V(x) = \sum_{n=1}^{\infty} \beta^n \exp\big(a_n + b_n(x-\mu)\big), \tag{C.2}$$

where

$$b_n = (1-\gamma)\rho\frac{1-\rho^n}{1-\rho},$$
$$a_n = (1-\gamma)\mu n + \sum_{k=1}^{n} \log M\left((1-\gamma)\frac{1-\rho^k}{1-\rho} \right),$$

and $M(\cdot)$ is the moment generating function of ε_t.

In general, the infinite series (C.1) or (C.2) have to be approximated. Burnside (1999) notes that truncating the series (C.1) may not be accurate when α is close to zero since each term would have order β^n, so for β close to 1 the truncation error is substantial. A better way is to use the exact terms up to some large number N, and for $n > N$ we can replace C_n, Ψ_n by their limits $C_\infty = B(I-B)^{-1}$, $\Psi_\infty = \sum_{k=1}^{\infty} B^k\tilde{\Psi}(B')^k$, and Ω_n by

$$n\tilde{\Psi} - C_\infty\tilde{\Psi} - \tilde{\Psi}C_\infty' + \Psi_\infty,$$

in which case the infinite sum can be calculated explicitly. The result is

$$\begin{aligned} V(x) \approx \sum_{n=1}^{N} \beta^n \exp\left(n\alpha'\mu + \alpha' C_n(x-\mu) + \frac{1}{2}\alpha'\Omega_n\alpha \right) \\ + \frac{r^{N+1}}{1-r} \exp\left(\alpha' C_\infty(x-\mu) + \frac{1}{2}\alpha'\big(\Psi_\infty - C_\infty\tilde{\Psi} - \tilde{\Psi}C_\infty'\big)\alpha \right), \end{aligned} \tag{C.3}$$

where $r = \beta\exp(\alpha'\mu + \frac{1}{2}\alpha'\tilde{\Psi}\alpha) < 1$. If $r \geq 1$, the price–dividend ratio is infinite. Proposition C.1 shows that the approximation error of (C.3) is $O((r\rho)^N)$, where ρ is the absolute value of the largest eigenvalue of B. On the other hand, if we simply truncate the series (C.1) at N, the error would be $O(r^N)$, which is much larger.

PROPOSITION C.1. *Consider the asset pricing formula* (C.2). *Let $V_N(x)$ be the value of $V(x)$, where ρ^n is replaced by 0 for $n > N$. Let a_n, b_n be as in* (C.2), *let $m_n = \log M((1-\gamma)(1-\rho^n)/(1-\rho))$, let $S_n = \sum_{k=1}^n m_n$, let $b = \lim b_n = \frac{1-\gamma}{1-\rho}\rho$, let $m = \lim m_n = \log M(\frac{1-\gamma}{1-\rho})$, and assume $r = \beta((1-\gamma)\mu + m) < 1$. Then*

$$V_N(x) = \sum_{n=1}^{N} \beta^n \exp\big(a_n + b_n(x-\mu)\big) + \frac{r^{N+1}}{1-r} \exp\big(S_N - mN + b(x-\mu)\big).$$

Furthermore, the approximation error $|V(x) - V_N(x)|$ is of order $(r\rho)^N$.

PROOF. Let a'_n be the value of a_n, where ρ^k is set to 0 for $k > N$. Since $a'_n = (1-\gamma)\mu n + S_N + m(n-N)$, we get

$$\begin{aligned}
&V_N(x) - \sum_{n=1}^{N} \beta^n \exp\big(a_n + b_n(x-\mu)\big) \\
&= \sum_{n=N+1}^{\infty} \beta^n \exp\big(a'_n + b(x-\mu)\big) \\
&= \sum_{n=N+1}^{\infty} \beta^n \exp\big((1-\gamma)\mu n + S_N + m(n-N) + b(x-\mu)\big) \\
&= \sum_{n=N+1}^{\infty} r^n \exp\big(S_N - mN + b(x-\mu)\big) = \frac{r^{N+1}}{1-r} \exp\big(S_N - mN + b(x-\mu)\big).
\end{aligned}$$

If we replace ρ^n with 0 for $n > N$, since $\log M(\cdot)$ is differentiable and the domain of M for the asset pricing formula is bounded (hence $\log M$ is Lipschitz continuous), $|m_n - m|$ and $|b_n - b|$ are both of the order ρ^n. Since a_n contains the sum of m_n's, we have $|a_n - a'_n| \approx \sum_{k=N+1}^n \rho^k = O(\rho^N)$. Since $|\rho| < 1$, letting $c_n = a_n + b_n(x-\mu)$ and $c'_n = a'_n + b(x-\mu)$, we have $|c_n - c'_n| < 1$ eventually, so by the mean value theorem, $|\exp(c_n - c'_n) - 1| \le \mathrm{e}|c_n - c'_n| = O(\rho^N)$. Therefore,

$$\begin{aligned}
\big|V(x) - V_N(x)\big| &\le \sum_{n=N+1}^{\infty} \beta^n \big|\exp\big(a_n + b_n(x-\mu)\big) - \exp\big(a'_n + b(x-\mu)\big)\big| \\
&= \sum_{n=N+1}^{\infty} \beta^n \exp\big(a'_n + b(x-\mu)\big)\big|\exp\big(c_n - c'_n\big) - 1\big| \\
&\approx \sum_{n=N+1}^{\infty} r^n \rho^N = O\big((r\rho)^N\big).
\end{aligned}$$

□

C.2 *Discretizing the rare disasters model*

In this appendix we provide the details of the discretization of the resilience process (5.2). The discussion is partly based on footnote 9 in Gabaix (2012) and his online

appendix. First, for (5.2) to be stable, we need

$$\frac{1+H_*}{1+H_t}\mathrm{e}^{-\phi_H} \le 1 \quad \Longleftrightarrow \quad \widehat{H}_t \ge (1+H_*)\big(\mathrm{e}^{-\phi_H}-1\big). \tag{C.4}$$

Since in Gabaix (2012) $p_t = p$ and $B_{t+1} = B$ are constant, and by definition $0 \le F_{t+1} \le 1$, from (5.1) we obtain

$$-p \le H_* + \widehat{H}_t \le p\big(B^{-\gamma}-1\big). \tag{C.5}$$

We can take $H_* = p(B^{1-\gamma}-1)$ because Gabaix assumes that the average dividend recovery rate is the same as consumption. The inequalities (C.4) and (C.5) define bounds for $\widehat{H}_t$, which we denote by $[\widehat{H}_{\min}, \widehat{H}_{\max}]$. For the process to remain within this bound, Gabaix assumes that the conditional variance of ε^H_{t+1} shrinks to 0 as we approach the boundary. Namely, he assumes

$$\sigma^2(\widehat{H}) = 2K(1-\widehat{H}/\widehat{H}_{\min})^2(1-\widehat{H}/\widehat{H}_{\max})^2,$$

where $K = 0.2\phi_H|\widehat{H}_{\min}\widehat{H}_{\max}|$. See (59) in the online appendix of Gabaix (2012). We use the exact same functional form.

We define the grid of discretization to be $[\widehat{H}_{\min}+\epsilon, \widehat{H}_{\max}-\epsilon]$, where $\epsilon > 0$ is a small number that we set to be $\epsilon = 10^{-3} \times (\widehat{H}_{\max} - \widehat{H}_{\min})$. The reason for shrinking the interval slightly is because otherwise the conditional variance becomes exactly zero at the boundary points, which is impossible for a discrete Markov chain. Once we have defined the endpoints of the grid this way, we put grid points and discretize the beta distribution at each point by matching the conditional moments using our method. We consider the evenly spaced grid (trapezoidal formula), Clenshaw–Curtis quadrature (Clenshaw and Curtis (1960), Trefethen (2008)), and Gauss–Legendre quadrature, which are the most natural choices since the integration is over a bounded interval.

C.3 *Solving the rare disasters model*

In this appendix we explain how to numerically solve the variable rare disaster model using discretization. We follow the notation in Gabaix (2012).

The stochastic discount factor between time t and $t+1$ is

$$M_{t+1} = \mathrm{e}^{-\rho}(C_{t+1}/C_t)^{-\gamma} = \mathrm{e}^{-\delta} \times \begin{cases} 1, & \text{no disaster,} \\ B_{t+1}^{-\gamma}, & \text{disaster,} \end{cases}$$

where $\delta = \rho + \gamma g_C$. Letting P_t be the cum-dividend price of the stock and letting $V_t = P_t/D_t$ be the price–dividend ratio, it follows from the Euler equation that

$$\begin{aligned} P_t &= D_t + \mathrm{E}_t[M_{t+1}P_{t+1}] \quad \Longrightarrow \\ V_t &= 1 + \mathrm{E}_t\left[M_{t+1}\frac{D_{t+1}}{D_t}V_{t+1}\right] \\ &= 1 + \mathrm{e}^{-\delta+g_D}\big((1-p_t)\,\mathrm{E}_t^{\mathrm{ND}}[V_{t+1}] + p_t\,\mathrm{E}_t^{\mathrm{D}}\big[B_{t+1}^{-\gamma}F_{t+1}V_{t+1}\big]\big), \end{aligned}$$

where p_t is the disaster probability, and $\mathrm{E}_t^{\mathrm{ND}}$ and $\mathrm{E}_t^{\mathrm{D}}$ denote the expectation conditional on no disaster or disaster. By the structure of the model, V_{t+1} depends only on the resilience (5.1), which evolves independently from disasters. Therefore, $\mathrm{E}_t^{\mathrm{ND}}[V_{t+1}] = \mathrm{E}_t^{\mathrm{D}}[V_{t+1}] = \mathrm{E}_t[V_{t+1}]$. Using the definition of resilience, it follows that

$$V_t = 1 + \mathrm{e}^{-\delta+g_D}(1+H_t)\,\mathrm{E}_t[V_{t+1}].$$

To solve for the price–dividend ratio using discretization, suppose the state space of resilience H_t is discretized, and let $s = 1, \dots, S$ be the states. Since the disaster probability is constant, it follows that

$$v_s = 1 + \mathrm{e}^{-\delta+g_D}(1+h_s)\sum_{s'=1}^{S}\pi_{ss'}v_{s'},$$

where v_s is the price–dividend ratio in state s, h_s is the resilience in state s, and $\pi_{ss'}$ is the transition probability from state s to s'. Letting $v = (v_1, \dots, v_S)$ and $h = (h_1, \dots, h_S)$ be the vectors of those values, and letting $P = (\pi_{ss'})$ be the transition probability matrix, it follows that

$$v = 1 + \mathrm{e}^{-\delta+g_D}\operatorname{diag}(1+h)Pv \quad\Longleftrightarrow\quad v = \big(I - \mathrm{e}^{-\delta+g_D}\operatorname{diag}(1+h)P\big)^{-1}1.$$

The continuous solution is obtained by interpolating these values over the entire grid (see Proposition 4.1).

Appendix D: Asset pricing with Gaussian AR(1) shocks

In this appendix we solve the simple asset pricing model with Gaussian AR(1) shocks

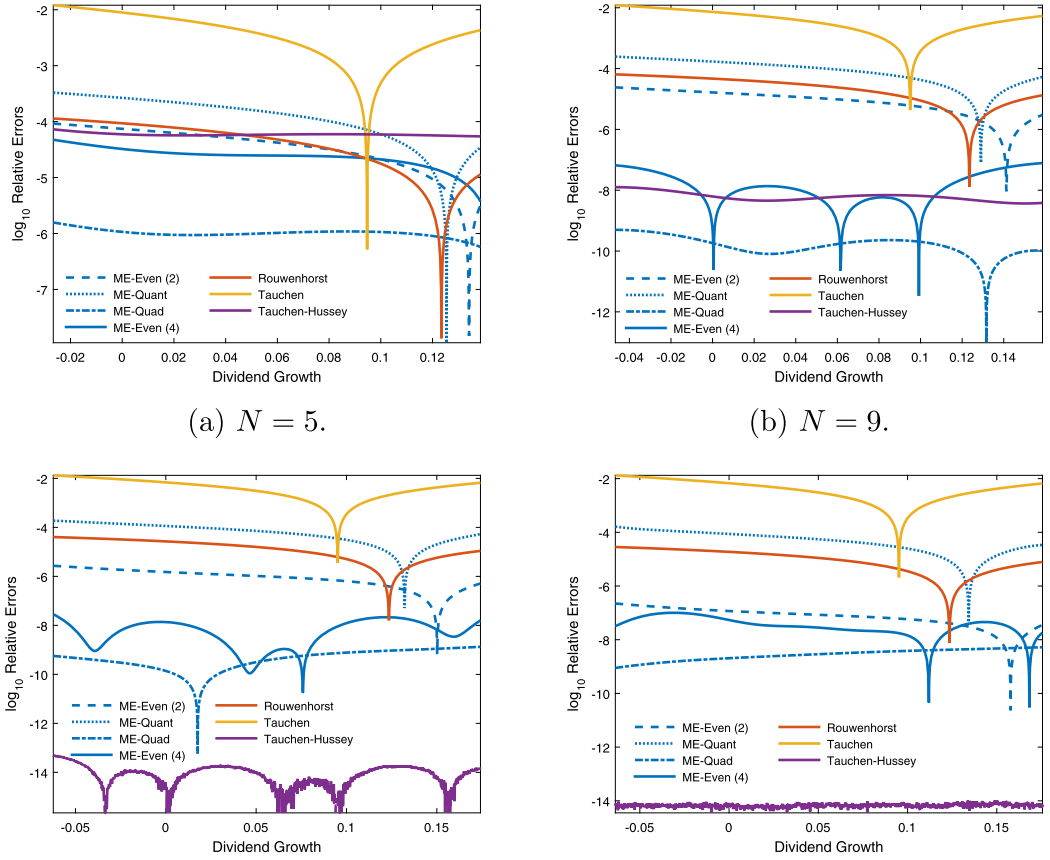
$$x_t = (1-\rho)\mu + \rho x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N\big(0, \sigma^2\big),$$

where x_t is log dividend growth. Using postwar data, the OLS estimates are $\mu = 0.0559$, $\rho = 0.405$, and $\sigma = 0.0589$. Preference parameters are risk aversion $\gamma = 2$ and discount factor $\beta = 0.95$. To avoid cherry-picking, we consider all major existing methods: Tauchen (1986),[3] Tauchen and Hussey (1991), and Rouwenhorst (1995). For the ME methods, we consider ME-Even, ME-Quant, and ME-Quad (all with two moments) as well as ME-Even with four moments.[4] We consider two robustness checks: (i) changing the number of grid points N and (ii) changing the persistence of dividend growth ρ.[5] The number of grid points is always $N = 9$ unless otherwise stated.

[3] For the Tauchen method, we need to specify the grid spacing. To give it the best chance, following Kopecky and Suen (2010) we choose the grid spacing so as to match the unconditional variance exactly. We also experimented with $\sqrt{N-1}$ standard deviations (as in ME-Even and Rouwenhorst) or $1.2\log N$ (as in Flodén (2008)), but the performance was worse.

[4] As discussed below, ME-Quant is uniformly dominated by other ME methods, so there is no point in considering ME-Quant with four moments. The results for ME-Quad with four moments are similar to two moments. We also considered matching six moments, but the performance is similar to four moments.

[5] Collard and Juillard (2001) perform robustness checks across other parameters such as the discount factor, risk aversion, and volatility. They find that the solution accuracy is most susceptible to turning up the persistence.



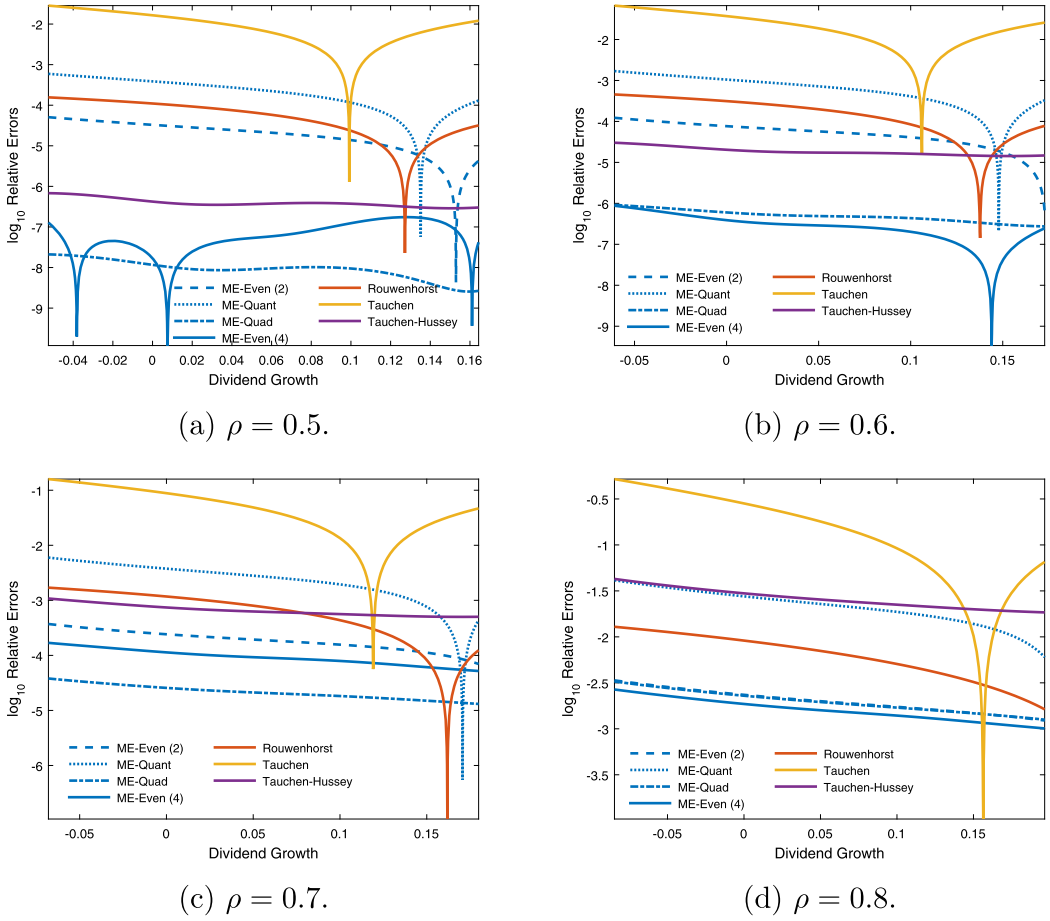
(a) $N = 5$. (b) $N = 9$.

(c) $N = 15$. (d) $N = 21$.

FIGURE D.1. The $\log_{10}$ relative errors of the price–dividend ratio with various discretization methods and numbers of points for the Gaussian AR(1) model. ME-Even (L) shows the result with L moments.

Figure D.1 shows the $\log_{10}$ relative errors of the price–dividend ratio with various discretization methods and numbers of points N. We can make a few observations. First, as we increase N, all methods become more accurate, as expected. This is especially true for Tauchen–Hussey, whose performance is sensitive to N. Second, for methods other than Tauchen–Hussey, the order of the performance is generally ME-Quad > ME-Even (4) > ME-Even (2) > Rouwenhorst > ME-Quant > Tauchen. ME-Quad and ME-Even (four moments) give a solution accuracy of order 10^{-4}–10^{-9}. Third, the performance of ME-Quad does not improve beyond $N = 9$. This is because since ME methods involve a numerical optimization, in which we set the error tolerance to 10^{-10}, the theoretical lower bound for the $\log_{10}$ errors is about -10.

Figure D.2 shows the $\log_{10}$ relative errors when we increase the persistence ρ, fixing the number of points at $N = 9$. Not surprisingly, the performance worsens for all methods as we make the dividend process more persistent. However, the performance of the Tauchen–Hussey method deteriorates quickly, as is well known. ME-Quad, which uses



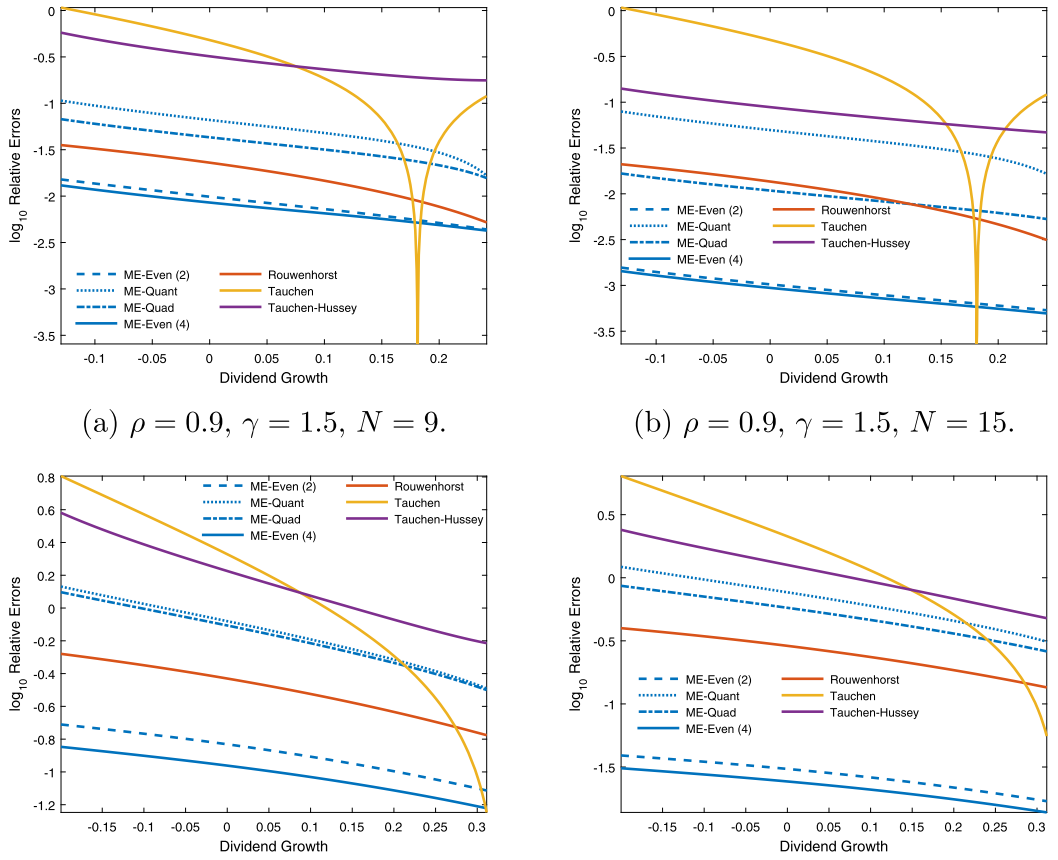
(a) $\rho = 0.5$. (b) $\rho = 0.6$.

(c) $\rho = 0.7$. (d) $\rho = 0.8$.

FIGURE D.2. The $\log_{10}$ relative errors of the price–dividend ratio with various discretization methods and persistence for the Gaussian AR(1) model.

the same Gauss–Hermite quadrature grid as Tauchen–Hussey, also gets poorer, but it is still the best performer along with ME-Even (four moments). The performance of the Rouwenhorst method is robust, although it is uniformly dominated by ME-Even (2 or 4 moments) and ME-Quad.

It is well known that existing methods except Rouwenhorst are poor when the process is persistent (Flodén (2008), Kopecky and Suen (2010)). However, since the price–dividend ratio is infinite (i.e., the series (C.1) diverges) beyond $\rho = 0.8$ with the baseline specification $\gamma = 2$ and $\beta = 0.95$, the performance of the ME methods when persistence is high is still unanswered. To see what happens when the AR(1) process is very persistent, we set $(\rho, \gamma) = (0.9, 1.5), (0.95, 1.3)$, for which the price–dividend ratio is finite. Figure D.3 shows the results. With $\rho = 0.9$, Tauchen–Hussey is one of the worst performers. ME-Quad also deteriorates, and is slightly worse (better) than Rouwenhorst with $N = 9$ ($N = 15$) grid points. The best performers are ME-Even, with comparable performance with two or four moments.



(a) $\rho = 0.9$, $\gamma = 1.5$, $N = 9$.

(b) $\rho = 0.9$, $\gamma = 1.5$, $N = 15$.

(c) $\rho = 0.95$, $\gamma = 1.3$, $N = 15$.

(d) $\rho = 0.95$, $\gamma = 1.3$, $N = 21$.

FIGURE D.3. The $\log_{10}$ relative errors of the price–dividend ratio with various discretization methods for the highly persistent Gaussian AR(1) model with $(\rho, \gamma) = (0.9, 1.5), (0.95, 1.3)$.

To get a better idea of the solution accuracy, consider an investor purchasing 1 million dollar's worth of an asset. If the investor uses each discretization method to compute the fair price of the asset, what is the mistake in dollar amounts? Table D.1 shows the mispricing using the average $\log_{10}$ relative errors. With the baseline specification ($N = 9$, $\rho = 0.405$), the mispricing for the 1 million dollar investment is only 1 cent with ME-Even (four moments). With ME-Quad and Tauchen–Hussey, the pricing error is virtually zero. Even with the Rouwenhorst method, the mispricing is only 18 dollars, so it does not make a material difference across methods except the Tauchen method, which is off by more than 3000 dollars. However, the choice of the discretization method matters as we increase the persistence of the dividend process. With $\rho = 0.8$, the Tauchen method is off by 12%, Tauchen–Hussey is off by 2.6%, and Rouwenhorst is off by 0.6%, as opposed to 0.16% with ME-Even (four moments). The result is even more stark with $\rho = 0.9, 0.95$.

In summary, we find that for discretizing a Gaussian AR(1) process, (i) Tauchen–Hussey is best if there are many points ($N \geq 15$) and the process is not so persistent ($\rho \leq$

TABLE D.1. Mispricing in dollars when investing 1 million dollars.

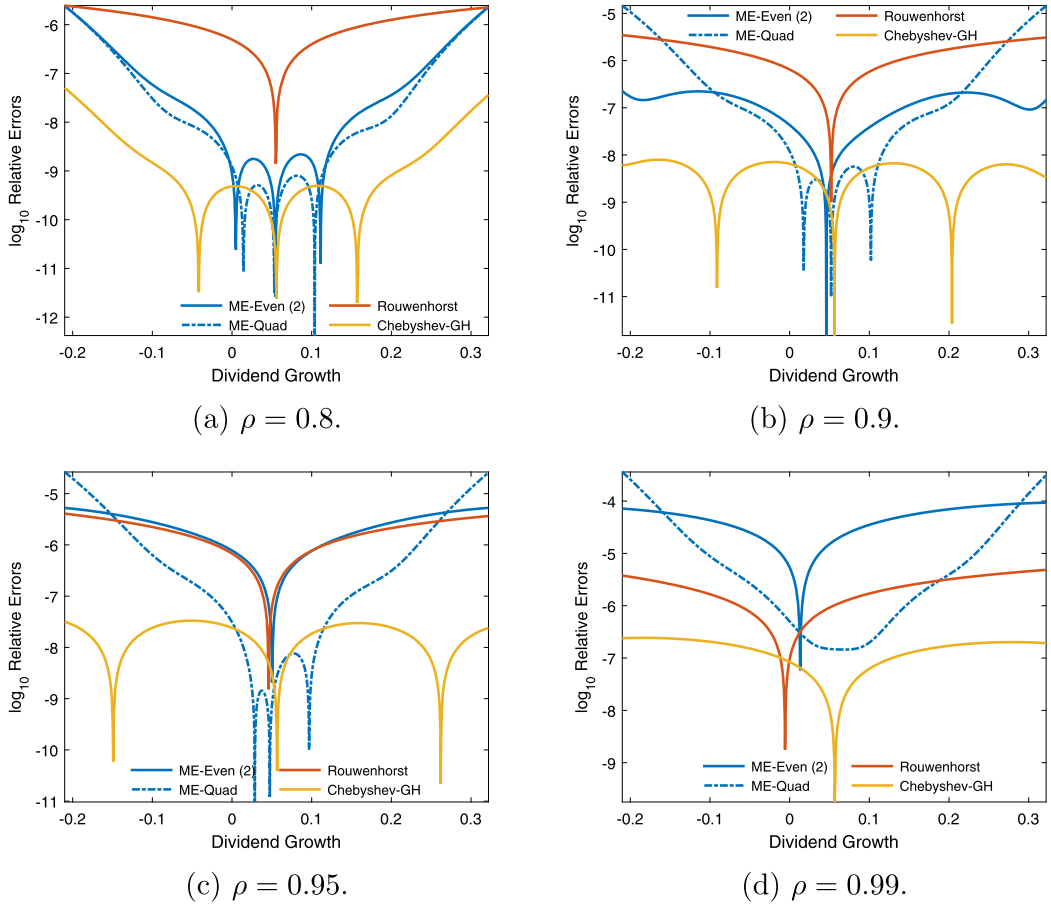
		ME Methods				Existing Methods		
N	ρ	Even (2)	Quant	Quad	Even (4)	R	Tauchen	TH
Changing number of grid points ($\gamma = 2$)								
5	0.405	31.6	103	10.1	23.3	33.8	3161	58.9
9		7.27	71.1	0	0.011	18.1	3136	0.006
15		0.767	51.7	0	0.005	11.2	3380	0
21		0.065	39.8	0	0.03	7.89	3363	0
Changing persistence ($\gamma = 2$)								
9	0.5	16.1	172	0.009	0.051	43.6	7.2K	0.393
	0.6	46.6	507	0.491	0.235	127	17K	18.3
	0.7	185	2.1K	21.4	92.3	501	39K	652
	0.8	2.0K	21K	2.0K	1.6K	6.1K	120K	26K
Highly persistent case ($\gamma = 1.5$)								
9	0.9	8.3K	53K	36K	7.4K	17K	218K	280K
15		0.89K	41K	9.3K	0.82K	9.9K	218K	77K
Highly persistent case ($\gamma = 1.3$)								
15	0.95	13K	70K	67K	9.8K	32K	1.3M	1.4M
21		2.7K	65K	50K	2.2K	25K	1.3M	1.1M

Note: Even (L) denotes the ME-Even method with L moments, R denotes the Rouwenhorst (1995) method, and TH denotes the Tauchen and Hussey (1991) method; K and M denote thousands and millions of dollars.

0.4), (ii) ME-Quad is best if the process is moderately persistent ($0.4 \leq \rho \leq 0.8$), with ME-Even (four moments) comparable, (iii) ME-Even and Rouwenhorst perform well over all choices of grid points N and persistence ρ (especially $\rho > 0.8$), with solution accuracy ME-Even (4) > ME-Even (2) > Rouwenhorst, and (iv) ME-Quant is poor.

Finally, one may be interested in how the discretization solution fares against conventional methods such as projection, and how the performance of discretization deteriorates as the persistence increases. To address this issue, we fix the preference parameters at $\beta = 0.2$ and $\gamma = 1.3$, number of points $N = 9$, and consider the autocorrelation $\rho = 0.8, 0.9, 0.95, 0.99$. (It is necessary to reduce the discount factor β to an unrealistically small number so that the analytical solution exists even for high persistence.) For this exercise, we only consider ME-Even (2), ME-Quad, Rouwenhorst, and the projection method. For the projection method, we make the Euler equation errors zero at the Chebyshev collocation points, and the conditional expectation is computed using a highly accurate Gauss–Hermite quadrature (see Pohl, Schmedders, and Wilms (2015) for details). Figure D.4 shows the results.

Unsurprisingly, the projection method is extremely accurate, since a highly accurate Gauss–Hermite quadrature nodes are chosen for each Chebyshev collocation point. The performance of discretization methods deteriorates as we increase the persistence. The maximum entropy methods are more accurate for persistence less than 0.95, but beyond that the Rouwenhorst method becomes more accurate. This is probably because the Rouwenhorst method does not involve any numerical optimization.



(a) $\rho = 0.8$. (b) $\rho = 0.9$.

(c) $\rho = 0.95$. (d) $\rho = 0.99$.

FIGURE D.4. The $\log_{10}$ relative errors of the price–dividend ratio with discretization and projection methods for the highly persistent Gaussian AR(1) model with $\beta = 0.2$, $\gamma = 1.3$, and $N = 9$. "Chebyshev-GH" refers to the projection method with Chebyshev collocation and Gauss–Hermite quadrature.

APPENDIX E: MATLAB FILES

We implement the discretization of various stochastic processes in the Matlab files posted in a supplementary file on the journal website, http://qeconomics.org/supp/737/code_and_data.zip.[6]

E.1 *Subroutines*

The subroutine `entropyObjective.m` computes the objective function (D′) for minimizing the Kullback–Leibler information and its gradient (2.4a) as well as the Hessian (2.4b). The subroutine `discreteApproximation.m` solves the minimization problem

[6]Also available on our website, https://sites.google.com/site/discretevar/.

(D′) and computes the moment error (2.5). Writing a code for discretizing a particular process is straightforward by using these subroutines and imitating the files listed below.

E.2 VAR

The file `discreteVAR.m` requires four input arguments: the parameters b,B, and Ψ in (3.1), and N, the number of discrete points in each dimension. It outputs the grid and the transition probability matrix. There are three optional arguments: `nMoments`, `method`, and `nSigmas`. The argument `nMoments` specifies the number of conditional moments to target (default: 2). The argument `method` specifies the method for choosing the grid, which has to be either `"even"` (evenly spaced grid), `"quadrature"` (Gauss–Hermite quadrature grid and weights as in Tauchen and Hussey (1991)), or `"quantile"` (quantile grid as in Adda and Cooper (2003)). The default is `"even"`. If the method is `"even"`, then the optional argument `nSigmas` specifies the number of unconditional standard deviations over which the grid points span around the unconditional mean (default: $\sqrt{N-1}$). The file `discreteVAR.m` tries to match the first `nMoments` conditional moments of the VAR, so `nMoments` $= 2$ means the conditional mean and variance. If a solution to the dual problem (D'_n) fails to exist (which sometimes happens when the VAR is highly persistent and the process is close to a boundary point), then it tries to match low order moments. Furthermore, since the discretization of highly persistent VAR is poor with the quadrature grid, when the method `"quadrature"` is chosen, the file returns a warning message if B has an eigenvalue with absolute value exceeding 0.9.

E.3 *Stochastic volatility model*

The file `discreteSV.m` discretizes the stochastic volatility model in (3.4). It requires seven input arguments: `lambda`, `rho`, `sigmaU`, `sigmaE`, `Ny`, `Nx`, and `method`. The arguments `lambda`, `rho`, and `sigmaE` are λ, ρ, and σ in (3.4). The equality $\sigma_u =$ `sigmaU` is the unconditional volatility of the y_t process, so $\sigma_u^2 = \mathrm{E}[\mathrm{e}^{x_t}]$. `Ny` and `Nx` are the number of grid points for the y_t and x_t processes; `method` specifies the method to discretize the AR(1) x_t process, which has to be either `"even"`, `"quadrature"`, or `"quantile"`. (The y_t process is discretized using an evenly spaced grid that spans $\sqrt{N_y - 1}$ unconditional standard deviations because the explicit density is unknown.)

E.4 AR(1) *with non-Gaussian shocks*

The file `discreteARGM.m` discretizes the AR(1) process with Gaussian mixture shocks, $x_t = (1-\rho)\mu + \mu x_{t-1} + \varepsilon_t$. It requires four input arguments, `mu`, `rho`, `gmObj`, and `Nm`, and three optional arguments, `nMoments`, `method`, and `nSigmas`. The arguments `mu` and `rho` are the AR(1) parameters μ and ρ; `gmObj` is the Matlab Gaussian mixture object,[7] typically obtained by running `fitgmdist.m` on the OLS residuals; `nMoments` and `nSigmas` are the same as in the VAR; `method` must be either `"even"`, `"gauss-legendre"`, `"clenshaw-curtis"`, or `"gauss-hermite"`; `"even"` is the usual evenly

[7] http://www.mathworks.com/help/stats/gmdistribution-class.html

spaced grid (trapezoidal formula). The others are quadrature formulas corresponding to each name. (In the paper we discuss only the evenly spaced and Gauss–Hermite quadrature grid because the solution accuracy using the Gauss–Legendre and Clenshaw–Curtis quadrature are about the same as evenly spaced grid.)

References

Adda, J. and R. W. Cooper (2003), *Dynamic Economics: Quantitative Methods and Applications*. MIT Press, Cambridge, MA. [19]

Burnside, C. (1998), "Solving asset pricing models with Gaussian shocks." *Journal of Economic Dynamics and Control*, 22 (3), 329–340. [9]

Burnside, C. (1999), "Discrete state-space methods for the study of dynamic economies." In *Computational Methods for the Study of Dynamic Economies* (R. Marimon and A. Scott, eds.), 95–113, Oxford University Press, Oxford. [10]

Caldara, D., J. Fernándes-Villaverde, J. F Rubio-Ramírez, and W. Yao (2012), "Computing DSGE models with recursive preferences and stochastic volatility." *Review of Economic Dynamics*, 15 (2), 188–206. [8]

Clenshaw, C. W. and A. R. Curtis (1960), "A method for numerical integration on an automatic computer." *Numerische Mathematik*, 2 (1), 195–207. [12]

Collard, F. and M. Juillard (2001), "Accuracy of stochastic perturbation methods: The case of asset pricing models." *Journal of Economic Dynamics and Control*, 25 (6-7), 979–999. [13]

Flodén, M. (2008), "A note on the accuracy of Markov-chain approximations to highly persistent AR(1) processes." *Economics Letters*, 99 (3), 516–520. [13, 15]

Gabaix, X. (2012), "Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance." *Quarterly Journal of Economics*, 127 (2), 645–700. [11, 12]

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. [5, 6, 7, 8]

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. [6, 13, 15]

W. Pohl, K. Schmedders, and O. Wilms (2015), "Higher-order dynamics in asset-pricing models with recursive preferences." Available at https://ssrn.com/abstract=2540586. [17]

Rouwenhorst, K. G. (1995), "Asset pricing implications of equilibrium business cycle models." In *Frontiers of Business Cycle Research* (T. F. Cooley, ed.), 294–330, Princeton University Press. [13, 17]

Tanaka, K. and A. A. Toda (2013), "Discrete approximations of continuous distributions by maximum entropy." *Economics Letters*, 118 (3), 445–450. [1]

Tauchen, G. (1986), "Finite state Markov-chain approximations to univariate and vector autoregressions." *Economics Letters*, 20 (2), 177–181. [5, 6, 13]

Tauchen, G. and R. Hussey (1991), "Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models." *Econometrica*, 59 (2), 371–396. [6, 7, 13, 17, 19]

Trefethen, L. N. (2008), "Is Gauss quadrature better than Clenshaw-Curtis?" *SIAM Review*, 50 (1), 67–87. [12]

Tsionas, E. G. (2003), "Exact solution of asset pricing models with arbitrary shock distributions." *Journal of Economic Dynamics and Control*, 27 (5), 843–851. [10]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 21 June, 2016; final version accepted 26 October, 2016; available online 31 October, 2016.