Discretizing Nonlinear, Non-Gaussian Markov Processes with Exact Conditional Moments

Farmer, Leland and Toda, Alexis Akira

University of California, San Diego, University of California, San Diego

1 November 2016

Discretizing Nonlinear, Non-Gaussian Markov Processes with Exact Conditional Moments*

Leland E. Farmer† Alexis Akira Toda‡

This version: November 1, 2016

Abstract

Approximating stochastic processes by finite-state Markov chains is useful for reducing computational complexity when solving dynamic economic models. We provide a new method for accurately discretizing general Markov processes by matching low order moments of the conditional distributions using maximum entropy. In contrast to existing methods, our approach is not limited to linear Gaussian autoregressive processes. We apply our method to numerically solve asset pricing models with various underlying stochastic processes for the fundamentals, including a rare disasters model. Our method outperforms the solution accuracy of existing methods by orders of magnitude, while drastically simplifying the solution algorithm. The performance of our method is robust to parameters such as the number of grid points and the persistence of the process.

Keywords: asset pricing models, duality, Kullback-Leibler information, numerical methods, solution accuracy.

JEL codes: C63, C68, G12.

1 Introduction

Many nonlinear dynamic economic models such as dynamic stochastic general equilibrium (DSGE) models, asset pricing models, or optimal portfolio problems imply a set of integral equations that do not admit explicit solutions. Finite-state Markov chain approximations of stochastic processes are a useful way of reducing computational complexity when solving and estimating such models because

*We thank Roy Allen, Jinhui Bai, Brendan Beare, Craig Burnside, Nikolay Gospodinov, Jim Hamilton, Ivana Komunjer, Alisdair McKay, Juan Rubio-Ramírez, Andres Santos, Rosen Valchev, Gianluca Violante, and seminar participants at Duke, McGill, UCSD, University of Technology Sydney, and 2016 Computing in Economics and Finance Conference for helpful comments and feedback. We are especially grateful to four anonymous referees for constructive comments and suggestions that significantly improved the paper. Matlab codes are posted on our website `https://sites.google.com/site/discretevar/`.

†Department of Economics, University of California San Diego. Email: lefarmer@ucsd.edu

‡Department of Economics, University of California San Diego. Email: atoda@ucsd.edu

integration is replaced by summation.[1] However, existing methods only work on a limited case by case basis, and apply mostly to linear Gaussian autoregressive processes.

In this paper, we provide a new method for accurately discretizing general nonlinear, non-Gaussian Markov processes. The dynamics of any Markov process are characterized by its transition kernel, which summarizes the conditional distribution of the subsequent state for all possible current states. We construct a discrete approximation to the underlying Markov process by approximating a finite set of its conditional distributions.[2] Given a set of discrete points in the state space, we construct a transition matrix, where each row corresponds to a discrete probability measure which mimics the dynamics of the continuous process in that particular state. This is accomplished by starting from a coarse approximation of the underlying process and modifying the transition probabilities so as to exactly match a set of conditional moments, such as the mean and variance. Because there are typically more grid points than there are conditional moments of interest, there are infinitely many candidates for the approximate conditional distribution. To deal with this underdetermined system, we obtain the discrete approximation by minimizing the relative entropy (Kullback-Leibler information) of the conditional distribution from an initial approximation, subject to the given moment constraints. Although this primal problem is a high dimensional constrained optimization problem, its dual is a computationally tractable, *low dimensional unconstrained* optimization problem. We provide recommendations for how to choose the initial approximation and the moments to match.

The two ingredients of our method—matching conditional moments to approximate a Markov process and using the maximum entropy principle to match moments—have already been proposed separately in the literature. Our main contribution is that we combine these two ingredients and show that this idea can be used to discretize a wide variety of nonlinear, non-Gaussian Markov processes, for which there is currently no systematic way of discretizing. Furthermore, we provide sufficient conditions for the existence of a discretization with exact moments and study economic applications to which existing methods do not apply.

The closest papers to ours are Tanaka and Toda (2013, 2015) and Gospodinov and Lkhagvasuren (2014). Tanaka and Toda (2013) construct discrete approximations of continuous probability distributions (as opposed to stochastic processes) by modifying an initial discretization so as to exactly match low order moments using the maximum entropy principle. While they briefly discuss how to apply their method to discretize vector autoregressive processes (VARs), because they need a closed-form expression for the ergodic distribution—which is not available

[1]Examples include heterogeneous-agent incomplete markets models (Aiyagari, 1994; Heaton and Lucas, 1996), optimal taxation (Aiyagari, 1995; Dávila et al., 2012), portfolio problems (Haliassos and Michaelides, 2003; Judd et al., 2011), asset pricing (Zhang, 2005; Guvenen, 2009), DSGE models (Aruoba et al., 2006; Caldara et al., 2012), estimating dynamic games (Aguirregabiria and Mira, 2007), inflation dynamics and monetary policy (Vavra, 2014), among many others.

[2]For the remainder of the paper, "discrete" should be understood to refer to the state space of the Markov process. Time is always discrete.

in most situations—their method cannot be directly used for discretizing general Markov processes. Tanaka and Toda (2015) prove that their approximation method weakly converges to the true distribution as the number of grid points tends to infinity. They also show that the integration error diminishes by a factor proportional to the error when the integrand is approximated using the functions defining the moments of interest as basis functions. Therefore, the approximation quality of the Tanaka-Toda method depends on two factors, (i) the quality of the initial discretization, and (ii) how well the moment defining functions approximate the integrand.

Gospodinov and Lkhagvasuren (2014) (henceforth GL) propose a discretization method of VARs that targets the first and second conditional moments. According to their numerical results, the GL method seems to be the most accurate finite-state Markov chain approximation for VARs currently available in the literature. As in GL, we target the conditional moments in order to discretize VARs. However, our method improves upon theirs in three important ways.

First, unlike the GL method, our approach is not limited to the approximation of VARs. It applies to *any* Markov process for which we can compute conditional moments and thus has a much wider range of applicability. For instance, we can discretize stochastic processes with interesting nonlinear and non-Gaussian conditional dynamics. Additionally, we do not require a parametric specification of the Markov process to use our approach. Given sufficient data, we can estimate the conditional moments and transition kernel nonparametrically, and use these to construct our discrete approximation.

Second, GL adjust the transition probabilities to match moments directly, whereas we solve the dual problem, which is a low dimensional unconstrained convex minimization problem. The gradient and Hessian of the objective function can be computed in closed form, which allows us to use a standard Newton-type algorithm to find the minimum. Consequently, our method is computationally tractable even when the number of grid points is large. This is an important property, particularly for the case of high dimensional processes.

Finally, for general VARs (which may even feature stochastic volatility), under certain regularity conditions we prove that our method matches all k-step ahead conditional mean, variance, and covariance as well as the unconditional ones. This property has been known only for the Rouwenhorst (1995) method for discretizing univariate AR(1) processes. We further discuss the relation of our method to the existing literature in Section 3.3.

In order to illustrate the general applicability of our method, we solve for the price-dividend ratio in Lucas-tree asset pricing models, under different assumptions about the stochastic processes driving consumption and dividend growth, including more standard AR(1) and VAR(1) processes with Gaussian shocks, an AR(1) model with non-Gaussian shocks, and the variable rare disasters model of Gabaix (2012), whose underlying stochastic process is highly nonlinear and non-Gaussian. In each case, we show that our method produces more accurate solutions than all existing discretization methods,[3] often by several orders of mag-

[3]Several papers such as Aruoba et al. (2006) and Caldara et al. (2012) compare the accuracy of

nitude, requiring only minor modifications between specifications and trivial computing time. We also show that solving general asset pricing models (*e.g.*, with recursive utility and complicated dynamics) using discretization and projection (Judd, 1992) is actually equivalent to solving a discrete-state model (which is a matter of inverting a matrix) and interpolating. Therefore our method provides a simple but systematic way for solving asset pricing models.

We emphasize that our method has many potential applications beyond the asset pricing models considered here. For example, our method can be used to facilitate the estimation of nonlinear state space models. In parallel work, Farmer (2016) shows that by discretizing the dynamics of the state variables, one can construct an approximate state space model with closed-form expressions for the likelihood and filtering recursions, as in Hamilton (1989). The parameters of the model can then be estimated using standard likelihood or Bayesian techniques. This procedure offers an alternative to computationally expensive, simulation-based methods like the particle filter, and simple but often inaccurate linearization approaches like the extended Kalman filter. Our paper provides a computationally tractable method for discretizing general nonlinear Markov processes governing the state dynamics.

2 Maximum entropy method for discretizing Markov processes

In this section we review the maximum entropy method for discretizing probability distributions proposed by Tanaka and Toda (2013, 2015) and apply it to discretize general Markov processes.

2.1 Discretizing probability distributions

2.1.1 Description of method

Suppose that we are given a continuous probability density function $f : \mathbb{R}^K \to \mathbb{R}$, which we want to discretize. Let X be a random vector with density f, and $g : \mathbb{R}^K \to \mathbb{R}$ be any bounded continuous function. The first step is to pick a quadrature formula

$$\mathrm{E}[g(X)] = \int_{\mathbb{R}^K} g(x)f(x)\,\mathrm{d}x \approx \sum_{n=1}^{N} w_n g(x_n)f(x_n), \tag{2.1}$$

where N is the number of integration points, $\{x_n\}_{n=1}^N$, and $w_n > 0$ is the weight on the integration point x_n.[4] Let $D_N = \{x_n \,|\, n = 1, \ldots, N\}$ be the set of grid points.

various solution techniques (log-linearization, value function iteration, perturbation, projection, etc.), *given the discretization method*. To the best of our knowledge, Kopecky and Suen (2010) is the only paper that compares the solution accuracy across various discretization methods, fixing the solution technique. However, they consider only Gaussian AR(1) processes.

[4]Since the grid points $\{x_n\}$ and weights $\{w_n\}$ may depend on the number of grid points N, a more precise notation might be $x_{n,N}$ and $w_{n,N}$. Since there is no risk of confusion, we keep

For example, if we let

$$D_N = \{(m_1 h, \ldots, m_K h) \,|\, m_1, \ldots, m_K = 0, \pm 1, \ldots, \pm M\},$$

which consists of $N = (2M+1)^K$ lattice points with grid size h, setting the weight $w_n = h^K$ in quadrature formula (2.1) gives the trapezoidal formula.

For now, we do not take a stance on the choice of the initial quadrature formula, but take it as given. Given the quadrature formula (2.1), a coarse but valid discrete approximation of the density f would be to assign probability q_n to the point x_n proportional to $w_n f(x_n)$, so

$$q_n = \frac{w_n f(x_n)}{\sum_{n=1}^N w_n f(x_n)}. \tag{2.2}$$

However, this is not necessarily a good approximation because the moments of the discrete distribution $\{q_n\}$ do not generally match those of f.

Tanaka and Toda (2013) propose exactly matching a finite set of moments by updating the probabilities $\{q_n\}$ in a particular way. Let $T : \mathbb{R}^K \to \mathbb{R}^L$ be a function that defines the moments that we wish to match and let $\bar{T} = \int_{\mathbb{R}^K} T(x) f(x)\,\mathrm{d}x$ be the vector of exact moments. For example, if we want to match the first and second moments in the one dimensional case ($K = 1$), then $T(x) = (x, x^2)'$. Tanaka and Toda (2013) update the probabilities $\{q_n\}$ by solving the optimization problem

$$\begin{aligned} \underset{\{p_n\}}{\text{minimize}} \quad & \sum_{n=1}^N p_n \log \frac{p_n}{q_n} \\ \text{subject to} \quad & \sum_{n=1}^N p_n T(x_n) = \bar{T}, \ \sum_{n=1}^N p_n = 1, \ p_n \geq 0. \end{aligned} \tag{P}$$

The objective function in the primal problem (P) is the Kullback and Leibler (1951) information of $\{p_n\}$ relative to $\{q_n\}$, which is also known as the relative entropy. This method matches the given moments exactly while keeping the probabilities $\{p_n\}$ as close to the initial approximation $\{q_n\}$ as possible in the sense of the Kullback-Leibler information.[5] Note that since (P) is a convex minimization problem, the solution (if one exists) is unique.

The optimization problem (P) is a constrained minimization problem with a large number (N) of unknowns ($\{p_n\}$) with $L + 1$ equality constraints and N inequality constraints, which is in general computationally intensive to solve. However, it is well-known that entropy-like minimization problems are computationally

the simpler notation x_n and w_n.

[5]The Kullback-Leibler information is not the only possible loss function. One may also use other criteria such as the L^2 norm or other generalized entropies. However, the Kullback-Leibler information has the unmatched feature that (i) the domain of the dual function is the entire space, so the dual problem becomes unconstrained, and (ii) the constraint $p_n \geq 0$ never binds, so the dual problem becomes low dimensional. See Borwein and Lewis (1991) for more details on duality in entropy-like minimization problems and Owen (2001), Tsao (2004), Kitamura (2007), and Tsao and Wu (2013) for discussions on the computational aspects of empirical likelihood methods, which is mathematically related.

tractable by using duality theory (Borwein and Lewis, 1991). Tanaka and Toda (2013) convert the primal problem (P) to the dual problem

$$\max_{\lambda \in \mathbb{R}^L} \left[\lambda' \bar{T} - \log \left(\sum_{n=1}^{N} q_n \mathrm{e}^{\lambda' T(x_n)} \right) \right], \tag{D}$$

which is a *low dimensional* (L unknowns) *unconstrained* concave maximization problem and hence computationally tractable. The following theorem shows how the solutions to the two problems (P) and (D) are related. Below, the symbols "int" and "co" denote the interior and the convex hull of sets.

Theorem 2.1. *1. The primal problem* (P) *has a solution if and only if* $\bar{T} \in \operatorname{co} T(D_N)$. *If a solution exists, it is unique.*

2. The dual problem (D) *has a solution if and only if* $\bar{T} \in \operatorname{int} \operatorname{co} T(D_N)$. *If a solution exists, it is unique.*

3. If the dual problem (D) *has a (unique) solution* λ_N, *then the (unique) solution to the primal problem* (P) *is given by*

$$p_n = \frac{q_n \mathrm{e}^{\lambda_N' T(x_n)}}{\sum_{n=1}^{N} q_n \mathrm{e}^{\lambda_N' T(x_n)}} = \frac{q_n \mathrm{e}^{\lambda_N' (T(x_n) - \bar{T})}}{\sum_{n=1}^{N} q_n \mathrm{e}^{\lambda_N' (T(x_n) - \bar{T})}}. \tag{2.3}$$

2.1.2 Practical implementation

Theorem 2.1 provides a practical way to implement the Tanaka-Toda method. After choosing the initial discretization $Q = \{q_n\}$ and the moment defining function T, one can numerically solve the unconstrained optimization problem (D). To this end, we can instead solve

$$\min_{\lambda \in \mathbb{R}^L} \sum_{n=1}^{N} q_n \mathrm{e}^{\lambda' (T(x_n) - \bar{T})} \tag{D$'$}$$

because the objective function in (D$'$) is a monotonic transformation (-1 times the exponential) of that in (D). Since (D$'$) is an unconstrained convex minimization problem with a (relatively) small number (L) of unknowns (λ), solving it is computationally simple. Letting $J_N(\lambda)$ be the objective function in (D$'$), its gradient and Hessian can be analytically computed as

$$\nabla J_N(\lambda) = \sum_{n=1}^{N} q_n \mathrm{e}^{\lambda' (T(x_n) - \bar{T})} (T(x_n) - \bar{T}), \tag{2.4a}$$

$$\nabla^2 J_N(\lambda) = \sum_{n=1}^{N} q_n \mathrm{e}^{\lambda' (T(x_n) - \bar{T})} (T(x_n) - \bar{T})(T(x_n) - \bar{T})', \tag{2.4b}$$

respectively. In practice, we can quickly solve (D′) numerically using optimization routines by supplying the analytical gradient and Hessian.[6]

If a solution to (D′) exists, it is unique, and we can compute the updated discretization $P = \{p_n\}$ by (2.3). If a solution does not exist, it means that the regularity condition $\bar{T} \in \operatorname{int} \operatorname{co} T(D_N)$ does not hold and we cannot match moments. Then one needs to select a smaller set of moments. Numerically checking whether moments are matched is straightforward: by (2.3), (D′), and (2.4a), the error is

$$\sum_{n=1}^{N} p_n T(x_n) - \bar{T} = \frac{\sum_{n=1}^{N} q_n \mathrm{e}^{\lambda_N'(T(x_n)-\bar{T})}(T(x_n)-\bar{T})}{\sum_{n=1}^{N} q_n \mathrm{e}^{\lambda_N'(T(x_n)-\bar{T})}} = \frac{\nabla J_N(\lambda_N)}{J_N(\lambda_N)}. \tag{2.5}$$

2.1.3 Error estimate and convergence

Tanaka and Toda (2015) prove that whenever the quadrature approximation (2.1) converges to the true value as the number of grid points N tends to infinity, the discrete distribution $\{p_n\}$ in (2.3) also weakly converges to the true distribution f and improves the integration error as follows. Let g be the integrand in (2.1) and consider approximating g using $T = (T_1, \ldots, T_L)$ as basis functions:

$$g(x) \approx \widehat{g}_T(x) = \sum_{l=1}^{L} b_l T_l(x),$$

where $\{b_l\}_{l=1}^{L}$ are coefficients. Let $r_{g,T} = \frac{g-\widehat{g}_T}{\|g-\widehat{g}_T\|_\infty}$ be the normalized remainder term, where $\|\cdot\|_\infty$ denotes the supremum norm. Letting

$$E_{g,N}^{(Q)} = \left| \int_{\mathbb{R}^K} g(x) f(x) \, \mathrm{d}x - \sum_{n=1}^{N} q_n g(x_n) \right|$$

be the integration error under the initial discretization $Q = \{q_n\}$ and $E_{g,N}^{(P)}$ be the error under $P = \{p_n\}$, Tanaka and Toda (2015) prove the error estimate

$$E_{g,N}^{(P)} \le \|g - \widehat{g}_T\|_\infty \left(E_{r_{g,T},N}^{(Q)} + \frac{2}{\sqrt{C}} E_{T,N}^{(Q)} \right), \tag{2.6}$$

where C is a constant explicitly given in the paper. Equation (2.6) says that the integration error improves by the factor $\|g - \widehat{g}_T\|_\infty$, which is the approximation error of the integrand g by the basis functions $\{T_l\}_{l=1}^{L}$ that define the targeted moments. It is clear from (2.6) that the approximation quality of the Tanaka-Toda method depends on two factors, (i) the quality of the initial discretization (how small $E_{g,N}^{(Q)}$ is), and (ii) how well the moment defining functions approximate the integrand (how small $\|g - \widehat{g}_T\|_\infty$ is).

[6] Since the dual problem (D) is a concave maximization problem, one may also solve it directly. However, according to our experience, solving (D′) is numerically more stable. This is because the objective function in (D) is close to linear when $\|\lambda\|$ is large, so the Hessian is close to singular and not well-behaved. On the other hand, since the objective function in (D′) is the sum of exponential functions, it is well-behaved.

2.2 Discretizing general Markov processes

Next we show how to extend the Tanaka-Toda method to the case of time-homogeneous Markov processes.

2.2.1 Description of method

Consider the time-homogeneous first-order Markov process

$$P(x_t \le x' | x_{t-1} = x) = F(x', x),$$

where x_t is the vector of state variables and $F(\cdot, x)$ is a cumulative distribution function (CDF) that determines the distribution of $x_t = x'$ given $x_{t-1} = x$. The dynamics of any Markov process are completely characterized by its Markov transition kernel. In the case of a discrete state space, this transition kernel is simply a matrix of transition probabilities, where each row corresponds to a conditional distribution. We can discretize the continuous process x by applying the Tanaka-Toda method to each conditional distribution separately.

More concretely, suppose that we have a set of grid points $D_N = \{x_n\}_{n=1}^N$ and an initial coarse approximation $Q = (q_{nn'})$, which is an $N \times N$ probability transition matrix. Suppose we want to match some conditional moments of x, represented by the moment defining function $T(x)$. The exact conditional moments when the current state is $x_{t-1} = x_n$ are

$$\bar{T}_n = \mathrm{E}\left[T(x_t) \,|\, x_n\right] = \int T(x) \,\mathrm{d}F(x, x_n),$$

where the integral is over x, fixing x_n. (If these moments do not have explicit expressions, we can use highly accurate quadrature formulas to compute them.) By Theorem 2.1, we can match these moments exactly by solving the optimization problem

$$\begin{aligned} \underset{\{p_{nn'}\}_{n'=1}^N}{\text{minimize}} \quad & \sum_{n'=1}^N p_{nn'} \log \frac{p_{nn'}}{q_{nn'}} \\ \text{subject to} \quad & \sum_{n'=1}^N p_{nn'} T(x_{n'}) = \bar{T}_n, \;\; \sum_{n'=1}^N p_{nn'} = 1, \;\; p_{nn'} \ge 0 \end{aligned} \tag{P_n}$$

for each $n = 1, 2, \ldots, N$, or equivalently the dual problem

$$\min_{\lambda \in \mathbb{R}^L} \sum_{n'=1}^N q_{nn'} \mathrm{e}^{\lambda'(T(x_{n'}) - \bar{T}_n)}. \tag{D'_n}$$

(D'_n) has a unique solution if and only if the regularity condition

$$\bar{T}_n \in \operatorname{int} \operatorname{co} T(D_N) \tag{2.7}$$

holds. We summarize our procedure in Algorithm 2.2 below.

Algorithm 2.2 (Discretization of Markov processes)**.**

1. Select a discrete set of points $D_N = \{x_n\}_{n=1}^N$ and an initial approximation $Q = (q_{nn'})$.
2. Select a moment defining function $T(x)$ and corresponding exact conditional moments $\{\bar{T}_n\}_{n=1}^N$. If necessary, approximate the exact conditional moments with a highly accurate numerical integral.
3. For each $n = 1, \ldots, N$, solve minimization problem (D'_n) for λ_n. Check whether moments are matched using formula (2.5), and if not, select a smaller set of moments. Compute the conditional probabilities corresponding to row n of $P = (p_{nn'})$ using (2.3).

The resulting discretization of the process is given by the transition probability matrix $P = (p_{nn'})$. Since the dual problem (D'_n) is an unconstrained convex minimization problem with a typically small number of variables, standard Newton type algorithms can be applied. Furthermore, since the probabilities (2.3) are strictly positive by construction, the transition probability matrix $P = (p_{nn'})$ is a strictly positive matrix, so the resulting Markov chain is stationary and ergodic.

2.2.2 The regularity condition

How stringent is the regularity condition (2.7)? Note that $\operatorname{co} T(D_N)$ is the convex hull of the image of the grid D_N under the moment defining function T, so any element of $\operatorname{co} T(D_N)$ has the form $\sum_n \alpha_n T(x_n)$, where $\alpha_n \geq 0$, $\sum_n \alpha_n = 1$, and $x_n \in D_N$. Also, by definition $\bar{T}_n = \mathrm{E}\left[T(x_t) \,|\, x_{t-1} = x_n\right]$, which is a weighted average of $T(x)$'s. Therefore in practice it is not hard to meet the regularity condition $\bar{T}_n \in \operatorname{int} \operatorname{co} T(D_N)$. The only case difficulty arises is when x_n is close to the boundary of (the convex hull of) D_N and the stochastic process is highly persistent. Then $\bar{T}_n$ also tends to be close to the boundary of $\operatorname{co} T(D_N)$, and it may happen to be outside the set, violating (2.7). But since the boundary of a convex set has measure zero, for the vast majority of the grid points we are able to match moments exactly. A practical solution to the potential failure of the regularity condition is thus to match moments whenever we can by solving the minimization problem (D'_n), and if a solution fails to exist (which can be checked by computing the error (2.5)), we can match only a subset of the moments $T = (T_1, \ldots, T_L)$.

2.2.3 How to choose the grid

In order to implement our method in practice, we need to overcome two issues: (i) the choice of the grid, and (ii) the choice of the targeted moments.

According to the convergence analysis in Tanaka and Toda (2015), the grid D_N should be chosen as the integration points of the quadrature formula (2.1),

which is used to obtain the initial coarse approximation in (2.2). For simplicity we often choose the trapezoidal formula and therefore even-spaced grids. Alternatively, we can place points using the Gaussian quadrature nodes as in Tauchen and Hussey (1991), or, for that matter, any quadrature formula with positive weights such as Simpson's rule, low-degree Newton-Cotes type formulas, or the Clenshaw-Curtis quadrature (see Davis and Rabinowitz (1984) for quadrature formulas); or quantiles as in Adda and Cooper (2003).

Although tensor grids work well in low dimensional problems, in higher dimensions they are not computationally tractable because the number of grid points increases exponentially with the dimension.[7] In such cases, one needs to use sparse grids (Krueger and Kubler, 2004; Heiss and Winschel, 2008) or select the grid points to delimit sets that the process visits with high probability (Maliar and Maliar, 2015).

In practice, we find that the even-spaced grid (trapezoidal formula) works very well and is robust across a wide range of different specifications. However, if there is some special structure to the conditional distribution, such as normality, a Gaussian quadrature approximation can result in better solution accuracy for dynamic models.

2.2.4 How to choose the moments to match

Our method approximates a continuous Markov process by a discrete transition matrix. A good approximation is one for which the integral of any bounded continuous function using the discrete measure is close to the integral using the original continuous measure. The quality of this approximation depends on how accurately the integrand can be approximated by the moment defining functions (see $\|g - \widehat{g}_T\|_\infty$ in (2.6)).

In the case of a single probability distribution, we can choose a grid over a set with high probability and therefore match as many moments as we wish, up to 1 fewer than the number of grid points. In the case of stochastic processes, the situation is more restrictive. As an illustration, consider the AR(1) process

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

with ρ close to 1.

Let $D_N = \{\bar{x}_1, \ldots, \bar{x}_N\}$ be the grid, with $\bar{x}_1 < \cdots < \bar{x}_N$. When $x_{t-1} = \bar{x}_N$, the conditional distribution of x_t is $N(\rho\bar{x}_N, 1)$. But when ρ is close to 1, this (true) distribution has nearly 1/2 of its probability mass on the interval $(\bar{x}_N, \infty)$, which lies outside the grid. Since there is such a discrepancy between the location of the grid points and the probability mass, we do not have the flexibility to match many moments, because the regularity condition $\bar{T}_n \in \operatorname{int} \operatorname{co} T(D_N)$ may fail to hold near the boundary. In the examples below, we consider matching up to 4 conditional moments whenever we can.

[7] Note that with our method, having a large number of grid points is not an issue for solving the dual problem (D'_n). The number of unknowns is equal to the number of targeted moments, which is fixed. The issue with tensor grids is that the *number of dual problems* we need to solve grows exponentially with the dimension.

3 Discretizing VAR(1)s and stochastic volatility models

Applied researchers often specify vector autoregressive processes (VARs) to describe the underlying shocks in their models. In this section we explain how our method can be used to discretize general VARs and stochastic volatility models, and prove some theoretical properties.

3.1 VAR(1)

Suppose we want to discretize a VAR(1) process

$$x_t = (I - B)\mu + Bx_{t-1} + \eta_t, \quad \eta_t \sim N(0, \Psi), \tag{3.1}$$

where all vectors are in $\mathbb{R}^K$, μ is the unconditional mean of x_t, Ψ is the conditional variance matrix, and B is a $K \times K$ matrix with all eigenvalues smaller than 1 in absolute value in order to guarantee stationarity. Using the Cholesky decomposition, without loss of generality, we can rewrite (3.1) as

$$y_t = Ay_{t-1} + \varepsilon_t, \tag{3.2}$$

where $y_t = C^{-1}(x_t - \mu)$, $A = C^{-1}BC$, $\varepsilon_t = C^{-1}\eta_t \sim N(0, D)$, C is lower triangular, D is diagonal (typically $D = I$), and $\Psi = CDC'$.[8] Once we have a discretization for y_t, we have one for $x_t = \mu + Cy_t$.

3.1.1 Description of method

First we introduce some additional notation. Let $y_t = (y_{1t}, \dots, y_{Kt})$ and assume that the discrete approximation of y_{kt} takes N_k values denoted by $D_{k,N_k} = \{\bar{y}_{kn}\}_{n=1}^{N_k}$. In total, there are $J = N_1 \times \cdots \times N_K$ states.[9] Let $j = 1, \dots, J$ be an index of the state, corresponding to a particular combination of points

[8] Clearly there are infinitely many such decompositions. Experience tells that the quality of discretization is best when each compoment of the y_t process in (3.2) has the same unconditional variance. We can do as follows to construct such a decomposition. First, take $\tilde{C}$ such that $\Psi = \tilde{C}\tilde{C}'$, so $D = I$. Define $\tilde{y}_t = \tilde{C}^{-1}(x_t - \mu)$, $\tilde{A} = \tilde{C}^{-1}B\tilde{C}$, and $\tilde{\varepsilon}_t = \tilde{C}^{-1}\eta_t \sim N(0, I)$. Let $\tilde{\Sigma}$ be the unconditional variance of the $\tilde{y}$ process. Let $y_t = U'\tilde{y}_t$ for some orthogonal matrix U, and define $A = U'\tilde{A}U$, $\varepsilon_t = U'\tilde{\varepsilon}_t$, and $C = \tilde{C}U'$. Then $\operatorname{Var}[\varepsilon_t] = U'IU = I$. The unconditional variance of the y process is then $\Sigma = U'\tilde{\Sigma}U$. Since $\operatorname{tr}\Sigma = \operatorname{tr}\tilde{\Sigma}$, the diagonal elements of Σ become equal if $\Sigma_{kk} = (U'\tilde{\Sigma}U)_{kk} = \frac{1}{K}\operatorname{tr}\tilde{\Sigma}$. We can make this equation (approximately) true by solving the optimization problem

$$\begin{aligned} &\text{minimize} && \sum_{k=1}^{K}\left((U'\tilde{\Sigma}U)_{kk} - \frac{1}{K}\operatorname{tr}\tilde{\Sigma}\right)^2 \\ &\text{subject to} && U'U = I. \end{aligned}$$

With this choice of U, the unconditional variances of the compoments of $\{y_t\}$ are close to each other, and in fact equal if the objective function takes the value zero.

[9] In practice, we take $N_1 = N_2 = \cdots = N_K = N$, so $J = N^K$.

$(\bar{y}_{1n}(j), \ldots, \bar{y}_{Kn}(j))$. Let $p_{kn}(j)$ be the probability that $y_{kt} = \bar{y}_{kn}$ conditional on being in state j. Define the conditional mean and variance of y_{kt} given state j as $\mu_k(j)$ and $\sigma_k(j)^2$, respectively. We outline the procedure in Algorithm 3.1. (Although we describe it for the case of two conditional moments, the case with higher order moments is similar.)

Algorithm 3.1 (Discretization of VAR(1) processes)**.**

1. For each component of $y_t = (y_{1t}, \ldots, y_{Kt})$, select a discrete set of points $D_{k,N_k} = \{\bar{y}_{kn}\}_{n=1}^{N_k}$.
2. For $j = 1, \ldots, J$,
   (a) For $k = 1, \ldots, K$ (note that we can treat each component k separately because the variance-covariance matrix D is diagonal),
      i. Define the moment defining function and exact moments by
      $$T_{kj}(x) = \begin{bmatrix} x \\ (x - \mu_k(j))^2 \end{bmatrix} \quad \text{and} \quad \bar{T}_{kj} = \begin{bmatrix} \mu_k(j) \\ \sigma_k(j)^2 \end{bmatrix}.$$
      ii. Select an initial approximation $\{q_{kn}(j)\}_{n=1}^{N_k}$, where $q_{kn}(j)$ is the probability of moving to point $\bar{y}_{kn}$ conditional on being in state j.
      iii. Solve minimization problem (D'_n) for λ_{kj} and compute the conditional probabilities $\{p_{kn}(j)\}_{n=1}^{N_k}$ using (2.3).
   (b) Compute the conditional probabilities $\{p_{jj'}\}_{j'=1}^{J}$ by multiplying together the conditional probabilities $p_{kn}(j)$ that make up transitions to elements of state j'.
3. Collect the conditional probabilities $\{p_{jj'}\}_{j'=1}^{J}$ into a matrix $P = (p_{jj'})$.

In order to determine $\{p_{kn}(j)\}$ using Algorithm 3.1, we need an initial coarse approximation $\{q_{kn}(j)\}$. The simplest way is to take the grid points $\{\bar{y}_{kn}\}_{n=1}^{N_k}$ to be evenly spaced and assign $q_{kn}(j)$ to be proportional to the conditional density of y_{kt} given state j, which corresponds to choosing the trapezoidal rule for the initial quadrature formula. Alternatively, we can use the nodes and weights of the Gauss-Hermite quadrature as in Tauchen and Hussey (1991),[10] or take the grid points $\{\bar{y}_{kn}\}_{n=1}^{N_k}$ as quantiles of the unconditional distribution and assign probabilities according to the cumulative distribution function, as in Adda and Cooper (2003).[11]

[10]Following the original paper by Tauchen and Hussey (1991), we always use the conditional variance matrix D to construct the Gauss-Hermite quadrature. This is the most logical way since dynamic economic models involve conditional expectations (*e.g.*, Euler equations), which are integrals that use the conditional distributions.

[11]The specific procedure is as follows. Let the stationary distribution of y_{kt} be $N(0, \sigma_k^2)$. Since

Which grid/quadrature formula is best is a practical problem and we explore this issue in subsequent sections.

This method can be generalized to VAR(p) processes, although the dimension of the state space would grow exponentially in p unless we use a sparse grid.

3.1.2 Theoretical properties of the discretization

If a solution to the dual problem (D'_n) exists, by construction our method generates a finite-state Markov chain approximation of the VAR with exact 1-step ahead conditional moments. But how about k-step ahead conditional moments and unconditional moments? The following theorem provides an answer.

Theorem 3.2. *Consider the VAR(1) process in* (3.2)*, with grid D_N. Suppose that the regularity condition $\bar{T}_n \in \operatorname{int}\operatorname{co} T(D_N)$ holds, and hence our method matches the conditional mean and variance. Then the method also matches any k-step ahead conditional mean and variance, as well as the unconditional mean and all autocovariances (hence spectrum).*

This result holds even for a certain class of stochastic volatility models (Theorem A.1). According to its proof, there is nothing specific to the choice of the grid, the normality of the process, or the diagonalization. Therefore the result holds for any non-Gaussian linear process.

So far, we have assumed that the regularity condition (2.7) holds, so that a discrete approximation with exact conditional moments using our method exists. As we see in the numerical examples below, such a discretization exists most of the time, but not always. Therefore it is important to provide easily verifiable conditions that guarantee existence. For general VARs, the following proposition shows that it is always possible to match conditional means.

Proposition 3.3. *Consider the VAR(1) process in* (3.2) *with coefficient matrix $A = (a_{kk'})$. Let $|A| = (|a_{kk'}|)$ be the matrix obtained by taking the absolute value of each element of A. If the spectral radius of $|A|$ is less than 1 (*i.e.*, all eigenvalues are less than 1 in absolute value), then there exists a tensor grid such that we can match all conditional means.*

How about the conditional mean and variance? Since addressing this issue for general VAR processes is challenging, we restrict our analysis to the case of an AR(1) process. The following proposition shows that a solution exists if the grid is symmetric, sufficiently fine, and the grid points span more than one unconditional standard deviation around 0.

---

there are N_k discrete points for y_{kt}, we divide the real line $\mathbb{R}$ into N_k intervals using the n-th N_k-quantile ($n = 1, \ldots, N_k - 1$), which we denote by $I_{k1}, \ldots, I_{kN}$. The discrete points are then the median of each interval, so $\bar{y}_{kn} = F^{-1}((2n-1)/2N_k)$ ($n = 1, 2, \ldots, N_k$), where F is the CDF of $N(0, \sigma_k^2)$. When the $t-1$ state is j, since the conditional distribution of y_{kt} is $N(\mu_k(j), \sigma_k^2(j))$, we assign initial probability $q_{kn}(j) = P(I_{kn})$ to the point $\bar{y}_{kn}$ under the conditional distribution $N(\mu_k(j), \sigma_k^2(j))$.

Proposition 3.4. *Consider the AR(1) process*

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim (0, 1),$$

where $0 \leq \rho < 1$. *Suppose that (i) the grid* $\{\bar{x}_n\}_{n=1}^N$ *is symmetric and spans more than one unconditional standard deviation around 0, so* $\max_n |\bar{x}_n| > 1/\sqrt{1-\rho^2}$, *and (ii) either the maximum distance between two neighboring grid points is less than 2, or for each positive grid point* $\bar{x}_n > 0$ *there exists a grid point* $\bar{x}_{n'}$ *such that*

$$\rho\bar{x}_n - \frac{1}{(1-\rho)\bar{x}_n} < \bar{x}_{n'} \leq \rho\bar{x}_n. \tag{3.3}$$

Then (D'_n) *has a unique solution for all* n.

When the grid $\{\bar{x}_n\}$ is even-spaced, we can obtain a simple sufficient condition for existence.

Corollary 3.5. *Let the grid points* $\{\bar{x}_n\}_{n=1}^N$ *be symmetric and even-spaced,* $\sigma = 1/\sqrt{1-\rho^2}$ *be the unconditional standard deviation, and* $M = \max_n \bar{x}_n$. *Suppose that either*

1. $\rho \leq 1 - \frac{2}{N-1}$ *and* $\sigma < M \leq \sqrt{2}\sigma\sqrt{N-1}$, *or*
2. $\rho > 1 - \frac{2}{N-1}$ *and* $\sigma < M \leq \sigma\sqrt{N-1}$.

Then (D'_n) *has a unique solution for all* n.

Interestingly, Kopecky and Suen (2010) show that the Rouwenhorst (1995) method matches the first and second conditional moments when the grid span is $M = \sigma\sqrt{N-1}$, the upper bound in Corollary 3.5 for the case $\rho > 1 - \frac{2}{N-1}$. Choosing a grid span of order $\sqrt{N}$ can also be theoretically justified. In that case, the grid spacing is of order $N/\sqrt{N} = 1/\sqrt{N}$. Since the grid gets finer while the grid span tends to infinity, the trapezoidal formula converges to the true integral. Therefore the approximation error can be made arbitrarily small by increasing N. For general VARs, we do not have theoretical results for the existence of a discretization that matches second moments. However, we recommend using a grid span $M = \sigma\sqrt{N-1}$ in each dimension, where σ is the square root of the smallest eigenvalue of the unconditional variance of the VAR.

Theorem 3.2, Proposition 3.4, and Corollary 3.5 are significant. Note that among all existing methods, the Rouwenhorst (1995) method for discretizing Gaussian AR(1) processes is the only one known to match the first and second conditional moments exactly.[12]

[12]Kopecky and Suen (2010) prove that the 1-step ahead conditional moments are exact. By Theorem 3.2, all k-step ahead conditional moments are also exact.

3.2 AR(1) with stochastic volatility

Consider an AR(1) process with stochastic volatility of the form

$$y_t = \lambda y_{t-1} + u_t, \qquad u_t \sim N(0, \mathrm{e}^{x_t}), \tag{3.4a}$$
$$x_t = (1-\rho)\mu + \rho x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2), \tag{3.4b}$$

where x_t is the unobserved log variance process and y_t is the observable, *e.g.*, stock returns. We assume that y_t is mean zero without loss of generality.

Since the log variance process x_t evolves independently of the level y_t as an AR(1) process, we can discretize it using Algorithm 3.1. For y_t, note that the unconditional variance is given by

$$\sigma_y^2 = \mathrm{E}[y_t^2] = \frac{\mathrm{E}[\mathrm{e}^{x_t}]}{1-\lambda^2}.$$

Since the unconditional distribution of x_t is $N\left(\mu, \frac{\sigma^2}{1-\rho^2}\right)$, we have

$$\mathrm{E}[\mathrm{e}^{x_t}] = \exp\left(\mu + \frac{\sigma^2}{2(1-\rho^2)}\right)$$

using the properties of lognormal random variables. We can then construct an even-spaced grid for y_t spanning some number of unconditional standard deviations around 0.

With some more algebra, we can show that

$$y_t | x_{t-1}, y_{t-1} \sim N\left(\lambda y_{t-1}, \exp\left((1-\rho)\mu + \rho x_{t-1} + \sigma^2/2\right)\right).$$

We discretize these conditional distributions for each (x_{t-1}, y_{t-1}) pair using our method and combine them with the discretization obtained for $x_t | x_{t-1}$ above, to come up with a joint transition matrix for the state (x_t, y_t).

3.3 Relation to the existing literature

In this section we discuss the existing literature in detail.

The standard method for approximating an AR(1) process is that of Tauchen (1986), which divides the state space into evenly spaced intervals, with the grid chosen as the midpoints of those intervals. Tauchen constructs each approximate conditional distribution by matching the probabilities of transitioning from a particular point to each interval. The Tauchen method is intuitive, simple, and reasonably accurate when the number of grid points is large enough. It is easily generalized and widely used for the approximation of VAR processes. Variants of the Tauchen method have been developed in the literature by using Gauss-Hermite quadrature (Tauchen and Hussey, 1991), placing grid points using quantiles instead of even-spaced intervals (Adda and Cooper, 2003), and using multivariate normal integration techniques (Terry and Knotek, 2011). Rouwenhorst (1995) proposes an alternative discretization method of a Gaussian AR(1) process that

matches the unconditional first and second moments exactly. His idea is to approximate a normal distribution by binomial distributions.

VARs are highly persistent in typical macroeconomic applications. It has been recognized that the Tauchen and Tauchen-Hussey methods often fail to give accurate approximations to such processes (Zhang, 2005; Flodén, 2008),[13] which has spurred a renewed research interest in accurately discretizing autoregressive processes. Kopecky and Suen (2010) prove that for a certain choice of the grid, the Rouwenhorst method actually matches the autocorrelation and the *conditional* mean and variance. This means that the Rouwenhorst method is suitable for discretizing highly persistent Gaussian AR(1) processes, for which earlier methods failed. Applying it to typical macroeconomic models such as stochastic growth and income fluctuation models, they show that the relative error in the solution accuracy is less than 1% with the Rouwenhorst method, compared with 10–20% with earlier methods.

Galindev and Lkhagvasuren (2010) generalize the Rouwenhorst method to the multivariate case by transforming a VAR into a set of cross-correlated AR(1) processes. However, their method works only when the AR(1) processes are equally persistent (a knife-edge case), for otherwise the state space is not finite.

Gospodinov and Lkhagvasuren (2014) propose an alternative discretization method of VARs by first discretizing independent AR(1) processes using the Rouwenhorst method and then targeting the first and second conditional moments to mimic the conditional distributions of the actual VAR process. Solving a stochastic growth model with a highly persistent bivariate VAR, they find that the relative error in the solution accuracy is about 1–3% with their method, compared with 10–30% with the Tauchen method.

Since our method matches conditional moments, it is similar in spirit to Rouwenhorst (1995) (AR(1)) and Gospodinov and Lkhagvasuren (2014) (VAR(1)), though our method is not limited to VARs. Here we contrast our method to these two in more details. According to Proposition 3 in Kopecky and Suen (2010), the ergodic distribution of the resulting Markov chain of the Rouwenhorst method is a standardized binomial distribution with parameter $N-1$ and $s = 1/2$, so by the central limit theorem it converges to $N(0,1)$ as $N \to \infty$. This argument suggests that the Rouwenhorst method is designed to discretize a Gaussian AR(1). It immediately follows that neither our method (for AR(1)) nor the Rouwenhorst method is a special case of the other: our method is not limited to Gaussian AR(1) processes (Proposition 3.4 and Corollary 3.5 do not assume normality), and generally has a different grid.

With regard to VARs, both the Gospodinov and Lkhagvasuren (2014) (GL) method and ours target the first and second conditional moments. The GL method uses the Rouwenhorst method to obtain a preliminary discretization and then targets the moments. As GL acknowledge in their paper, the GL method has fewer free variables than the number of targeted moments, and hence it is generally

[13] In the original paper, Tauchen (1986) himself admits that "[e]xperimentation showed that the quality of the approximation remains good except when λ [the persistence parameter] is very close to unity."

impossible to match all moments. While we do not have a proof that our method matches all first and second conditional moments (Proposition 3.3 shows that it is possible to match conditional means), according to our experience it seems that for most applications we can indeed match all first two conditional moments when we use the even-spaced grid. Again neither of the two methods is a special case of the other.

We do not claim that our method is always preferable, although we emphasize that our method is not limited to the discretization of linear Gaussian processes. Whether our method is superior or not can only be answered by studying the accuracy in specific problems. The Online Appendix compares the accuracy of discretization and shows that our method outperforms existing ones by several orders of magnitude. However, discretization is not an end in itself. A more important question is whether different discretization methods lead to substantial differences in the solution accuracy of dynamic economic models, and whether these differences matter economically. We provide answers to these questions in the next sections.

4 Solution accuracy of asset pricing models

Whenever one proposes a new numerical method for solving dynamic models, it must be evaluated by two criteria: (i) Does the new method improve the solution accuracy of well-known, standard dynamic economic models? (ii) Can the new method be applied to solve more complicated models for which existing methods are not readily available? In order for a new method to be useful, it must meet at least one (preferably both) of these two criteria.

This section addresses these questions by solving simple asset pricing models with or without Gaussian shocks. We use the closed-form solutions obtained by Burnside (1998) for Gaussian shocks and Tsionas (2003) for non-Gaussian shocks as comparison benchmarks.[14]

4.1 Model and numerical solution

Consider a representative agent with additive CRRA utility function

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma},$$

where C_t is consumption, $\beta > 0$ is the discount factor, and $\gamma > 0$ is the coefficient of relative risk aversion. The agent is endowed with aggregate consumption $\{C_t\}_{t=0}^{\infty}$, and can trade assets in zero net supply. Let D_t be the dividend to an asset and P_t be its price. When log consumption and dividend growth

$$x_t = (\log(C_t/C_{t-1}), \log(D_t/D_{t-1}))$$

[14]Collard and Juillard (2001) and Schmitt-Grohé and Uribe (2004) also use this model in order to evaluate the solution accuracy of the perturbation method.

follow a VAR(1) process with i.i.d. shocks, it is possible to obtain a closed-form solution for the price-dividend ratio $V_t = P_t/D_t$, which depends only on x_t. See the Online Appendix for details.

We obtain numerical solutions as follows. By the Euler equation, we have

$$P_t = \mathrm{E}_t[\beta(C_{t+1}/C_t)^{-\gamma}(P_{t+1} + D_{t+1})]. \tag{4.1}$$

Dividing (4.1) by D_t, we obtain

$$V_t = \beta\,\mathrm{E}_t[\exp(\alpha' x_{t+1})(V_{t+1} + 1)], \tag{4.2}$$

where $\alpha = (-\gamma, 1)'$. Suppose that the process for consumption and dividend growth is discretized. Let $s = 1, \ldots, S$ be the states, x_s be the vector of log consumption/dividend growth in state s, and $P = (\pi_{ss'})$ be the transition probability matrix. Then the discrete analog of (4.2) is

$$v_s = \beta \sum_{s'=1}^{S} \pi_{ss'} \mathrm{e}^{\alpha' x_{s'}}(v_{s'} + 1), \tag{4.3}$$

where v_s is the price-dividend ratio in state s. Let $v = (v_1, \ldots, v_S)'$ $(S \times 1)$ and $X = (x_1', \ldots, x_S')'$ $(S \times 2)$ be the matrices of those values. Then (4.3) is equivalent to the linear equation

$$v = \beta P \,\mathrm{diag}(\mathrm{e}^{X\alpha})(v + 1) \iff v = (I - \beta P \,\mathrm{diag}(\mathrm{e}^{X\alpha}))^{-1}\beta P \mathrm{e}^{X\alpha}. \tag{4.4}$$

This formula gives the price-dividend ratio only at the grid points, and one might be interested in computing the value at any point. In this case, we can use the projection method (Judd, 1992). The idea of the projection method with Chebyshev collocation is to approximate the unknown policy function using Chebyshev polynomials as a basis.[15] Suppose we approximate $V(x)$ as

$$\widehat{V}(x; b) = \sum_{s=1}^{S} b_s \Psi_s(x),$$

where $\{\Psi_s\}_{s=1}^{S}$ is a set of basis functions (Chebyshev polynomials) and $b = \{b_s\}_{s=1}^{S}$ is the vector of coefficients to be determined. We can solve for b that sets the Euler equation (4.2) to exactly zero at each of the S grid points implied by each discretization method, which leads to an exactly identified system. The equation becomes

$$\widehat{V}(x_s; b) = \beta \sum_{s'=1}^{S} \pi_{ss'} \mathrm{e}^{\alpha' x_{s'}} \left(\widehat{V}(x_{s'}; b) + 1\right). \tag{4.5}$$

However, if we set $v_s = \widehat{V}(x_s; b)$, then (4.5) becomes identical to (4.3)! Therefore finding coefficients $\{b_s\}$ that solve (4.5) is equivalent to first solving the linear equation (4.3) (whose solution is given by (4.4)) and then finding an interpolating polynomial. We summarize the above discussion in the following proposition.

[15] Unlike standard Chebyshev collocation, we are constrained to solve for coefficients that set the Euler equation residuals equal to 0 at the discretization points rather than the zeroes of the Chebyshev polynomial. This in general means we are only guaranteed pointwise convergence of our approximation rather than uniform convergence.

Proposition 4.1. *Solving an asset pricing model with a continuous state space using discretization and projection is equivalent to solving a model with a discrete state space, which can be done by inverting a matrix as in* (4.4)*. The continuous solution can be obtained by interpolating the discrete solution.*

Proposition 4.1 is quite powerful. Note that there is nothing specific to the preferences of the agent or the underlying stochastic process needed to apply the proposition. For example, suppose that the agent has a general recursive utility of the form

$$U_t = f(C_t, \mathcal{M}_t(U_{t+1})), \tag{4.6}$$

where U_t is the utility at time t, C_t is consumption, f is the aggregator, and $\mathcal{M}_t$ is the certainty equivalent of the continuation utility U_{t+1}.[16] Suppose that $f, \mathcal{M}$ are homogeneous of degree 1 (which is true for almost all applications) and the underlying stochastic process is discretized. Dividing (4.6) by C_t, we can solve for the S nonlinear equations in S unknowns

$$u_s = f(1, \mathcal{M}_s(\mathrm{e}^{x_{ss'}} u_{s'})), \tag{4.7}$$

where $x_{ss'}$ is log consumption growth from state s to s' and $u_s = (U_t/C_t)(s)$ is the utility-consumption ratio in state s. After solving for these values $\{u_s\}$, one can compute the pricing kernel and price any assets by inverting a matrix as in (4.4). In practice, solving (4.7) and inverting a matrix to compute asset prices take only a fraction of a second to carry out.[17]

4.2 Calibration

We calibrate the model at annual frequency. We select the preference parameters $\beta = 0.95$ and $\gamma = 2$, which are relatively standard in the macro literature. We consider three specifications for the law of motion of x_t: Gaussian AR(1), Gaussian VAR(1), and AR(1) with non-Gaussian shocks. We estimate the parameters of each of these models using data on real personal consumption expenditures per capita of nondurables from FRED, and 12-month moving sums of dividends paid on the S&P 500 obtained from the spreadsheet in Welch and Goyal (2008).[18] For the two univariate specifications, we assume that $C_t = D_t$, *i.e.*, $x_{1,t} = x_{2,t} = x_t$, and use the data on dividends to estimate the parameters.

The reason why we use dividend data instead of consumption data for the univariate models is as follows. Given the mean μ and persistence ρ of the AR(1) process, according to Tsionas (2003) the price-dividend ratio depends only on the moment generating function (MGF) $M(s)$ of the shock distribution in the range

[16] A typical example is $f(c, v) = ((1-\beta)c^{1-1/\psi} + \beta v^{1-1/\psi})^{\frac{1}{1-1/\psi}}$ (CES aggregator with elasticity of intertemporal substitution ψ) and $\mathcal{M}_t(X) = \mathrm{E}_t[X^{1-\gamma}]^{\frac{1}{1-\gamma}}$ (CRRA certainty equivalent with relative risk aversion γ) in which case we obtain the Epstein-Zin preference.

[17] The idea of using discretization to solve asset pricing models is not particularly new: see, for example, Mehra and Prescott (1985), Cecchetti et al. (1993), and Bonomo et al. (2011), among others. The point is that there have been no systematic ways to accurately discretize the underlying stochastic process in the literature to make discretization a viable option.

[18] `http://www.hec.unil.ch/agoyal/`

$\frac{1-\gamma}{1-\rho} \leq s \leq 1-\gamma$ (assuming $\gamma > 1$ and $\rho > 0$). But if two shock distributions have identical mean and variance, then the Taylor expansion of their MGF around $s = 0$ will coincide up to the second order term. Therefore, in order to make a difference for asset pricing, we either need to (i) move away from $s = 0$ by increasing γ, (ii) make the domain of the MGF larger by increasing ρ, or (iii) make the MGF more nonlinear by increasing the variance or skewness. Since dividend growth is more persistent, volatile, and skewed than consumption growth, using dividend growth will make the contrasts between methods more stark.

4.3 Solution accuracy

After computing the numerical and closed-form solutions as described in the Online Appendix, we evaluate the accuracy by the $\log_{10}$ relative errors

$$\log_{10} \left| \widehat{V}(x)/V(x) - 1 \right|,$$

where $V(x)$ is the true price-dividend ratio at x and $\widehat{V}(x)$ is the approximate (numerical) solution corresponding to each method obtained by the interpolating polynomial as in Proposition 4.1. To compare the relative errors of each method, we first take the largest common support across all discretization methods so that the approximation is well defined, and then compute the relative errors on a fine grid (say 1,001 points in each dimension) on this support. All methods beginning with "ME" refer to the maximum entropy method developed in this paper with different choices of the underlying grid and quadrature formula. For example, "ME-Even" refers to the maximum entropy method using an even-spaced grid.

4.3.1 Gausian AR(1)

Modeling the dynamics of dividend growth by a Gaussian AR(1) is straightforward and we relegate the details to the Online Appendix.

4.3.2 Gaussian VAR(1)

We next consider specifying the joint dynamics of dividend growth and consumption growth as a Gaussian VAR(1)

$$x_t = (I - B)\mu + Bx_{t-1} + \eta_t, \quad \eta_t \sim N(0, \Psi)$$

where μ is a 2×1 vector of unconditional means, B is a 2×2 matrix with eigenvalues less than 1 in absolute value, η is a 2×1 vector of shocks, and Ψ is a 2×2 variance covariance matrix. The estimated parameters of the VAR(1) model are

$$\mu = \begin{bmatrix} 0.0128 \\ 0.0561 \end{bmatrix}, \quad B = \begin{bmatrix} 0.3237 & -0.0537 \\ 0.2862 & 0.3886 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 0.000203 & 0.000293 \\ 0.000293 & 0.003558 \end{bmatrix}.$$

The eigenvalues of B are $0.3561 \pm 0.1196i$, with spectral radius $\rho(B) = 0.3757$, so the VAR is moderately persistent.

We consider eight different discretization methods. For our method, we consider the even-spaced grid with 2 or 4 moments (ME-Even (2,4)), the quantile grid (ME-Quant), and the Gauss-Hermite quadrature grid (ME-Quad). For existing methods, we consider those of Tauchen (1986)(Tau), Tauchen and Hussey (1991) (TH), and Gospodinov and Lkhagvasuren (2014) with (GL) and without (GL0) moment matching. Figure 4.1 shows the graphs of $\log_{10}$ relative errors for the VAR(1) model. Table 4.1 shows the mean and maximum $\log_{10}$ relative errors over the entire grid.

Table 4.1: Mean and maximum $\log_{10}$ relative errors for the asset pricing model with VAR(1) consumption/dividend growth.

	ME methods				Existing methods			
N	Even (2)	Quant	Quad	Even (4)	Tau	TH	GL0	GL
Mean $\log_{10}$ *errors*								
5	-3.381	-2.963	-5.028	-3.570	-1.463	-2.964	-3.439	-2.191
7	-3.667	-3.066	-6.758	-5.134	-1.520	-4.920	-2.586	-2.618
9	-3.949	-3.146	-8.563	-6.739	-1.546	-6.900	-2.449	-3.106
Maximum $\log_{10}$ *errors*								
5	-3.292	-2.865	-4.975	-3.485	-1.327	-2.890	-2.365	-1.982
7	-3.566	-2.954	-6.717	-4.891	-1.360	-4.838	-2.125	-2.140
9	-3.838	-3.022	-8.451	-5.730	-1.370	-6.581	-2.212	-2.471

For all choices of N, the Gaussian quadrature based methods, ME-Quad and TH, perform the best, with ME-Quad being always about two orders of magnitude more accurate than TH. For even-spaced methods, the order of accuracy is always ME-Even (4) > ME-Even (2) > GL0, GL > Tauchen, and ME-Even (4) is as accurate as Tauchen-Hussey. ME-Quant is not particularly accurate but its performance is similar to the GL methods. According to Table 4.1, the conclusions drawn from Figure 4.1 are robust.

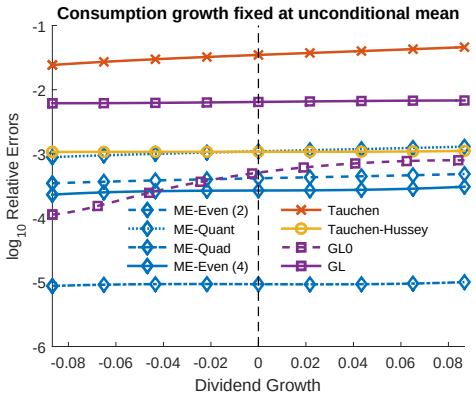
4.3.3 AR(1) with non-Gaussian shocks

Researchers often assume normality of the conditional shock distributions for analytical and computational convenience. However, there is much evidence of non-normality in financial data. One might prefer to specify a parametric distribution with fatter tails, or refrain from parametric specifications altogether. For this reason, we consider an AR(1) with i.i.d., but non-Gaussian shocks:

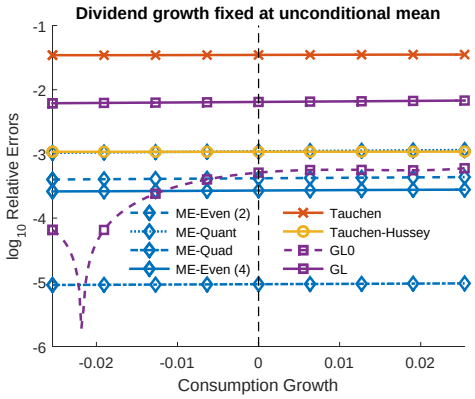
$$x_t = (1-\rho)\mu + \rho x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim F.$$

We model the shock distribution F by a Gaussian mixture, because it is flexible yet analytically tractable (all moments and moment generating function have closed-form expressions). Table 4.2 shows the parameter estimates.

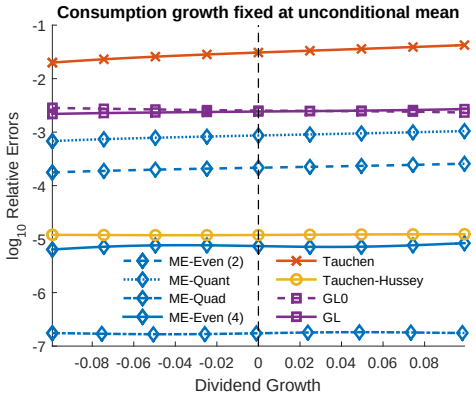
Figure 4.2 plots the PDFs of ε_t fit to the dividend growth data under the assumptions of normal and Gaussian mixture shocks, as well as the nonparametric



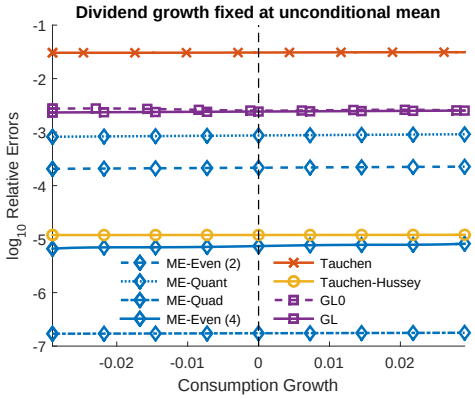
(a) $N = 5$, fixed consumption.



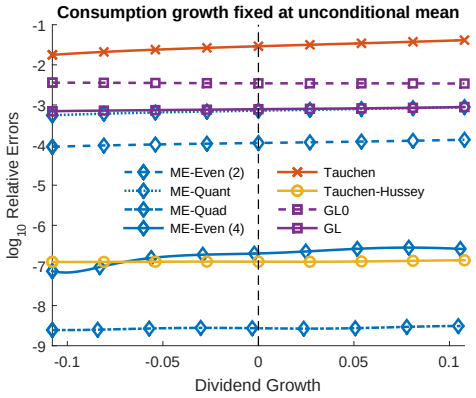
(b) $N = 5$, fixed dividend.



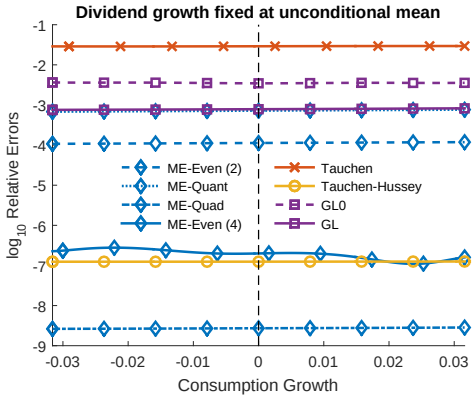
(c) $N = 7$, fixed consumption.



(d) $N = 7$, fixed dividend.



(e) $N = 9$, fixed consumption.



(f) $N = 9$, fixed dividend.

Figure 4.1: $\log_{10}$ relative errors of price-dividend ratio with various discretization methods and number of points for the VAR(1) model.

Note: each row corresponds to a certain number of grid points ($N = 5, 7, 9$). The left panels show the accuracy along the dividend growth dimension, fixing consumption growth at its unconditional mean. The right panels fix dividend growth at its unconditional mean and vary consumption growth. The grids are demeaned so that the unconditional mean corresponds to 0 in the figures.

kernel density estimate. The Gaussian mixture with three components appears to capture the skewness and kurtosis lacking in the normal specification by placing

Table 4.2: Parameters of the AR(1) process with Gaussian mixture shocks.

Parameter	Symbol	Value
Mean dividend growth	μ	0.0559
Persistence of dividend growth	ρ	0.4049
Volatility of dividend growth	σ	0.0589
Proportion of mixture components	w_j	0.0304, 0.8489, 0.1207
Mean of mixture components	μ_j	-0.2282, -0.0027, 0.0766
S.D. of mixture components	σ_j	0.0513, 0.0316, 0.0454

Note: this table shows the parameter estimates of the AR(1) process with Gaussian mixture shocks $x_t = (1-\rho)\mu + \rho x_{t-1} + \varepsilon_t$, where $x_t = \log(D_t/D_{t-1})$ is log dividend growth and $\varepsilon_t \sim N(\mu_j, \sigma_j^2)$ with probability w_j, $j = 1, \ldots, J$. μ, ρ are estimated by OLS. $\sigma = \sqrt{\text{Var}[\varepsilon_t]}$ is computed from the squared sum of residuals. The Gaussian mixture parameters are estimated by maximum likelihood from the residuals, and the number of components $J = 3$ is chosen to minimize the Akaike Information Criterion (AIC).

more weight on large negative realizations of the shock as well as ones close to zero.

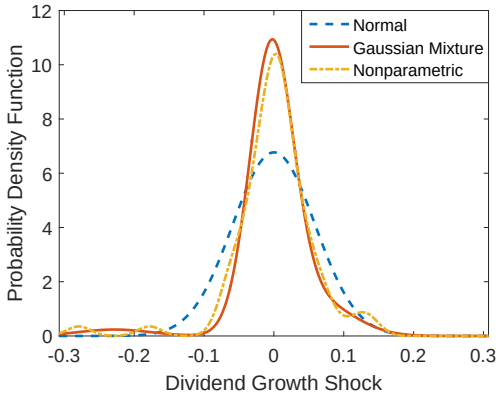


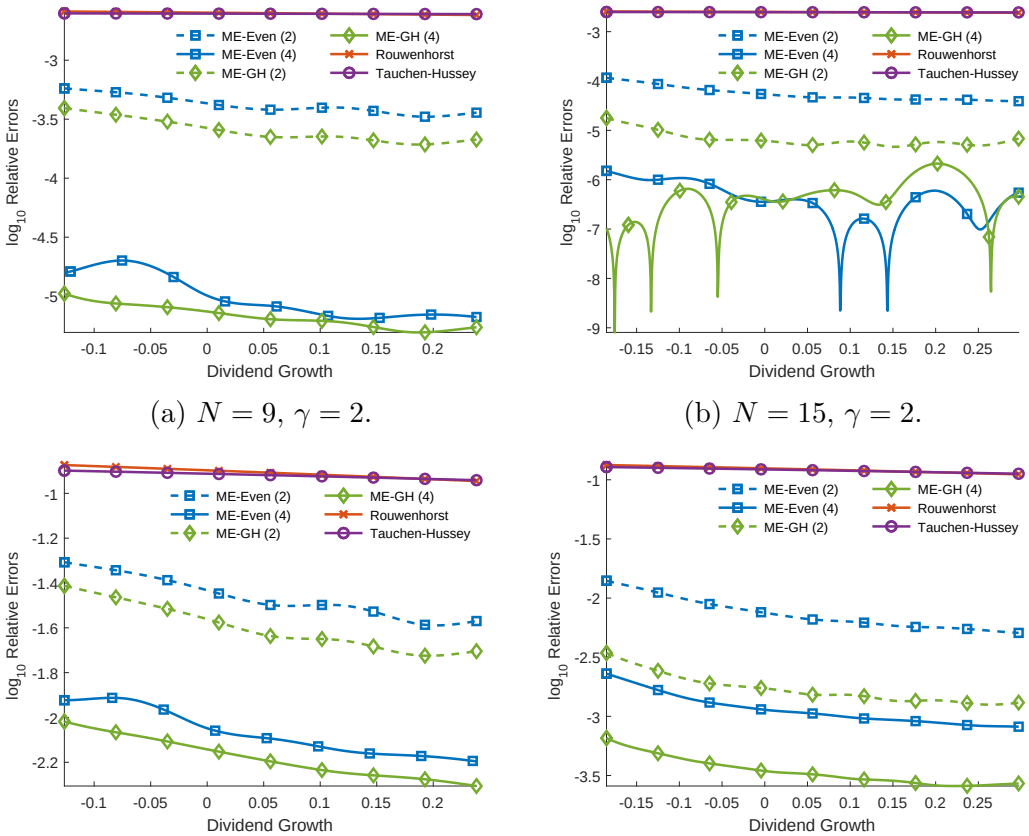
Figure 4.2: Densities fitted to AR(1) OLS residuals.

We consider six different discretizations for the log dividend growth process. The first two are the Rouwenhorst (1995) and the Tauchen and Hussey (1991) methods, which can be thought of as a case where the researcher incorrectly believes the conditional density to be Gaussian. The other four methods are the ME methods with even-spaced (ME-Even) or Gauss-Hermite quadrature grid (ME-GH), each with 2 or 4 moments matched. For ME-Even, we implement the discretization exactly as in Algorithm 3.1, except that we use the Gaussian mixture density instead of the normal density. We choose the grid spacing as the upper bound in Corollary 3.5. For ME-GH, we take the following approach. Suppose the true (Gaussian mixture) density at a given grid point is $f(x)$. Let $\phi(x)$ be the normal density with mean 0 and the same standard deviation as $f(x)$. Then the

expectation of a function $g(x)$ is

$$\int g(x)f(x)\,\mathrm{d}x = \int g(x)\frac{f(x)}{\phi(x)}\phi(x)\,\mathrm{d}x \approx \sum_{n=1}^{N} w_n \frac{f(x_n)}{\phi(x_n)} g(x_n),$$

where $\{x_n\}$ and $\{w_n\}$ are nodes and weights for the Gauss-Hermite quadrature corresponding to $\phi(x)$. This argument suggests that we can use the Gauss-Hermite quadrature grid with weights $w'_n = w_n \frac{f(x_n)}{\phi(x_n)}$ in order to discretize $f(x)$. Figure 4.3 plots the $\log_{10}$ relative errors of the AR(1) model with Gaussian mixture shocks. Table 4.3 shows the mean and maximum $\log_{10}$ relative errors.



(a) $N = 9$, $\gamma = 2$.

(b) $N = 15$, $\gamma = 2$.

(c) $N = 9$, $\gamma = 5$.

(d) $N = 15$, $\gamma = 5$.

Figure 4.3: $\log_{10}$ relative errors with various discretization methods and number of points for the Gaussian mixture model.

Note: the top panels show the accuracy for approximations to the benchmark model with risk aversion $\gamma = 2$ and different number of grid points $N = 9, 15$. The bottom panels show the results for an alternative specification in which the risk aversion is higher at $\gamma = 5$.

As we can see from the figure and the table, the order of accuracy is always ME-GH $\approx$ ME-Even $>$ Rouwenhorst $\approx$ Tauchen-Hussey, and matching 4 moments instead of 2 increases the solution accuracy by about 1 to 2 orders of magnitude. For low risk aversion ($\gamma = 2$), even the misspecified models (Rouwenhorst and Tauchen-Hussey) have relative errors less than 10^{-2} or 1%, so the choice of

Table 4.3: Mean and maximum $\log_{10}$ relative errors for the AR(1) asset pricing model with Gaussian mixture shocks.

		ME methods				Existing methods	
N	γ	Even (2)	Even (4)	GH (2)	GH (4)	R	TH
Mean $\log_{10}$ *errors*							
9	2	-3.381	-5.013	-3.602	-5.176	-2.602	-2.606
15	2	-4.264	-6.445	-5.189	-6.414	-2.604	-2.606
9	5	-1.466	-2.071	-1.602	-2.182	-0.909	-0.919
15	5	-2.137	-2.948	-2.774	-3.467	-0.913	-0.919
Maximum $\log_{10}$ *errors*							
9	2	-3.239	-4.698	-3.406	-4.978	-2.587	-2.603
15	2	-3.935	-5.821	-4.748	-5.673	-2.591	-2.602
9	5	-1.307	-1.913	-1.413	-2.018	-0.874	-0.900
15	5	-1.854	-2.639	-2.464	-3.184	-0.875	-0.892

Note: Even (L): even-spaced grid with L moments; GH (L): Gauss-Hermite quadrature grid with L moments; R: Rouwenhorst (1995) method; TH: Tauchen and Hussey (1991) method.

the discretization method does not matter so much. However, with higher risk aversion ($\gamma = 5$), the misspecified models are off by more than 10^{-1} (10%), while ME methods with 4 moments has errors less than 10^{-2} (1%) with 9 points and 10^{-3} (0.1%) with 15 points. Hence the choice of the discretization method makes an economically significant difference when risk aversion is moderately high, which is often the case for many asset pricing models in the literature.

5 Solution accuracy of a rare disasters model

To illustrate the general applicability of our method, in this section we solve an asset pricing model with variable rare disasters (Gabaix, 2012). There are several good reasons to consider this model. First, the dynamics of the underlying stochastic process are nonlinear and non-Gaussian, which makes our method more useful. Second, Gabaix's model admits closed-form solutions, which makes the accuracy comparison particularly simple. Finally, since rare disaster models have recently become quite popular in the literature (Rietz, 1988; Barro, 2006; Gourio, 2012; Wachter, 2013), providing a simple yet accurate solution algorithm seems to be useful, especially for the purpose of calibration and estimation.

5.1 Model

Gabaix (2012) considers a representative-agent asset pricing model in an endowment economy. The representative agent has CRRA preferences

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \mathrm{e}^{-\rho t} \frac{C_t^{1-\gamma}}{1-\gamma},$$

where $\rho > 0$ is the discount rate and $\gamma > 0$ is relative risk aversion. Disasters occur with probability p_t at time $t+1$. The consumption growth is given by

$$\frac{C_{t+1}}{C_t} = \mathrm{e}^{g_C} \times \begin{cases} 1, & \text{(no disaster)} \\ B_{t+1}, & \text{(disaster)} \end{cases}$$

where g_C is the growth rate in normal times and $B_{t+1} \in (0, 1]$ is the consumption recovery rate after a disaster. Similarly, the dividend growth is

$$\frac{D_{t+1}}{D_t} = \mathrm{e}^{g_D} \times \begin{cases} 1, & \text{(no disaster)} \\ F_{t+1}, & \text{(disaster)} \end{cases}$$

where g_D is the growth rate in normal times and $F_{t+1} \in (0, 1]$ is the dividend recovery rate after a disaster. Gabaix (2012) defines the following quantity, which he calls "resilience":

$$H_t = p_t \, \mathrm{E}_t^{\mathrm{D}}[B_{t+1}^{-\gamma} F_{t+1} - 1], \tag{5.1}$$

where $\mathrm{E}_t^{\mathrm{D}}$ denotes the expectation conditional on *d*isaster. Instead of specifying the dynamics of the fundamentals p_t, B_t, F_t individually, Gabaix directly specifies the dynamics of $H_t = H_* + \widehat{H}_t$ as follows:

$$\widehat{H}_{t+1} = \frac{1 + H_*}{1 + H_t} \mathrm{e}^{-\phi_H} \widehat{H}_t + \varepsilon_{t+1}^H, \tag{5.2}$$

where H_* is a constant, $\phi_H > 0$ is the speed of mean reversion at $H_t = H_*$, and ε_{t+1}^H is an innovation. Since $1 + H_t$ appears in the denominator of the right-hand side, (5.2) is a highly nonlinear process. It turns out that the price-dividend ratio at time t depends only on $\widehat{H}_t$ independent of the distribution of ε_{t+1}^H, and Gabaix obtains a closed-form solution (see Eq. (13) in his paper).

5.2 Solution accuracy

To compare numerical solutions obtained by our method to the exact solution, we need to discretize the process (5.2). Since the distribution of the innovation ε_{t+1}^H does not matter, and since Gabaix shows that the process $\left\{\widehat{H}_t\right\}$ must be bounded, we assume that the distribution of $\widehat{H}_{t+1}$ given $\widehat{H}_t$ is a beta distribution (properly rescaled) with mean and variance implied by (5.2). Once we specify the conditional distribution this way, it is straightforward to discretize the Markov process using our method. See the Online Appendix for the details on discretization and the computation of the numerical solution. Although there are no accepted standard ways for solving the rare disasters model, we also compare the solution accuracy of our method to the perturbation method proposed in Levintal (2014).[19]

For the parameter values, following Gabaix (2012) we set the discount rate $\rho = 0.0657$, relative risk aversion $\gamma = 4$, consumption and dividend growth rate $g_C = g_D = 0.025$, disaster probability $p = 0.0363$, consumption recovery rate

[19] `https://sites.google.com/site/orenlevintal/5th-order-perturbation`

$B = 0.66$, and the speed of mean reversion $\phi_H = 0.13$. The implied value for the constant H_* in (5.2) is 0.09. Figure 5.1 shows the ergodic distribution of the variable part of resilience $\widehat{H}$ computed from the discrete approximation with $N = 201$ points. The distribution is bimodal.

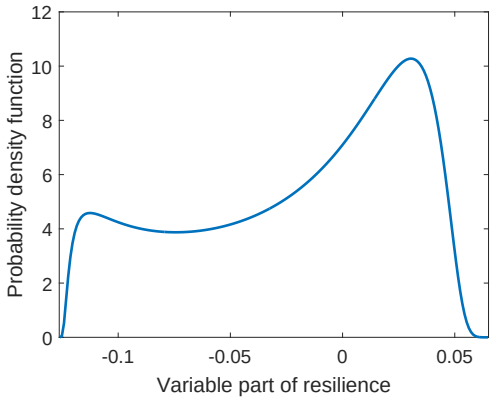
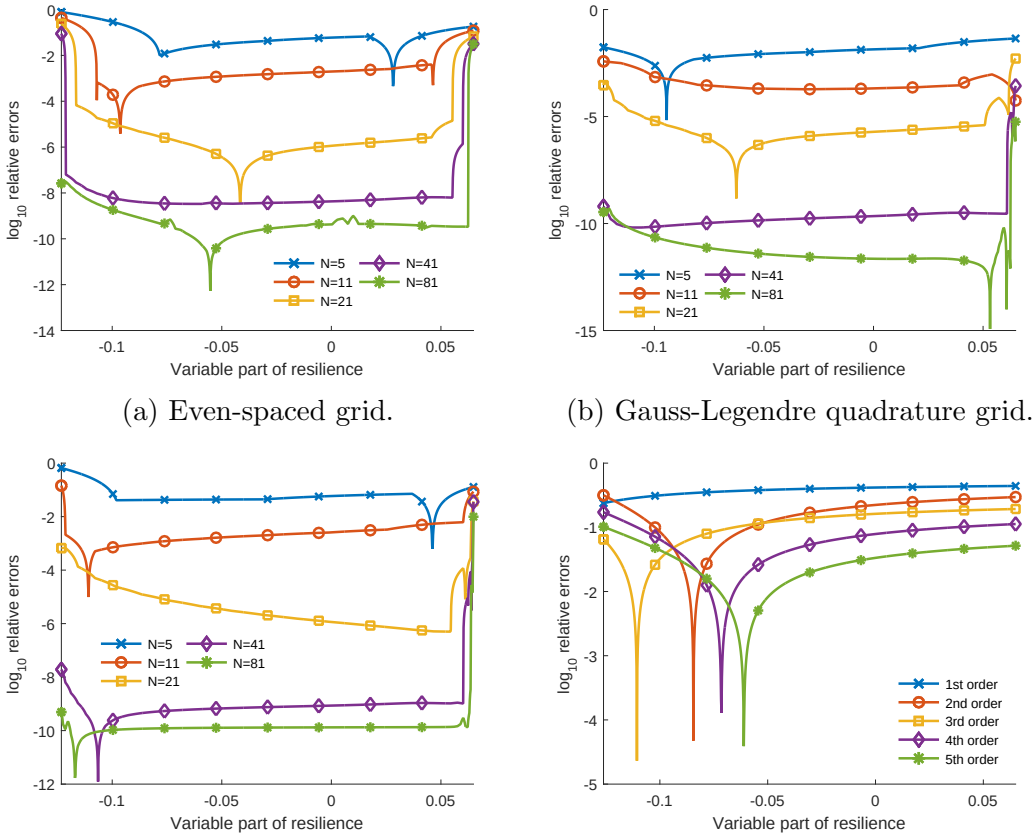


Figure 5.1: Ergodic distribution of the variable part of resilience $\widehat{H}$.

For our method, we consider the even-spaced grid, Gauss-Legendre quadrature grid, and the Clenshaw-Curtis quadrature grid, which are the most natural choices since the integration is over a bounded interval. The number of points are $N = 5, 11, 21, 41, 81$. For the perturbation method in Levintal (2014), we consider up to the fifth-order approximation (the maximum allowed). In order to apply the perturbation method, we need to supply the unconditional standard deviation of the innovation in resilience, ε^H_{t+1}. We compute this number using the ergodic distribution in Figure 5.1, which is 0.0174. We also simulated the true process (5.2) for a long time and verified that we obtain the same number up to four decimal places. Figure 5.2 shows the $\log_{10}$ relative errors of the price-dividend ratio. Table 5.1 shows the mean and maximum $\log_{10}$ relative errors over the entire grid.

Because the resilience process (5.2) is highly nonlinear, we need many grid points in order to obtain an accurate solution. Overall using the Gauss-Legendre quadrature grid (Figure 5.2b) is the most accurate, with relative errors about 10^{-3} with $N = 11$ points, 10^{-5} with $N = 21$ points, and 10^{-10} with $N = 41$ points. Hence for practical purposes 11 points are enough. Clenshaw-Curtis quadrature (Figure 5.2c) is similar to Gauss-Legendre, as documented in Trefethen (2008). The performance of the even-spaced grid (Figure 5.2a) is worse near the boundary points. This is because the conditional variance of the resilience process (5.2) approaches zero near the boundary, which makes it hard to match the conditional variance. Since there are many grid points near the boundary for Gauss-Legendre and Clenshaw-Curtis, a low variance is not a problem. The perturbation method (Figure 5.2d) is not so accurate, with about 10% error with 3rd-order approximation and 2.6% error with 5th-order. Even the 5-point Gauss-Legendre discretization is more accurate than the 5th-order perturbation in terms of both mean and maximum $\log_{10}$ errors.



(a) Even-spaced grid.

(b) Gauss-Legendre quadrature grid.

(c) Clenshaw-Curtis quadrature grid.

(d) Perturbation method.

Figure 5.2: $\log_{10}$ relative errors of price-dividend ratio with various methods and number of points or order of approximation for the variable rare disaster model.

Table 5.1: Mean and maximum $\log_{10}$ relative errors for the variable disaster model.

	ME methods			Perturbation	
N	Even	Gauss-Legendre	Clenshaw-Curtis	Order	
Mean $\log_{10}$ *errors*					
5	-1.187	-1.982	-1.218	1	-0.422
11	-2.582	-3.451	-2.676	2	-0.856
21	-5.383	-5.560	-5.354	3	-1.007
41	-8.007	-9.679	-9.040	4	-1.268
81	-9.228	-11.23	-9.873	5	-1.590
Maximum $\log_{10}$ *errors*					
5	-0.107	-1.353	-0.182	1	-0.356
11	-0.365	-2.422	-0.841	2	-0.501
21	-0.628	-2.291	-1.430	3	-0.715
41	-1.053	-3.567	-1.447	4	-0.765
81	-1.503	-5.245	-2.003	5	-0.992

Do these differences in solution accuracy economically matter? To address this question, we simulate the resilience process (5.2) for $T = 100,000$ periods and compute some financial moments from the true solution as well as the numerical solutions. Table 5.2 shows the results. As expected from Figure 5.2 and Table 5.1, the 11-point Gauss-Legendre discretization gives accurate results up to the third significant digit (0.1%). The perturbation method does not fare well: with the 1st-order approximation, the stock return is 4 percentage points higher than the true value; the 3rd order approximation is off by 10–20%, and the 5th-order approximation is off by about 10% for the standard deviation.

Table 5.2: Financial moments.

	ME methods			Perturbation	
N	Even	Gauss-Legendre	Clenshaw-Curtis	Order	
Mean P/D				True = 16.7330	
5	17.5101	16.9876	17.8134	1	9.9614
11	16.8498	16.7268	16.6894	2	13.6059
21	16.7523	16.7330	16.7329	3	14.2745
41	16.7351	16.7330	16.7330	4	15.6998
81	16.7331	16.7330	16.7330	5	16.3267
Standard deviation of $\log(P/D)$				True = 0.3366	
5	0.2432	0.3467	0.2955	1	0.2640
11	0.3129	0.3371	0.3342	2	0.1859
21	0.3309	0.3366	0.3366	3	0.2718
41	0.3359	0.3366	0.3366	4	0.2717
81	0.3366	0.3366	0.3366	5	0.3020
Mean stock returns (%)				True = 6.9574	
5	6.2558	6.9003	6.3332	1	11.4419
11	6.7882	6.9627	6.9637	2	7.9205
21	6.9187	6.9575	6.9577	3	7.8651
41	6.9527	6.9574	6.9574	4	7.1212
81	6.9572	6.9574	6.9574	5	6.9676
Standard deviation of stock returns (%)				True = 11.8058	
5	10.2217	12.1749	11.3956	1	9.9833
11	11.5335	11.8175	11.7561	2	6.7575
21	11.7549	11.8062	11.8069	3	9.7367
41	11.8003	11.8058	11.8058	4	9.6381
81	11.8055	11.8058	11.8058	5	10.6445

Note: this table shows the financial moments from $T = 100,000$ simulations. "True" indicates the values from the exact solution. The numbers are slightly different from Table III of Gabaix (2012) because (i) we simulate at the annual frequency, while he simulates at the monthly frequency, and (ii) in Gabaix's calibration, the stock resilience volatility is $\sigma_H = 0.019$ while we have $\sigma_H = 0.0174$ because we specify beta distributions for the conditional dynamics.

Based on the numerical results in the last two sections, we provide some rec-

ommendations to allow the reader to make an informed decision on what kind of computational strategy to adopt. The perturbation method is fast but it is inherently a local approximation. When the model is highly nonlinear and shocks are large, the solution accuracy can be poor. Discretization is easy to implement and seems to be accurate enough for most problems. For Gaussian VARs, our method (with even-spaced or quadrature grid) seems best. Numerical results in the appendix suggest that for univariate Gaussian AR(1) process, ME-Quad is most accurate for persistence less than 0.8, ME-Even is most accurate for persistence between 0.8 and 0.99, and the Rouwenhorst method is best for persistence 0.99 and beyond (because the Rouwenhorst method is error-free, *i.e.*, it does not involve any numerical optimization). However, for persistence beyond 0.99, it may be better to use the projection method. Pohl et al. (2015) suggest that for solving the long run risk model (Bansal and Yaron, 2004), which features very persistent processes, using the projection method makes an economically meaningful difference in the solution accuracy. For nonlinear or non-Gaussian processes, as in the rare disasters model, our discretization method would be the first choice since there may not be any readily available quadrature formulas to use along with the projection method.

6 Conclusion

In this paper, we provide a new method for discretizing a general class of stochastic processes by matching low order conditional moments. Our method is computationally tractable and allows researchers to approximate a wide variety of nonlinear non-Gaussian Markov processes. We demonstrate that our method produces discrete approximations which are often several orders of magnitude more accurate than existing methods for both linear and nonlinear stochastic processes. This is the case whether we consider the relative bias of unconditional moments implied by the discretization or the accuracy of solutions to asset pricing models.

Our maximum entropy procedure has a wide range of potential applications beyond asset pricing models. It is common in the quantitative macro literature to use an AR(1) specification for technology or income. We believe that researchers use AR(1) specifications because existing methods do not easily allow for more realistic assumptions. Recent work on the dynamics of the income distribution has shown that while income shocks have roughly constant variance, skewness and kurtosis display significant time-variation (Guvenen et al., 2014). Our method can be used to solve a life cycle model with a realistic income process by matching the dynamics of these higher order moments. Our method can also be used for estimating nonlinear, non-Gaussian state space models (Farmer, 2016). In this paper we considered only tensor grids since our applications involved only one or two state variables. An interesting and important future research topic is to explore the performance of our method in conjunction with sparse grids for solving dynamic models with many state variables.

References

Jérôme Adda and Russel W. Cooper. *Dynamic Economics: Quantitative Methods and Applications*. MIT Press, Cambridge, MA, 2003.

Victor Aguirregabiria and Pedro Mira. Sequential estimation of dynamic discrete games. *Econometrica*, 75(1):1–53, January 2007. doi:10.1111/j.1468-0262.2007.00731.x.

S. Rao Aiyagari. Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics*, 109(3):659–684, 1994. doi:10.2307/2118417.

S. Rao Aiyagari. Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting. *Journal of Political Economy*, 103 (6):1158–1175, December 1995.

S. Borağan Aruoba, Jesús Fernándes-Villaverde, and Juan F. Rubio-Ramírez. Comparing solution methods for dynamic equilibrium economies. *Journal of Economic Dynamics and Control*, 30(12):2477–2508, December 2006. doi:10.1016/j.jedc.2005.07.008.

Ravi Bansal and Amir Yaron. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance*, 59(4):1481–1509, August 2004. doi:10.1111/j.1540-6261.2004.00670.x.

Robert J. Barro. Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, 121(3):823–866, 2006. doi:10.1162/qjec.121.3.823.

Marco Bonomo, René Garcia, Nour Meddahi, and Roméo Tédongap. Generalized disappointment aversion, long-run volatility risk, and asset prices. *Review of Financial Studies*, 24(1):82–122, 2011. doi:10.1093/rfs/hhq116.

Jonathan M. Borwein and Adrian S. Lewis. Duality relationships for entropy-like minimization problems. *SIAM Journal on Control and Optimization*, 29(2): 325–338, March 1991. doi:10.1137/0329017.

Craig Burnside. Solving asset pricing models with Gaussian shocks. *Journal of Economic Dynamics and Control*, 22(3):329–340, March 1998. doi:10.1016/S0165-1889(97)00075-4.

Craig Burnside. Discrete state-space methods for the study of dynamic economies. In Ramon Marimon and Andrew Scott, editors, *Computational Methods for the Study of Dynamic Economies*, chapter 5, pages 95–113. Oxford University Press, Oxford, 1999. doi:10.1093/0199248273.003.0005.

Dario Caldara, Jesús Fernándes-Villaverde, Juan F. Rubio-Ramírez, and Wen Yao. Computing DSGE models with recursive preferences and stochastic volatility. *Review of Economic Dynamics*, 15(2):188–206, April 2012. doi:10.1016/j.red.2011.10.001.

Stephen G. Cecchetti, Pok-sang Lam, and Nelson C. Mark. The equity premium and the risk-free rate: Matching moments. *Journal of Monetary Economics*, 31 (1):21–45, February 1993. doi:10.1016/0304-3932(93)90015-8.

C. W. Clenshaw and A. R. Curtis. A method for numerical integration on an automatic computer. *Numerische Mathematik*, 2(1):195–207, 1960. doi:10.1007/BF01386223.

Fabrice Collard and Michel Juillard. Accuracy of stochastic perturbation methods: The case of asset pricing models. *Journal of Economic Dynamics and Control*, 25(6-7):979–999, June 2001. doi:10.1016/S0165-1889(00)00064-6.

Julio Dávila, Jay H. Hong, Per Krusell, and José-Víctor Ríos-Rull. Constrained efficiency in the neoclassical growth model with uninsurable idiosyncratic shocks. *Econometrica*, 80(6):2431–2467, November 2012. doi:10.3982/ECTA5989.

Philip J. Davis and Philip Rabinowitz. *Methods of Numerical Integration*. Academic Press, Orlando, FL, second edition, 1984.

Leland E. Farmer. The discretization filter: A simple way to estimate nonlinear state space models. 2016. URL `https://ssrn.com/abstract=2780166`.

Martin Flodén. A note on the accuracy of Markov-chain approximations to highly persistent AR(1) processes. *Economics Letters*, 99(3):516–520, June 2008. doi:10.1016/j.econlet.2007.09.040.

Xavier Gabaix. Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *Quarterly Journal of Economics*, 127(2):645–700, 2012. doi:10.1093/qje/qjs001.

Ragchaasuren Galindev and Damba Lkhagvasuren. Discretization of highly persistent correlated AR(1) shocks. *Journal of Economic Dynamics and Control*, 34(7):1260–1276, July 2010. doi:10.1016/j.jedc.2010.02.006.

Nikolay Gospodinov and Damba Lkhagvasuren. A moment-matching method for approximating vector autoregressive processes by finite-state Markov chains. *Journal of Applied Econometrics*, 29(5):843–859, August 2014. doi:10.1002/jae.2354.

François Gourio. Disaster risk and business cycles. *American Economic Review*, 102(6):2734–2766, October 2012. doi:10.1257/aer.102.6.2734.

Fatih Guvenen. A parsimonious macroeconomic model for asset pricing. *Econometrica*, 77(6):1711–1750, November 2009. doi:10.3982/ECTA6658.

Fatih Guvenen, Serdar Ozkan, and Jae Song. The nature of countercyclical income risk. *Journal of Political Economy*, 122(3):621–660, June 2014. doi:10.1086/675535.

Michael Haliassos and Alexander Michaelides. Portfolio choice and liquidity constraints. *International Economic Review*, 44(1):143–177, February 2003. doi:10.1111/1468-2354.t01-1-00065.

James D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, March 1989. doi:10.2307/1912559.

John Heaton and Deborah J. Lucas. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy*, 104(3):443–487, June 1996.

Florian Heiss and Viktor Winschel. Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics*, 144(1):62–80, May 2008. doi:10.1016/j.jeconom.2007.12.004.

Kenneth L. Judd. Projection methods for solving aggregate growth models. *Journal of Economic Theory*, 58(2):410–452, December 1992. doi:10.1016/0022-0531(92)90061-L.

Kenneth L. Judd, Felix Kubler, and Karl Schmedders. Bond ladders and optimal portfolios. *Review of Financial Studies*, 24(12):4123–4166, 2011. doi:10.1093/rfs/hhr074.

Yuichi Kitamura. Empirical likelihood methods in econometrics: Theory and practice. In Richard Blundell, Whitney Newey, and Torsten Persson, editors, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, volume 3 of *Econometric Society Monographs*, chapter 7, pages 174–237. Cambridge University Press, New York, 2007.

Karen A. Kopecky and Richard M. H. Suen. Finite state Markov-chain approximations to highly persistent processes. *Review of Economic Dynamics*, 13(3): 701–714, July 2010. doi:10.1016/j.red.2010.02.002.

Dirk Krueger and Felix Kubler. Computing equilibrium in OLG models with stochastic production. *Journal of Economic Dynamics and Control*, 28(7):1411–1436, April 2004. doi:10.1016/S0165-1889(03)00111-8.

Solomon Kullback and Richard A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, March 1951. doi:10.1214/aoms/1177729694.

Oren Levintal. Fifth order perturbation solution to DSGE models. 2014. URL `http://ssrn.com/abstract=2364989`.

Lilia Maliar and Serguei Maliar. Merging simulation and projection approaches to solve high-dimensional problems with an application to a new Keynesian model. *Quantitative Economics*, 6(1):1–47, March 2015. doi:10.3982/QE364.

Rajnish Mehra and Edward C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161, March 1985. doi:10.1016/0304-3932(85)90061-3.

Art B. Owen. *Empirical Likelihood*. Number 92 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, 2001.

Walter Pohl, Karl Schmedders, and Ole Wilms. Higher-order dynamics in asset-pricing models with recursive preferences. 2015. URL `https://ssrn.com/abstract=2540586`.

Thomas A. Rietz. The equity risk premium: A solution. *Journal of Monetary Economics*, 22(1):117–131, July 1988. doi:10.1016/0304-3932(88)90172-9.

K. Geert Rouwenhorst. Asset pricing implications of equilibrium business cycle models. In Thomas F. Cooley, editor, *Frontiers of Business Cycle Research*, chapter 10, pages 294–330. Princeton University Press, 1995.

Stephanie Schmitt-Grohé and Martín Uribe. Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control*, 28(4):755–775, January 2004. doi:10.1016/S0165-1889(03)00043-5.

Ken'ichiro Tanaka and Alexis Akira Toda. Discrete approximations of continuous distributions by maximum entropy. *Economics Letters*, 118(3):445–450, March 2013. doi:10.1016/j.econlet.2012.12.020.

Ken'ichiro Tanaka and Alexis Akira Toda. Discretizing distributions with exact moments: Error estimate and convergence analysis. *SIAM Journal on Numerical Analysis*, 53(5):2158–2177, 2015. doi:10.1137/140971269.

George Tauchen. Finite state Markov-chain approximations to univariate and vector autoregressions. *Economics Letters*, 20(2):177–181, 1986. doi:10.1016/0165-1765(86)90168-0.

George Tauchen and Robert Hussey. Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica*, 59(2): 371–396, March 1991. doi:10.2307/2938261.

Stephen J. Terry and Edward S. Knotek, II. Markov-chain approximations of vector autoregressions: Application of general multivariate-normal integration techniques. *Economics Letters*, 110(1):4–6, January 2011. doi:10.1016/j.econlet.2010.09.008.

Lloyd N. Trefethen. Is Gauss quadrature better than Clenshaw-Curtis? *SIAM Review*, 50(1):67–87, 2008. doi:10.1137/060659831.

Min Tsao. Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Annals of Statistics*, 32(3):1215–1221, 2004. doi:10.1214/009053604000000337.

Min Tsao and Fan Wu. Empirical likelihood on the full parameter space. *Annals of Statistics*, 41(4):2176–2196, 2013. doi:10.1214/13-AOS1143.

Efthymios G. Tsionas. Exact solution of asset pricing models with arbitrary shock distributions. *Journal of Economic Dynamics and Control*, 27(5):843–851, March 2003. doi:10.1016/S0165-1889(02)00017-9.

Joseph Vavra. Inflation dynamics and time-varying volatility: New evidence and an Ss interpretation. *Quarterly Journal of Economics*, 129(1):215–258, 2014. doi:10.1093/qje/qjt027.

Jessica A. Wachter. Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance*, 68(3):987–1035, June 2013. doi:10.1111/jofi.12018.

Ivo Welch and Amit Goyal. A compehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508, 2008. doi:10.1093/rfs/hhm014.

Lu Zhang. The value premium. *Journal of Finance*, 60(1):67–103, February 2005. doi:10.1111/j.1540-6261.2005.00725.x.

Online Appendix

A Proofs

Proof of Theorem 2.1. 1. 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.

2. 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 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 $\bar{T} \in \operatorname{int} \operatorname{co} T(D_N)$.

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

Proof of Theorem 3.2. Special case of the following theorem by setting $\Sigma_t = D$ (constant). □

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

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

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

$$\begin{aligned} \mathrm{E}\left[y_t^d \,\middle|\, z_{t-1}^d\right] &= \mathrm{E}\left[y_t \,\middle|\, z_{t-1}^d\right] = A y_{t-1}^d, \\ \mathrm{Var}\left[y_t^d \,\middle|\, z_{t-1}^d\right] &= \mathrm{Var}\left[y_t \,\middle|\, z_{t-1}^d\right] = \Sigma_{t-1}. \end{aligned}$$

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

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

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 $y_t^d = Ay_{t-1}^d + \varepsilon_t^d$, we get $\mathrm{E}[y_t^d] = 0$. Therefore the unconditional mean is exact. To compute the variance, note that

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

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

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

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

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

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 k-th order autocovariance matrix and $\Gamma^d(k) = \mathrm{E}[y_{t+k}^d(y_t^d)']$ be that of the discretized process. Multiplying $(y_t^d)'$ from the right to both sides of $y_{t+k+1}^d = Ay_{t+k}^d + \varepsilon_{t+k+1}^d$ 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

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

for any $j \geq 1$. Therefore $\mathrm{E}\left[y_{t+k}^d \,\middle|\, z_t^d\right] = A^k y_t^d$, 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 $\mathrm{E}\left[\Sigma_t \,\middle|\, \Sigma_{t-1}\right]$.

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

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

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

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

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 $M = \max_n |\bar{x}_n|$. 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 $x = M$, in order for $X' \in \operatorname{int}\operatorname{co} T(D)$ it is necessary that X' lies below the segment PQ, so we need

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

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_1A_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

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

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

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

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

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

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 even-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 $0, \pm M$, so $N \geq 3$.

Case 1: $\boldsymbol{\rho \leq 1 - \frac{2}{N-1}}$. By Proposition 3.4, it suffices to show $d < 2 \iff M < N-1$. Since $M \leq \sqrt{2}\sigma\sqrt{N-1}$ by assumption, it suffices to show

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

But this inequality is trivial because $\rho^2 < \rho \leq 1 - \frac{2}{N-1}$.

Case 2: $\boldsymbol{\rho > 1 - \frac{2}{N-1}}$. 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 $\left(\rho\bar{x}_n - \frac{1}{(1-\rho)\bar{x}_n}, \rho\bar{x}_n\right)$ 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

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

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

$$\frac{\sqrt{N-1}}{\sqrt{1-\rho^2}} < \frac{N-1}{\sqrt{2(1-\rho)(N-3)}} \iff 1+\rho > \frac{2(N-3)}{N-1} \iff \rho > 1 - \frac{4}{N-1},$$

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

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

which is trivial. The left inequality is equivalent to

$$\begin{aligned}\rho\bar{x}_N - \frac{1}{(1-\rho)\bar{x}_N} < \bar{x}_{N-1} &\iff \rho M - \frac{1}{(1-\rho)M} < M\frac{N-3}{N-1} \\ &\iff M^2\left(\rho - \frac{N-3}{N-1}\right) < \frac{1}{1-\rho}.\end{aligned}$$

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

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

which is trivial. □

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 1-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

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

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

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

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

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

where

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

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 $\log_{10}\left|\theta^d/\theta - 1\right|$ for various number of nodes in each dimension, $N = 5, 9, 15, 21$. For our method, we consider the even-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).[20] 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.

Table B.1: $\log_{10}$ 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: number of discrete points in each dimension; TH: Tauchen and Hussey (1991) method; GL, GL0: Gospodinov and Lkhagvasuren (2014) methods with or without moment targeting; ME: 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 $\{y_t\}$.

We can make a few observations from Table B.1. First, as is well-known,

[20]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 in order to make the unconditional variance as close to the true VAR as possible.

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.

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 to 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}.[21] 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, 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 $1-\zeta_1, 1-\zeta_2$ 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 $1-\zeta_1, 1-\zeta_2$ 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 number 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 which 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 and AR(1) process (for which the Rouwenhorst method is accurate)

[21] 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 2.2GHz Intel Core i5 processor. The Tauchen method matches the unconditional variance. The codes for the ME methods are available on our website discussed in Appendix E. The GL methods use the codes supplied in the online appendix of Gospodinov and Lkhagvasuren (2014).

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 even-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 $\lambda = 0.95$, $\rho = 0.9$, $\sigma = 0.06$, 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 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.

Table B.3: $\log_{10}$ 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
	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
9	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
	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
15	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
	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
21	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

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. In order 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

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

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

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

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

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

where $r = \beta\exp\left(\alpha'\mu + \frac{1}{2}\alpha'\tilde{\Psi}\alpha\right) < 1$. 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), $m_n = \log M((1-\gamma)(1-\rho^n)/(1-\rho))$, $S_n = \sum_{k=1}^{n} m_n$, $b = \lim b_n = \frac{1-\gamma}{1-\rho}\rho$, $m = \lim m_n = \log M(\frac{1-\gamma}{1-\rho})$, *and assume $r = \beta((1-\gamma)\mu + m) < 1$. Then*

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

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

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

If we replace ρ^n by 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

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

□

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, in order for (5.2) to be stable, we need

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

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

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

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 $[\widehat{H}_{\min}, \widehat{H}_{\max}]$. In order 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 $K = 0.2\phi_H \left|\widehat{H}_{\min}\widehat{H}_{\max}\right|$. See Eq. (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 which 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 end points 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 even-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 $V_t = P_t/D_t$ be the price-dividend ratio, it follows from the Euler equation that

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

where p_t is the disaster probability and $\mathrm{E}_t^{\mathrm{ND}}, \mathrm{E}_t^{\mathrm{D}}$ 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 ration using discretization, suppose the state space of resilience H_t is discretized, and let $s = 1, \ldots, S$ 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 $v = (v_1, \ldots, v_S)$ and

$h = (h_1, \ldots, h_S)$ be the vectors of those values, and $P = (\pi_{ss'})$ be the transition probability matrix, it follows that

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

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

D Asset pricing with Gaussian AR(1) shocks

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

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

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 $\beta = 0.95$. In order to avoid cherry-picking, we consider all major existing methods, Tauchen (1986),[22] Tauchen and Hussey (1991), and Rouwenhorst (1995). For the ME methods, we consider ME-Even, ME-Quant, ME-Quad (all with two moments) as well as ME-Even with 4 moments.[23] We consider two robustness checks, (i) changing the number of grid points N, and (ii) changing the persistence of dividend growth ρ.[24] The number of grid points is always $N = 9$ unless otherwise stated.

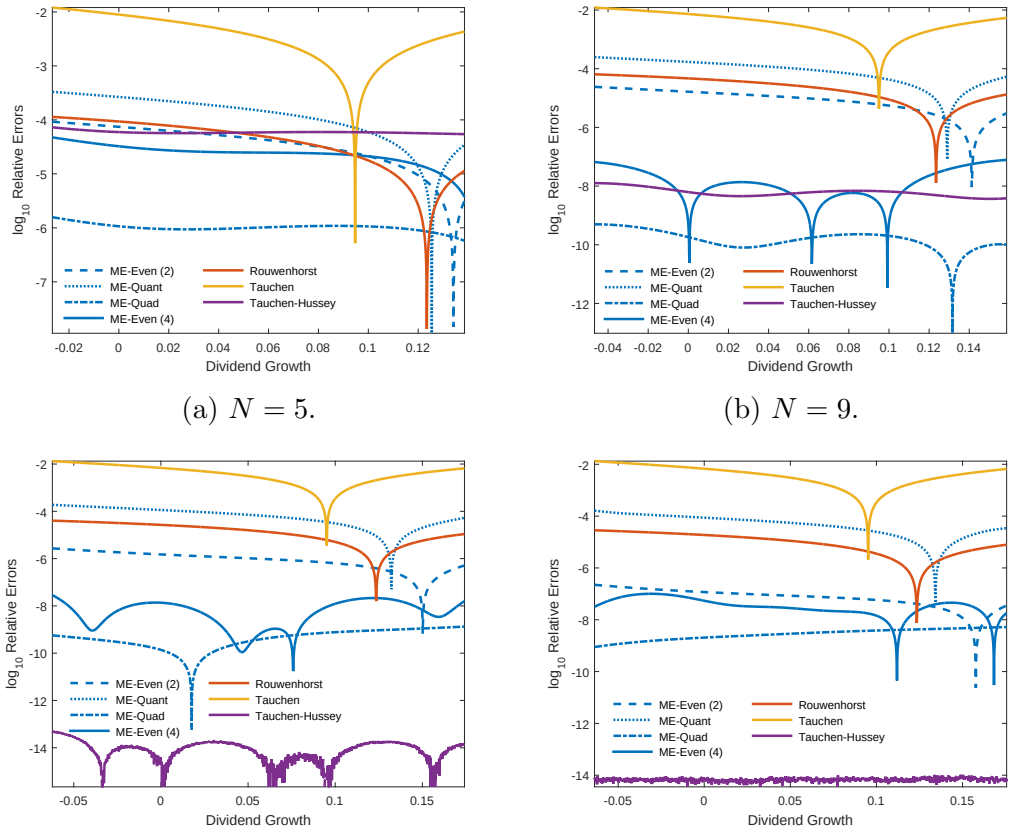
Figure D.1 shows the $\log_{10}$ relative errors of the price-dividend ratio with various discretization methods and number of points N. We can make a few observations. First, as we increase N, all methods become more accurate, as expected. This is especially true for Tauchen-Hussey, whose performance 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 (4 moments) give a solution accuracy of order 10^{-4} to 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

[22] 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 in order 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.

[23] As discussed below, ME-Quant is uniformly dominated by other ME methods, so there is no point in considering ME-Quant with 4 moments. The results for ME-Quad with 4 moments are similar to 2 moments. We also considered matching 6 moments, but the performance is similar to 4 moments.

[24] 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) $N = 9$.

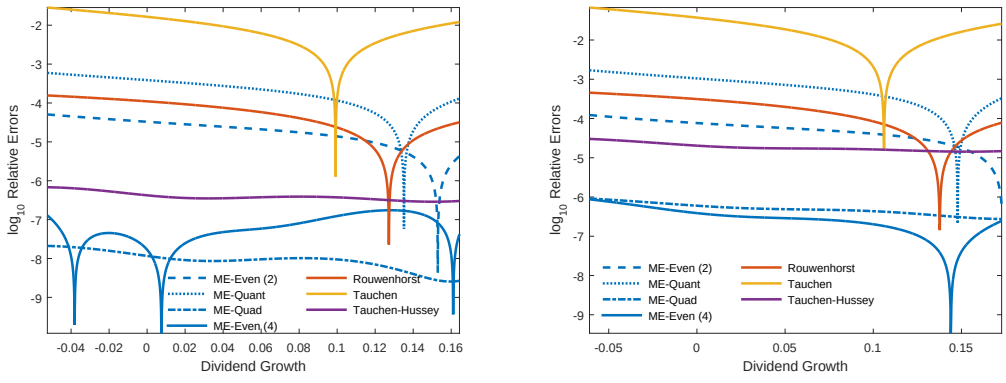
(c) $N = 15$. (d) $N = 21$.

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

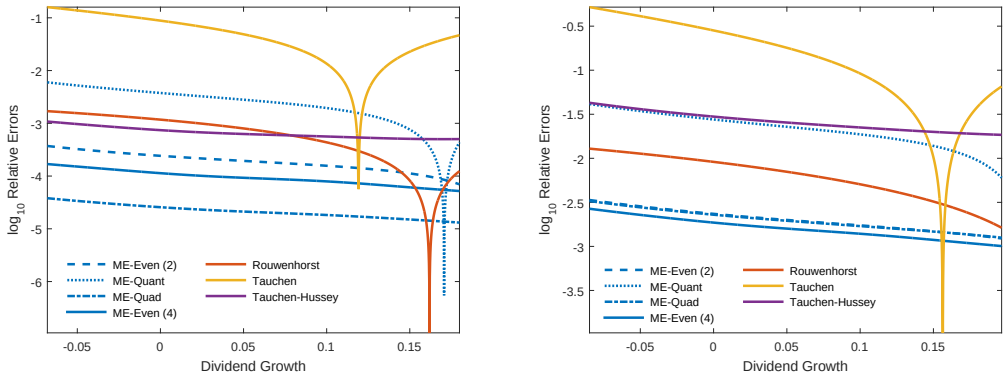
worsens for all methods as we make the dividend process more persistent. However, the performance of the Tauchen-Hussey method deteriorates quickly, as is well-known. ME-Quad, which uses the same Gauss-Hermite quadrature grid as Tauchen-Hussey, also gets poorer, but it is still the best performer along with ME-Even (4 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. In order 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 2 or 4 moments.

To get a better idea of the solution accuracy, consider an investor purchasing



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

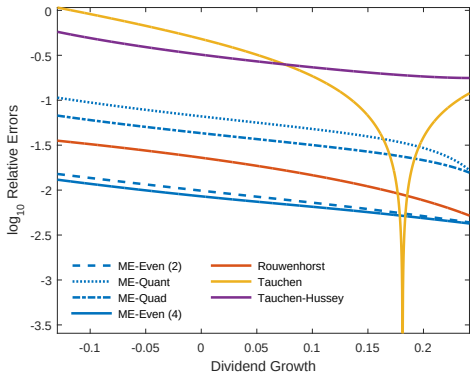


(c) $\rho = 0.7$. (d) $\rho = 0.8$.

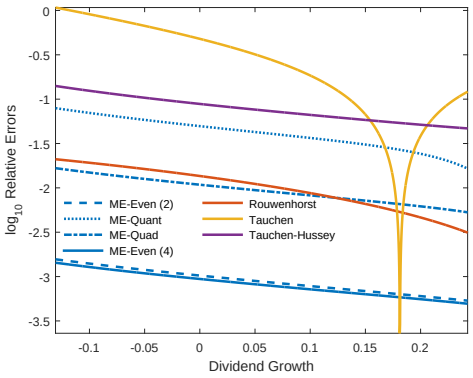
Figure D.2: $\log_{10}$ relative errors of price-dividend ratio with various discretization methods and persistence for the Gaussian AR(1) model.

\$1 Million worth of the 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 \$1M investment is only 1 cent with ME-Even (4 moments). With ME-Quad and Tauchen-Hussey, the pricing error is virtually zero. Even with the Rouwenhorst method, the mispricing is only \$18, so it does not make a material difference across methods except the Tauchen method, which is off by more than \$3,000. 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 by 2.6%, Rouwenhorst by 0.6%, as opposed to 0.16% with ME-Even (4 moments). The result is even more stark with $\rho = 0.9, 0.95$.

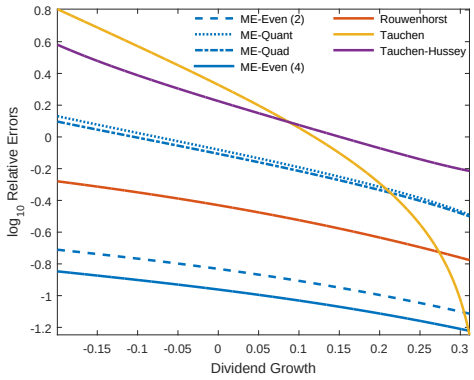
In summary, we find that for discretizing a Gaussian AR(1) process, (i) Tauchen-Hussey is best if there are many points ($N \geq 15$) and the process is not so persistent ($\rho \leq 0.4$), (ii) ME-Quad is best if the process is moderately persistent ($0.4 \leq \rho \leq 0.8$), with ME-Even (4 moments) 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,



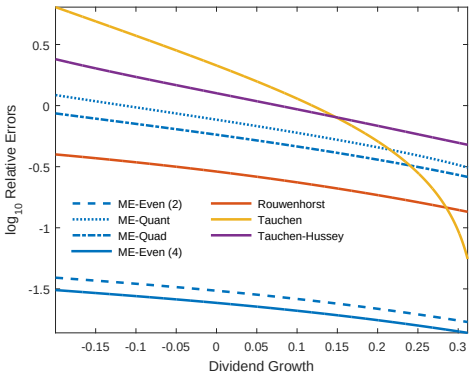
(a) $\rho = 0.9$, $\gamma = 1.5$, $N = 9$.



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

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 et al. (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.

Table D.1: Mispricing in dollars when investing $1 Million.

		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	3,161	58.9
9		7.27	71.1	0	0.011	18.1	3,136	0.006
15		0.767	51.7	0	0.005	11.2	3,380	0
21		0.065	39.8	0	0.03	7.89	3,363	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): ME-Even method with L moments; R: Rouwenhorst (1995) method; TH: Tauchen and Hussey (1991) method. K, M denote thousands and millions of dollars.

This is probably because the Rouwenhorst method does not involve any numerical optimization.

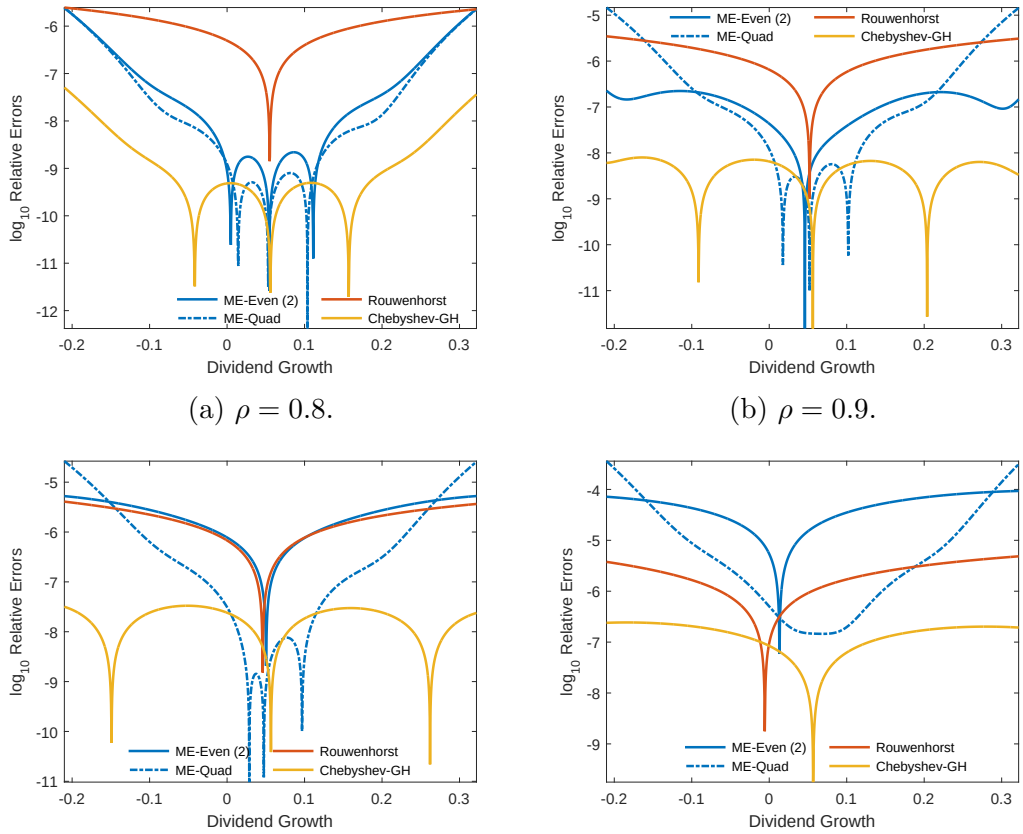
E Matlab files

We implement the discretization of various stochastic processes in the Matlab files posted on our website.[25]

E.1 Subroutines

`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). `discreteApproximation.m` solves the minimization problem (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.

[25] `https://sites.google.com/site/discretevar/`



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

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

Figure D.4: $\log_{10}$ relative errors of 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.

E.2 VAR

`discreteVAR.m` requires four input arguments, the parameters b, B, 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'` (even-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: $\sqrt{N-1}$). `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

`discreteSV.m` discretizes the stochastic volatility model in (3.4). It requires seven input arguments, `lambda`, `rho`, `sigmaU`, `sigmaE`, `Ny`, `Nx`, `method`. `lambda`, `rho`, `sigmaE` are λ, ρ, σ in (3.4). $\sigma_u =$ `sigmaU` is the unconditional volatility of the y_t process, so $\sigma_u^2 = \mathrm{E}[\mathrm{e}^{x_t}]$. `Ny`, `Nx` are the number of grid points for the y_t, 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 even-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

`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`, `Nm`, and three optional arguments `nMoments`, `method`, `nSigmas`. `mu`, `rho` are the AR(1) parameters μ, ρ. `gmObj` is the Matlab Gaussian mixture object,[26] 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 even-spaced grid (trapezoidal formula). The others are quadrature formulas corresponding to each name. (In the paper we discuss only the even-spaced and Gauss-Hermite quadrature grid because the solution accuracy using the Gauss-Legendre and Clenshaw-Curtis quadrature are about the same as even-spaced grid.)

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